The Micro–Aggregated Profit Share[*]

Thomas Hasenzagl

University of Minnesota

Luis Pérez

JOB MARKET PAPER

University of Minnesota

Click here for the most recent version.

This version: October 9, 2023.

First version: May 12, 2023.

Abstract

How much has market power increased in the United States in the last fifty years? Using micro-level data from U.S. Compustat, we find that several indicators of market power have steadily increased since 1970. In particular, the aggregate markup has gone up from 10% of price over marginal cost in 1970 to 23% in 2020, and aggregate returns to scale have risen from 1.00 to 1.13. We connect these market-power indicators to profitability by showing that the aggregate profit share can be expressed in terms of the aggregate markup, aggregate returns to scale, and a sufficient statistic for production networks that captures double marginalization in the economy. We find that despite the rise in market power, the profit share has been constant at 18% of GDP because the increase in monopoly rents has been completely offset by rising fixed costs and changes in technology. Our empirical results have subtle implications for policymakers: overly aggressive enforcement of antitrust law could decrease firm dynamism and paradoxically lead to lower competition and higher market power.

JEL Codes: E25, D22, L16, L40.
Keywords: Market Power, Profit Share, Aggregation, Markups, Networks.

1 Introduction

How much has market power increased in the United States over the last fifty years? We use micro-level data from U.S. Compustat to address this question and find that several indicators of market power have steadily increased since 1970. We document that the aggregate markup, measured as the ratio of price to marginal cost, has gone up from 10% in 1970 to 23% in 2020 and that aggregate returns to scale have risen from 1.00 to 1.13. We map these market-power indicators to the aggregate profit share, which is the fraction of aggregate value added that is not used to compensate factors of production or cover fixed operating costs. More specifically, we show that the aggregate profit share can be expressed in terms of the aggregate markup, aggregate returns to scale, and a sufficient statistic for production networks that captures double marginalization in the economy (i.e., how profits propagate from downstream sellers to upstream suppliers.)

By connecting indicators of aggregate market power to the profit share, we can understand the determinants of the profit share. When firms gain market power, monopoly rents increase, but if fixed costs of production rise simultaneously, profits may remain unchanged. We argue that this narrative describes the experience of the United States between 1970 and 2020. Over that period, monopoly rents increased from 18% to 41% of GDP while the profit share remained constant at 18% of GDP. We reconcile the increase in market power with a constant profit share by showing that the rise in monopoly rents associated with the increase in markups was counteracted by a proportional increase in returns to scale. Rising returns to scale reflect changes in technology and increases in fixed costs, such as R&D expenses, advertising, regulatory compliance costs, and costs associated with information and communication technologies.

Our findings are consistent with a view of the economy in which rising fixed costs lower competition by encouraging firm exit and discouraging firm entry.[1,2] Rising fixed costs increase entry barriers.[3] Monopoly rents are only sustainable in the presence of entry barriers since such rents will otherwise attract entrants and will be competed away. At the same time, firms that do not have enough market power to earn the monopoly rents required to cover rising fixed costs will exit. This lowers the degree of competition in the economy and allows surviving firms to increase their monopoly rents. However, firms use this increase in monopoly rents to cover the increasing fixed costs, and profits remain unchanged.

---

[1]This narrative is consistent with models of monopolistic competition *à la* Dixit and Stiglitz (1977) and models of oligopolistic competition *à la* Atkeson and Burstein (2008).

[2]Business Dynamics Statistics from the US Census Bureau indicate that entry and exit rates of establishments and firms have both declined since 1980.

[3]De Ridder (2019), Aghion et al. (2019), Gutiérrez et al. (2021), De Loecker et al. (2022), and others have pointed out the important role of rising fixed- and entry costs.

Our empirical findings have subtle implications for policymakers. The increasing monopoly rents reflect increases in market power that could be counteracted by more aggressive enforcement of antitrust law. Such aggressive enforcement could, however, lower monopoly rents, making it unsustainable for some firms to operate given the high fixed costs. Therefore, overly aggressive enforcement of antitrust law could decrease firm dynamism and paradoxically lead to lower competition and higher market power.

Our empirical analysis is built on three novel theoretical results. First, we show that the aggregate profit share can be constructed by weighing individual profit rates (i.e., profits over sales) using Domar weights (i.e., producer sales divided by GDP.)[4] Second, we use economic theory to derive a general expression of the profit rate in terms of a producer's markup, markdowns, and returns to scale. Third, we show that the profit share can be expressed in terms of several measures of aggregate market power—the aggregate markup, aggregate returns to scale, an aggregate monopsony term—and a sufficient statistic for production networks that captures double marginalization.

Our first theoretical result establishes that the aggregate profit share can be constructed from micro-level data by weighting producers' profit rates—defined as profits over sales—using Domar weights (i.e., producer sales over GDP.) Domar weights are sufficient statistics for production networks that capture the influence of each producer on aggregate value added. If there are no production networks, Domar weights reduce to sales weights, sum up to one, and reflect the influence of producers on aggregate value added through the sale of final goods. With production networks, Domar weights differ from sales shares, their sum exceeds unity, and they reflect both the direct- and the indirect influence of producers on aggregate value added. Producers have a direct influence on aggregate value added through the sale of final consumption goods, and an indirect influence through the sale of intermediate goods to other producers.

Our second theoretical result provides a general expression for a producer's *economic* profit rate in terms of its markup, markdowns, and returns to scale.[5,6] We put emphasis on economic profits to distinguish them from accounting profits. Economic profits, unlike accounting profits, are unobservable and capture the opportunity cost of using factors of production. We are able to recover economic profits, cast in the form of an economic profit rate, under very mild assumptions on producer behavior and technology. All that is needed to establish our second result is cost minimization and

---

[4]It is worth emphasizing that Domar aggregation extends to the construction of other factor shares. When expenditures on factors of production are expressed relative to sales, labor- and capital shares can also be constructed from micro data using Domar weights.

[5]A similar profit rate expression is provided by Basu (2019) and Syverson (2019). Our expression is more general than theirs since we allow for fixed costs and market power in factor markets.

[6]Implicit in our formulation is the assumption that the firm's optimality condition for capital holds in every period, which also allows us to back out the user-specific cost of capital.

the existence of a production technology that satisfies standard regularity conditions (i.e., differentiability, quasiconcavity, and Inada conditions.)

Why do economic profit rates depend on markups, markdowns, and returns to scale? A producer exerts market power if it charges a price that is above marginal cost—that is, if its markup is greater than unity—or if it compensates factors of production (say, workers) at a rate below their marginal revenue product—that is if its markdowns are lower than unity. Despite exerting market power, a producer only profits from market power when its rents from monopoly and monopsony power exceed its fixed costs of production, which are captured by the returns to scale.

Our third theoretical result is an aggregation theorem that expresses the aggregate profit share in terms of several indicators of aggregate market power—the aggregate markup, aggregate returns to scale, an aggregate monopsony term—and a sufficient statistic for production networks that captures the degree of vertical integration in the economy. Production networks are important because of double marginalization: sales of upstream producers do not only capture the market-power rents of these producers but also those of downstream suppliers.

This aggregation theorem is our main theoretical contribution, and we use it to guide our empirical analysis. It is a powerful result and presents several advantages over existing formulas for computing the aggregate profit share. First and foremost, our theorem links aggregate indicators of market power to the aggregate profit share, allowing us to understand the determinants of the profit share, as well as the origins of aggregate profits. Does an increase in market power translate into an increase in profits? If not, why? Do aggregate profits originate from monopoly, monopsony, or both? And to what extent do each of these forces drive the level of aggregate profits?

Apart from allowing us to answer these questions, our theorem implies aggregate measures of markups, markdowns, and returns to scale. Aggregation of firm-level markups is a topic that has generated substantial debate among academic economists. We contribute to this debate by suggesting that the aggregate markup is the harmonic sales-weighted markup, that aggregate markdowns are sales-weighted markdowns, and that aggregate returns to scale are sales-weighted returns to scale.

Our theorem can also be used to assess the external validity of micro-level estimates of markups, markdowns, and returns to scale. Do firm-level estimates of these objects imply reasonable profit shares? In Appendix C, we use our theorem to elucidate the back-and-forth discussion between De Loecker et al. (2020) and Basu (2019) on whether De Loecker et al.'s micro-level estimates have unreasonable macroeconomic implications. Under additional assumptions on technology, our theorem can also be used to easily calibrate economic models with monopolistic or monopsonistic wedges (e.g., obtaining a markup shock series for a New Keynesian model.)

Related Literature

The study of market power has been a subject of extensive research, with a rich and expansive literature dating back to the works of Pigou (1920), Chamberlin (1933), and Robinson (1933).

Our paper mainly relates to a large literature on market power and macroeconomics. A very influential paper in this literature is De Loecker, Eeckhout and Unger (2020). They estimate markups in Compustat data and find that the aggregate markup has increased from 21% of price over marginal cost in 1980 to 61% in 2016. Instead, we find that the aggregate markup has increased from 12% to 23% over the same period. There are three reasons why our series differ. The main reason is that we calculate the aggregate markup as the harmonic sales-weighted average of firm-level markups while De Loecker, Eeckhout and Unger use the sales-weighted average markup. Our aggregation scheme is based on theory and, unlike theirs, has a solid theoretical foundation.[7] Had we used the sales-weighted average markup instead of the harmonic sales-weighted one, we would have found an increase in price over marginal cost of 22 (rather than 13) percentage points over the last 50 years. The second reason is that we categorize selling, general, and administrative expenses (SG&A,) as well as the costs of goods sold (COGS,) as variable costs, following Traina (2018).[8] The different categorization of variable costs explains less than half of our discrepancies. A third reason why our series of aggregate markup differs—one that does not matter much in practice—is that we use a measure of the capital stock that includes both physical- and intangible capital.[9]

While we are interested in the evolution of the aggregate markup and other indicators of aggregate market power, our main objective is to study the connection between market power and the profit share. We establish and study this connection through our three theoretical results. In doing so, our work relates and contributes to the literatures on production networks and the functional distribution of income. Here, we briefly review some of the most pertinent papers in those literatures and discuss our contributions in relation to them.

---

[7]Other studies in which the harmonic sales-weighted markup emerges as a natural measure of aggregate markup include Baqaee and Farhi (2020), Edmond et al. (2023), and Smith and Ocampo (2022).

[8]Markup estimates are sensitive to the choice of variable costs, but estimates of the profit share are not because profit rates are the same regardless of whether inputs are categorized as variable or fixed as long as all costs are accounted for.

[9]A recent literature relates the rise in market power to the increasing importance of intangible capital, such as software and intellectual property products (see, for example, Crouzet et al., 2022; De Ridder, 2019). On the one hand, investment in intangibles increases fixed costs. On the other hand, the presence of intangibles in the production process lowers marginal costs, allowing firms to charge higher markups. Our empirical analysis captures these two mechanisms. We do so by including intangibles in the stock of capital in our production function estimation.

Production Networks. Research on production networks, which builds upon the pioneering work of François Quesnay and Wassily Leontief, has significantly increased in recent decades.[10] Within this literature, our work relates most closely to Baqaee and Farhi (2020). Our aggregation theorem, which expresses the aggregate profit share in terms of a sufficient statistic for production networks and aggregate measures of market power, can be seen as a generalization of the production side of that framework. It is possible to show that in their framework, the profit share can be expressed as the input-output multiplier (i.e., the ratio of sales to GDP) times one minus the inverse of the harmonic sales-weighted markup. Thus, our theorem nests the production side of Baqaee and Farhi (2020)'s economy when there is no monopsony, no fixed costs, and returns to scale are constant for each producer.

The Functional Distribution of Income. Income shares are important summary statistics for understanding many macroeconomic phenomena, including the global decline in the labor share, the stability of the Kaldor facts, and economic inequality. One such statistic that has historically caught the eye of both economists and policymakers is the profit share.

Studying the profit share is important for several reasons. First, understanding the evolution of the profit share may shed light on the global decline in the labor share documented by Karabarbounis and Neiman (2014), the fall in business dynamism reported by Decker et al. (2014), and the TFP slowdown noted by Gordon (2012).[11,12,13] More generally, income shares (i.e., labor-, capital- and profit shares) are important for assessing the extent of wealth and income inequality (Atkinson et al., 2011; Piketty and Saez, 2003), and for calibrating economic models, whether it is in the areas of growth, development, or monetary policy (e.g., see Boppart et al., 2023; Gali and Monacelli, 2005). Finally, the profit share is informative for economic policies on corporate taxation, redistribution, and antitrust.

Our theoretical results unify micro and macro approaches for computing income shares. Consistent aggregation of micro data permits obtaining income shares that will be identical to those obtained using macro data. That is, it does not matter whether one computes the aggregate profit share either as a residual of one minus the sum of labor and capital shares or by aggregating individual profit rates (defined as profits over sales) using Domar weights as long as one uses representative micro-level data. If the

---

[10]A list of non-exhaustive papers in this literature is: Long and Plosser (1983), Acemoglu et al. (2016, 2012, 2015), Bigio and La'o (2020), Ghassibe (2021), and the work of Baqaee and Farhi (2019a,b, 2020).

[11]Income shares—that is, labor-, capital- and profit shares—add up to unity. Thus, if the labor share has declined, that necessarily means that either the capital or the profit share must have increased.

[12]In a world where technological change results from innovative activity, economic profits provide incentives for inventors and entrepreneurs to pursue new ideas and business projects.

[13]Economic profits result from monopolistic- and monopsonistic distortions, both of which may induce misallocation of production factors and thus affect aggregate productivity.

aggregate user cost of capital used to calculate the capital share via the macroeconomic approach is consistent with the underlying heterogeneity in producers' user costs of capital, then income shares would be identical regardless of the procedure—micro or macro—employed to compute these shares.

Thus, an advantage of our theoretical results is that they can be used to make a direct connection between labor, capital, and profit shares using micro-level data. Two examples where our results can help include studying the role of the profit share in the decline of the labor share (Elsby, Hobijn and Şahin, 2013; Karabarbounis and Neiman, 2014) and the stability of the Kaldor facts (Eggertsson, Robbins and Wold, 2021). In Appendix E.7, we calculate the labor share from US National Accounts and use our estimates of the micro-aggregated profit share to back out the capital share. In line with the famous Kaldor facts, we conclude that factor shares are relatively stable over time. In 2020, we estimate profit-, capital- and labor shares to be around 18%, 27%, and 55% of GDP, respectively.

Layout of the paper. The rest of the paper is organized as follows. Section 2 presents our main theoretical results and connects them to the existing literature. Section 3 discusses data sources and the methodology. Section 4 offers our main empirical results. Section 5 concludes. Appendices A–E provide omitted proofs, summary statistics for our empirical application, additional results, and robustness exercises.

2   Aggregation Results on Market Power

In this section, we provide two aggregation results on market power. First, we show that the aggregate profit share can be constructed by weighting individual profit rates, defined as profits over sales, using Domar weights (i.e., sales divided by GDP.) Second, we show that under additional assumptions, a producer's profit rate can be written as a function of its markup, markdowns, and returns to scale. Third, we express the aggregate profit share in terms of the aggregate markup, an aggregate monopsony term, aggregate returns to scale, and a sufficient statistic for production networks. Along the way, we provide intuition for these results and connect them to the existing literature.

2.1   Aggregate Profit Shares: Two Approaches

In principle, one can compute aggregate profit shares using two different approaches: the macroeconomic- and the microeconomic approach. We use *macro approach* to refer to the procedure employed to compute the profit share using aggregate data from

National Accounts. And we use *micro approach* to refer to the procedure that constructs the profit share from micro-level data.

The Macro Approach. The macroeconomic approach starts by recognizing that the aggregate value added of an economy is equal to the sum of compensation to factors of production and a residual term, here labeled "wedge income," that captures all other sources of income. Under the usual assumption that there are only two factors of production—capital and labor—we can write

$$\text{GDP} = WL + RK + \text{wedge income,} \tag{1}$$

where GDP is aggregate value added, WL is labor compensation, and RK are (gross-of-depreciation) capital rents. In theory, the term "wedge income" does not only capture profit income but also any income that accrues to other wedges such as taxes on production and sales, tariffs, etc.[14] In the absence of wedges other than monopolistic or monopsonistic wedges, we can write

$$\text{GDP} = WL + RK + \text{Profits,} \tag{2}$$

and then compute the aggregate profit share as

$$\Lambda_{\Pi}^{\text{Macro}} = 1 - \Lambda_L - \Lambda_K, \tag{3}$$

where Λ_{Π} is the profit share, Λ_L is the labor share, and Λ_K is the capital share.

Existing measures of the aggregate profit share have been constructed using this procedure. That is, one first obtains value added, labor compensation, and the stock of capital from National Accounts. Next, one imputes or estimates the aggregate user cost of capital to compute the capital share. Finally, equipped with labor- and capital shares, one infers the profit share using equation (3).

The Micro Approach. In this paper, we develop the microeconomic approach, which consists of aggregating individual producers' profit rates according to

$$\Lambda_{\Pi}^{\text{Micro}} = \sum_{i \in \mathcal{I}} \omega_i s_{\pi_i}, \tag{4}$$

where ω_i is producer i's weight, and s_{π_i} is its profit rate.

---

[14] Karabarbounis and Neiman (2019) refer to the term that we call wedge income as "factorless income." Since, in practice, such a term also reflects measurement error caused by mismeasuring capital or its user cost, as well as by incorrectly imputing payments to labor, this term could contain factor income, and thus we find "wedge income" to be more appropriate terminology.

Next, we show that to construct the profit share from micro-level data one needs to use Domar weights—that is, producer sales over aggregate value added.

Lemma 1 (**The Micro–Aggregated Profit Share**). *If profit rates, defined as profits over sales, are aggregated using Domar weights (i.e., sales over GDP,) then micro- and macro approaches both yield the aggregate profit share.*

Proof. Let Π denote aggregate profits. For each producer $i \in \mathcal{I}$, π_i denotes profits, $p_i y_i$ are sales, and $s_{\pi_i} := \pi_i/(p_i y_i)$ is i's profit rate. Then, we have

$$\Lambda_\Pi^{\text{Macro}} = \frac{\Pi}{\text{GDP}} = \frac{\sum_{i \in \mathcal{I}} \pi_i}{\text{GDP}} = \frac{\sum_i s_{\pi_i} p_i y_i}{\text{GDP}} = \sum_{i \in \mathcal{I}} \frac{p_i y_i}{\text{GDP}} \times s_{\pi_i} \equiv \Lambda_\Pi^{\text{Micro, DW}}.$$

$\square$

The intuition for using Domar weights to arrive at the aggregate profit share is that Domar weights, defined as producer sales over GDP, are "summary statistics" for input-output linkages that capture the influence of each producer on aggregate value added. If there are no production networks, Domar weights reduce to sales shares, sum up to one, and reflect the influence of producers on aggregate value added through the sale of final goods. With production networks, Domar weights differ from sales shares, their sum exceeds unity, and reflect both the direct- and the indirect influence of producers on aggregate value. Producers have a direct influence on aggregate value added through the sale of final consumption goods, and an indirect influence through the sale of intermediate inputs to other producers.

Why do production networks matter in the construction of the aggregate profit share? In the presence of intermediate goods, because of double marginalization, the income generated by downstream producers (i.e., those directly selling to consumers) can be split into three terms: own profits, profits to others, and pure costs. Profits to others are accounted for in the costs of downstream producers that buy materials from other producers which may also charge positive markups and thus earn profits. In order to arrive at the aggregate profit share from micro-level data, one must take these input-output linkages into account, which is precisely what Domar weights do.

Three important remarks are in order. First, establishing Lemma 1 requires no modeling assumptions, only fundamental accounting principles. If economic profit rates were observable at the individual level, the profit share could be easily constructed from micro data. Second, alternative aggregation schemes are possible. If individual profit rates are alternatively defined as profits over value added rather than sales, then the aggregate profit share can be constructed from micro-level data by weighting profit rates using value-added weights (See Lemma 2 in Appendix A.) Finally, although we

have defined the profit share economy-wide, Lemma 1 applies at any desired level of aggregation as long as we make the necessary adjustments in terms of value added.

2.2 Profit Rates in Terms of Monopoly and Monopsony

While Lemma 1 is a useful aggregation result, it may not be so practical when confronted with the data. This is because profit rates are typically not observed, and, even when they are, they do not measure economic- but accounting profits.[15] To overcome this, we show that the *economic* profit rate of any given producer can be written in terms of its markup, markdowns, and returns to scale.

Proposition 1 (Profit Rates, Monopoly, and Monopsony). *Under the assumptions of cost-minimizing behavior, the existence of a continuously differentiable and quasiconcave production function, and monopsony power in factor markets, a producer's profit rate, defined as profits over sales, can be written as*

$$s_\pi = 1 - \frac{RS}{\mu} = 1 - \frac{SE^{adj}}{\mu} + \frac{\mathcal{M}}{\mu}, \tag{5}$$

where $RS = SE^{adj} - \mathcal{M}$ are the returns to scale, SE^{adj} is the scale elasticity of the production function adjusted for fixed costs, $\mathcal{M}$ is a monopsony term capturing market power in factor markets, and μ is the markup of price over marginal cost.

The scale elasticity adjusted for fixed costs is defined as

$$SE^{adj} := SE \times \left(\frac{TC}{TC - FC} \right), \tag{6}$$

where SE is the scale elasticity, TC are total costs, and FC are fixed costs.

The monopsony term is given by

$$\mathcal{M} := \left(\frac{TC}{TC - FC} \right) \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j), \tag{7}$$

where $\theta_j \equiv \partial F / \partial x_j \times x_j / y$ is the elasticity of output with respect to input j, and ν_j is the markdown on factor $j \in \mathcal{F}$, where the markdown is defined as the ratio of input j's rental rate to its marginal revenue product; that is, $\nu_j := w_j(x_j)/MRP_j$.

Proof. See Appendix A.2. □

---

[15]Economic profits reflect the opportunity costs of using factors of production, whereas accounting profits do not. Businesses may have incentives to overstate capital depreciation when reporting accounting profits in balance-sheet data. The classic reference shedding light on the distinction between economic- and accounting profits is Fisher and McGowan (1983).

Remarks. Under minimal assumptions on producer behavior and technology—namely, cost minimization and a continuously differentiable and quasiconcave production technology—equation (5) states that a producer's profit rate can be written in terms of its markup μ and returns to scale RS. Alternatively, the profit rate can be written in terms of the markup, μ, the scale elasticity of the production technology adjusted for fixed costs, SE^{adj}, and a monopsony term $\mathcal{M}$ that depends on markdowns ν_j and captures market power in factor markets.

The economic intuition for equation (5) is the following. Profits originate from two sources: monopoly and monopsony. The term SE^{adj}/μ captures how monopoly power translates into profits. A producer that exerts monopoly power makes profits when it charges a markup above the scale elasticity adjusted for fixed costs; that is, if $\mu > SE^{adj}$. If the markup were to equal the scale elasticity adjusted for fixed costs and there were no markdowns, a producer would use his market power to cover fixed costs and compensate for scale economies. The term $\mathcal{M}/\mu$, captures how monopsony power in factor markets translates into profits. When a producer has market power in a factor market, the rental rate of that factor is below its marginal revenue product so that the producer earns monopsony rents from using that factor. Monopsony power translates into profits if monopsonistic rents exceed fixed costs.[16]

Equation (5) is a generalization of other expressions of profit rates in the literature. Basu (2019) and Syverson (2019) both show that the profit rate (defined as profits over sales) can be written as

$$s_\pi = 1 - \frac{\text{RS}}{\mu}. \tag{8}$$

Although equations (8) and (5) look identical when expressed in terms of returns to scale, our expression is more general. This is because of two reasons. First, we explicitly allow for fixed costs. Second, we allow for monopsony power in factor markets. Both of these generalizations affect the profit rate through the returns to scale. When there are no explicit fixed costs (i.e., FC = 0) and no monopsony power (i.e., $\nu_j = 1, \forall j \in \mathcal{F}$,) the returns to scale equal the scale elasticity of the production technology and our expression and that of Basu and Syverson are identical.[17]

When there are explicit fixed costs (i.e., FC > 0,) returns to scale are no longer equal to the scale elasticity and must be adjusted to incorporate fixed costs. That explains why in the absence of monopsony power, our formula equates returns to scale to the

---

[16]Notice that the term $\mathcal{M}$ features a fixed-cost adjustment, $TC/(TC-FC)$, and input elasticities for each factor j. For simplicity, consider the case in which there are no fixed costs and focus on a particular factor j. If there is no monopoly power and ν_j is lower than unity, monopsony power translates into profits if $\theta_j(1-\nu_j) \neq 1$. With fixed costs, market power in input j translates into profits if $\frac{TC}{TC-FC}\theta_j(1-\nu_j) \neq 1$.

[17]Imposing constant returns to scale, equations (5) and (8) both yield the familiar profit rate $s_\pi = 1-\mu^{-1}$.

scale elasticity of the production function adjusted for fixed costs. Further allowing for monopsony power in factor markets (i.e., $\nu_j \in (0,1)$ for some $j \in \mathcal{F}$) requires making one additional adjustment to the returns to scale. With both explicit fixed costs and monopsony power, returns to scale depend on the scale elasticity, explicit fixed costs, and a monopsony term. Returns to scale, defined as the ratio of average- to marginal cost, depend on the monopsony term because a producer can affect its marginal cost of production by exerting monopsony power in factor markets. This explains why, according to our formula, returns to scale equal the scale elasticity adjusted for fixed costs minus the monopsony term; that is, $\mathrm{RS} = \mathrm{SE}^{\mathrm{adj}} - \mathcal{M}$.

An important clarification is that the scale elasticity of the production function adjusted for fixed costs captures all fixed costs.[18] In the data, it is not always possible to separate inputs that are directly used in the production of output from those that are not. If available data combines expenditures on production factors without separating those expenditures into variable and fixed costs, the estimated scale elasticity will capture all fixed costs. If, however, a direct measure of fixed costs is observable in the data, it can be beneficial to estimate the scale elasticity only using variable inputs. This is because the estimated scale elasticity can be adjusted to incorporate fixed costs not included in the estimation procedure, and the scale elasticity can rarely be estimated at the firm level. By using a fixed-cost adjustment factor, we can exploit firm-level heterogeneity in scale elasticities even within particular industries. We use this theoretical insight in our empirical exercise and estimate scale elasticities at the firm level. We do so by constructing a fixed-cost adjustment factor for each firm with observable R&D expenditures, and by estimating the scale elasticity at the industry level.

Remarks. In Appendix A, section A.4, we show that Lemma 1 combined with equation (5) can be used to identify the aggregate user cost of capital. Finally, we note that expressing profit rates as profits divided by sales rather than value added is standard practice since value added is not observed in many datasets, but sales are.

2.3 Linking the Aggregate Profit Share to Market-Power Indicators

Our next result states that the profit share can be expressed in terms of a sufficient statistic for production networks that captures double marginalization in the economy and two indicators of aggregate market power—the aggregate markup and an aggregate monopsony term that captures market power in factor markets.

---

[18]See Appendix B for a discussion on the relationship between returns to scale and fixed costs related to production factors. A related point is that changes in returns to scale can in principle capture both increasing fixed costs and technological change. Throughout the paper, we interpret rising returns to scale as increases in fixed costs for the ease of exposition.

Theorem 1. *With cost-minimizing producers, continuously differentiable and quasiconcave production functions, fixed costs, and market power in factor- and output markets, the profit share can be expressed as*

$$\Lambda_\Pi = \chi \left(1 - \mathbb{E}_\omega \left[\frac{SE^{adj}}{\mu} \right] + \mathbb{E}_\omega \left[\frac{\mathcal{M}}{\mu} \right] \right) \tag{9}$$

$$= \chi \left(1 - \frac{\overline{SE}^{adj}}{\overline{\mu}_{hsw}} + \frac{\overline{\mathcal{M}}}{\overline{\mu}_{hsw}} - Cov_\omega \left[SE^{adj}, \frac{1}{\mu} \right] + Cov_\omega \left[\mathcal{M}, \frac{1}{\mu} \right] \right), \tag{10}$$

where $\chi = \sum_{k \in \mathcal{I}} \frac{p_k y_k}{GDP}$ is the input-output multiplier, $\mathbb{E}_\omega [\cdot]$ denotes a sales-weighted average where ω indexes sales weights, SE^{adj} is the scale elasticity adjusted for fixed costs given by (6), μ denotes the markup, and $\mathcal{M}$ is a monopsony term given by (7).

The term $\overline{X}$ is the sales-weighted average of X, $\overline{\mu}_{hsw}$ is the harmonic sales-weighted markup, and $Cov_\omega(X, Y)$ is the sales-weighted covariance of variables X and Y.

Proof. See Appendix A.5. □

Theorem 1 links several indicators of aggregate market power to the profit share. Since these market-power indicators are obtained from micro data, Theorem 1 can be used to assess the macroeconomic implications of micro-level estimates of markups, markdowns, and returns to scale in terms of the aggregate profit share. This theorem also permits decomposing the aggregate profit share into different sources of market power and, hence, assessing to what extent aggregate profits originate from monopoly and monopsony. An important clarification is that while we have stated our theorem in terms of the economy-wide profit share, Theorem 1 can be adapted with ease to study profit shares at any desired level of aggregation, such as industry-level profit shares. By quantifying the sources of market power—monopoly *vis-à-vis* monopsony—for particular industries, empirical applications of Theorem 1 can inform policy discussions on antitrust and help draft industry-specific regulations. We agree with Berry, Gaynor and Morton (2019) in that detailed industry analysis is needed, and believe that our theorem is useful in that respect as it bridges macro and industrial-organization approaches.

The second line in Theorem 1, equation (10), states that the profit share can be computed using the input-output multiplier (i.e., the ratio of total sales to aggregate value added,) the aggregate markup, an aggregate monopsony term, and covariance terms. Since the aggregate profit share can be expressed as a function of the harmonic sales-weighted markup, this theorem provides a natural argument for using the harmonic sales-weighted markup as our measure of aggregate markup. We acknowledge that we are not the first to point to this measure as a natural choice for the aggregate markup. Baqaee and Farhi (2020) and Edmond, Midrigan and Xu (2023), among others, have

noted this earlier. However, we show that the harmonic sales-weighted markup continues to be an appropriate choice for the aggregate markup in more general environments. While Baqaee and Farhi allow for arbitrary production networks, as we do, neither they nor Edmond, Midrigan and Xu allow for increasing returns to scale, fixed costs, or monopsony. In the same manner that we established a measure of aggregate markup, we note that sales-weighted average returns to scale constitute a natural measure of aggregate returns to scale. Similarly, one could refer to the sales-weighted monopsony term, $\overline{\mathcal{M}}$, as the aggregate monopsony term. Finally, we note that the aggregate monopsony term captures market power on several factor markets through factor markdowns. For those interested in computing aggregate markdowns, our theorem suggests that the aggregate markdown of a particular factor is the sales-weighted markdown.

The following corollaries are special cases to Theorem 1 and are useful for inferring profit shares and markups, as well as for assessing the validity of a researcher's own micro estimates or those of existing studies, as we demonstrate in the sequel.

Corollary 1 (No Monopsony). *Assuming price-taking behavior in input markets (i.e., $v_{ij} = 1$ for all i, j), Theorem 1 reduces to*

$$\Lambda_\Pi = \chi \left(1 - \frac{\overline{RS}}{\overline{\mu}_{hsw}} - Cov_\omega \left[RS, \frac{1}{\mu} \right] \right), \tag{11}$$

where $RS = SE^{adj}$.

Corollary 2 (No Monopsony and No Fixed Costs). *Assuming price-taking behavior in input markets (i.e., $v_{ij} = 1$ for all i, j) and no fixed costs (i.e., $FC_i = 0$ for all i,) Theorem 1 reduces to*

$$\Lambda_\Pi = \chi \left(1 - \frac{\overline{RS}}{\overline{\mu}_{hsw}} - Cov_\omega \left[RS, \frac{1}{\mu} \right] \right), \tag{12}$$

where $RS = SE$.

Corollary 3 (No Monopsony, No Fixed Costs, and CRS). *Assuming price-taking behavior in input markets (i.e., $v_{ij} = 1$ for all i, j,) no fixed costs (i.e., $FC_i = 0$ for all i,) and constant returns to scale (i.e., $RS = 1$ for all i,) Theorem 1 reduces to*

$$\Lambda_\Pi = \chi \left(1 - \frac{1}{\overline{\mu}_{hsw}} \right). \tag{13}$$

Equation (13) in Corollary 3 states that under the assumptions of cost minimization, price-taking behavior in input markets, no fixed costs, and continuously-differentiable production functions of homogeneity degree one, the aggregate profit share can be computed as the input-output multiplier—the ratio of sales to GDP—times one minus

the inverse of the harmonic sales-weighted markup. This expression of the aggregate profit share obtains in the general framework of Baqaee and Farhi (2020).[19] Such an expression can be used for inference provided one is willing to make strong assumptions on returns to scale. Equipped with markup estimates for all producers in the economy, one can compute the harmonic sales-weighted average markup, read the input-output multiplier from National Accounts, and then infer the aggregate profit share. Similarly, provided knowledge on the aggregate profit share, one can rearrange equation (13) and back out the aggregate markup as

$$\overline{\mu}_{hsw} = \left(1 - \frac{\Lambda_\Pi}{\chi}\right)^{-1}. \tag{14}$$

Importantly, notice that taking input-output networks into account is crucial for inferring the aggregate profit share from micro data and for backing out the aggregate markup from macro data. Neglecting production networks would lead to underestimating the profit share by a factor of χ, which is around two in the United States. Similarly, ignoring production networks would lead to overstating the aggregate markup by a factor of

$$\frac{(\chi - 1)\Lambda_\Pi}{(1 - \Lambda_\Pi)(\chi - \Lambda_\Pi)}.$$

This bias can be substantial. To see this, suppose that $\chi = 2$ and $\Lambda_\Pi = 0.2$. Then, ignoring production networks would lead us to conclude that the aggregate markup is 113% larger than it actually is. Finally, notice that if there is no vertical integration in the economy, $\chi = 1$, and neither inference of the aggregate markup nor that of the profit share are biased.

Equation (12) in Corollary 2 presents a more interesting case by allowing for decreasing and increasing returns to scale in production. In that case, the aggregate profit share can be constructed as the input-output multiplier times one minus sales-weighted returns to scale divided by the harmonic sales-weighted markup and a covariance term.

Finally, Corollary 1 relaxes the assumption of no fixed costs. Under a no-monopsony assumption, the aggregate profit share can be constructed as the input-output multiplier times one minus sales-weighted returns to scale divided by the harmonic sales-weighted markup and a covariance term. Returns to scale equal the scale elasticity times a fixed-cost adjustment factor. In Appendix C, we use Corollary 1 to elucidate the back-and-forth discussion between Basu (2019) and De Loecker et al. (2020) on how to map micro-levels estimates of market power to the aggregate profit share.

---

[19]This profit share would be recorded in the first row and f^*-th column of the Leoentief inverse matrix augmented with a fictitious factor called "profits" and denoted f^*.

3 Data and Methodology

In this section, we review our data sources and the methodology employed for estimating markups and inferring the aggregate profit share.

3.1 Compustat

Our main data comes from Compustat North America Fundamentals Annual, which we retrieve through the Wharton Research Data Services (WRDS.) Compustat provides balance-sheet data for US-incorporated publicly-traded firms.

Sample selection. We download annual data for all available firms between 1966 and 2020. We deflate all financial variables using appropriate price deflators from US NIPA, and classify firms according to two-digit NAICS industries to estimate input elasticities. The NAICS industry classification is adapted to conform with the Bureau of Economic Analysis (BEA) when classifying industries.[20] We drop observations from our sample when firms have missing fiscal years, no NAICS code, or multiple NAICS codes. We also drop observations that register negative or missing values for sales, costs of goods sold, selling, general and administrative expenses, R&D, or investment since all these variables are essential for estimating markups. Following De Loecker et al. (2020), we drop observations with the ratio of sales to COGS in the 1st and 99th percentiles. We construct R&D following Peters and Taylor (2017), and treat it as a fixed cost contributing to the future stock of intangible capital. We measure a firm's total capital stock as the sum of physical and intangible capital.[21] We use the K_INT variable from Peters and Taylor (2017) to measure intangible capital. K_INT is constructed by applying the perpetual inventory method to Compustat firms. We measure physical capital as Property, Plant, and Equipment (PPEGT.) We provide basic summary statistics in Appendix D, Table 3.

3.2 NIPA Tables and BEA

We retrieve data on sectoral sales, value added, and GDP from the US National Income and Product Accounts (NIPA.) We use these data to compute Domar weights and input-output multipliers for the different industries. In essence, this procedure assumes the representativeness of Compustat data for the overall population of US firms. While this is a stark assumption, it allows us to develop a proxy for the aggregate profit share from micro-level data.

---

[20]This essentially means that 2-digit NAICS industries 31-33, 44-45, and 48-49 are grouped together.

[21]Intangible capital, such as patents, trademarks, and software, has become an increasingly important input in firms' production processes over the course of our sample period (see Crouzet et al., 2022).

3.3 Production Function Estimation

One popular approach for studying market power relies on the production function approach pioneered by Hall (1988).[22] Hall's seminal contribution was to derive an expression for the markup by taking the first-order condition of a cost-minimizing producer with respect to a variable input, assuming the producer acts as a price taker in that input's market. Formally, he showed that the markup can be computed as

$$\mu = \frac{\theta_j}{\alpha_j}, \tag{15}$$

where $\theta_j \equiv (\partial y / \partial x_j)$ is the elasticity of output with respect to variable input j, and $\alpha_j \equiv p_j x_j / (py)$ is the revenue share of input j.

From equation (15), it is clear that in order to obtain an empirical counterpart of the markup, one needs data on revenue and input spending $\{py, p_j x_j\}$, as well as on a variable input's output elasticity θ_j. The main empirical challenge in markup estimation is obtaining reliable estimates of the variable input's output elasticity, $\hat{\theta}_j$.[23] In this respect, the literature has benefited from the work of a great number of economists who have developed specialized econometric methods tailored to this purpose. Some of the most influential papers in this area are Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), Loecker and Warzynski (2012), Ackerberg et al. (2015), Bond et al. (2021), and De Ridder et al. (2022).

Control Function Approach. We estimate output elasticities using a control function approach. We assume that producers generate output y using a Cobb-Douglas technology that takes as arguments one variable input ℓ and one quasi-fixed input k. The associated log regression is

$$\log y_{it} = \theta_\ell \log \ell_{it} + \theta_k \log k_{it} + \omega_{it} + \varepsilon_{it}, \tag{16}$$

where (θ_ℓ, θ_k) are output elasticities, ω_{it} is the idiosyncratic productivity, which is observed by the producer (*potentially* prior to making input choices) but not by the econometrician, and ε_{it} is an unanticipated shock to output or productivity which is observed by neither the econometrician nor the producer and may simply capture measurement error.

---

[22]There is an alternative approach to obtain markups based on demand estimation. The classic reference in this literature is Berry et al. (1995). The vast majority of papers in this literature focus on market power in a narrowly defined market (see, for example, Berto Villas-Boas, 2007; Nevo, 2001). One exception in this literature is Döpper et al. (2023), who estimate markups for more than 100 distinct product categories in the United States from 2006 to 2019 using Kilts Nielsen Retail Scanner Data.

[23]More recently, the production function approach has been extended by Brooks et al. (2021) and Yeh et al. (2022) to allow for market power in input markets. The novel insight of these studies is to show that markdowns can be estimated using a two-step procedure where markups are obtained in the first step.

The presence of time-varying firm-level productivity, ω_{it}, raises standard concerns regarding simultaneity bias. To address this source of bias, we follow Ackerberg et al. (2015) and estimate the production function (16) using a two-step procedure based on a control function for the productivity process.[24] In the first stage, we clean measurement error and the unanticipated productivity shock from output using a non-parametric regression. Then, we estimate production-function parameters in the second stage using appropriate moment conditions via GMM.

More specifically, we estimate output elasticities using Olley and Pakes (1996)'s methodology coupled with Ackerberg et al. (2015)'s correction. Our baseline estimates rely on Cobb–Douglas technologies, allow output elasticities to be time-varying, and impose common elasticities across 2-digit NAICS industries, which we operationalize by running separate regressions for each industry. In each case, in the first stage of the estimation procedure, we clean out measurement errors and unanticipated shocks by regressing the outcome variable on a third-degree polynomial of the state, free- and proxy variables. In the second stage, we recover the underlying productivity measure and use an AR(1) process for productivity to estimate output elasticities using investment as a proxy variable to instrument productivity. When we run our regression (16), we use the capital stock from the previous period since Compustat records end-of-period capital assets. Table 1 summarizes our estimation procedure.[25]

Table 1: Estimation Details in the Application of the Control–Function Approach.

Technology	Cobb–Douglas
Elasticities	Time-varying, 9-year rolling windows
Method	Olley and Pakes (1996)
Productivity process	AR(1)
Degree of polynomial	3rd
Ackerberg et al. (2015)'s correction	✓
Deflated variables	✓
Outcome, y	SALE
State, k	PPEGT + K_INT
Free, ℓ	OPEX (= COGS + SG&A)
Proxy, x	ICAPT

Table Notes. Elasticities are estimated using 9-year rolling windows. That is, for year t, we use information from years $t-4, t-3, t-2, t-1, t+1, t+2, t+3, t+4$. Hence, to estimate elasticities for the period 1970–2020, we need data starting from 1966 to 2024. Data from 1966 to 2020 is available, but data from 2024 is not. Hence, for years 2016–2020, we set the elasticities to be those estimated in 2016–the last year for which a 9-year panel regression can be run.

---

[24] Ackerberg et al. (2015) argue that the first stage of the estimation procedures proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) may suffer from a functional-dependence problem that prevents identification of the variable input's elasticity. To solve this issue, they propose inverting conditional (rather than unconditional) demand functions for the proxy variable and then estimating the variable input's elasticity (along with all other production-function parameters) in the second stage.

[25] In our baseline specification, the stock of capital is the sum of physical- and intangible capital. In Appendix E.8, we show that none of our main empirical results are driven by the inclusion of intangibles.

Our estimation procedure provides estimates of output elasticities for each industry. We remove outliers from our sample of estimated output elasticities. We consider an elasticity to be an outlier if it is more than one standard deviation away from its group mean, where a group is a 2-digit NAICS industry. We set outliers to missing values and then replace them by linearly interpolating each group's input elasticities. Lastly, we winsorize output elasticities at the 5th and 95th percentile.

Equipped with output elasticities, we compute scale elasticities at the industry level as the sum of output elasticities. We obtain firm-specific returns to scale by multiplying the industry-level scale elasticity with a firm-specific fixed-cost adjustment factor. We compute markups for each producer as $\mu_i = \theta_{j\ell}\alpha_{i\ell}^{-1}$, where $\theta_{j\ell}$ is the output elasticity of the variable input for producer i which belongs to industry j. In section 4, Table 2, we report markup and return to scale estimates at the sectoral level for 2019. The evolution of variable input elasticities, returns to scale, and markups are provided for each industry in Appendices E.3–E.5.

4 Empirical Results

In this section, we compute the micro-aggregated profit share using the theoretical results presented in Section 2. We benchmark our estimates of the micro-aggregated profit share to macro estimates and discuss their macroeconomic implications. We also provide some results at the industry level.

Figure 1 plots the time series of the components that make up the micro-aggregated profit share as given by equation (11); that is, assuming no monopsony power.[26] Panel A displays three series: the aggregate markup, aggregate returns to scale, and the aggregate scale elasticity. Panel B depicts the sales-weighted covariance of firm-level returns to scale with inverse markups.

The aggregate markup has increased from 1.10 in 1970 to 1.23 in 2020; that is, from 10% of price over marginal cost to 23%. Trends are the same as in De Loecker et al. (2020), but levels are much lower. As noted earlier, this is mostly because the harmonic sales-weighted markup is much lower than the sales-weighted markup. Another reason why our markup estimates differ from theirs is that we categorize selling, general and administrative expenditures (SG&A,) as well as the costs of goods sold (COGS,) as variable inputs, in line with Traina (2018), while De Loecker, Eeckhout and Unger treat

---

[26]As noted earlier, future versions of the paper will include empirical results with monopsony power. The reason why we cannot offer such results with Compustat is that we have no data on materials and, hence, we cannot estimate markdowns. However, we do not expect very different estimates for the profit share since our current markup estimates are a composite of monopoly and monopsony power. Nevertheless, we recognize that accounting for monopsony can potentially affect our markup estimates.

Figure 1: Aggregate Markup, Aggregate Returns to Scale, and its Covariance.
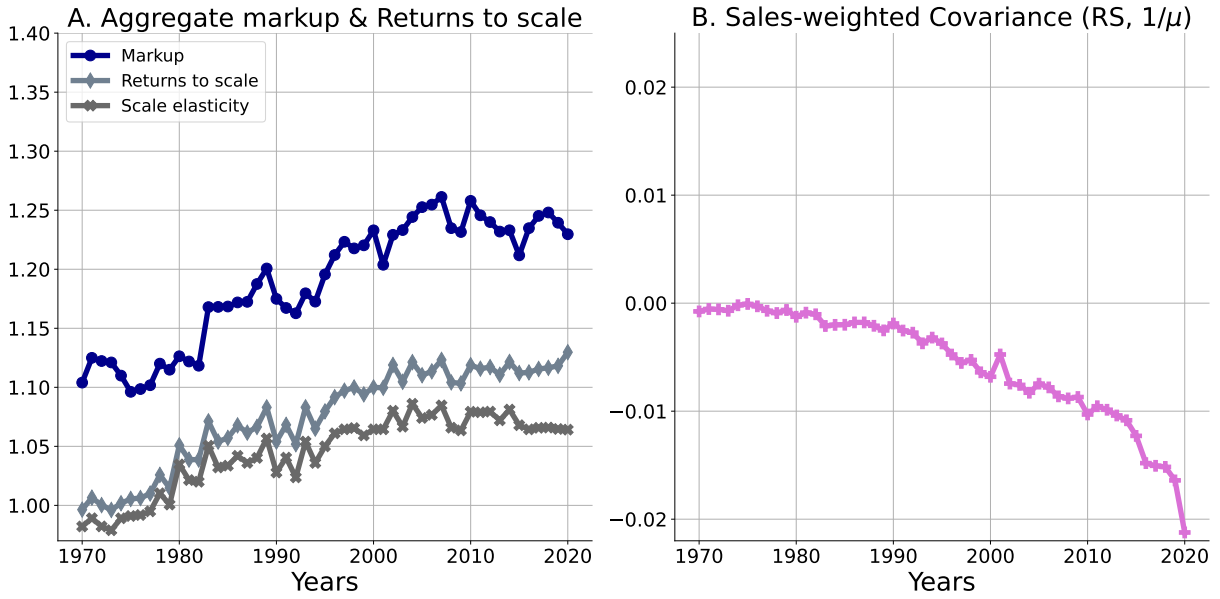

Figure Notes. Firm-level data comes from US Compustat, 1970–2020.

SG&A as fixed costs. We also define the stock of capital to be a composite of physical- and intangible capital while they only consider physical capital.[27] Looking at returns to scale, we can see that returns to scale have increased from 1.00 in 1970 to 1.13 in 2020. Half of this increase is explained by a rising scale elasticity, which has gone up from 0.98 in 1970 to 1.06 in 2020, and the remaining part by the fixed-cost adjustment factor, $(TC - FC)/FC$, which has risen from 1.02 in 1970 to 1.06 in 2020. The divergence between the two series, returns to scale and the scale elasticity, thus reflects the increase in R&D spending, which we treat as an explicit fixed cost.

The covariance between returns to scale and inverse markups is slightly negative at the beginning of the sample and becomes more negative as time passes. We find this pattern to be reasonable. A negative covariance between inverse markups and returns to scale means that industries with higher fixed costs tend to charge higher markups.

Equipped with the input-output multiplier—that is, the ratio of sales to GDP—and the measures provided in the previous figure, we compute the micro-aggregated profit share.[28] The results are depicted in Figure 2, which also shows the profit share that would obtain under zero covariance, and zero covariance and constant returns to scale.

---

[27]In Appendix E.1 we show that making the same data choices as De Loecker et al., we replicate their results. We also compare our markup series to theirs, and explain the differences between them.

[28]In Appendix E.2 we show that the input-output multiplier in the US—the ratio of sales to GDP—is mildly procyclical and roughly stable at around 1.8.

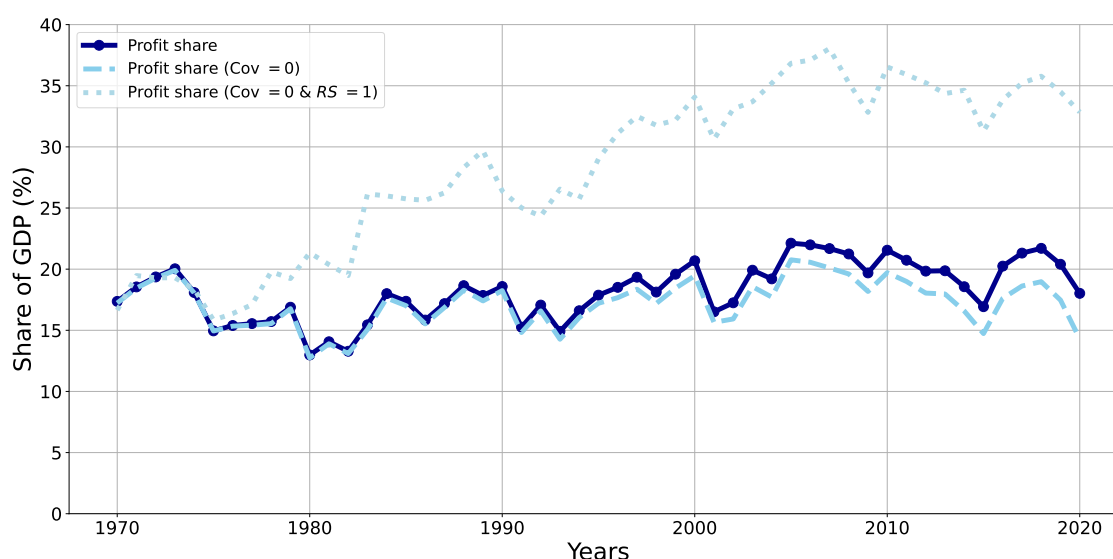**Figure 2:** The Micro–Aggregated Profit Share in the United States, 1970–2020.

Figure Notes. Firm-level data comes from US Compustat, and data on value-added and GDP comes at the aggregate level from the BEA. The micro-aggregated profit share is expressed in terms of profits divided by value added.

Figure 2 indicates that the profit share in the United States has been constant at around 18% of GDP over the past fifty years.[29] We note that our series likely provides an upper bound for the profit share for several reasons. First, Compustat data covers only publicly-listed firms, which are not representative of the US population of firms and likely have higher profit rates than smaller firms. Second, if intangibles or fixed costs are underestimated, the profit share will be overestimated. We have good reasons to believe that fixed costs, as well as organizational- and intangible capital, are underestimated as we will argue shortly on. The series in Figure 2 which impose zero covariance and constant returns to scale stress the importance of accounting for both of these terms, especially non-constant returns to scale. If we had estimated the profit share imposing constant returns to scale, we would have concluded that the profit share in 2020 was twice as large as we report.

Decomposing the Profit Share in Several Market-Power Forces. An advantage of using the microeconomic approach (instead of the macroeconomic one) to compute the profit share is that it allows us to understand the determinants of the profit share. More specifically, the micro-aggregated profit share can be decomposed into fractions

---

[29]Because of double marginalization, a profit share of 18% implies that the average firm in the United States has a profit rate of about 10%.

of value added pertaining to monopoly rents, fixed costs, and non-linearities. That is,

$$\Lambda_\Pi = \chi \left(1 - \frac{\overline{RS}}{\overline{\mu}_{hsw}} - \text{Cov}_\omega \left[RS, \frac{1}{\mu} \right] \right) \tag{11}$$

$$= \underbrace{\chi \left(1 - \frac{1}{\overline{\mu}_{hsw}} \right)}_{\text{Monopoly rents}} + \underbrace{\chi \left(1 - \overline{RS} \right)}_{\substack{\text{Fixed costs and} \\ \text{changing technology}}} + \underbrace{\chi \left\{ \left(\frac{1}{\overline{\mu}_{hsw}} - 1 \right) \left(1 - \overline{RS} \right) - \text{Cov}_\omega \left[RS, \frac{1}{\mu} \right] \right\}}_{\text{Non-linearities}} \tag{17}$$

In Figure 3, we offer the results of this decomposition. The term "monopoly rents" has increased from 17% to 33% of aggregate value added because producers have increased their markups over time. The term "fixed costs and changing technology" has decreased from 0% to −23% of value added because of increasing fixed costs and a rising scale elasticity. And the term "non-linearities" has increased from 0% to 8% because the aggregate markup and aggregate returns to scale are both positive and increasing over time, and also because the covariance term is increasingly negative. As pointed out earlier, the profit share has been roughly constant at around 18% of GDP, with movements around the trend reflecting cyclical variation. Thus, although market power increased, profitability did not because the increase in monopoly rents was entirely offset by rising fixed costs and changing technology.

Figure 3: US Profit Share Decomposition into Sources of Market-Power, 1970–2020.
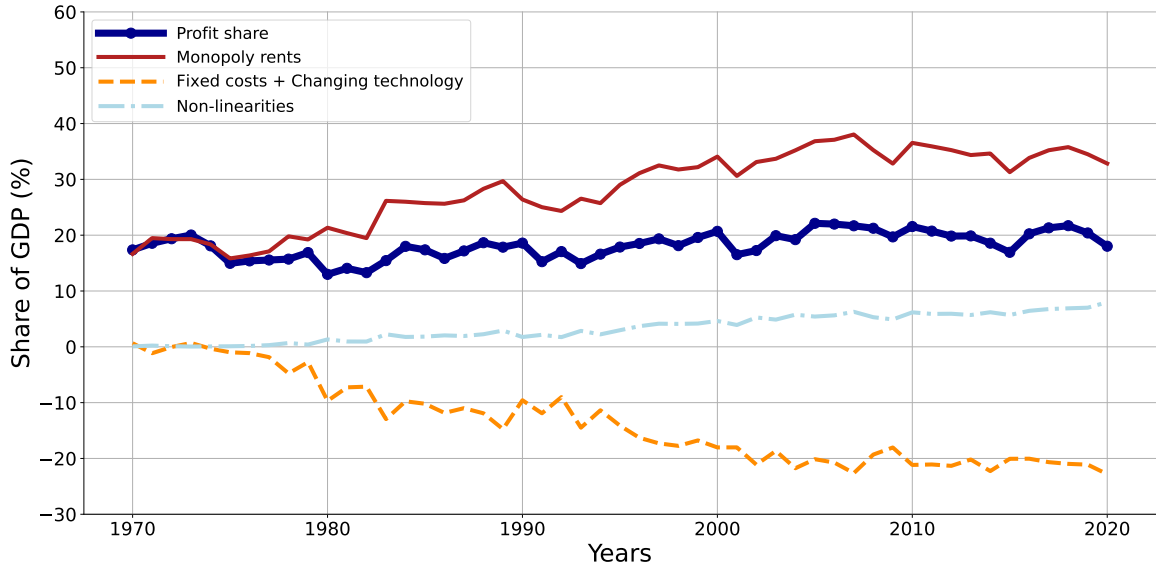
Benchmarking our results. Figure 4 benchmarks our estimates of the profit share to the profit share of Karabarbounis and Neiman (2019) using the macro approach.[30]

---

[30]We take the profit share from "Case Π" in their paper. To be clear, they refer to this series as the "factorless income" share.

Figure 4: The Profit Share in the United States: Micro vs. Macro Estimates.
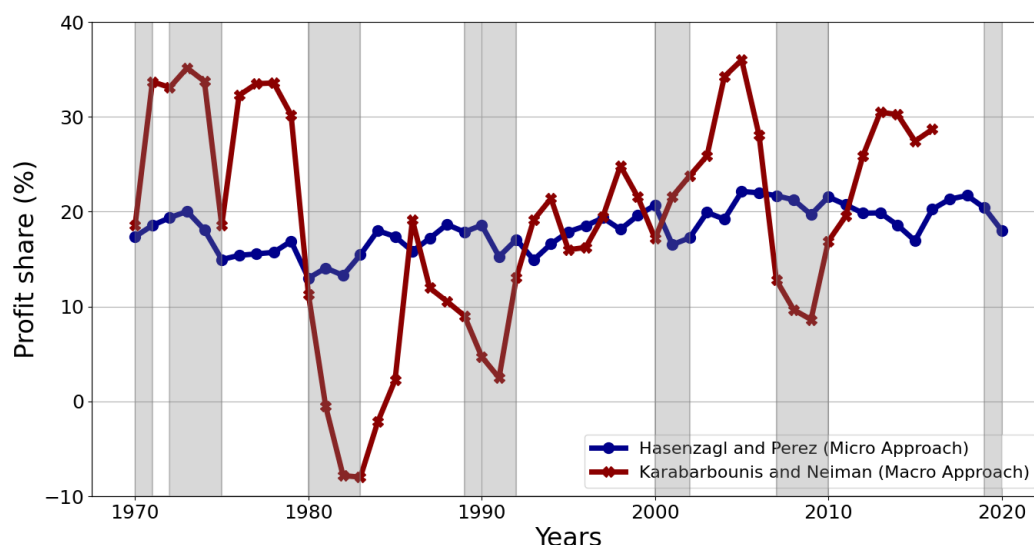

Figure Notes. The micro-aggregated profit share is computed using firm-level data from Compustat, assuming representativeness of these data, and using sales and GDP from the BEA. The Karabarbounis and Neiman's series is a non-smoothed version of the one reported in their paper and has been constructed using their replication files. They construct the profit share using aggregate data from National Accounts and imputing an aggregate user-cost of capital.

Our series of the profit share is a better measure of pure profits than the factorless-income series of Karabarbounis and Neiman. First, their series is extremely volatile, and second, it implies unreasonably low capital shares. Their estimates of factorless-income in the early 1970s do, for example, suggest a capital share of 0%, which is not realistic. Our series is more stable and implies more plausible capital shares. In line with Karabarbounis and Neiman, and consistent with our countercyclical aggregate markup estimates, we find procyclical profit shares. Although our profit share series also fluctuates because of cyclical variation, we find these fluctuations to be within reasonable bounds. Finally, we follow Karabarbounis and Neiman and compute the correlation between the profit share and the aggregate user cost of capital. With their series, we find a large negative correlation of 0.83; with our series, the correlation is also negative, but about a third in magnitude. As they note, these negative correlations suggest that both profit share series capture some form of unmeasured capital.

We can also compare our results to those of Barkai (2020), who also estimates capital- and profit shares for the United States using data from National Accounts. Barkai estimates that the profit share has grown from –5% in 1985 to 8% in 2014. We find his profit share estimates substantially less reasonable than ours. Our profit share estimates in 2014 are consistent with a labor share of 53%, as estimated by Karabarbounis and Neiman (2014), and a capital share of 29%, which aligns with Barkai's capital share estimate. This means that the implied labor share in Barkai (2020) differs by almost ten percentage points from that of Karabarbounis and Neiman. In Appendix E.7, we discuss the implications of our profit share estimates for capital income shares.

Assessing De Loecker et al. (2020)'s Estimates. De Loecker et al.'s estimates of markups and profitability have generated substantial debate in academic and policy forums. Scholars have scrutinized their methodology and their estimates. Here, we discuss the most relevant issues in calculating aggregate markups and the implications of their estimates in terms of profitability.

Weighting firm-level markups. An important choice is how to aggregate markups. Karabarbounis and Neiman (2019) noted that aggregating markups using cost (rather than sales) weights dramatically changes the magnitude of the markup series. This goes back to the discussion on which notion of aggregate markup is appropriate to use. In this paper, we provide an argument for using the harmonic sales-weighted markup based on our aggregation theorems. As pointed out by Edmond, Midrigan and Xu (2023), the harmonic sales-weighted markup is equivalent to a cost-weighted markup when cost elasticities are homogeneous across producers, in which case the point raised by Karabarbounis and Neiman remains valid.

Measuring variable costs. Traina (2018) argues that De Loecker et al.'s estimates substantially overstate true markups because they do not correctly account for variable costs.[31] Failing to properly account for variable costs may affect estimates of the variable input's elasticity and thus impact markup estimates, which are sensitive to these elasticities. Using a more appropriate categorization of variable and fixed costs, Traina finds a more modest increase in markups. Following a dichotomy of fixed and variable costs similar to Traina, we find much lower levels for aggregate markups compared to De Loecker, Eeckhout and Unger. Our results point to an increase in the aggregate markup from 1.13 in 1980 to 1.23 in 2016, whereas De Loecker et al. report a hike in markups from 1.21 to 1.61.

In Figure 5, we plot aggregate markups, returns to scale, and the sales-weighted covariance of firm-level inverse markups with returns to scale using two different categorizations of variable costs, OPEX and COGS. As argued earlier, correctly categorizing variable costs is important for estimating input elasticities, markups, and returns to scale. This can be seen in panels A and B. Panel C displays the covariance term, which we use to compute micro-aggregated profit shares.

In this paper, we have dealt with both of these issues, categorizing variable costs as suggested by Traina (2018) and aggregating markups as implied by our aggregation theorem, and have shown that the aggregate markup in the US economy has increased from 1.10 in 1970 to 1.23 in 2020. Discrepancies in markups arising from different categorizations of variable inputs can be seen in Figure 5.

---

[31]De Loecker et al. (2020) measure variable costs as the costs of goods sold (COGS). Traina (2018) and others argue that variable costs do not only include costs of goods sold but also selling, general, and administrative (SG&A) expenses. The latter represent indirect inputs to production, such as sales or marketing expenditures, management expenditures, and accounting or legal expenditures.

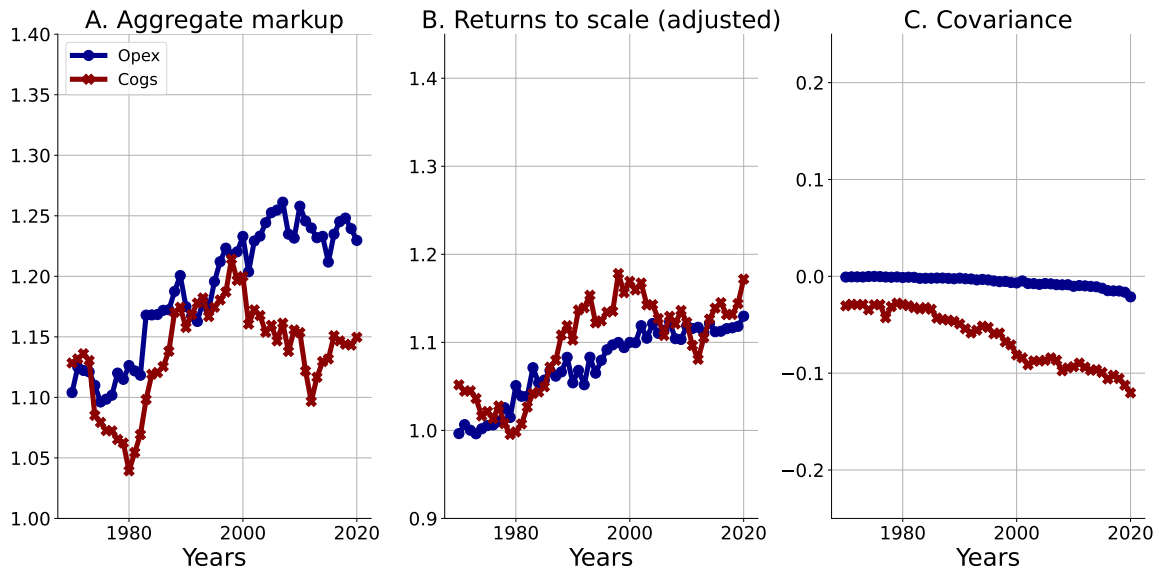**Figure 5:** Markups, Returns to Scale, and Variable Costs (OPEX vs. COGS).

Figure Notes. Aggregate markups are computed as harmonic, sales-weighted averages of individual firms' markups. The COGS series treats the costs of goods sold as the only source of variable costs. In contrast, the OPEX series considers costs of goods sold and selling, general and administrative expenses, as variable costs. Returns to scale are sales-weighted averages of firm-level returns to scale adjusted for fixed costs. The covariance is a sales-weighted covariance of firm-level inverse markups with returns to scale adjusted for fixed costs.

Finally, we touch upon the implications of De Loecker et al.'s estimates of aggregate markups on the aggregate profit share, since this has been an area of debate in the literature. Basu, Barkai, and others have argued that the micro estimates of De Loecker, Eeckhout and Unger have unreasonable implications for aggregate income shares. Specifically, Basu has argued that De Loecker, Eeckhout and Unger's markup estimates imply an unrealistically high profit share of about 70% (See Appendix C.) Consistently with what we argued in Section 2, we have shown that firm-level estimates of markups, once properly measured and aggregated, no longer have unreasonable implications when fixed costs are accounted for. It is also easy to verify that while the categorization of variable costs does matter for markup estimates, it does not matter for estimates of the aggregate profit share, simply because profit rates are the same regardless of whether inputs are categorized as variable or fixed as long as all costs are accounted for.

4.1 The Profit Share and The User Cost of Capital

A critical input for obtaining profit rates and, thus, the profit share is the user cost of capital. This is true regardless of whether the profit share is computed using the macroeconomic or the microeconomic approach. In this section, we construct the profit share using three different user costs of capital: ours (henceforth, Hasenzagl–Perez,) De Loecker et al. (2020)'s, and Karabarbounis and Neiman (2019)'s.

Theorem 1 implicitly assumes that we can compute the user cost of capital for each individual producer using its first-order condition for capital. The first-order condition of producer i at time t permits obtaining i's user cost of capital as

$$r_{it} = \frac{\theta_{jt}^k}{\mu_{it}} \times \frac{p_{it}y_{it}}{k_{it}} \tag{18}$$

where θ_j^k is the output elasticity of capital for the industry j in which producer i operates, μ is the markup, py are sales, and k is the capital stock.

We believe that allowing for heterogeneous user costs of capital is an attractive feature of our method. First, if the composition of capital is different across producers and different types of capital generate heterogeneous returns, producers' user cost of capital will be different. Second, in the presence of financial frictions, rental rates of capital may be different across producers. Third, risk premia and other wedges may differ across individual producers or sectors. We capture these sources of heterogeneity by backing out the user cost of capital for each producer using its first-order condition.

Equipped with heterogeneous user costs of capital given by equation (18), we can compute the aggregate profit share using either Lemma 1 and equation (20) or, alternatively, using equation (11).

De Loecker et al. (2020)'s user cost of capital is assumed to be equal across *all* producers in the economy and computed as

$$r_t = i_t - \pi_t + \delta, \tag{19}$$

where i is the nominal interest rate, π is the inflation rate, and δ is the depreciation rate.

De Loecker et al. (2020) compute the user cost of capital using the federal funds effective rate (FF) as a proxy for the nominal interest rate, the growth rate of the GDP implicit price deflator (GDPDEF) as a proxy for the inflation rate, and an exogenous depreciation rate which they set to 0.12 for the entire sample period. Given this user cost, they compute producers' profit rates as

$$s_{\pi_{it}} = 1 - \frac{\theta_{jt}^v}{\mu_{it}} - \frac{r_t k_{it}}{p_{it}y_{it}} - \frac{\text{FC}_{it}}{p_{it}y_{it}}, \tag{20}$$

where θ_j^v is the elasticity of the variable input for the industry j in which producer i operates, μ is the markup, py are sales, k is capital, FC are fixed costs, and r is the common user-cost of capital. Once we have obtained profit rates for all producers, we can compute the aggregate profit share using Domar weights as shown in Lemma 1.

Karabarbounis and Neiman (2019) compute several user costs of capital. We focus on the user cost of capital calculated under what they label "Case R." To calculate the user cost of capital, they use the equation

$$P^Q Q = WN + \tilde{R}^I K^I + \tilde{R}^N K^N + \Pi^Q, \tag{21}$$

where $P^Q Q$ is the business sector's value added, WN are labor payments, $\tilde{R}^I K^I$ are capital payments to IT capital, $\tilde{R}^N K^N$ are capital payments to non-IT capital, and Π^Q are business profits. They use $\tilde{R}^j$ for $j \in \{I, N\}$ to denote the "revised" rental rates of IT and non-IT capital and compute these taking $\{P^Q Q, WN, K^I, K^N, \Pi^Q\}$ as given so that equation (21) holds while being consistent with the Hall–Jorgenson formula for the measured (or non-revised) user costs of capital. They use the word "revised" because this way of computing the user cost of capital creates a wedge with measured rental rates. As is standard, they attribute the difference between measured- and revised rental rates to time-varying risk premia, adjustment costs, financial frictions, and the like. We borrow the time series of the revised rental rate from their paper and compute the profit share using this user cost for all producers using Lemma 1 and equation (20).

While we recognize that it is difficult to argue for one particular way of computing the user cost of capital, we believe that using the firm's first-order condition with respect to capital presents several advantages over other methods. First, and in contrast to both De Loecker et al. and Karabarbounis and Neiman, we allow for heterogeneous user costs of capital across producers. Second, we do not impose a homogeneous and constant depreciation rate for all producers as De Loecker et al..[32] Third, an advantage of estimating user costs of capital as we do is that these should be gross of wedges, meaning that risk premia, adjustment costs, financial frictions, and the like should be factored in at the level of each producer.

Figure 6 depicts the three different users costs of capital. The Hasenzagl–Perez series corresponds to the aggregate user cost of capital resulting from aggregating individual users costs.[33] Our series registers abrupt jumps around recessions and times of financial stress (e.g., 1973, 1980, 2000, 2007, 2012,) which is consistent with higher risk premia, adjustment costs, and heightened financial frictions during these times.[34] The level of our series is close to that of Karabarbounis and Neiman, which should also capture the

---

[32]There is substantial evidence that depreciation rates are heterogeneous across capital types (e.g., IT vs. non-IT) and that industries differ in the composition of their capital stocks. There is also evidence that rates of depreciation have increased over time as industries have become more reliant on technology.

[33]One can think of this user cost of capital as the user cost of capital that a representative firm would face for the total capital stock in the economy.

[34]While we find evidence in favor of higher risk premia and heightened financial frictions during recessions, in line with Gilchrist and Zakrajšek (2012), Duarte and Rosa (2015), Caballero et al. (2017), Jordà et al. (2019) and others, and consistently with many theories of financial frictions, there is also evidence on the contrary (see, for example Brinca et al., 2016; Chari et al., 2007).

aforementioned wedges. The level of the De Loecker et al.'s series is generally higher than those of the Karabarbounis–Neiman and Hasenzagl–Perez series because they impose a constant depreciation rate of 0.12.

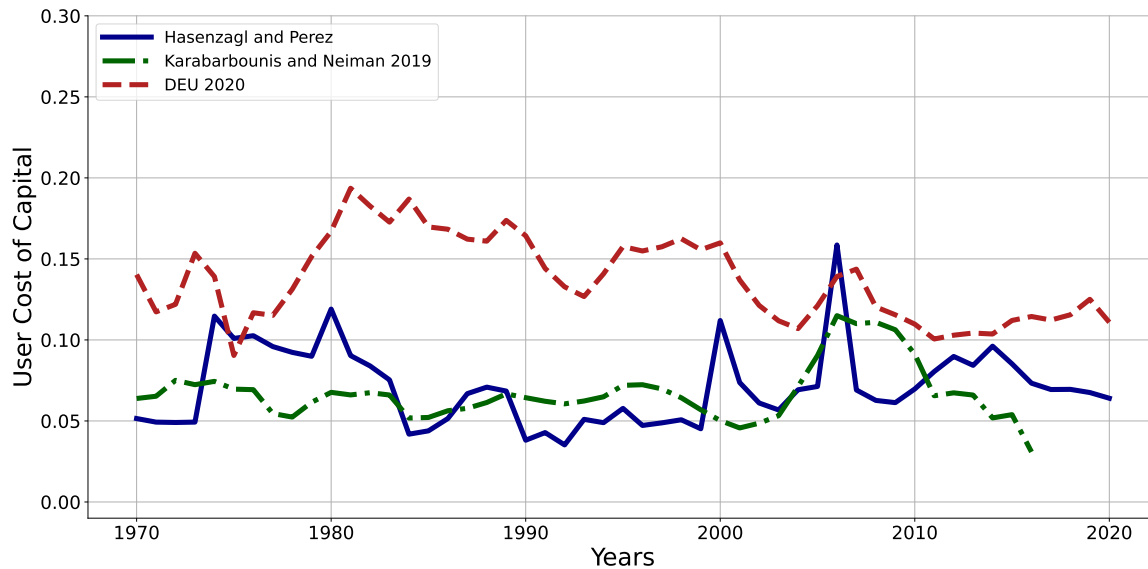**Figure 6:** Comparison of Aggregate User Costs of Capital Across Studies.

Figure 7 displays three different profit shares. We constructed these series by first computing firm-level profit rates using the different measures of the user cost of capital from figure 6 and then aggregating these profit rates. The series of Hasenzagl–Perez and Karabarbounis–Neiman have similar trends, except for the period 2000–2010.[35] The profit share implied by De Loecker et al.'s user cost of capital exhibits a different trend: the profit share increases dramatically from the early 1980s to 2020.

Figures 6 and 7 together shed light on the controversy surrounding the profit share: what is the most appropriate user cost of capital to use? While we believe that our aggregate user cost of capital displays a reasonable level and presents important advantages over other user cost measures (e.g., it captures risk premia, financial frictions, adjustment costs, and heterogeneity in financing opportunities across producers,) we acknowledge that any measure of the user cost of capital is subject to controversy.

---

[35]Even if the user cost of capital of Hasenzagl–Perez and Karabarbounis–Neiman are similar, profit shares may not be because of the underlying heterogeneity in user costs of capital in Hasenzagl–Perez.

Figure 7: Profit Shares and User Costs of Capital.
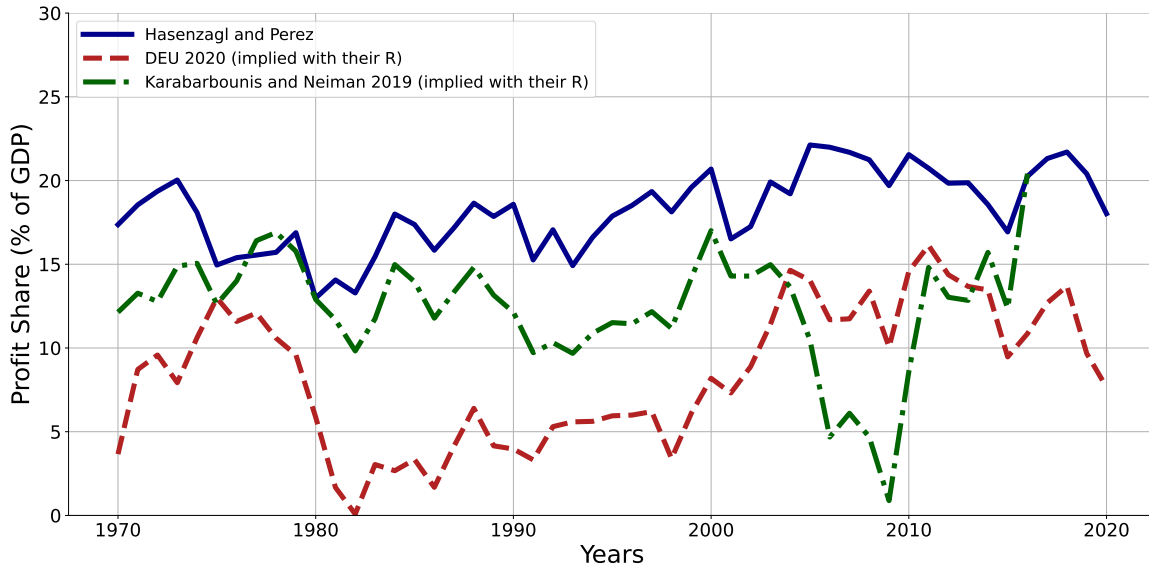
4.2 Industry-Level Heterogeneity

Next, we investigate the existing heterogeneity in markups, scale elasticities, returns to scale, and profitability across US industries. Table 2 reports our estimates of aggregate markups and scale elasticities at the level of 2-digit NAICS industries for 2019.[36] Markups are harmonic sales-weighted averages of firm-level markups, and scale elasticities are the sum of industries' output elasticities. In line with our expectations, these numbers indicate that industries with higher scale elasticities charge higher markups.

Figure 8 offers a scatter plot of industry-level markups and returns to scale, where larger dots reflect higher industry profit shares. We again see the relationship documented in Table 2, only that we now report aggregate returns to scale instead of scale elasticities. Recall that returns to scale are scale elasticities adjusted for explicit fixed costs. Thus, a positive correlation exists between returns to scale and industry-level markups. This figure also shows that industries with higher markups and returns to scale (e.g., IT, manufacturing, mining, utilities) generally have higher profit shares. While our results suggest that industries with higher fixed costs charge markups way above and beyond what is necessary to cover those costs, these results should be interpreted with care. It is possible that these industries charge higher markups today to cover fixed costs that were incurred in the past.

In Appendices E.3–E.5 we report time series of input elasticities, returns to scale, and markups for each industry.

---

[36]We have purposely chosen not to report markup estimates in 2020 because of Covid-19.

Table 2: Markups and Scale Elasticity Estimates from US Compustat, 2019.

Sector	NAICS code	MARKUPS		SCALE ELASTICITY	
		Mean	Std. dev.	Mean	Std. dev.
Agriculture & Fishing	11	1.09	0.20	1.01	–
Mining & Quarrying	21	1.26	0.62	1.11	–
Utilities	22	1.25	1.14	1.07	–
Construction	23	1.11	0.25	1.06	–
Manufacturing	31–33	1.34	0.47	1.09	–
Wholesale	42	0.98	0.48	1.00	–
Retail	44–45	1.07	0.23	1.00	–
Transportation	48–49	0.98	0.66	1.00	–
IT	51	1.51	0.49	1.08	–
Financial Services	52	1.19	1.07	1.05	–
Real Estate	53	1.19	1.08	1.03	–
Professional Services	54	1.17	0.38	1.06	–
Administration	56	1.23	0.27	1.07	–
Education	61	1.16	0.58	1.01	–
Health Care	62	1.17	0.38	1.11	–
Arts & Entertainment	71	1.18	0.36	1.04	–
Hospitality	72	1.14	0.25	0.99	–
Other	81	1.15	0.36	1.03	–

Table Notes. Reported markups are harmonic sales-weighted markups. Standard deviations come from individual-level markups. Scale elasticities are the sum of output elasticities. Since scale elasticities are estimated at the industry level, we cannot report standard deviations.
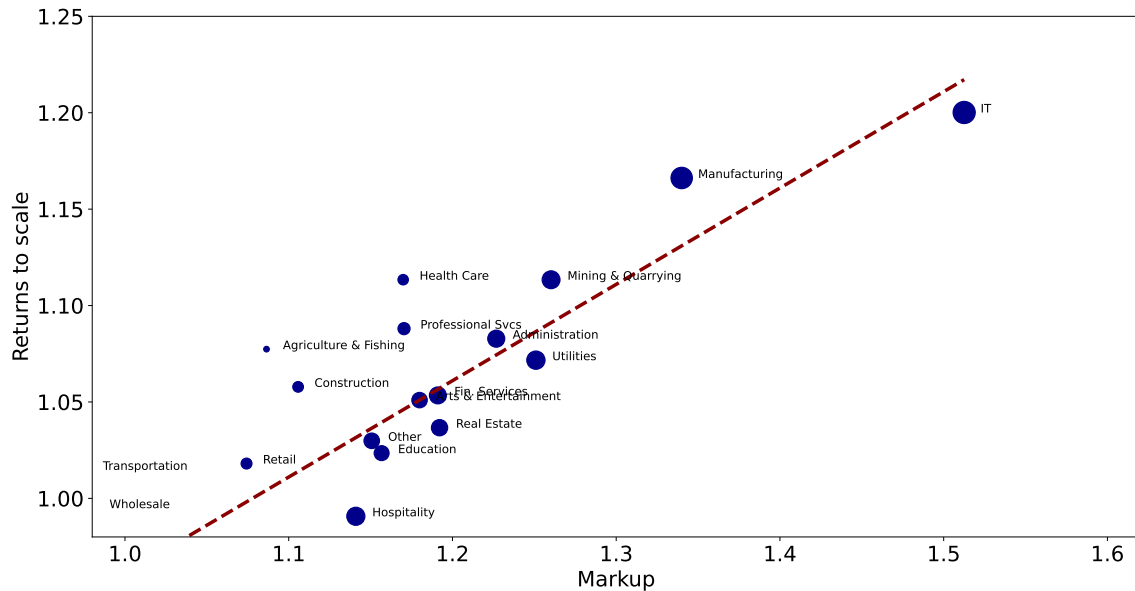
Figure 8: Industry-level Markups, Returns to Scale, and Profitability in the US, 2019.

Figure Notes. Markups are harmonic sales-weighted markups. Returns to scale are sales-weighted scale elasticities adjusted for fixed costs. Larger dots are associated with higher profit shares.

4.3 Decomposing the Rise in the Aggregate Markup

To understand the economic forces behind the increase in the aggregate markup, we provide a novel statistical decomposition for the harmonic sales-weighted markup.

Proposition 2 (Aggregate Markup Decomposition)*. The aggregate markup—the harmonic sales-weighted markup $\overline{\mu}_{hsw}$—can be decomposed according to*

$$\Delta\overline{\mu}_{hsw} = -\underbrace{\frac{\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\Delta\mu_i^{-1}}{\overline{\mu}_{hsw,t}\times\overline{\mu}_{hsw,t-1}}}_{within} - \underbrace{\frac{\sum_{i\in\mathcal{C}_t}\Delta\omega_i\overline{\mu_i^{-1}}}{\overline{\mu}_{hsw,t}\times\overline{\mu}_{hsw,t-1}}}_{between}$$
$$\underbrace{-\frac{\sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1}-\overline{\mu}_{hsw}^*\right)}{\overline{\mu}_{hsw,t}\times\overline{\mu}_{hsw,t-1}} + \frac{\sum_{i\in X_{t-1}}\omega_{it-1}\left(\mu_{it-1}^{-1}-\overline{\mu}_{hsw}^*\right)}{\overline{\mu}_{hsw,t}\times\overline{\mu}_{hsw,t-1}}}_{net\ entry} \quad (22)$$

where i and t index producers and time, respectively, $\mathcal{C}_t := \left\{i\in\mathcal{I}_t\ \wedge\ i\in\mathcal{I}_{t-1}\right\}$ is the set of incumbents, $\mathcal{E}_t := \left\{i\in\mathcal{I}_t\ \wedge\ i\notin\mathcal{I}_{t-1}\right\}$ is the set of entrants, $\mathcal{X}_{t-1} := \left\{i\in\mathcal{I}_{t-1}\ \wedge\ i\notin\mathcal{I}_t\right\}$ is the set of exiting firms, $\mathcal{I}_t = \mathcal{C}_t\cup\mathcal{E}_t$, $\mathcal{I}_{t-1} = \mathcal{C}_t\cup\mathcal{X}_{t-1}$, $\overline{\mu}_{hsw,t}$ is the harmonic sales-weighted markup in period t, $\Delta X := X_t - X_{t-1}$, $\overline{X} := \frac{1}{2}\left(X_t + X_{t-1}\right)$, and $\overline{\mu}_{hsw}^* = \frac{1}{2}\left(\overline{\mu}_{hsw,t} + \overline{\mu}_{hsw,t-1}\right)$.

Proof. See Appendix A.3. □

The results of this novel statistical decomposition are depicted in Figure 9.[37] Here, we decompose the change in the harmonic sales-weighted markup for 1980–1981, 1980–1982, and so on. As mentioned before, the aggregate markup in the United States has increased from around 1.10 in 1970 to 1.24 in 2019; that is, from 10% of price over marginal cost to 24%. Firm entry and exit accounts for about 50% of that increase (i.e., 7% percentage points,) and the remaining part is accounted for in equal parts by changes in markups within firms and reallocation of economic activity between firms.

The fact that the aggregate markup has increased mostly because of net entry is consistent with our narrative that rising fixed costs increase entry barriers, discouraging firm entry and encouraging firm exit. Importantly, we note that the increasing importance of net entry in the markup decomposition is not informative about the number of firms that enter or exit. This term should be interpreted as new entrants having to charge relatively high markups in order to operate without losses, whereas exiting firms are those that charge relatively lower markups.

---

[37]Since the last year of our sample is 2020, 2019 is the last year we use for the markup decomposition. This is because we identify exiting firms in year t using information from years t and $t+1$. Hence, we cannot identify exiting firms in 2020.

Figure 9: Cumulative Decomposition of the Rise in the Aggregate Markup, 1970–2019.
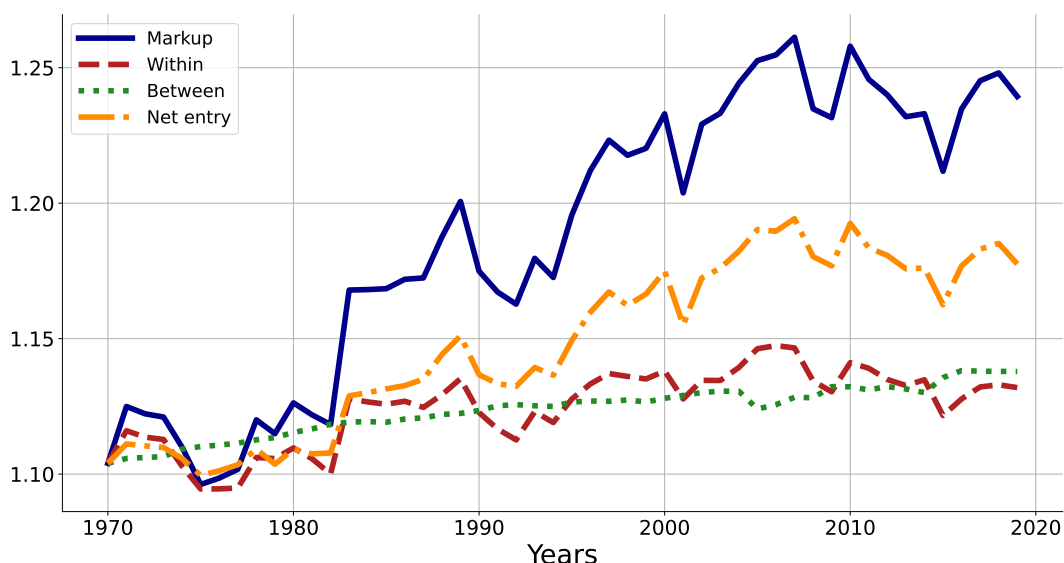

Figure Notes. The aggregate markup is the harmonic, sales-weighted markup. The within term captures the aggregate effect of changes in markups within firms, the between term captures how the reallocation of economic activity between firms has contributed to the change in the aggregate markup, and the net-entry component captures the effect of entry and exit.

5 Conclusion

Income shares are important summary statistics for understanding many macroeconomic phenomena, including the global decline in the labor share, the stability of the Kaldor facts, and economic inequality, to name just a few. One such statistic that has historically caught the eye of both economists and policymakers is the profit share.

In this paper, we show that the economy-wide profit share can be constructed by weighting individual profit rates using Domar weights (i.e., producer sales over GDP.) We refer to the profit share so constructed as the *micro-aggregated profit share* and note that no modeling assumptions—only fundamental accounting principles—are needed to establish this result. This procedure for calculating the aggregate profit share brings micro data and macro analysis closer, potentially allowing researchers to leverage interesting features of micro-level datasets for gaining further insights.

Since profits or profit rates are rarely observable—and even when they are, they reflect accounting- rather than economic profits—economic theory is often needed to compute economic measures of profits. Imposing minimal assumptions on producer behavior and technology, we provide a generic expression for a producer's profit rate—defined as profits over sales—in terms of monopoly and monopsony terms. Our formula is a generalization of standard formulas in the literature because it allows for explicit fixed costs and relaxes the assumption of price-taking behavior in factor

markets. Apart from standard regularity conditions on production functions, our result only requires producers to minimize costs.

The main theoretical contribution of this paper is to show that the profit share can be expressed in terms of several indicators of aggregate market power—the aggregate markup and an aggregate monopsony term—and a sufficient statistic for production networks that captures double marginalization (i.e., how profits propagate from downstream sellers to upstream suppliers.) Our theorem, which can be applied at any desired level of aggregation, is useful for (i) understanding the determinants of the profit share; (ii) quantifying different sources of profits—that is, monopoly *vis-à-vis* monopsony; (iii) assessing the plausibility of micro-level estimates of markups, markdowns, and returns to scale; (iv) using certain indicators of aggregate market power (e.g., the harmonic sales-weighted markup, sales-weighted markdowns, sales-weighted returns to scale); and (v) calibrating economic models with monopolistic and monopsonistic wedges.

We use the aforementioned theoretical results to guide our empirical analysis and reassess the extent of market power and profitability in the United States from 1970 to 2020 using firm-level data from US Compustat. Our empirical results are the following. First, we find that an economically meaningful notion of aggregate markup—the harmonic, sales-weighted markup—has increased from 10% of price over marginal cost in 1970 to 23% in 2020. Second, we find that the rise in the aggregate markup is mostly explained by the appearance of firms that charge higher markups. Reallocation of economic activity towards firms with higher markups and markup increases within firms each explain about 25% of the rise in the aggregate markup. Third, we find that aggregate returns to scale have risen from 1.00 in 1970 to 1.13 in 2020, reflecting both rising fixed costs and changes in technology. Fourth, despite of the rise in market power, we find that the aggregate profit share in the United States has been constant at 18% of GDP over the past five decades. We reconcile the increase in market power with the constancy of the aggregate profit share by decomposing the profit share into monopoly rents and fixed costs, both as shares of GDP, and showing that the increase in monopoly rents has been entirely offset by rising fixed costs and changing technology.

Our empirical findings have subtle policy implications. In particular, they suggest that a rise in market power does not necessarily call for stricter enforcement of antitrust law. If firms use their increase in market power to cover rising fixed costs, aggregate profits as a share of GDP remain stable. Although stricter enforcement of antitrust law could lower monopoly rents, it could also cause firms that cannot cover the rising fixed costs to exit the market. Fewer firms means less competition, which translates into higher market power for surviving firms. Thus, stricter enforcement of antitrust law can paradoxically lead to higher market power.

Our current empirical analysis is not complete and presents several limitations. We see our estimates of the profit share as providing an upper bound for the profit share since certain forms of capital are not correctly measured, and Compustat data are not representative of the US population of firms. In future versions of the paper, we plan to use data from the US Census of Manufactures to provide a decomposition of this industry's profit share into monopoly and monopsony rents, as well as to assess the representativeness assumption of Compustat firms for this particular sector.

References

Acemoglu, D., Akcigit, U. and Kerr, W. (2016). Networks and the macroeconomy: An empirical exploration, *NBER Macroeconomics Annual* **30**(1): 273–335.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations, *Econometrica* **80**(5): 1977–2016.

Acemoglu, D., Ozdaglar, A. and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks, *American Economic Review* **105**(2): 564–608.

Ackerberg, D. A., Caves, K. and Frazer, G. (2015). Identification properties of recent production function estimators, *Econometrica* **83**(6): 2411–2451.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J. and Li, H. (2019). A theory of falling growth and rising rents, *Technical report*, National Bureau of Economic Research.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices, *American Economic Review* **98**(5): 1998–2031.

Atkinson, A. B., Piketty, T. and Saez, E. (2011). Top incomes in the long run of history, *Journal of Economic Literature* **49**(1): 3–71.

Baqaee, D. and Farhi, E. (2019a). Networks, barriers, and trade, *Technical report*, National Bureau of Economic Research.

Baqaee, D. R. and Farhi, E. (2019b). The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem, *Econometrica* **87**(4): 1155–1203.

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

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

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

Berry, S., Gaynor, M. and Morton, F. S. (2019). Do increasing markups matter? lessons from empirical industrial organization, *Journal of Economic Perspectives* **33**(3): 44–68.

Berry, S., Levinsohn, J. and Pakes, A. (1995). Automobile prices in market equilibrium, *Econometrica* **63**(4): 841–890.

Berto Villas-Boas, S. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data, *The Review of Economic Studies* **74**(2): 625–652.

Bigio, S. and La'o, J. (2020). Distortions in production networks, *The Quarterly Journal of Economics* **135**(4): 2187–2253.

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

Boppart, T., Kiernan, P., Krusell, P. and Malmberg, H. (2023). The macroeconomics of intensive agriculture, *Technical report*, National Bureau of Economic Research.

Brinca, P., Chari, V. V., Kehoe, P. J. and McGrattan, E. (2016). Accounting for business cycles, *Handbook of Macroeconomics*, Vol. 2, Elsevier, pp. 1013–1063.

Brooks, W. J., Kaboski, J. P., Li, Y. A. and Qian, W. (2021). Exploitation of labor? classical monopsony power and labor's share, *Journal of Development Economics* **150**: 102627.

Caballero, R. J., Farhi, E. and Gourinchas, P.-O. (2017). The safe assets shortage conundrum, *Journal of Economic Perspectives* **31**(3): 29–46.

Chamberlin, E. (1933). *The Theory of Monopolistic Competition*, Harvard University Press.

Chari, V. V., Kehoe, P. J. and McGrattan, E. R. (2007). Business cycle accounting, *Econometrica* **75**(3): 781–836.

Crouzet, N., Eberly, J. C., Eisfeldt, A. L. and Papanikolaou, D. (2022). The economics of intangible capital, *Journal of Economic Perspectives* **36**(3): 29–52.
URL: *https://www.aeaweb.org/articles?id=10.1257/jep.36.3.29*

De Loecker, J., Eeckhout, J. and Mongey, S. (2022). Quantifying market power and business dynamism in the macroeconomy, *Technical report*, Working Paper.

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

De Ridder, M. (2019). Market power and innovation in the intangible economy.

De Ridder, M., Grassi, B., Morzenti, G. et al. (2022). The hitchhiker's guide to markup estimation.

Decker, R., Haltiwanger, J., Jarmin, R. and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism, *Journal of Economic Perspectives* **28**(3): 3–24.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity, *The American Economic Review* **67**(3): 297–308.

Döpper, H., MacKay, A., Miller, N. and Stiebale, J. (2023). Rising markups and the role of consumer preferences, *Working Paper* .

Duarte, F. and Rosa, C. (2015). The equity risk premium: a review of models, *Economic Policy Review* (2): 39–57.

Edmond, C., Midrigan, V. and Xu, D. Y. (2023). How costly are markups?, *Journal of Political Economy* **131**(7): 000–000.

Eggertsson, G. B., Robbins, J. A. and Wold, E. G. (2021). Kaldor and piketty's facts: The rise of monopoly power in the united states, *Journal of Monetary Economics* **124**: S19–S38.

Elsby, M. W., Hobijn, B. and Şahin, A. (2013). The decline of the us labor share, *Brookings Papers on Economic Activity* **2013**(2): 1–63.

Fisher, F. M. and McGowan, J. J. (1983). On the misuse of accounting rates of return to infer monopoly profits, *The American Economic Review* **73**(1): 82–97.

Gali, J. and Monacelli, T. (2005). Monetary policy and exchange rate volatility in a small open economy, *The Review of Economic Studies* **72**(3): 707–734.

Ghassibe, M. (2021). Endogenous production networks and non-linear monetary transmission, *Technical report*, Tech. rep.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.

Gordon, R. J. (2012). Is us economic growth over? faltering innovation confronts the six headwinds, *Technical report*, National Bureau of Economic Research.

Gutiérrez, G., Jones, C. and Philippon, T. (2021). Entry costs and aggregate dynamics, *Journal of Monetary Economics* **124**: S77–S91.

Hall, R. E. (1988). The relation between price and marginal cost in us industry, *Journal of Political Economy* **96**(5): 921–947.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M. and Taylor, A. M. (2019). The rate of return on everything, 1870–2015, *The Quarterly Journal of Economics* **134**(3): 1225–1298.

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

Karabarbounis, L. and Neiman, B. (2019). Accounting for factorless income, *NBER Macroeconomics Annual* **33**(1): 167–228.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables, *The Review of Economic Studies* **70**(2): 317–341.

Loecker, J. D. and Warzynski, F. (2012). Markups and firm-level export status, *American Economic Review* **102**(6): 2437–2471.

Long, J. B. J. and Plosser, C. I. (1983). Real business cycles, *Journal of Political Economy* **91**(1): 39–69.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry, *Econometrica* **69**(2): 307–342.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry, *Econometrica* **64**(6): 1263–1297.

Peters, R. H. and Taylor, L. A. (2017). Intangible capital and the investment-q relation, *Journal of Financial Economics* **123**(2): 251–272.

Pigou, A. C. (1920). The economics of welfare, *London, United Kingdom* .

Piketty, T. and Saez, E. (2003). Income inequality in the united states, 1913–1998, *The Quarterly Journal of Economics* **118**(1): 1–41.

Robinson, J. (1933). *The Economics of Imperfect Competition*, Palgrave Macmillan.

Smith, D. A. and Ocampo, S. (2022). The evolution of us retail concentration, *arXiv preprint arXiv:2202.07609* .

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

Traina, J. (2018). Is aggregate market power increasing, *Production trends using financial statements* .

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables, *Economics Letters* **104**(3): 112–114.

Yeh, C., Macaluso, C. and Hershbein, B. (2022). Monopsony in the us labor market, *American Economic Review* **112**(7): 2099–2138.

Appendices

"The Micro–Aggregated Profit Share"

Thomas Hasenzagl   Luis Pérez

October 9, 2023

List of Appendices

A Omitted Proofs

A.1 Proof of Lemma 2

Lemma 2 (The Micro–Aggregated Profit Share). *The aggregate profit share can be constructed from micro-level data by aggregating individual producers' profit rates, defined as profits over value added, using valued-added weights.*

Proof. Let Π denote aggregate profits. For each producer $i \in \mathcal{I}$, π_i denotes profits, $p_i y_i$ are sales, and $s_{\pi_i} := \pi_i/(\text{va}_i)$ is i's profit rate, where VA_i is the value added generated by producer i—that is, sales minus material costs. Then, we have

$$
\begin{aligned}
\Lambda_\Pi^{\text{Macro}} &= \frac{\Pi}{\text{GDP}} \\
&= \frac{\sum_{i \in \mathcal{I}} \pi_i}{\text{GDP}} \\
&= \sum_{i \in \mathcal{I}} \frac{\text{VA}_i}{\text{GDP}} \times \frac{\pi_i}{\text{VA}_i} \equiv \Lambda_\Pi^{\text{Micro, VA}}.
\end{aligned}
$$

$\square$

A.2 Proof of Proposition 1

Proposition 1 (Profit Rates, Monopoly, and Monopsony). *Under the assumptions of cost-minimizing behavior, the existence of a continuously differentiable and quasiconcave production function, and monopsony power in factor markets, a producer's profit rate, defined as profits over sales, can be written as*

$$
s_\pi = 1 - \frac{\text{RS}}{\mu} = 1 - \frac{\text{SE}^{\text{adj}}}{\mu} + \frac{\mathcal{M}}{\mu}, \tag{5}
$$

where RS = SE^{adj} – $\mathcal{M}$ are the returns to scale, SE^{adj} is the scale elasticity of the production function adjusted for fixed costs, $\mathcal{M}$ is a monopsony term capturing market power in factor markets, and μ is the markup of price over marginal cost.

The scale elasticity adjusted for fixed costs is defined as

$$
\text{SE}^{\text{adj}} := \text{SE} \times \left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right),
$$

where SE is the scale elasticity, TC are total costs, and FC are fixed costs.

The monopsony term is given by

$$\mathcal{M} := \left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j), \tag{7}$$

where $\theta_j \equiv \partial F / \partial x_j \times x_j / y$ is the elasticity of output with respect to input j, and ν_j is the markdown on factor $j \in \mathcal{F}$, where the markdown is defined as the ratio of input j's rental rate to its marginal revenue product; that is, $\nu_j := w_j(x_j) / MRP_j$.

Proof. The cost minimization problem of a producer may be written as

$$\min_{x_j \geq 0} \quad \text{TC} \equiv \sum_{j \in \mathcal{N}} w_j(x_j) x_j + \text{FC}$$

$$\text{s.t.} \quad y = F\left(\{x_j\}_{j \in N}; A \right) \geq \overline{y},$$

where TC is the total cost function, $w_j(x_j)$ is the rental rate of input j, which depends on the quantity demanded by producer of input j, x_j, FC are fixed costs, y is gross output, $F : \mathbb{R}_+^N \to \mathbb{R}_+$ is a continuously differentiable and quasiconcave production technology which satisfies Inada conditions and is parametrized by productivity parameter A, and $\overline{y}$ is the minimum required amount of output. $\mathcal{N}$ denotes the set of total inputs, which can be either materials or production factors, and $\mathcal{F} \subseteq \mathcal{N}$ denotes the set of factors.

The Lagrangian of this programming problem is

$$\mathcal{L}(x; \lambda) = \sum_j w_j(x_j) x_j + \text{FC} + \lambda \left\{ \overline{y} - F(x; A) \right\},$$

where λ, the Lagrange multiplier, is the shadow value of relaxing the output constraint.

A generic first-order condition for an interior demand reads as

$$w_j(x_j) + w_j'(x_j) x_j - \lambda \frac{\partial F}{\partial x_j} = 0,$$

which implies

$$\left(1 + \frac{w_j'(x_j) x_j}{w_j(x_j)} \right) w_j(x_j) x_j = \lambda \frac{\partial F}{\partial x_j} \frac{x_j}{y} y,$$

where $\varepsilon_{Sj}^{-1} \equiv \frac{w_j'(x_j) x_j}{w_j(x_j)}$ is the producer's (perceived) inverse elasticity of supply of input j.

Using standard duality arguments, we can establish that $1 + \varepsilon_{Sj}^{-1}$ is the inverse markdown on input j—that is, the marginal revenue product of the producer with respect to j divided by the rental rate of input j. We denote the markdown on j by

v_j, and note that $v_j \in (0, 1]$. By the Envelope Theorem, it follows that the Lagrange multiplier λ is the marginal cost. Thus, we can write:

$$w_j(x_j)x_j = \text{mc} \times y\theta_j v_j,$$

where $\theta_j \equiv \partial F/\partial x_j \times x_j/y$ is the output elasticity of input j.

Summing over all inputs $j \in \mathcal{N}$,

$$\sum_{j \in \mathcal{N}} w_j(x_j)x_j = \text{mc} \times y \left(\sum_{j \in \mathcal{N}} \theta_j v_j \right)$$

$$= \text{mc} \times y \left(\text{SE} - \sum_{j \in \mathcal{F}} \theta_j\{1 - v_j\} \right),$$

where SE is the scale elasticity of the production function and the second equality follows by imposing the restriction that there is monopsony power only in factor markets, i.e., $v_j = 1$ for $j \notin \mathcal{F}$.

Using the definition of the markup, we can write the producer's profit rate s_π, defined as profits π over sales py as

$$s_\pi = 1 - \frac{\text{SE}}{\mu}\left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right) + \frac{1}{\mu}\left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j(1 - v_j).$$

Next, we show that returns to scale, defined as the ratio of average- to marginal cost, are given by $\text{RS} = \text{SE}^{\text{adj}} - \mathcal{M}$. Recall that

$$\sum_{j \in \mathcal{N}} w_j(x_j)x_j = \text{mc} \times y \left(\text{SE} - \sum_{j \in \mathcal{F}} \theta_j\{1 - v_j\} \right).$$

Doing simple algebraic manipulations,

$$\frac{\text{AC}}{\text{mc}} \frac{\text{TC} - \text{FC}}{\text{AC} \times y} = \text{SE} - \sum_{j \in \mathcal{F}} \theta_j\{1 - v_j\},$$

where $\text{AC} := \text{TC}/y$ is the average cost.

The last expression implies

$$\text{RS} = \text{SE}\left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right) - \left(\frac{\text{TC}}{\text{TC} - \text{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j\{1 - v_j\},$$

which establishes the result. $\square$

A.3 Proof of Proposition 2

Proposition 2. *The aggregate markup—the harmonic sales-weighted markup $\overline{\mu}_{hsw}$—can be decomposed according to*

$$\Delta\overline{\mu}_{\text{hsw}} = \underbrace{-\frac{\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\Delta\mu_i^{-1}}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}}}_{\text{within}}\underbrace{-\frac{\sum_{i\in\mathcal{C}_t}\Delta\omega_i\overline{\mu_i^{-1}}}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}}}_{\text{between}}$$

$$\underbrace{-\frac{\sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1}-\overline{\mu}_{hsw}^*\right)}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}}+\frac{\sum_{i\in X_{t-1}}\omega_{it-1}\left(\mu_{it-1}^{-1}-\overline{\mu}_{hsw}^*\right)}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}}}_{\text{net entry}}$$

where i and t index producers and time, respectively, $\mathcal{C}_t := \left\{i\in\mathcal{I}_t \ \wedge \ i\in\mathcal{I}_{t-1}\right\}$ is the set of incumbents, $\mathcal{E}_t := \left\{i\in\mathcal{I}_t \ \wedge \ i\notin\mathcal{I}_{t-1}\right\}$ is the set of entrants, $\mathcal{X}_{t-1} := \left\{i\in\mathcal{I}_{t-1} \ \wedge \ i\notin\mathcal{I}_t\right\}$ is the set of exiting firms, $\mathcal{I}_t = \mathcal{C}_t\cup\mathcal{E}_t$, $\mathcal{I}_{t-1} = \mathcal{C}_t\cup\mathcal{X}_{t-1}$, $\overline{\mu}_{hsw,t}$ is the harmonic sales-weighted markup in period t, $\Delta X := X_t - X_{t-1}$, $\overline{X} := \frac{1}{2}\left(X_t + X_{t-1}\right)$, and $\overline{\mu}_{hsw}^* = \frac{1}{2}\left(\overline{\mu}_{hsw,t} + \overline{\mu}_{hsw,t-1}\right)$.

Proof.

$$\Delta\overline{\mu}_{\text{hsw}} = \overline{\mu}_{\text{hsw},t} - \overline{\mu}_{\text{hsw},t-1}$$

$$= \left(\sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1}\right)^{-1} - \left(\sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right)^{-1} \qquad \text{(by definition)}$$

$$= \frac{\sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1} - \sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1}}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}} \qquad \text{(taking common denominator)}$$

$$= -\frac{1}{\overline{\mu}_{\text{hsw},t}\times\overline{\mu}_{\text{hsw},t-1}}\left\{\sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right\} \qquad \text{(rearranging)}$$

Noting that

$$\sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1} = \sum_{i\in\mathcal{C}_t}\omega_{it}\mu_{it}^{-1} + \sum_{i\in\mathcal{E}_t}\omega_{it}\mu_{it}^{-1}$$

and

$$\sum_{i\in\mathcal{I}_{t-1}}\omega_{it-1}\mu_{it-1}^{-1} = \sum_{i\in\mathcal{C}_t}\omega_{it-1}\mu_{it-1}^{-1} + \sum_{i\in\mathcal{X}_{t-1}}\omega_{it-1}\mu_{it-1}^{-1},$$

43

we can write

$$\Delta\overline{\mu}_{\mathrm{hsw}} = -\frac{1}{\overline{\mu}_{\mathrm{hsw},t} \times \overline{\mu}_{\mathrm{hsw},t-1}} \left\{ \sum_{i \in \mathcal{C}_t} \omega_{it} \mu_{it}^{-1} - \sum_{i \in \mathcal{C}_t} \omega_{it-1} \mu_{it-1}^{-1} + \sum_{i \in \mathcal{E}_t} \omega_{it} \mu_{it}^{-1} - \sum_{i \in \mathcal{X}_{t-1}} \omega_{it-1} \mu_{it-1}^{-1} \right\}.$$

The result follows from algebraic manipulations to this expression. $\qquad\square$

A.4 Identification Proof for the Aggregate User Cost of Capital

Proposition 3. *Suppose labor compensation, the capital stock, and value added are observable at the aggregate level. Further suppose that representative micro-level data on sales, gross output, variable inputs, and capital stocks are available. Then, under the assumptions of cost-minimizing behavior, homogeneous, continuously differentiable, and quasiconcave production functions, and no monopsony power for one variable input, we can identify the aggregate user cost of capital, R.*

Proof. By Lemma 1, the aggregate profit share can be computed as Domar-weighted profit rates. By Proposition 1, economic profit rates are given by equation (5). Markups, markdowns, elasticities, and returns to scale for each producer, $\{\mu_i, \nu_i, \theta_i, \mathrm{RS}_i\}_i$, can be estimated using the production function approach and the control function approach explained in Section 3. Importantly, notice that individual user costs of capital are not required in this estimation procedure. Markups are computed in a first step, using the first-order condition for the variable input no subject to monopsony, and markdowns for all the remaining inputs are estimated in a second step using markup estimates and first-order conditions. This is basically the procedure of Yeh et al. (2022).

Using the macro approach, we can obtain the aggregate user cost of capital as

$$R = \frac{\mathrm{GDP}}{K}(1 - \Lambda_L - \Lambda_\Pi),$$

where GDP is aggregate value added, K is the aggregate capital stock, Λ_L is labor compensation over value, and Λ_Π is the Domar-weighted profit share, computed using Lemma 1 and Proposition 1. □

A.5   Proof of Theorem 1

Theorem 1. *With cost-minimizer producers, continuously differentiable and quasiconcave production functions, fixed costs, and market power in factor- and output markets, the profit share can be expressed as*

$$\Lambda_\Pi = \chi \left(1 - \mathbb{E}_\omega \left[\frac{\text{SE}^{\text{adj}}}{\mu} \right] + \mathbb{E}_\omega \left[\frac{\mathcal{M}}{\mu} \right] \right) \tag{9}$$

$$= \chi \left(1 - \frac{\overline{SE}^{\text{adj}}}{\overline{\mu}_{hsw}} + \frac{\overline{\mathcal{M}}}{\overline{\mu}_{hsw}} - \text{Cov}_\omega \left[\text{SE}^{\text{adj}}, \frac{1}{\mu} \right] + \text{Cov}_\omega \left[\mathcal{M}, \frac{1}{\mu} \right] \right), \tag{10}$$

where $\chi = \sum_{k \in \mathcal{I}} \frac{p_k y_k}{\text{GDP}}$ is the input-output multiplier, $\mathbb{E}_\omega [\cdot]$ denotes a sales-weighted average where ω indexes sales weights, SE^{adj} is the scale elasticity adjusted for fixed costs given by (6), μ denotes the markup, and $\mathcal{M}$ is a monopsony term given by (7).

The term $\overline{X}$ is the sales-weighted average of X, $\overline{\mu}_{hsw}$ is the harmonic sales-weighted markup, and $Cov_\omega(X, Y)$ is the sales-weighted covariance of variables X and Y.

Proof. By Lemma 1, the aggregate profit share can be computed as

$$\Lambda_\Pi = \sum_{i \in \mathcal{I}} \frac{p_i y_i}{\text{GDP}} \times s_{\pi_i}.$$

Proposition 1 can be used to establish

$$s_{\pi_i} = 1 - \frac{\text{SE}_i^{\text{adj}}}{\mu_i} + \frac{\mathcal{M}_i}{\mu_i}.$$

Hence,

$$\Lambda_\Pi = \sum_{i \in \mathcal{I}} \frac{p_i y_i}{\text{GDP}} \left(1 - \frac{\text{SE}_i^{\text{adj}}}{\mu_i} + \frac{\mathcal{M}_i}{\mu_i} \right)$$

$$= \underbrace{\left(\sum_{k \in \mathcal{I}} \frac{p_k y_k}{\text{GDP}} \right)}_{\equiv \chi} \sum_{i \in \mathcal{I}} \underbrace{\frac{p_i y_i}{\sum_k p_k y_k}}_{\equiv \omega_i} \left(1 - \frac{\text{SE}_i^{\text{adj}}}{\mu_i} + \frac{\mathcal{M}_i}{\mu_i} \right)$$

$$= \chi \left(1 - \mathbb{E}_\omega \left[\frac{\text{SE}^{\text{adj}}}{\mu} \right] + \mathbb{E}_\omega \left[\frac{\mathcal{M}}{\mu} \right] \right)$$

$$= \chi \left(1 - \frac{\overline{SE}^{\text{adj}}}{\overline{\mu}_{hsw}} + \frac{\overline{\mathcal{M}}}{\overline{\mu}_{hsw}} - \text{Cov}_\omega \left[\text{SE}^{\text{adj}}, \frac{1}{\mu} \right] + \text{Cov}_\omega \left[\mathcal{M}, \frac{1}{\mu} \right] \right)$$

$$\square$$

B Example Linking Increasing Returns and Fixed Costs

We provide a simple example to support our claim that increasing returns to scale can be thought of as fixed costs arising from employing factors of production.

Suppose there is a firm with production technology

$$y = \begin{cases} A(\ell - \bar{\ell})^\alpha, & \ell > \bar{\ell} \\ 0, & \text{otherwise} \end{cases}, \tag{B.1}$$

where y is output, $A \in \mathbb{R}_{++}$ is a productivity shifter, $\ell \in \mathbb{R}_+$ is labor, $\bar{\ell} \in \mathbb{R}_{++}$ is a minimum requirement on labor to produce positive output, and $\alpha \in \mathbb{R}_{++}$ is a parameter governing the returns to scale in production.[38]

Assuming the production function is continuously differentiable and no monopsony power, we can write the derivative of output with respect to labor ℓ as

$$\frac{\mathrm{d}y}{\mathrm{d}\ell} = \alpha A(\ell - \bar{\ell})^{\alpha-1}. \tag{B.2}$$

Multiplying both sides of (B.2) by $(\ell - \bar{\ell})$ we get

$$\frac{\mathrm{d}y}{\mathrm{d}\ell}(\ell - \bar{\ell}) = \alpha A(\ell - \bar{\ell})^\alpha = \alpha y.$$

Rearranging this expression and doing simple algebraic manipulations yields

$$\frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y} = \alpha + \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\bar{\ell}}{y}$$

$$= \alpha + \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y} \times \frac{\bar{\ell}}{\ell},$$

which implies

$$\frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y} = \alpha\frac{\ell}{\ell - \bar{\ell}} > \alpha.$$

Thus, if $\alpha = 1$ and $\ell > \bar{\ell} > 0$, there are increasing returns to scale because of fixed costs.

Next, we argue that the returns to scale, RS, that we would estimate for this production function are

$$\mathrm{RS} = \alpha\frac{\ell}{\ell - \bar{\ell}}.$$

---

[38]An example of a firm with production function (B.1) could be a pharmaceutical company that needs to hire a minimum amount of labor hours $\bar{\ell}$ (say, hours provided by attorneys) in order to obtain approval by the US Food and Drug Administration (FDA).

To see this, suppose we have cross-sectional data on $\{y, (\ell + \bar{\ell})\}_i$, where y is output, $(\ell + \bar{\ell})$ is total labor, and i indexes firms. Then, consider the following regression

$$\log y_i = \beta \log(\ell + \bar{\ell})_i + \tilde{\omega}_i,$$

where β, the returns to scale, is the parameter of interest, and $\tilde{\omega}$ is white noise.

Note that

$$\begin{aligned} \beta &= \frac{\mathrm{d}\log y}{\mathrm{d}\log(\ell + \bar{\ell})} \\ &= \frac{\mathrm{d}\log y}{\mathrm{d}\log \ell} \\ &= \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y}. \end{aligned}$$

Given production function (B.1), it follows that

$$\beta = \alpha \frac{\ell}{\ell - \bar{\ell}},$$

which is the parameter we would recover from a regression.

From this simple example, we argue that increasing returns to scale may capture increasing fixed costs. We use this logic to interpret the returns to scale we estimate in our empirical analysis.

C The DEU–Basu Controversy

Basu (2019) was the first to suggest that De Loecker et al. (2020)'s markup estimates had implausible macroeconomic implications. Through an informal argument, he noted that a markup of 1.61 in 2016 and returns to scale of 1.05 imply a profit share of about 70% of value added in the United States once we recognize that the ratio of sales to value added is approximately two. He argued that such a profit share was implausible since the labor share, which can be read from NIPA tables, was around 55% of GDP. Barkai (2020) made Basu's argument more formally by explicitly using expression

$$\Lambda_\Pi = \chi \left(1 - \frac{\text{RS}}{\mu} \right), \tag{C.1}$$

where Λ_Π is the profit share, χ is the ratio of sales to aggregate value added, RS are returns to scale, and μ is the markup.

De Loecker, Eeckhout and Unger's response to Basu (and indirectly to Barkai) was to say that such an argument could not be made because it incorrectly relied on a representative-firm assumption and imposed no fixed costs. We contribute to this discussion by providing an exact aggregation result that clarifies where the arguments of Basu and Barkai fail and why the aggregate sales-weighted markup of De Loecker, Eeckhout and Unger cannot be used to infer the aggregate profit share.

Corollary 1 provides the following expression for the profit share under the assumption of no monopsony:

$$\Lambda_\Pi = \chi \left(1 - \frac{\overline{\text{RS}}}{\overline{\mu}_{hsw}} - \text{Cov}_\omega \left[\text{RS}, \frac{1}{\mu} \right] \right), \tag{11}$$

where $\overline{\text{RS}} = \overline{\text{SE}}^{\text{adj}}$ are sales-weighted returns to scale, which are equal to the sales-weighted scale elasticity adjusted for fixed costs, $\overline{\mu}_{hsw}$ is the harmonic sales-weighted markup, and $\text{Cov}_\omega(\cdot, \cdot)$ is the sales-weighted covariance operator.

Inspection of equation (11) reveals several problems with Basu and Barkai's arguments. First, as noted by De Loecker, Eeckhout and Unger, if there is heterogeneity in markups and returns to scale across producers, aggregation is nonlinear as by Jensen's inequality. That explains why the covariance term appears in equation (11). Second, our aggregation result demands the harmonic sales-weighted markup and sales-weighted returns to scale, not the sales-weighted markup. If there is dispersion in markups, the sales-weighted markup is larger than the harmonic sales-weighted markup, and the discrepancy between the two markup notions depends on the variance of the cross-sectional distribution of markups. Thus, using the sales-weighted markup in equation

(11) would generally lead to overstating the aggregate profit share. While Basu and Barkai used what was available to them—that is, the estimates provided by De Loecker, Eeckhout and Unger—the sales-weighted markup is not an appropriate measure of aggregate markup to draw conclusions about the aggregate profit share; instead, the harmonic sales-weighted markup is. A third problem with Basu and Barkai's arguments in inferring the profit share from De Loecker et al.'s estimates is to not account for fixed costs. What they use as returns to scale are not returns to scale, but rather the scale elasticity of the production function. In the absence of monopsony power, returns to scale are given by the scale elasticity times a fixed-cost adjustment factor as our formula shows. The fixed-cost adjustment factor is important because De Loecker et al. treat SG&A as an explicit fixed cost. Once we aggregate correctly and account for all fixed costs, we get equation (11), which is well suited to assess the validity of micro estimates of markups and returns to scale under the assumption of no monopsony.[39]

In Figure 10, we map micro-level estimates of markups and returns to scale to the aggregate profit share using equation (11). To relate to the discussion between Basu and De Loecker, Eeckhout and Unger, we use the production function parameters that replicate De Loecker et al. (2020)'s results. That is, we use the cost of goods sold (COGS) as a variable input, treat selling, general, and administrative expenses (SG&A) as fixed costs, and only consider physical capital when defining the capital stock. The Hasenzagl–Perez series maps micro-level estimates of markups and returns to scale to an aggregate profit share that is constant at around 15% of GDP. The other three series map firm-level estimates of markups to the aggregate profit share using flawed aggregation schemes and ignoring relevant data features such as fixed costs.

Figure 10: Mapping Micro-Level Estimates of Markups to the US Profit Share.
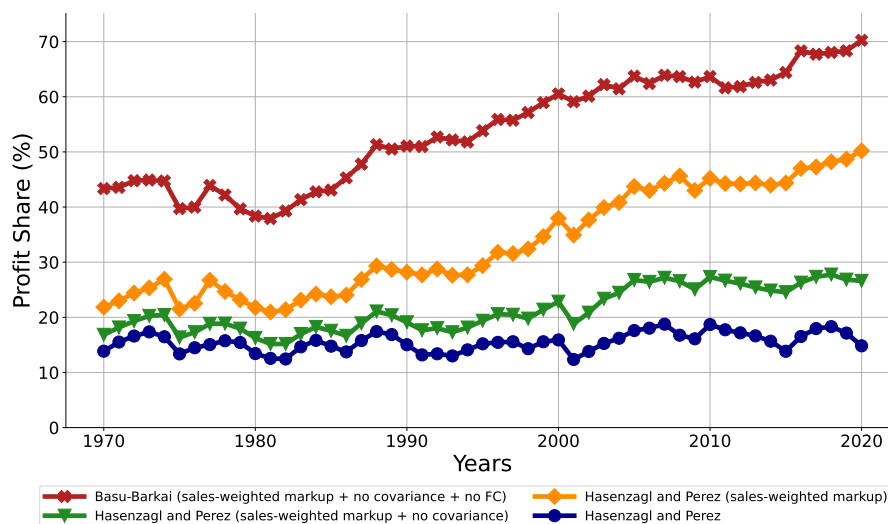
[39]Equation (11) nests Basu's and Barkai's expression (C.1) as the special case in which there are no explicit fixed costs and no heterogeneity in markups.

D Data Appendix

Table 3 provides key summary statistics for the variables we use in our estimation procedure. Although the relevant years of analysis are 1970–2020, we provide summary statistics for 1966–2020 given that information from 1966 onward is used to estimate input elasticities, as noted in Section 3.3.

Table 3: Summary Statistics from US Compustat, 1966–2020.

Variable	Mean	Median	Min	Max	Std. dev.
SALE	2261.51	157.67	0	489529.70	11930.17
COGS	1544.21	92.69	0	428240.60	9118.36
SG&A	332.07	28.22	0	98210.41	1619.11
R&D	51.64	0	0	37567.35	436.57
PPEGT	1658.22	54.94	0	542207.80	10979.76
K_INT	1013.56	45.12	0	334349.90	6091.17
CAPX	161.80	6.58	0	50038.86	1042.36
Observations	261,945				

E Additional Results

E.1 Replication of De Loecker, Eeckhout and Unger (2020)

In this section, we show that computing and aggregating markups as De Loecker et al. (2020) do, we almost perfectly replicate their results. The replication of markups is shown in Figure 11.[40] Small quantitative discrepancies result from updates in US Compustat and minor differences in sample selection, data cleaning, and input elasticities.

Figure 11: Replication of De Loecker et al.'s Markup Series, 1970–2020.
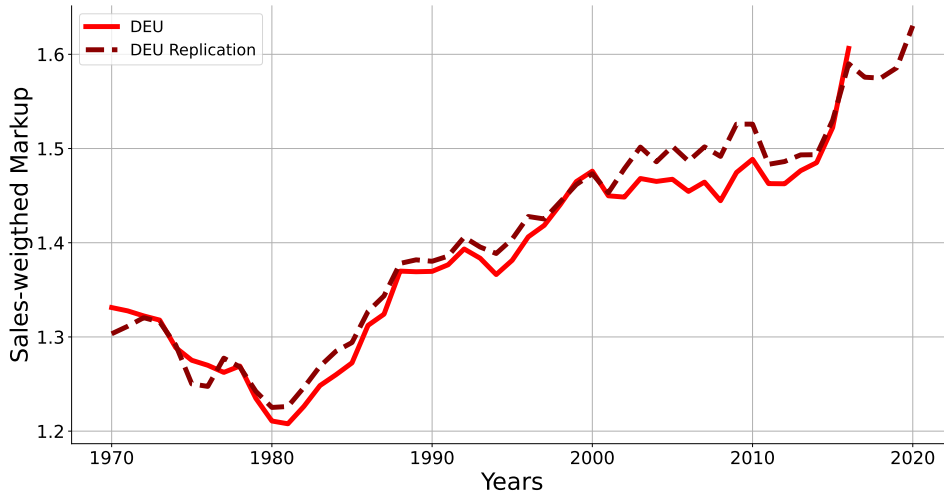
Table Notes. Firm-level data comes from US Compustat. Markups are computed using the costs of goods sold (COGS) as variable costs, time-varying elasticities, and aggregating markups using sales weights.

Figure 12 shows four different aggregate markup series: the one from De Loecker, Eeckhout and Unger (DEU replication) and three other aggregate markup series which rely on the aggregation scheme implied by our theorem. The difference between cross-marked and diamond-marked series illustrates that more than half of the differences between De Loecker et al.'s markup series and ours arises due to different aggregation schemes. We use the harmonic sales-weighted markup rather than the sales-weighted markup. The difference between cross-marked and triangle-marked series indicates in light of the above that the remaining difference between De Loecker et al.'s markup series and ours is due to a different categorization of variable inputs. We treat the cost of goods sold (COGS) and selling, general, and administrative expenses (SG&A) as variable inputs, while De Loecker et al. (2020) categorize only COGS as variable inputs. In theory, the markup could be recovered from any variable input so it should not matter whether one uses COGS or OPEX (= COGS + SG&A) to estimate markups. In

---

[40]We find very similar results when we estimate time-varying input elasticities which are stable and oscillate around 0.84–0.86.
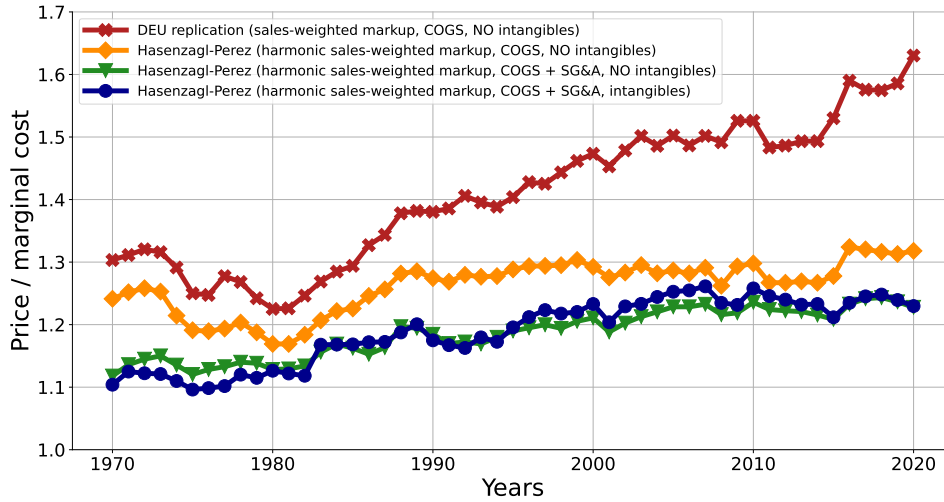
Figure 12: Differences Between De Loecker et al. (2020)'s Markup Series and Ours.

practice, however, the categorization of variable inputs matters for markup estimation. We follow Traina (2018) and most of the subsequent literature and treat both COGS and SG&A as variable inputs. Finally, we note that the inclusion of intangible capital does not affect markup estimates, and that the discrepancies between triangle-marked and dot-marked lines reflect measurement error. This is to be expected since markup estimates do not depend on the elasticity of output with respect to capital.

Figure 13 shows the replication of De Loecker et al. (2020)'s sales-weighted average profit rate. Again, we almost perfectly replicate their series. This is not surprising in the light of the almost perfect replication of markups and input elasticity estimates, since we we also use their exogenous user cost of capital.

Figure 13: Replication of De Loecker et al. (2020) Sales-weighted Profit Rate, 1970–2020.
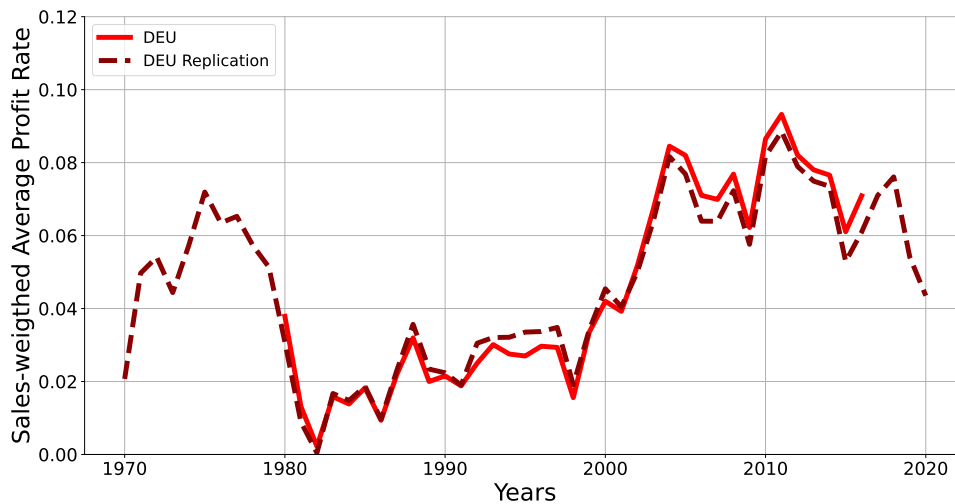
Table Notes. Firm-level data comes from US Compustat. Markups are computed using the costs of goods sold (COGS) as variable costs, time-varying elasticities, and aggregating markups using sales weights. Profit rates for each producer are computes as one minus the time-varying input elasticity of the variable input over the markup, minus fixed costs divded by sales.

## E.2 The Input–Output Multiplier in the United States

Figure 14 plots the time series for the input-output multiplier, which is the sum of Domar weights or, in other words, the ratio of economy-wide sales to GDP. Clearly from this figure, the IO multiplier for the United is roughly stable at around 1.8 and mildly procyclical. This implies that, absent changes on markups, markdowns and returns to scale during recessions, the aggregate profit share should decline during recessions.

Figure 14: The US Input–Output Multiplier, 1970–2020.
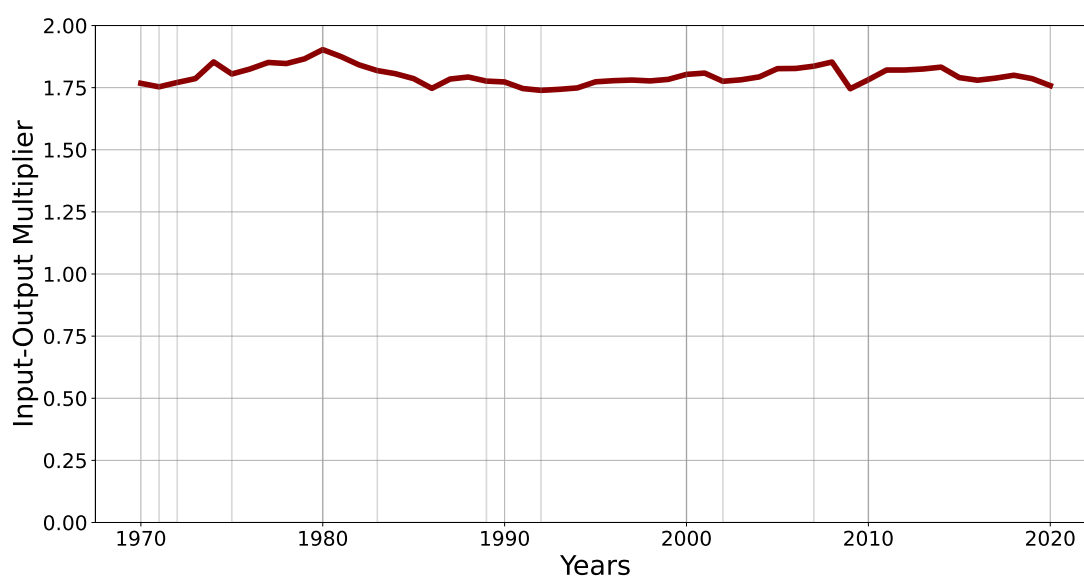
E.3 Heterogeneity in Variable-Input Elasticities

Figure 15: Variable Input Elasticities Across Industries, 1970–2020.
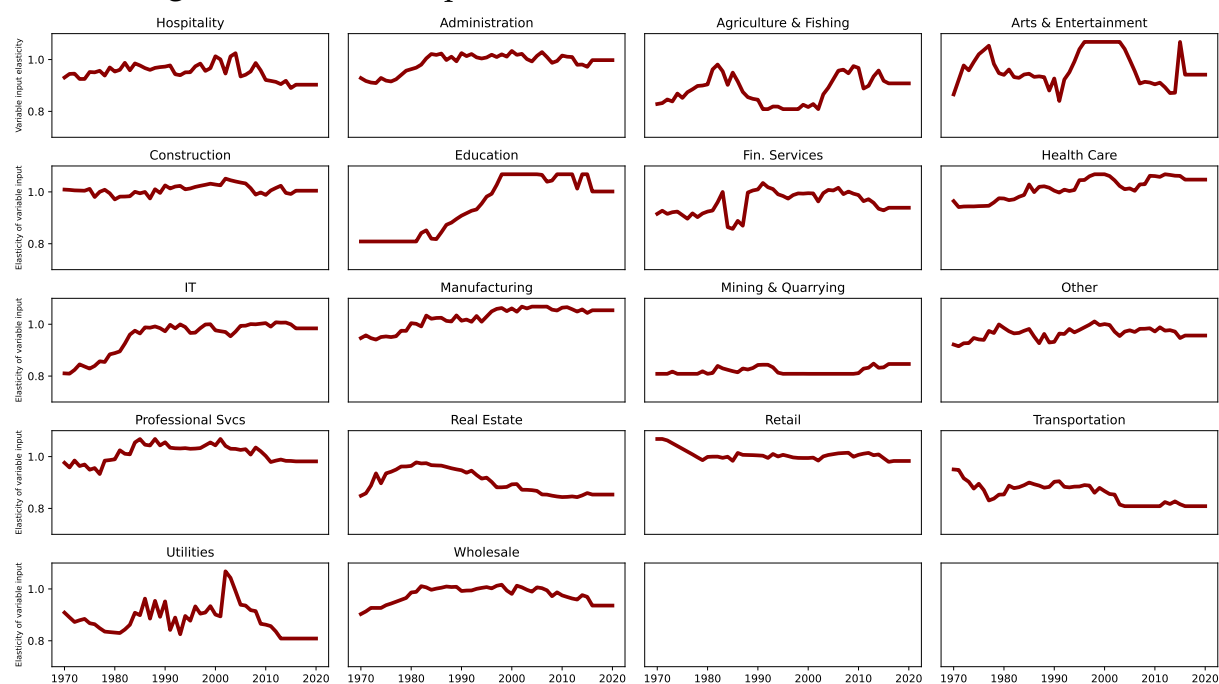
E.4 Heterogeneity in Returns to Scale

Figure 16: Returns to Scale Across Industries, 1970–2019.
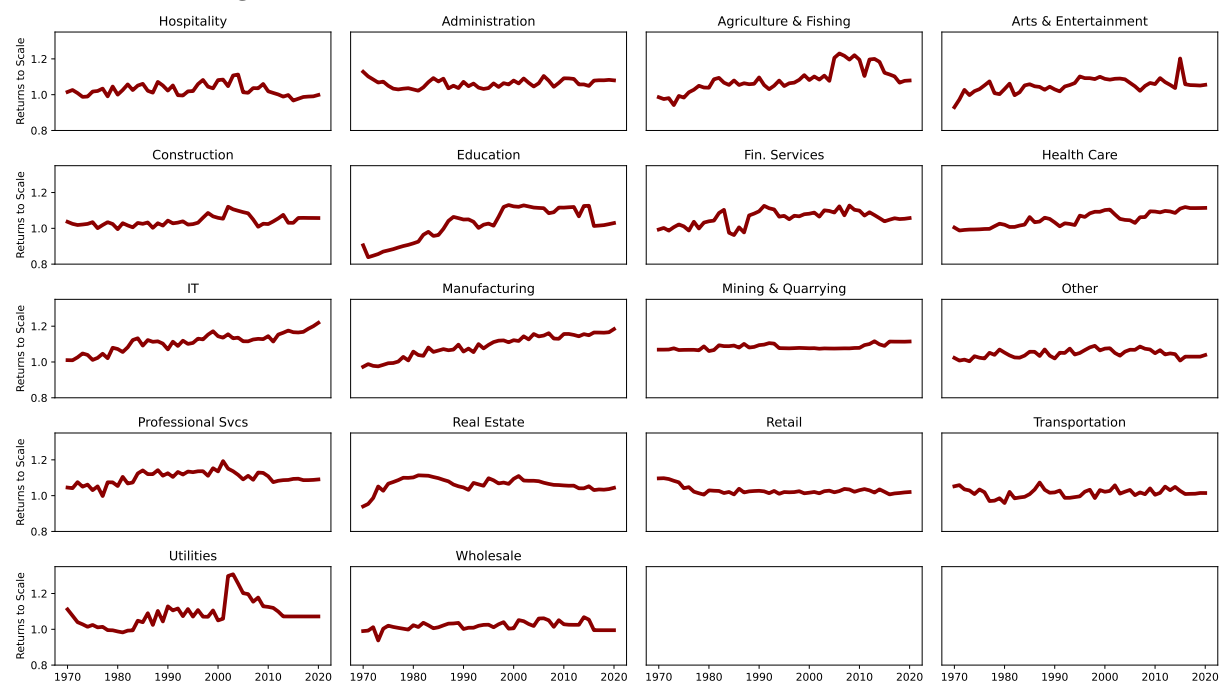
E.5 Heterogeneity in Markups

Figure 17 provides time series of industry-level markups. As the different panels show, industry-level markups have remained stable for a handful of industries—namely, construction, health care, mining and quarrying, professional services, transportation, arts and entertainment, and wholesale- and retail trade. Other industries such as hospitality, administration, agriculture and fishing, education, financial services, IT, manufacturing, real estate and utilities have experienced large increases in markups over the past fifty years.

Figure 17: Industry-level Markups, 1970–2020.
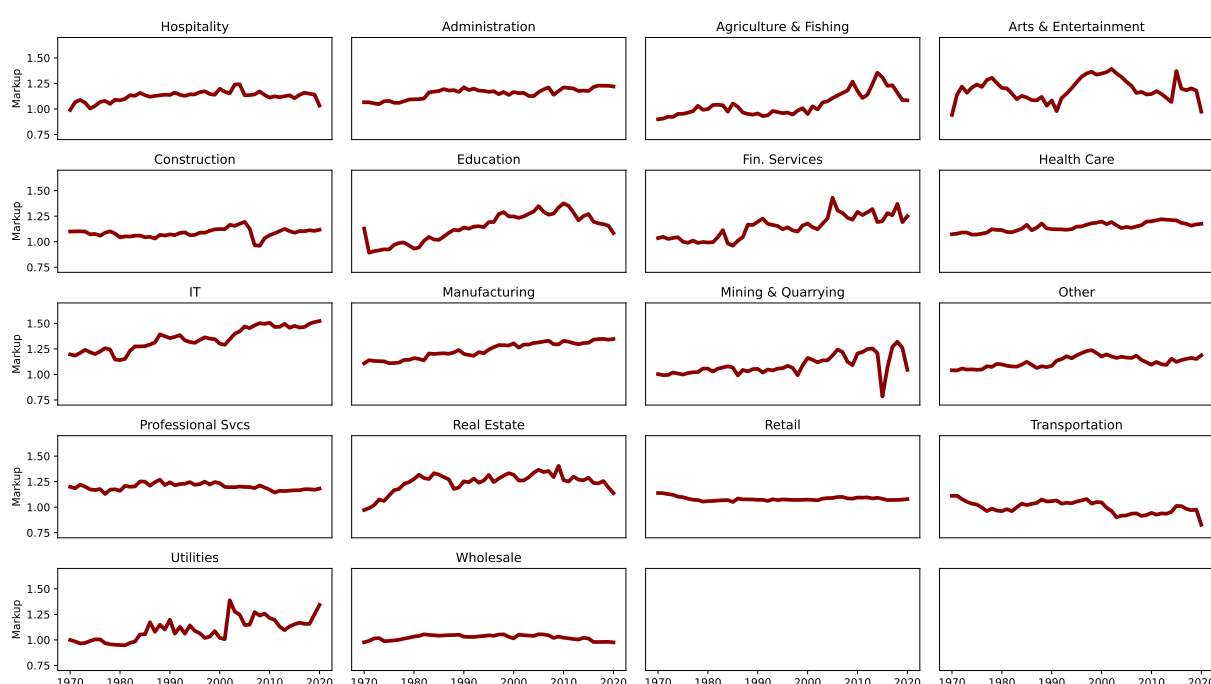
Figure 18 provides kernel densities of unweighted firm-level markups at different points in time. Clearly from this figure, the distribution of markups has become more spread over time, signaling an increase in variance. The flattening of the markup distribution is mostly driven by a fatter right tail. Figure 19 shows different percentiles of the distribution of firm-level markups, and how these have evolved over time. This figure reveals that the the aggregate markup has increased mostly because firms with markups above the median have sharply increased their markups since 1970. Interestingly, firms in the 10th percentile and below exhibit markups less than unity, which means they are setting prices below marginal costs. This may be the case if the are trying to liquidate inventories or engaging in predatory pricing.

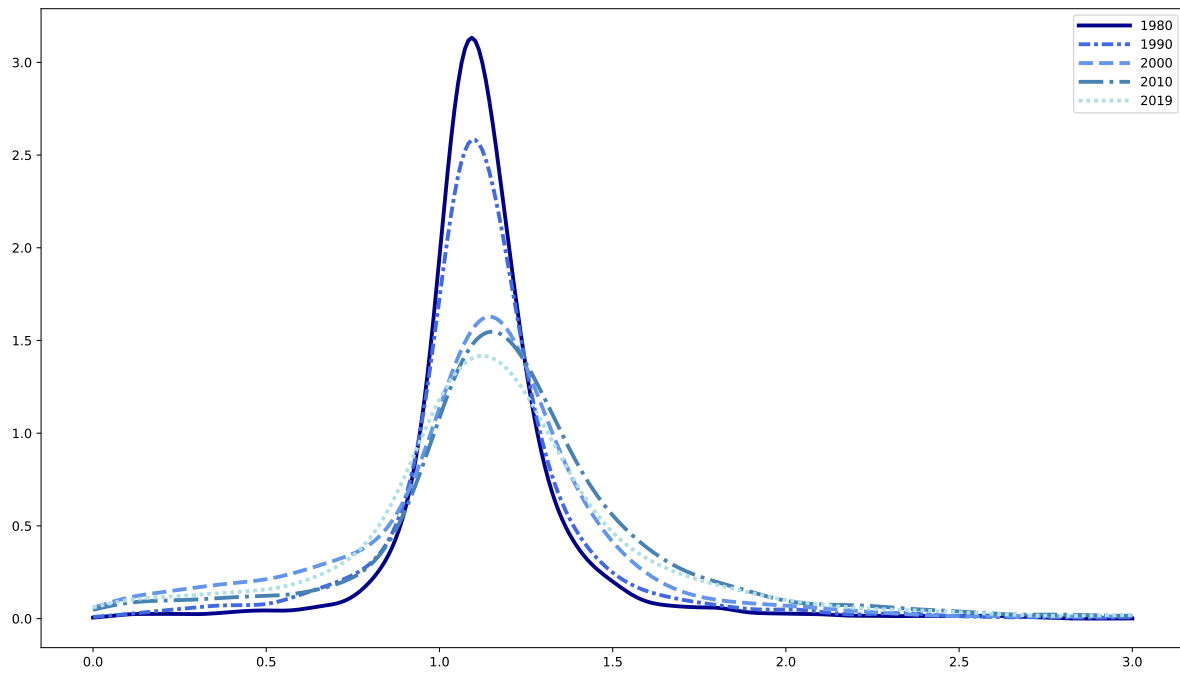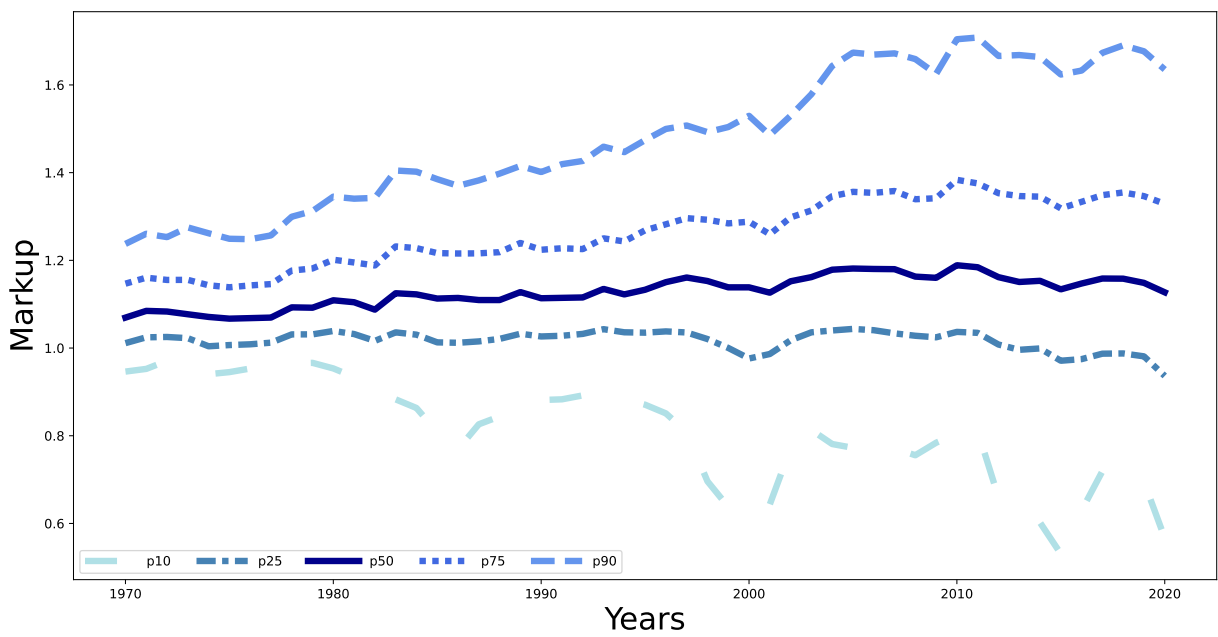**Figure 18:** Kernel Densities of Unweighted Firm-level Markups.

Figure 19: Percentiles of the Markup Distribution.

E.6 Concentration Measures

In this section, we report Herfindahl-Hirschman Index (HHI) measures of concentration, both at the national and the industry level, for the United States. We compute the HHI as the sum of squared sales shares. That is,

$$\text{HHI}_t = \sum_{i \in \mathcal{I}} \omega_{it}^2, \qquad \text{(National HHI)}$$

$$\text{HHI}_{jt} = \sum_{i \in \mathcal{I}_j} \omega_{it}^2, \qquad \text{(Industry-}j\text{ HHI)}$$

where ω_{it} is the sales share of producer i at time t (i.e., the sales of producer i over total sales), $\mathcal{I}$ is the set of all firms, and $\mathcal{I}_j$ is the set of all firms in industry j. Thus, total sales in the national HHI is the sum of sales for all firms in the country, whereas total sales in industry j's HHI is the sum of sales of all firms in that particular industry.

The national HHI for the United States, reported in Figure 20, indicates that market concentration has decreased from 1970 to 2020. More specifically, concentration has declined from 0.007 in 1970 to 0.004 in 2000, and then increased until 0.005 in 2020. To make sense of these numbers, we provide the interpretation of the HHI suggested by Smith and Ocampo (2022). Perhaps surprisingly, the probability that two dollars, chosen at random, were spent on the same firm was higher in 1970 than in 2020.

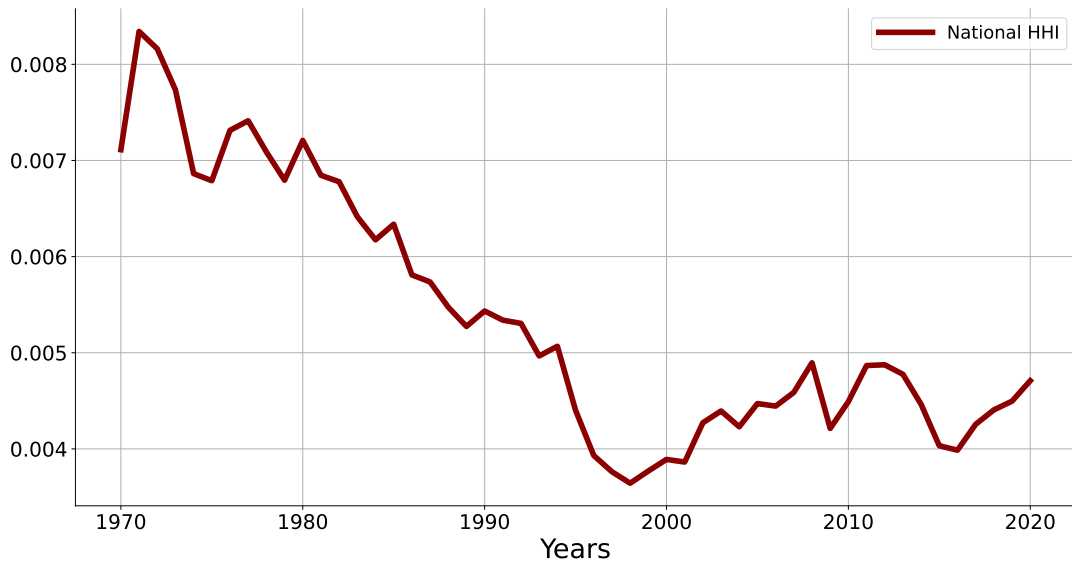**Figure 20:** National HHI in the United States, 1970–2020.

Figure Notes. The Herfindahl-Hirschman Index (HHI) is the sum of squared sales weights. Data comes from US Compustat.

Figure 21 shows the HHI for different US industries. While market concentration has increased in some industries, it has declined in the majority of industries: mining, manufacturing, transportation, IT, real estate, professional services, administration, education, arts & entertainment, and hospitality. Consistently with Smith and Ocampo (2022), we find that market concentration, as measured by sales, has sharply increased since the 1980s in the retail sector, which appears to be the exception to the norm.

Figure 21: Industry-level HHI in the United States, 1970–2020
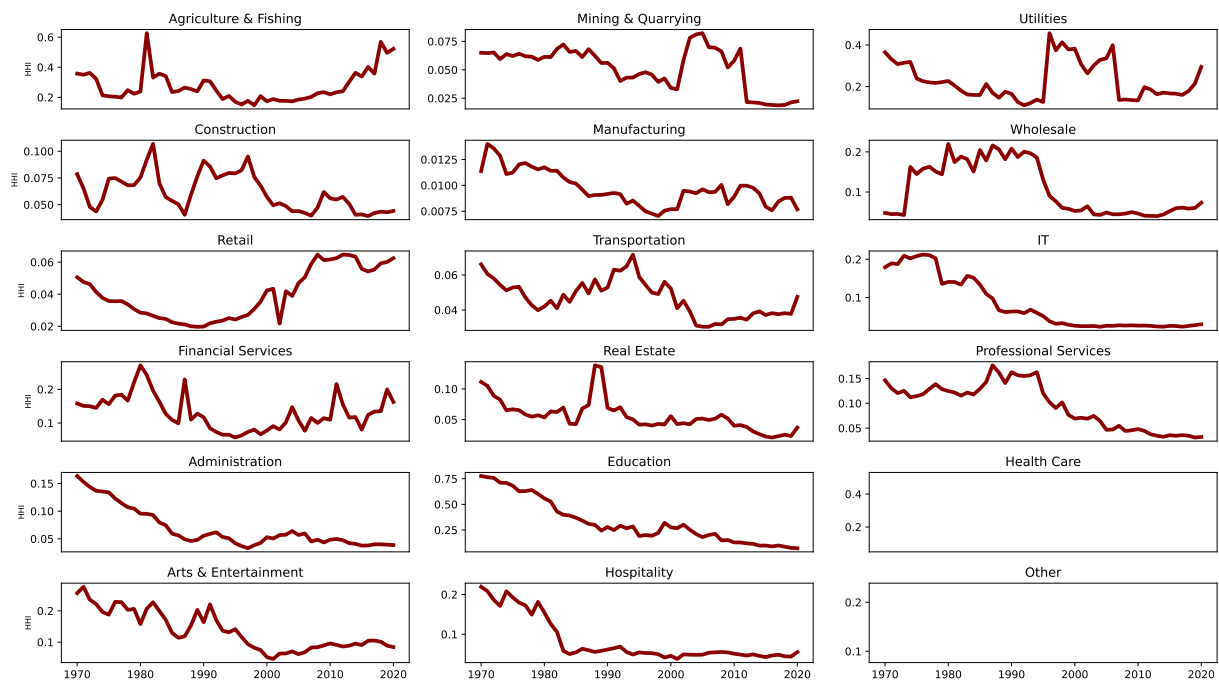
E.7 Income Shares

In this section, we discuss the implications of our profit share estimates for the capital share. Under the representativeness assumption of Compustat firms, we compute the labor share from the National Income and Product Income Accounts as paid compensation of employees (GDICOMPA) divided by GDP. We then compute the capital share as one minus the sum of the labor- and the micro-aggregated profit share. The results are displayed in Figure 22.

Figure 22: Income Shares in the United States, 1970–2020.

Taking the labor share as given and using our micro-aggregated profit share to back out the capital share, we conclude that factor shares are relatively stable over time in line with the famous Kaldor facts. The labor share decreased about 6 percentage points from 58% of GDP in 1970 to 52% in 2013, and then rebounded to 55% in 2020. Thus, over the entire sample period, we find that the labor share has decreased by about 3 percentage points. Given our estimate of a constant profit share at around 18% of GDP, we find that the capital share has increased from 24% to 27% of value added.

E.8 Robustness: No Intangible Capital

Figure 23: Aggregate Markup, Aggregate Returns to Scale, and its Covariance
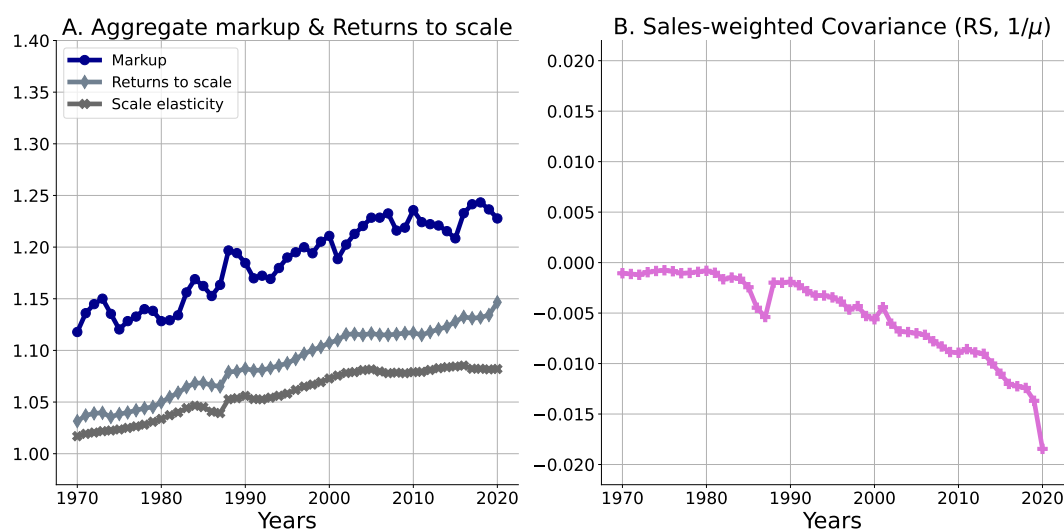

Figure Notes. Firm-level data comes from US Compustat, 1970–2020. In contrast to our benchmark results, where the capital stock is a composite of physical- and intangible capital, here we only consider the stock of physical capital.

Figure 24: The Micro–Aggregated Profit Share in the United States, 1970–2020.
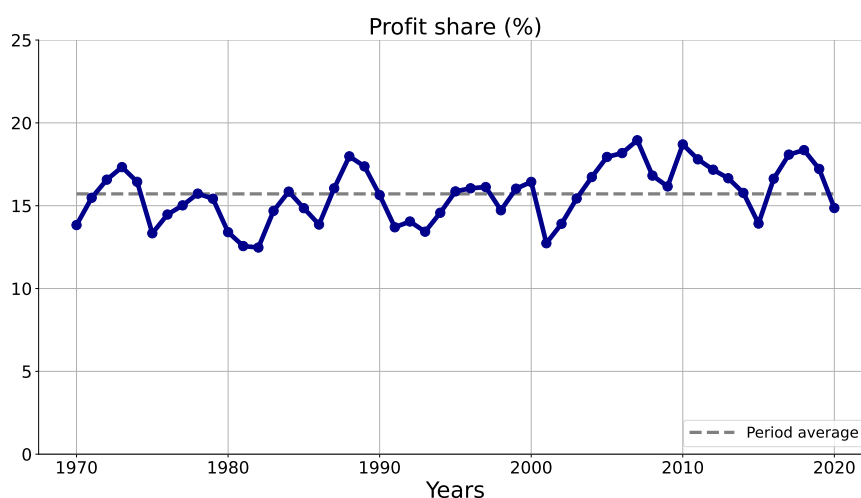

Figure Notes. Firm-level data comes from US Compustat, 1970–2020. In contrast to our benchmark results, where the capital stock is a composite of physical- and intangible capital, here we only consider the stock of physical capital.