The Micro–Aggregated Profit Share[*]

Thomas Hasenzagl
FRB of Richmond

Luis Pérez
SMU

This version: April 13, 2026.
First version: May 12, 2023.

Abstract

How much has market power increased in the United States in the last sixty years? And how did the rise in market power affect aggregate profits? Using micro-level data from U.S. Compustat, we find that several indicators of market power have steadily increased since 1960. In particular, the aggregate markup has gone up from roughly 10% of price over marginal cost in 1960 to 25% in 2020, and aggregate returns to scale have risen from about 1.03 to 1.15. 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 16% of GDP because the increase in monopoly rents (associated with increasing markups) 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.

[*]We are indebted to Fil Babalievsky, V.V. Chari, Jose-Elias Gallegos, Eugenia Gonzalez–Aguado, Fatih Guvenen, Kyle Herkenhoff, Larry E. Jones, Loukas Karabarbounis, Rishabh Kirpalani, Hannes Malmberg, Nathan Miller, Simon Mongey, Sergio Ocampo, Chris Phelan, Conor Ryan, Matt Rognlie, and James Traina for helpful comments and suggestions. This paper was conduced while the authors were Special Sworn Status researchers of the U.S. Census Bureau. Any views expressed are those of the authors and not those of the U.S. Census Bureau or the Federal Reserve System. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 3280. (CBDRB APPROVAL No. TBA). This was Pérez's JMP in AY 2023/24. Emails: thomas.hasenzagl@gmail.com, luisperez@smu.edu.

1 Introduction

How much has market power increased in the United States over the last sixty 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 1960. We document that the aggregate markup, measured as the ratio of price to marginal cost, has gone up from 10% in 1960 to 25% in 2020 and that aggregate returns to scale have risen from 1.03 to 1.15. 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. 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, that is, how profits propagate from downstream sellers to upstream suppliers.

By connecting indicators of aggregate market power to the profit share, we can not only quantify the profit share exploiting rich micro-level heterogeneity, but also understand its determinants. Studying the profit share is important for several reasons. First and foremost, profits, not markups, are the policy-relevant object for antitrust. Positive profits generally entail a loss of consumer surplus, while a markup above one, even if associated with a deadweight loss relative to perfect competition, may have a positive welfare effect. Thus, the markup is less informative for antitrust policy than the profit share is.[1] Leaving aside normative considerations, understanding the evolution of the profit share may also shed light on the decline of the labor share, the fall in business dynamism, the TFP slowdown, and other macroeconomic trends.

Our main goal is to understand whether the rise in market power that we document for the U.S. economy translated into higher profits. 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 1960 and 2020. Over that period, monopoly rents increased from 14% to 33% of GDP, while the profit share remained roughly constant at 16% of GDP. We reconcile the increase in market power with the stability of the profit share by showing that the rise in monopoly rents associated with higher markups was counteracted by a proportional increase in returns to scale. Rising returns to scale reflect changes in technology and increases in fixed costs, including overhead expenditures on labor and capital, such as accounting and advertising, as well as regulatory compliance costs.

[1] To see this, notice that in the presence of fixed costs, a firm that cannot charge a markup at least as large as price over average cost would exit the market, destroying all consumer surplus. Hence, a price cap that limits the ability of firms to price above marginal cost can have detrimental welfare effects. On the other hand, a regulation that leads firms to make zero profits (i.e., to price at average cost) may be welfare enhancing, as it is easy to show that for a given technological environment, under general conditions, consumer surplus is highest when the profit share is lowest.

While our Compustat analysis provides economy-wide evidence on the evolution of market power and profitability, it does not allow us to separate rents arising from output-market power from those arising from input-market power. To address this limitation, we complement the Compustat analysis with establishment-level data from the U.S. Census Bureau for the manufacturing sector. Because these data contain richer information on establishments' expenditures, they allow us to distinguish monopoly from monopsony rents and to decompose manufacturing profits into these two sources. We find that, in manufacturing, roughly XX% of profits are attributable to monopoly power and YY% to monopsony power in labor markets. This exercise serves a dual purpose: it provides direct evidence on the relative importance of output- and input-market power in shaping profitability, and it allows us to assess the representativeness of Compustat within manufacturing. Taken together, the two empirical exercises suggest that Compustat is well suited to document broad economy-wide trends in market power and profitability, while the Census data are essential for identifying the underlying sources of those rents.

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.[2,3] Rising fixed costs increase entry barriers.[4] Monopoly rents are only sustainable in the presence of entry barriers since such rents will otherwise attract entrants and 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 increasing fixed costs, and profits remain unchanged.

Our empirical findings have subtle implications for policymakers. The rise in monopoly rents reflects an increase in market power that could, in principle, be counteracted by more aggressive antitrust enforcement. Yet if firms must earn sufficient rents to cover high fixed costs, overly aggressive enforcement of antitrust law may make it unsustainable for some firms to operate, reducing firm dynamism and, paradoxically, leading to less competition and greater market power. At the same time, to the extent that part of aggregate rents reflects monopsony power rather than monopoly power, the policy implications extend beyond product-market antitrust to include labor-market regulation and competition policy in other input markets.

[2]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).

[3]Business Dynamics Statistics from the U.S. Census Bureau indicate that entry and exit rates of establishments and firms have both declined since 1980. See also Decker et al. (2016, 2014).

[4]De Ridder (2019), Aghion et al. (2019), Gutiérrez et al. (2021), De Loecker et al. (2022), Hsieh and Rossi-Hansberg (2023), and others have emphasized the important role of rising fixed- and entry costs.

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).[5] Second, we 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 aggregate profit share can be expressed in terms of aggregate market-power indicators, including the aggregate markup, a monopsony term, aggregate returns to scale, and a sufficient statistic for production networks that captures double marginalization. Together, these results provide a unified framework for measuring aggregate profits from micro data and for decomposing those profits into monopoly, monopsony, scale, and network effects.

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 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 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.[6] We emphasize economic profits to distinguish them from accounting profits. Economic profits, unlike accounting profits, are unobserved and exclude 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 the existence of a production technology that satisfies standard regularity conditions, including 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 above marginal cost—that is, if its markup is greater than unity—or if it compensates factors of production, such as workers, at a rate below their marginal revenue product—that is, if its markdowns

[5]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.

[6]A similar profit rate expression is provided by Basu and Fernald (1997, 2002), Basu (2019) and Syverson (2019). Our expression is more general than theirs because we allow for fixed costs and market power in factor markets.

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, 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. Our formula presents several advantages over existing ones 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? The Census analysis is particularly useful in answering these last two questions because it allows us to separately identify the contribution of output- and input-market power, whereas Compustat alone does not.

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 E, we use our theorem to elucidate the back-and-forth discussion between De Loecker, Eeckhout and Unger (2020) and Basu (2019) on whether the micro-level estimates of the former have unreasonable macroeconomic implications. Under additional assumptions on technology, our theorem can also be used to calibrate economic models with monopolistic or monopsonistic wedges, for example by obtaining a markup shock series for a New Keynesian model.

Related Literature

Our paper lies at the intersection of empirical industrial organization and the literature on the functional distribution of income. Its central question is how firm-level market-power—both monopoly and monopsony—maps into the aggregate profit share. It also connects to recent work on production networks through the role of input-output linkages in aggregation.

Our paper relates most directly to the growing literature on market power and macroeconomics. An influential contribution in this literature is De Loecker, Eeckhout and Unger (2020), who estimate markups in Compustat data and conclude that the aggregate markup rose sharply in the United States. While they find that price over marginal cost increased from 20% in 1980 to 60% in 2016, we find a substantially smaller increase, from 8% to 17% over the same period. There are three reasons why our series differ. The main reason is that we aggregate firm-level markups using the harmonic sales-weighted average implied by our framework, whereas they use the sales-weighted average. In our setting, the harmonic aggregation has a clear theoretical foundation.[7] A second reason is that we make different assumptions about variable costs. In particular, we classify both costs of goods sold and part of selling, general, and administrative expenses as variable costs, motivated by the analysis in Traina (2018).[8] A third reason, which matters less quantitatively, is that we measure capital using both physical and intangible capital.[9] More fundamentally, however, our focus is not the aggregate markup per se, but the aggregate profit share and its decomposition into monopoly power, monopsony power, returns to scale, and production-network effects. This distinction is important because rising markups need not imply higher aggregate profits when the increase in monopoly rents is offset by rising fixed costs.

Our paper also contributes to empirical industrial organization. A large literature has documented substantial heterogeneity in the evolution of market power across U.S. industries: markups have risen in settings such as cement, wholesale retail, brewing, and consumer packaged goods, but have declined in others, including automobiles and steel.[10] Researchers typically quantify market power using either demand systems or

[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 treatment of variable costs, but estimates of the profit share are not, provided all costs are ultimately 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). In our empirical analysis, intangible capital is incorporated into the firm's capital stock. Although markups are identified from the elasticity of the variable input, not from the elasticity of capital, including intangibles may still affect markup estimates indirectly through the estimation of the production function.

[10]See, for example, Miller et al. (2022), Ganapati (2021), De Loecker and Scott (2022), Brand (2021), Döpper et al. (2023), Grieco et al. (2023), and Collard-Wexler and De Loecker (2015).

production-function approaches, depending on the question and the available data.[11] We do not take a stand on one empirical strategy over the other. Instead, our contribution is to provide a common theoretical framework that links producer-level estimates of market power to the aggregate profit share. Conditional on data availability, the framework can be combined with either demand- or production-function methods.

Our paper also relates to the literature on the functional distribution of income. Income shares are important summary statistics for understanding many macroeconomic phenomena, including the global decline in the labor share (Elsby, Hobijn and Şahin, 2013; Karabarbounis and Neiman, 2014; Kehrig and Vincent, 2021), the fall in business dynamism (Decker, Haltiwanger, Jarmin and Miranda, 2014), the TFP slowdown (Gordon, 2012), the stability of the Kaldor facts (Eggertsson, Robbins and Wold, 2021), and economic inequality (Atkinson, Piketty and Saez, 2011; Piketty and Saez, 2003).[12] One such statistic that has historically attracted the attention of both economists and policymakers is the profit share. We contribute to this literature by showing how the aggregate profit share can be constructed consistently from micro data and how it relates to labor and capital shares within a unified framework. In particular, when micro data are representative and capital costs are measured consistently, aggregating firm-level profit rates using Domar weights yields the same aggregate profit share as the standard macroeconomic approach. This connection makes it possible to study the role of profits in changes in the functional distribution of income using micro data while preserving consistency with national accounts. In our application, the decline in the labor share is consistent with the national accounts, while the relative stability of the profit share implies that most of that decline is accounted for by a rise in the capital share rather than by higher profits.

Our framework also connects to the literature on production networks through the role of input-output linkages in aggregation. Research on production networks, building on the classic input-output tradition, has grown rapidly in recent decades.[13] 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, generalizes the production side of that framework. One can show that the profit share in their economy can be written as the ratio of sales to GDP times one minus the inverse of

[11] For classic demand-estimation approaches, see Bresnahan (1987) and Berry, Levinsohn and Pakes (1995). For production-function approaches, see Hall (1988) and De Loecker and Warzynski (2012). Collard-Wexler and De Loecker (2015) find that the two approaches deliver similar markup estimates in the U.S. brewing industry.

[12] Income shares are also important for calibrating economic models in macroeconomics, growth, development, and monetary economics (see, e.g., Boppart et al., 2023; Gali and Monacelli, 2005).

[13] For a non-exhaustive list, see Long and Plosser (1983), Acemoglu et al. (2016, 2012), Grassi (2017), Bigio and La'o (2020), and Baqaee and Farhi (2019, 2020).

the harmonic sales-weighted markup. Thus, our theorem nests their profit share when there is no monopsony, no fixed costs, and constant returns to scale for each producer.

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 presents the economy-wide evidence from U.S. Compustat. Section 4 presents the manufacturing evidence from the U.S. Census Bureau and uses it to decompose profits into monopoly and monopsony rents and to compare Census and Compustat within manufacturing. Section 5 concludes. Appendices A–G provide additional results, omitted proofs, and robustness exercises.

2 Theoretical Framework

In this section, we develop the structural foundations that connect micro-level market power to the aggregate profit share. We begin by showing that the aggregate profit share can be constructed from producer-level profit rates. We then define economic operating profits, including the user cost of predetermined inputs, characterize profit rates in terms of markups, markdowns, and returns to scale, and finally aggregate these objects to obtain the decomposition used in the empirical analysis.

2.1 The Aggregate Profit Share: Macro and Micro Approaches

The aggregate profit share can be constructed either from aggregate accounting data or from producer-level information. The macro approach derives profits residually from aggregate value added in the National Accounts. The micro approach builds the profit share from the bottom up by aggregating producer-level profit rates.

The Macro Approach. The macro approach rests on the national income identity that decomposes aggregate value added into measured payments to factors of production and residual income:

$$\text{GDP} = \sum_f W_f L_f + \text{residual income},$$

where L_f is the employed quantity of factor f and $W_f L_f$ its measured compensation.

In most applications, one works with two factors, labor and capital, and writes

$$\text{GDP} = WL + RK + \text{residual income},$$

where WL denotes labor compensation and RK denotes imputed (gross-of-depreciation) capital rents. The residual includes profits as well as production taxes, tariffs, and other regulatory distortions. It may also reflect measurement error arising from mismeasured capital services, incorrect user-cost imputation, or misclassifications of labor income.[14]

Under the maintained assumption that monopoly and monopsony are the only economically relevant distortions, residual income coincides with economic profits:

$$\text{GDP} = WL + RK + \text{Profits}.$$

In empirical implementations, it is standard practice to net out production and sales taxes from value added—or equivalently to normalize GDP appropriately—so that factor shares and the residual reflect economic returns rather than tax revenues.

The aggregate profit share is therefore

$$\Lambda_{\Pi}^{\text{Macro}} := \frac{\text{Profits}}{\text{GDP}} = 1 - \Lambda_L - \Lambda_K, \tag{1}$$

where Λ_L and Λ_K denote the labor and capital shares.

Empirically, existing measures of the aggregate profit share follow this procedure: aggregate value added, labor compensation, and capital stocks are obtained from the National Accounts; a user cost of capital is imputed or estimated to construct the capital share; and the profit share is inferred residually using equation (1).

The Micro Approach. We now construct the same object from producer-level data. Let producer $i \in \mathcal{I}$ earn profits π_i and have profit rate s_{π_i}. The aggregate profit share can then be written as

$$\Lambda_{\Pi}^{\text{Micro}} = \sum_{i \in \mathcal{I}} \omega_i s_{\pi_i}, \tag{2}$$

where ω_i denotes producer i's weight.

Expression (2) accommodates several aggregation schemes, depending on how profit rates and weights are defined. Different definitions of the profit rate naturally call for different weighting schemes. In most micro-level datasets, value added is not observed at the producer level, whereas revenues are. It is therefore convenient to define profit rates as profits over sales and to weight them by producer sales relative to aggregate value added—that is, by Domar weights. The following lemma shows that Domar aggregation recovers the macro profit share defined in equation (1).

[14] Karabarbounis and Neiman (2019) refer to a closely related residual as "factorless income."

Lemma 1 (**The Micro–Aggregated Profit Share**). *If producer-level profit rates are defined as profits over sales and aggregated using Domar weights, that is, producer sales divided by aggregate value added, then the micro and macro approach both yield the aggregate profit share:*

$$\Lambda_{\Pi}^{Micro} = \Lambda_{\Pi}^{Macro}.$$

Proof. Let $\Pi = \sum_{i \in \mathcal{I}} \pi_i$ denote aggregate profits. If $s_{\pi_i} = \pi_i/(p_i y_i)$, where $p_i y_i$ are sales,

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

□

Interpretation and scope. Domar weights are "summary statistics" for input-output linkages that capture the influence of each producer on aggregate value added. In an economy without intermediate goods, sales coincide with value added, and Domar weights reduce to sales shares, which sum to one. In that case, the aggregate profit share equals the sales-weighted average profit rate.

With production networks, gross sales exceed value added because intermediate transactions are recorded multiple times along production chains. Domar weights hence differ from sales shares and generally sum to more than one. They capture both a producer's direct contribution to aggregate value added through final sales and its indirect contribution through intermediate inputs embodied in downstream production. This distinction matters because profits earned by upstream producers are embedded in the costs of downstream producers and therefore do not appear in downstream profit rates. Aggregating profit rates using simple sales shares would therefore understate aggregate profits. Domar weights internalize these input-output linkages and ensure that each producer's total contribution to aggregate profits is properly accounted for.

Three remarks clarify the scope of Lemma 1. First, the result is purely an accounting identity. It does not rely on assumptions about pricing behavior, market structure, or technology. The equivalence between macro and micro approaches follows directly from the definition of aggregate profits and value added. Second, as mentioned earlier, alternative aggregation schemes are possible. If profit rates are defined as profits over value added rather than sales, then value-added weights recover the aggregate profit share. We present this alternative formulation as Lemma 5 in Appendix B.1. Third, although stated economy-wide, the result applies at any level of aggregation—industry, region, or firm group—provided that value added is defined consistently at that level.

2.2 Economic Profits and the User Cost of Predetermined Inputs

Lemma 1 shows that the aggregate profit share can be expressed as a Domar-weighted sum of producer-level profit rates. To connect aggregate profits to market power, we must therefore characterize the economic determinants of those profit rates. We begin by defining economic operating profits and the user cost of predetermined inputs.

Economic environment. Consider a producer with technology

$$y_t = F_t(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t),$$

where $\boldsymbol{x}_t = \{x_{jt}\}_{j\in\mathcal{N}}$ denotes variable inputs chosen within the period and $\boldsymbol{z}_t = \{z_{kt}\}_{k\in\mathcal{Z}}$ denotes predetermined or installed inputs. Variable inputs may include labor, materials, or service flows, while predetermined inputs may include physical capital, intangible capital, organizational capital, or other quasi-fixed factors. The productivity shifter A_t may incorporate both producer-specific and aggregate components. Throughout, we assume that the production function F_t is continuously differentiable, strictly increasing, and quasiconcave, and may exhibit arbitrary returns to scale. More generally, F_t should be thought of as an output constraint describing the set of feasible input-output combinations in a given period.[15]

We allow producers to exercise market power in both output and input markets. Output-market power is summarized by the markup, defined as the ratio of price to marginal cost. Input-market power is summarized by markdowns, defined as the ratio of an input's rental rate to its marginal revenue product. These wedges serve as reduced-form sufficient statistics for market power and will be the key objects linking producer-level profit rates to the decomposition developed below.

Profit concepts. For our purposes it is essential to distinguish accounting profits from economic operating profits. Firm-level datasets report accounting measures of income, but those objects need not coincide with economic profits because they do not generally deduct the full opportunity cost of predetermined factors. As emphasized by Fisher and McGowan (1983), accounting depreciation and related capital charges need not measure the true economic cost of using installed capital.

[15]This interpretation accommodates environments in which realized output may reflect richer dynamic or strategic considerations, including repeated-game settings in the spirit of Abreu (1986).

Accounting profits are defined as revenues minus recorded operating expenses and accounting depreciation:

$$\pi_t^{\text{acct}} = p_t y_t - \sum_{j \in \mathcal{N}} w_{jt} x_{jt} - \sum_{k \in \mathcal{Z}} \delta_{kt}^{\text{acct}} z_{kt} - \text{FC}_t,$$

where $\delta_{kt}^{\text{acct}}$ denotes accounting depreciation rates and FC_t fixed operating costs. Because accounting depreciation reflects historical cost allocation rules rather than economic opportunity costs, accounting profits need not measure the return to predetermined inputs at their true opportunity cost.

The object relevant for the profit share is instead *economic operating profits*, defined as revenues net of the opportunity cost of all inputs:

$$\pi_t = p_t y_t - \sum_{j \in \mathcal{N}} w_{jt} x_{jt} - \sum_{k \in \mathcal{Z}} r_{kt} z_{kt} - \text{FC}_t,$$

where r_{kt} denotes the user cost of predetermined input z_{kt}. This is the within-period shadow rental of installed input services: the reduction in minimum variable expenditure generated by an additional unit of the installed input, holding output fixed. Hence, $r_{kt} z_{kt}$ is the within-period opportunity cost of employing the observed stock of installed input z_{kt} in production during the period. Economically, it is the return the producer foregoes by allocating resources to this input rather than to alternative uses.

Although the user cost is not directly observed, it can be recovered from a restricted cost-minimization problem conditional on installed stocks. The key restriction for the analysis below is that, conditional on the realized state and installed stocks, producers choose variable inputs to minimize expenditure for a given level of output. This delivers the within-period shadow rental of installed inputs that is relevant for profit measurement. Appendix A.5 provides the formal derivation and shows that this object remains the relevant one when installed inputs are chosen in a dynamic environment with adjustment costs, discounting, and risk premia. In this sense, the user cost characterized as a static shadow rental reflects those forces because it is recovered empirically using the observed stock, which is itself the outcome of dynamic optimization.

2.3 Economic Profit Rates in Terms of Monopoly and Monopsony

We now characterize the producer-level economic profit rate, $s_{\pi_t} := \pi_t/(p_t y_t)$, in terms of output-market power, input-market power, and returns to scale. In particular, the profit rate can be expressed in terms of three components: the markup of price over marginal cost, the scale elasticity of the production function adjusted for fixed costs, and a monopsony term capturing factor-market distortions.

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

$$s_{\pi_t} = 1 - \frac{RS_t}{\mu_t} = 1 - \frac{SE_t^{adj}}{\mu_t} + \frac{\mathcal{M}_t}{\mu_t}, \tag{3}$$

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

$$RS_t = SE_t^{adj} - \mathcal{M}_t.$$

The adjusted scale elasticity is given by

$$SE_t^{adj} := SE_t \times \left(\frac{TC_t}{TC_t - FC_t}\right), \tag{4}$$

where $SE = \sum_{j\in\mathcal{N}} \theta_j$ is the scale elasticity of the production function, TC denotes total costs, and FC are fixed operating costs.

The monopsony term is given by

$$\mathcal{M}_t := \left(\frac{TC_t}{TC_t - FC_t}\right) \sum_{j\in\mathcal{F}} \theta_{jt}(1 - \nu_{jt}), \tag{5}$$

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 $j \in \mathcal{F}$, 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 B.2. □

Remarks. Equation (3) shows that a producer's profit rate s_π depends on three objects: its markup μ, the scale elasticity adjusted for fixed costs SE^{adj}, and a monopsony term $\mathcal{M}$ that depends on markdowns ν_j. Equivalently, it can be expressed in terms of the markup and the returns to scale, RS, the ratio of average to marginal cost.

Profits arise from two sources. First, monopoly power allows the producer to charge a markup above marginal cost. In the absence of monopsony power, profits arise when the markup is sufficiently high relative to the returns to scale or, equivalently, to the scale elasticity adjusted for fixed costs. Second, monopsony power generates rents by depressing factor payments relative to marginal revenue products. These rents contribute to profits to the extent that they offset the pressure of fixed costs and scale.

Equation (3) also clarifies the relationship between our formulation and the standard benchmark in the literature. Earlier work often writes the profit rate as

$$s_{\pi_t} = 1 - \frac{\mathrm{RS}_t}{\mu_t}, \tag{6}$$

as Basu and Fernald (1997, 2002), Basu (2019), and Syverson (2019), among others. In those environments, however, RS coincides with the scale elasticity SE, because fixed costs and monopsony distortions are ruled out by assumption. In our notation, when $\mathrm{FC} = 0$ and $\mathcal{M} = 0$, we have

$$\mathrm{RS}_t = \mathrm{SE}_t,$$

so the standard expression is recovered as a special case. Under constant returns to scale, the profit rate reduces further to

$$s_{\pi_t} = 1 - \frac{1}{\mu_t}.$$

For this reason, it is useful to distinguish between technological scale elasticity and the object that enters the profit-rate formula. In general, the relevant object is not the production-function scale elasticity itself, but the ratio of average cost to marginal cost. These two objects coincide only under the benchmark assumptions that eliminate fixed costs and factor-market distortions. Once fixed costs are present, average cost exceeds variable cost, so the profit-rate formula depends on the adjusted scale elasticity rather than on the technological scale elasticity alone. Once monopsony distortions are present, factor payments no longer equal marginal revenue products, which introduces an additional wedge. In our formulation,

$$\mathrm{RS}_t = \mathrm{SE}_t^{\mathrm{adj}} - \mathcal{M}_t,$$

so both fixed costs and markdowns affect profits through their impact on the relationship between average and marginal cost.

An important clarification is that the adjusted scale elasticity incorporates fixed-cost requirements. Empirically, however, when observed expenditures combine variable and fixed components, as is often the case in the data, the object estimated and labeled "scale elasticity" may reflect both technological curvature and fixed costs.[16] When all fixed costs are observed, the purely technological scale elasticity can be estimated using variable inputs alone and identified separately from fixed-cost requirements.

[16] Appendix C shows that empirically estimated scale elasticities may conflate technological curvature with fixed-cost requirements when fixed and variable inputs are not separately observed.

2.4 Linking the Profit Share to Aggregate Indicators of Market Power

Our next result shows that the aggregate profit share can be expressed in terms of a sufficient statistic for production networks and two aggregate indicators of market power: the aggregate markup and an aggregate monopsony term capturing distortions in factor markets.

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*

$$\Lambda_{\Pi t} = \underbrace{\chi_t}_{\text{IO multiplier}} \times \underbrace{\left(1 - \frac{\overline{SE}_t^{adj}}{\overline{\mu}_t^{hsw}} + \frac{\overline{\mathcal{M}}_t}{\overline{\mu}_t^{hsw}} - Cov_\omega\left[SE_t^{adj}, \frac{1}{\mu_t}\right] + Cov_\omega\left[\mathcal{M}_t, \frac{1}{\mu_t}\right]\right)}_{\text{sales-weighted profit rate}}, \tag{7}$$

where $\chi = \sum_{k \in \mathcal{I}} \frac{p_k y_k}{GDP}$ denotes the input-output multiplier, SE^{adj} the scale elasticity adjusted for fixed costs given by (4), μ the markup, and $\mathcal{M}$ the monopsony term given by (5). 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 B.4. □

This theorem provides a bridge between micro estimates of market power and the aggregate profit share. Because market-power indicators are obtained from micro data, Theorem 1 can be used to assess the macroeconomic implications of producer-level estimates of markups, markdowns, and returns to scale for the aggregate profit share. The theorem also decomposes the profit share into its underlying sources, allowing us to assess the extent to which profits originate from monopoly and from monopsony.

An important clarification is that, although we state Theorem 1 in terms of the economy-wide profit share, the same logic applies at any level of aggregation. By quantifying the sources of market power—monopoly *vis-à-vis* monopsony—within particular sectors, empirical applications of Theorem 1 can inform industry-level measurement and antitrust analysis. In this sense, the theorem links macroeconomic measurement to industrial-organization analysis, echoing the call in Berry, Gaynor and Morton (2019) for industry-level evidence.

Equation (7) shows that the aggregate profit share depends on the input-output multiplier and the sales-weighted average profit rate, which in turn is shaped by the sales-weighted scale elasticity adjusted for fixed costs, the harmonic sales-weighted markup, the sales-weighted monopsony term, and covariance terms. The theorem therefore identifies the natural aggregate counterparts of markup, monopsony terms, scale elasticities, and returns to scale. We acknowledge that we are not the first to point to the harmonic sales-weighted markup as a natural choice for the aggregate markup. Baqaee and Farhi (2020) and Edmond, Midrigan and Xu (2023), among others, have emphasized this earlier. Our result shows, however, that this object remains appropriate in a more general environment with arbitrary production networks, fixed costs, non-constant returns to scale, and monopsony distortions.

Theorem 1 also nests a number of familiar benchmark cases, including environments without monopsony distortions, without fixed costs, without production networks, and with constant returns to scale. These special cases are useful for both relating our framework to existing results in the literature and for clarifying the assumptions under which one may infer aggregate profit shares from micro estimates of market power, or conversely back out aggregate markups from macroeconomic data. They also make clear that production networks are central for measurement and inference: in benchmark cases commonly used in practice, the aggregate profit share is not simply the sales-weighted average profit rate, but that object multiplied the input-output multiplier—that is, the ratio of total sales to GDP. Because the input-output multiplier is empirically sizable (around 2 in the U.S.), ignoring production networks would lead to a substantial understatement of the aggregate profit share. Appendix E uses these benchmark cases to elucidate the discussion in Basu (2019) and De Loecker et al. (2020) on how micro-levels estimates of market power should be mapped into the profit share.

3 Economy-Wide Evidence from U.S. Compustat

In this section, we describe the data sources and the methodology used to estimate markups and returns to scale, and to infer the profit share for the United States.

3.1 Data Sources

Our main data source is Compustat North America Fundamentals Annual, which we access through the Wharton Research Data Services (WRDS). Compustat provides balance-sheet and income-statement data for U.S.-incorporated publicly traded firms.

Sample selection. We download annual data for all available firms beginning in 1956 and ending in 2024, restricting the sample to industrial format (INDL), standard data format (STD), domestic population (D), and consolidated accounts (C). We deflate all financial variables using the GDP deflator from FRED (base year 2010) and classify firms according to two-digit NAICS industries, adapted to conform with the Bureau of

Economic Analysis (BEA) grouping. In particular, we group together two-digit NAICS industries 31–33, 44–45, and 48–49. We drop observations with missing NAICS codes or with missing or non-positive values for sales, cost of goods sold (COGS), or selling, general and administrative expenses (SG&A), since all three variables are required for our main specification. We also require firms to have positive physical capital and positive intangible capital. Following De Loecker et al. (2020), we trim observations with the sales-to-COGS ratio in the 1st and 99th percentiles within each year.

Capital measurement. We measure a firm's capital stock as the sum of physical and intangible capital. Physical capital is measured using Property, Plant, and Equipment (PPEGT). Intangible capital is taken from the Peters and Taylor Total Q dataset (WRDS), and is the sum of knowledge capital and organizational capital, where each of these stocks is constructed using the perpetual inventory method with R&D as the investment in knowledge capital and 30% of SG&A net of R&D as the investment in organizational capital.[17] To preserve continuity in firm-level capital series, we interpolate missing capital values within each firm's time series.

SG&A decomposition. A key issue is that Compustat's SG&A measure may include R&D expenditures. Accordingly, we first subtract R&D from SG&A and define SG&A net of R&D. Following Peters and Taylor (2017), we capitalize 30% of this SG&A net of R&D as organizational capital investment. Of the remaining 70%, we treat a fraction $\alpha \in [0, 1]$ as a variable operating expense and the remaining share, $1 - \alpha$, as a fixed overhead cost. Accordingly, our measure of operating expenses is defined as OPEX = COGS + $\alpha \times 0.7 \times$ (SG&A net of R&D). The fixed portion enters our fixed-cost adjustment factor to the scale elasticity discussed in Section 2.

In our baseline specification, we set $\alpha = 0.5$. The non-capitalized component of SG&A net of R&D likely includes expenditures that vary with scale, such as commissions, delivery expenses, and some marketing and selling costs, and overhead items such as administrative and managerial expenses, lease rentals, and other headquarters-related costs. This ambiguity is also reflected in the literature: Traina (2018) emphasizes that SG&A contains variable costs relevant for markup estimation, whereas De Loecker et al. (2020) treat SG&A as overhead in their baseline specification. Karabarbounis and Neiman (2019) note that Compustat's SG&A category includes items whose classification as variable or fixed is not clear-cut. We therefore view $\alpha = 0.5$ as a neutral benchmark. Reassuringly, this choice implies a residual labor share that closely aligns with NIPA, and our main results are robust to alternative choices for this parameter.

[17] The 30% capitalization rate in Peters and Taylor (2017) follows the BEA methodology of Corrado et al. (2009), which is standard in the intangible capital literature. 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).

Labor expenditures. Compustat reports staff expenses (XLR) for only a subset of firms. In our sample, approximately 18% of observations (23% of firms) report this variable, with coverage relatively stable over time but never exceeding 30%. This limited coverage prevents us from separately identifying labor costs within COGS, a limitation we return to when discussing the interpretation of our markup estimates.

Table 1 provides summary statistics.

Table 1: Summary Statistics from US Compustat, 1956–2024.

Variable	Definition	Mean	Median	Min	Max	Std. dev.
Sales	SALE	2,304	162	0	484,752	12,109
Cost of goods sold (COGS)	COGS	1,552	95	0	412,099	9,102
SG&A	XSGA − XRD	344	29	0	109,887	1,713
Operating expenditures (OPEX)	COGS + 35% of SG&A	1,673	110	0	416,264	9,509
Fixed costs (FC)	35% of SG&A	121	10	0	38,461	600
Capital	PPEGT + K_INT	2,900	151	0	679,022	15,873
Physical	PPEGT	1,730	64	0	524,956	11,290
Intangible	K_INT (P&T)	1,170	54	0	343,012	6,779
Investment	CAPX + XRD + 30% of SG&A	321	24	0	155,554	1,741
Physical	CAPX	163	7	0	59,312	1,040
R&D	XRD	55	0	0	63,275	516
Organizational	30% of SG&A	103	9	0	32,966	514
Observations		281,707				
Firms		21,456				

Notes: All variables in millions of 2010 dollars.

BEA and NIPA Tables. We complement the Compustat data with the economy-wide input-output multiplier, defined as the ratio of gross output to value added, constructed from the Bureau of Economic Analysis (BEA) input-output tables. As shown in Section 2, this multiplier converts sales-weighted aggregation into Domar-weighted aggregation, which is the theoretically appropriate scheme for constructing the aggregate profit share from micro-level data. For the industry-level analysis, we construct sector-specific multipliers using BEA data on gross output and value added.

The economy-wide multiplier is available annually for 1970–2020 and is remarkably stable over time, fluctuating around 1.8. We extend the series forward to 2022 using the growth rate of aggregate gross output relative to GDP from FRED. For years prior to 1970, we hold the multiplier fixed at its 1970 value of 1.77, given its limited variation over the observed sample, where it ranges from 1.74 to 1.90.

We also use data from the National Income and Product Accounts (NIPA), retrieved through FRED, to construct a benchmark measure of the labor share for comparison with our Compustat-based estimates. Following Karabarbounis and Neiman (2014), we compute the NIPA labor share as the ratio of compensation of employees to GDP net of taxes on production, subsidies, and proprietor's income.

Compustat coverage. A limitation of our analysis is that Compustat covers only publicly traded firms, which tend to be larger, older, and more capital-intensive than the average firm. This is a well-known feature of the data. To the extent that publicly traded firms differ systematically from the broader population of firms, our estimates should be interpreted as a proxy for the aggregate profit share rather than a direct measure of it. Despite this limitation, Compustat remains the most suitable data source for studying economy-wide trends in market power in the United States, since it provides broad sectoral coverage at annual frequency and includes measures of both physical and intangible capital. By contrast, although Census micro data are available, they do not offer the same long-run and economy-wide coverage, and consistent capital measures are not broadly available across sectors and years.

3.2 Estimation and Identification

A standard approach to estimating markups relies on the production function approach pioneered by Hall (1988). Hall's key insight is that, for a cost-minimizing producer facing a competitive market for a variable input j, the markup can be recovered from the ratio of the output elasticity of that input to its revenue share:

$$\mu = \frac{\theta_j}{\alpha_j}, \tag{8}$$

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

As shown in Proposition 1, when the producer also exerts monopsony power in factor markets, the same ratio recovers not the markup alone but the product μ_j/ν_j, where $\nu_j \in (0, 1]$ denotes the markdown on input j. Since Compustat does not separately report labor cost for most firms, we cannot decompose COGS into labor and materials and therefore cannot separately identify markdowns. Accordingly, the markups we recover from Compustat may reflect distortions in both product and input markets.[18]

Equation (8) makes clear that empirical markup estimation requires data on revenues and input expenditures, $\{py, p_j x_j\}$, as well as an estimate of the variable input's output elasticity θ_j. The central empirical challenge is to obtain credible estimates of the variable input's output elasticity, $\hat{\theta}_j$, since revenues and input expenditures are

[18] More precisely, $\theta_j/\alpha_j = \mu/\nu_j$, so what we call the markup is an upper bound on the true output-market markup when firms exert monopsony power ($\nu_j < 1$). In Section 4, we use Census manufacturing data, which report input expenditures in greater detail, to decompose this wedge into markup and markdown components.

often observed. In this respect, the literature has benefited from the work of economists who have developed specialized econometric methods tailored to this purpose.[19]

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

$$\log y_{it} = \theta_\ell \log \ell_{it} + \theta_k \log k_{it} + \omega_{it} + \varepsilon_{it}, \tag{9}$$

where (θ_ℓ, θ_k) are output elasticities, ω_{it} is firm-level productivity observed by the producer 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.

Because firms choose inputs with knowledge of productivity, estimating this equation by OLS would generally suffer from simultaneity bias. To address this problem, we follow the control-function approach of Olley and Pakes (1996), with the Ackerberg, Caves and Frazer (2015)'s correction.[20] In the first stage, we approximate the conditional expectation of output using a flexible function of the state, free, and proxy variables, thus removing measurement error and the unexpected component of productivity. In the second stage, we impose a law of motion for productivity and estimate the production-function parameters using appropriate moment conditions via GMM.

Our baseline implementation assumes a Cobb–Douglas technology, allows elasticities to vary over time, and imposes common elasticities within two-digit NAICS industries by estimating the model separately for each industry. Elasticities are estimated using five-year rolling windows. In the first stage, we regress the outcome variable on a third-degree polynomial in the state, free, and proxy variables. In the second stage, we assume that productivity follows an AR(1) process and use lagged free variables and current state variables as instruments to form moment conditions. When estimating equation (9), we use the capital stock from the previous period, since Compustat records end-of-period capital assets. Table 2 summarizes our estimation procedure.[21]

[19] See, for instance, Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), De Loecker and Warzynski (2012), Gandhi et al. (2014), and Ackerberg et al. (2015).

[20] Ackerberg et al. (2015) show that the first-stage procedures in Olley and Pakes (1996) and Levinsohn and Petrin (2003) may suffer from a functional-dependence problem that prevents identification of the flexible-input 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.

[21] In our baseline specification, the capital stock includes both physical and intangible capital. Appendix F.5 shows that our main empirical results are robust to excluding intangibles.

Table 2: Estimation Details in the Application of the Control–Function Approach.

Technology	Cobb–Douglas
Elasticities	Industry specific, time-varying, 5-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	SALE
State	PPEGT + K_INT
Free	OPEX (= COGS + 35% of SG&A net of R&D)
Proxy	CAPX + R&D + 30% of SG&A net of R&D

Table Notes. Elasticities are estimated using 5-year rolling windows. That is, for year t, we use information from years $t-2, t-1, t, t+1, t+2$. Hence, to estimate elasticities for 1958–2022, we use data from 1956 to 2024.

Our estimation procedure yields elasticity estimates for each industry and year. We compute the scale elasticity at the industry-year level as the sum of input elasticities. We then recover firm-specific returns to scale by combining the industry-level scale elasticity with a firm-specific fixed-cost adjustment factor, which in the baseline specification treats 35% of SG&A net of R&D as overhead costs. We compute firm-level markups as

$$\mu_{it} = \frac{\hat{\theta}_{j\ell t}}{\alpha_{i\ell t}},$$

where $\hat{\theta}_{j\ell t}$ denotes the estimated elasticity of the variable input for industry j in year t. As discussed above, this object, although interpreted here as a "markup," it is a market-power wedge that may also reflect input-market distortions.

Revenue vs. Output Elasticities. Because only revenue data are available, we cannot estimate output elasticities directly and instead recover revenue elasticities. As first noted by Klette and Griliches (1996), using revenue elasticities generally biases the level of measured markups. In Appendix D, however, we show that although this issue may affect firm-level markup estimates, our estimates of profit rates and the aggregate profit share are unaffected. This appendix also discuss how the critiques in Doraszelski and Jaumandreu (2020), Bond et al. (2021) and De Ridder et al. (2022) bear on our results.

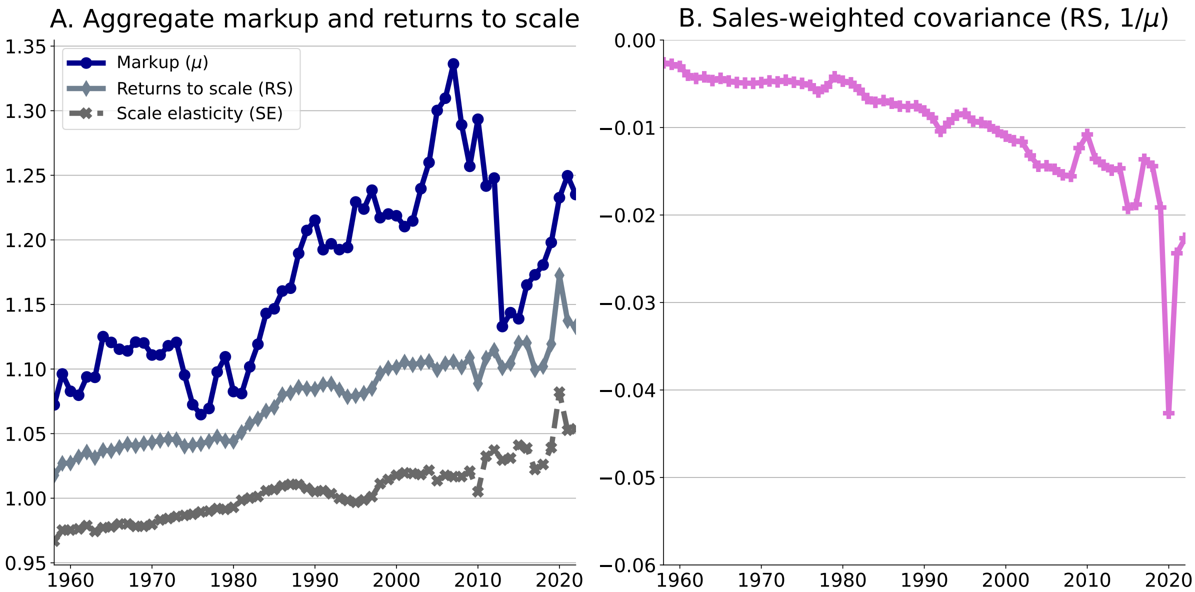
3.3 Aggregate Facts

We now use the the theoretical results of Section 2 and our micro-level estimates to construct the aggregate profit share for the United States. We proceed in five steps. We first document the aggregate behavior of markups, returns to scale, and their covariance.

We then map these objects into the profit share, and examine its implications for other income shares. Next, we decompose the profit share into its underlying components. Finally, we study the role of the user cost of capital in shaping aggregate profitability.

Aggregate markups and returns to scale. Figure 1 plots the time series of the components that determine the micro-aggregated profit share.[22] Panel A displays three series: the aggregate markup, aggregate returns to scale, and the aggregate scale elasticity. Panel B reports the covariance between firm-level returns to scale and inverse markups.

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



The aggregate markup has risen from roughly 1.10 in 1960 to 1.25 in 2020, that is, from about 10 to 25 percent of price above marginal cost. The series also displays a pronounced decline following the Great Recession, from 1.34 in 2007 to 1.13 in 2013, and then recovers gradually. The increase in markups that we document is meaningful but substantially more modest than in parts of the recent literature. In Appendix E, we discuss the sources of the discrepancy with De Loecker et al. (2020), and in Section 3.4 we examine the forces behind this secular increase in the aggregate markup.

Returns to scale also rise over the sample, from roughly 1.03 to 1.15. About 80 percent of this increase is explained by rising scale elasticities, and 20 percent by growing fixed costs. The scale elasticity rises from roughly 0.97 to 1.05, while the fixed-cost adjustment factor —that is, the ratio of total cost to variable costs—increases from 1.06 to 1.09, explaining the gap between returns to scale and the scale elasticity.

[22] As noted earlier, because Compustat reports staff expenses for only a limited subset of firms, we cannot separately identify monopsony power in labor markets. Thus, the markup series reported here may capture market power in both output and input markets.

Panel B shows that the covariance between returns to scale and inverse markups is slightly negative at the beginning of the sample and becomes more negative over time. This pattern implies that firms or sectors with higher returns to scale tend to charge higher markups. Economically, this pattern is intuitive: firms with larger fixed costs or increasing returns in production need higher markups to cover operating costs. The covariance term is also quantitatively important. A covariance of -0.02 contributes roughly four percentage points to the profit share once network effects are taken into account, as can be seen from equation (7). The strength of this amplification depends on the input-output multiplier—that is the ratio of sales to GDP—which in the U.S. economy is roughly stable at around two (see Appendix F.1).

The micro-aggregated profit share. Using the input-output multiplier together with the measures reported in the previous figure, we compute the micro-aggregated profit share, which we report in Figure 2 together with the profit share series estimated by Karabarbounis and Neiman (2019) using the macro approach.

Figure 2: The Profit Share in the United States: Micro vs. Macro Estimates.

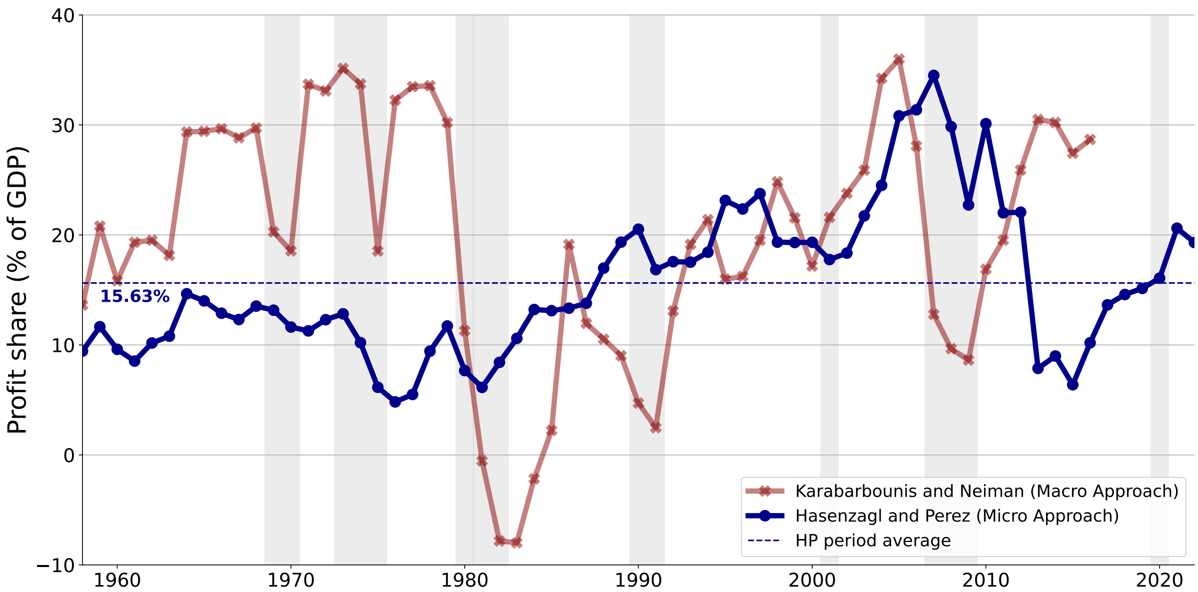


Figure Notes. The Hasenzagl–Perez profit share uses firm-level data from U.S. Compustat and the sales to GDP ratio from the BEA. The Karabarbounis–Neiman profit share reported here is a non-smoothed version of the "Case Π" reported in their paper and has been constructed using their replication files. They construct the profit share using the macro approach and NIPA data.

According to our estimates, the profit share in the United States has averaged around 16 percent of GDP over the past six decades, but this masks considerable variation: it was relatively stable around 10 percent through the 1980s, rose sharply in the 1990s and 2000s, peaking in 2007, and then declined to levels comparable to those at the beginning of the sample. Because of double marginalization, a profit share of 16 percent implies that the average firm in the United States has a profit rate of about 8 percent.

The aggregate profit share we estimate likely provides an upper bound for several reasons. First, Compustat covers only publicly listed firms, which are not representative of the full U.S. firm population and likely have higher profit rates than smaller private firms. Second, if fixed costs or intangible capital are underestimated, then the profit share will be overstated. As we discuss below, there are good reasons to believe that these components are indeed underestimated.

We argue that our estimates provide a more plausible measure of the profit share than the series of Karabarbounis and Neiman (2019), for two reasons. First, their series is extremely volatile. Second, it implies implausibly low capital shares. In particular, their estimates of factorless income in the 1970s or mid-2000s imply a capital share close to zero, which is difficult to reconcile with any reasonable view of aggregate production. Our series is more stable and, as we show below, it implies plausible capital shares. At the same time, our estimates agree with theirs on two important qualitative points. First, profit shares are procyclical. Second, both series are strongly negatively correlated with the aggregate user cost of capital. Following their approach, we compute this correlation. Using their series, we obtain a correlation of −0.83; using ours, the correlation is similarly large, at −0.85. As they note, these large negative correlations suggest that both series may capture some form of unmeasured capital.

The macroeconomic implications of the micro-aggregated profit share. Figure 3 shows the decomposition of valued added into four income shares—labor, physical capital, intangible capital, and profits—for both the Compustat sample (Panel A) and the aggregate economy using NIPA data (Panel B).

Figure 3: Income Shares in the United States.

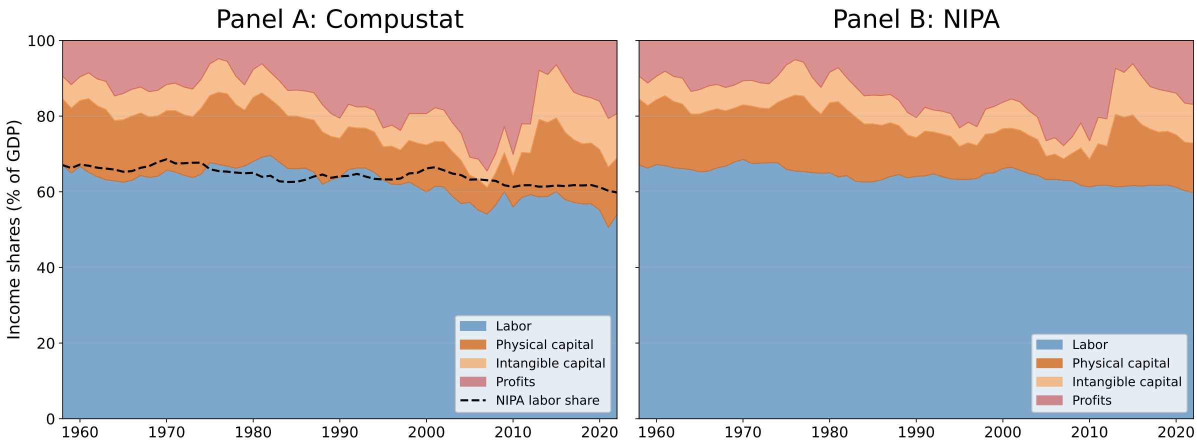


Figure Notes. Panel A decomposes value added into labor, physical capital, intangible capital, and profit shares using our production function estimates. The capital share is computed from the FOC-implied user cost of capital, split between physical and intangible capital in proportion to their shares of the total capital stock. The labor share is obtained as a residual. The dashed line shows the NIPA labor share, computed following Karabarbounis and Neiman (2014). Panel B applies our estimated capital-to-profit ratio to the NIPA non-labor share to obtain a comparable decomposition.

In Panel A, capital income is computed using the aggregate user cost of capital and then split between physical and intangible capital in proportion to their shares of the total capital stock, while the labor share is obtained residually. Panel B applies our estimated decomposition of non-labor income into physical capital, intangible capital, and profits to the NIPA non-labor share, thereby anchoring the exercise to national accounts data while preserving the micro-level information embedded in our estimates. Two features stand out. First, the Compustat-implied labor share closely tracks the NIPA labor share (dashed line in Panel A), providing external validation: although we do not target the labor share in estimation, the implied residual closely matches its independently measured counterpart. Second, the composition of capital income has shifted markedly over the sample: intangible capital accounted for about 25 percent of the total capital share in the early part of the sample but roughly 45 percent by the end, consistent with the secular rise in intangibles documented by Crouzet et al. (2022).

These comparisons also help benchmark our estimates against alternative decompositions of aggregate income. Barkai (2020) decomposes value added for the U.S. nonfinancial corporate sector over 1984–2014 and finds that the profit share increased from roughly −6 to 8 percent of gross value added. A notable feature of his estimates is that the profit share is negative for approximately two decades, from the early 1980s through the early 2000s. Our profit share, by contrast, remains positive throughout the sample and averages roughly 16 percent of GDP. Over Barkai's sample period, both approaches find that the profit share is higher in the 2000s than in the 1980s, but our estimates fluctuate within a positive range rather than crossing zero. The two approaches also agree on the magnitude of the decline in the labor share—about 7 percentage points over 1984–2014—but differ on the decomposition of non-labor income.

Profit share decomposition. A key advantage of the microeconomic approach is that it allows us to understand the determinants of the aggregate profit share by decomposing it into economically interpretable components, something that is not possible under the standard macroeconomic approach. In particular, the micro-aggregated profit share can be decomposed into fractions of value added pertaining to monopoly rents, fixed costs, and non-linearities. That is,

$$\Lambda_\Pi = \underbrace{\chi\left(1-\frac{1}{\overline{\mu}_{hsw}}\right)}_{\text{Monopoly rents}} - \underbrace{\chi\left(\overline{\text{RS}}-1\right)}_{\substack{\text{Fixed costs and}\\ \text{changing technology}}} + \underbrace{\chi\left\{\left(\frac{1}{\overline{\mu}_{hsw}}-1\right)\left(1-\overline{\text{RS}}\right)-\text{Cov}_\omega\left[\text{RS},\frac{1}{\mu}\right]\right\}}_{\text{Non-linearities}} \quad (10)$$

In Figure 4, we report the results of this decomposition. For expositional clarity, we group the "non-linearities" and the "fixed costs and changing technology" terms together and refer to the combined term simply as "fixed costs and changing technology."

Monopoly rents have increased from 12 to 35 percent of GDP due to the rise in the aggregate markup. The term "fixed costs and changing technology" has also increased, from 3 to 16 percent of GDP, due to increasing fixed costs and a rising scale elasticity.[23] This helps explain the pattern emphasized earlier: the profit share has been roughly constant at around 16 percent of GDP because the increase in monopoly rents was entirely offset by rising fixed costs and changing technology. We interpret movements around the 16% average profit share as reflecting cyclical variation and specific macroeconomic episodes of the U.S. economy.

Figure 4: Decomposition of the U.S. Profit Share.

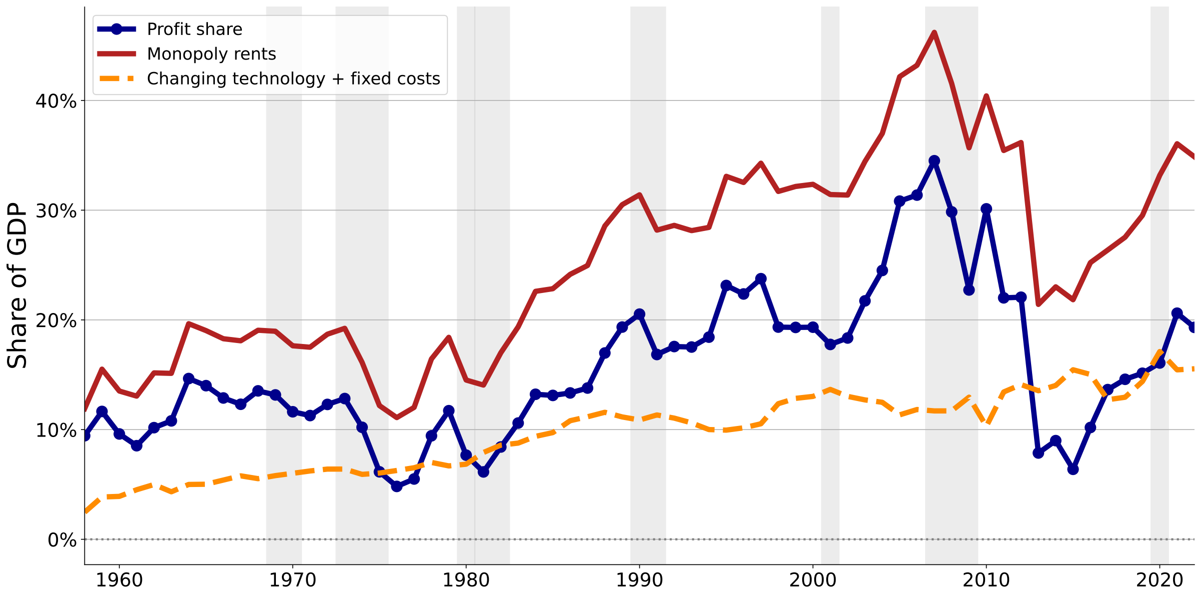


Figure Notes. The profit share is decomposed following equation (10). The series labeled "changing technology + fixed costs" is inclusive of non-linearities for expositional clarity. Shaded areas indicate NBER recession dates

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. Our approach backs out the user cost for each firm from its first-order condition for capital, and then aggregates these user cost using sales weights (guided by our theory):

$$R_t = \sum_i \omega_{it} r_{it}, \qquad \text{where} \qquad r_{it} = \frac{\theta^k_{jt}}{\mu_{it}} \times \frac{p_{it} y_{it}}{k_{it}} \tag{11}$$

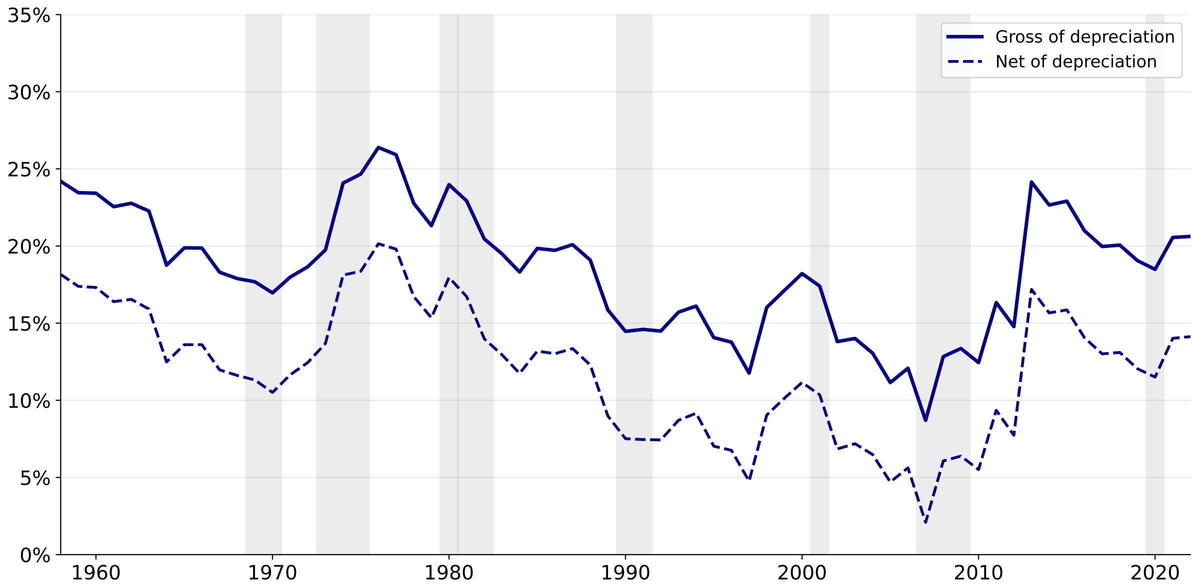
where ω_i is the sales weight of firm i, θ^k_j is the output elasticity of capital for industry j, μ is the markup, py are sales, and k is the capital stock.

[23]Nine percentage points of this increase are attributable to non-linearities, whose importance rose steadily from essentially zero to about eight percent of GDP over the sample period.

Thus, our aggregate user cost captures micro-level heterogeneity in user costs, which is an attractive feature of our method. Firms may have different user costs if there are differences in capital composition and depreciation rates, financing conditions, adjustments costs, risk premia, and other frictions. All these are embedded in our aggregate user cost of capital, as shown in Appendix A.

Figure 5 plots the aggregate user cost of capital, both gross and net of depreciation. The gross series averages roughly 19 percent over the sample, while the net-of-depreciation series averages roughly 12 percent, with the roughly 7 percentage point gap corresponding to the BEA aggregate depreciation rate. The net-of-depreciation return is well above risk-free rates throughout the sample, implying a substantial risk premium on productive capital. The series displays notable cyclical variation, driven primarily by movements in risk premia and financial frictions.[24]

Figure 5: The User Cost of Capital in the United States.



Notes. The user cost of capital is computed from the first-order condition for capital as $r_{it} = (\theta^k_{jt}/\mu_{it}) \times (p_{it}y_{it}/k_{it})$ and aggregated using sales weights. The net-of-depreciation series subtracts the BEA aggregate depreciation rate, computed as the ratio of consumption of fixed capital to the current-cost net stock of private fixed assets.

3.4 Heterogeneity and Industry-Level Results

We now study heterogeneity in markups, returns to scale, and profitability across firms and industries. We first examine firm-level heterogeneity in markups to understand the forces behind the increase in the aggregate markup, focusing on the evolution of the markup distribution and on the decomposition of the aggregate markup into within-

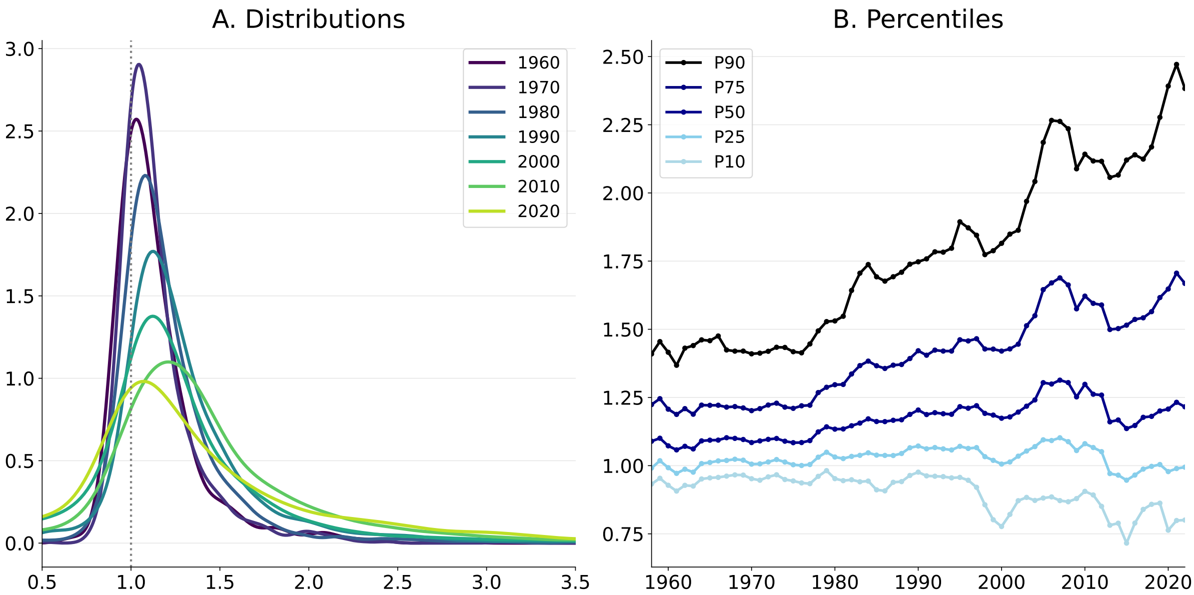
[24] The evidence of time-varying risk premia is consistent with Gilchrist and Zakrajšek (2012), Caballero et al. (2017), Farhi and François (2018), and Jordà et al. (2019), among others.

firm changes, between-firm reallocation, and net entry. We then present industry-level evidence on markups, returns to scale, and profitability.

Explaining the rise in the aggregate markup. Our empirical results so far have shown that the rise in markups has been counteracted by rising fixed costs and changing technologies. But, has the increase in markups been homogeneous across firms, or does the rise in the aggregate markup mask important micro-level heterogeneity? Did the aggregate markup increase because economic activity shifted towards firms with higher markups? Or because incumbent firms increased their markups over time? Or because entering firms charge relatively higher markups?

Figure 6 provides snapshots of the distribution and the evolution of percentiles of firm-level markups. Panel A plots kernel densities of unweighted firm-level markups at different points in time, showing that the distribution has become substantially more dispersed, consistent with the rise in variance emphasized by De Loecker et al. (2020). The flattening is driven primarily by a thicker right tail. Panel B quantifies this pattern: the 90th percentile rises from about 1.40 in 1960 to 2.40 in 2020, while the median rises only from 1.07 to 1.21. The increase in the aggregate markup is therefore driven mostly by firms at and above the median, whose markups have risen sharply since the 1980s. By contrast, firms in the lower part of the distribution have fared very differently. A less appreciated fact is that firms below the 25th percentile exhibit markups below unity, and that the lower tail deteriorates from the 1990s onward—the 10th percentile falls from 0.93 in 1960 to 0.76 in 2020. Such low measured markups may reflect firms on an exit margin, temporary losses, or predatory behavior.

Figure 6: Distributions and Percentiles of Firm-Level Markups in the United States.



Next, we study to what extent the rise in the aggregate markup is due to markup increases within incumbent firms, reallocation of economic activity toward firms with higher markups, and the appearance of firms that charge higher markups than those currently in place. To do so, we provide a statistical decomposition for the harmonic sales-weighted markup, as formalized in the following proposition.

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

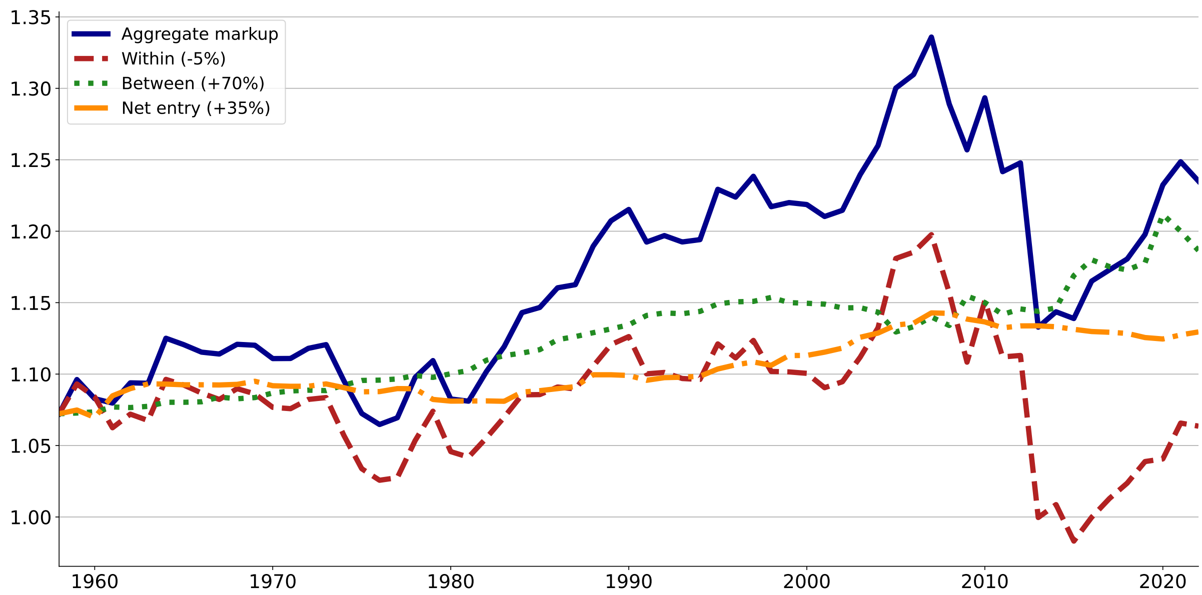
$$\Delta\overline{\mu}_t^{hsw} = -\overline{\mu}_t^{hsw}\,\overline{\mu}_{t-1}^{hsw}\Bigg[\underbrace{\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\,\Delta\mu_{it}^{-1}}_{within} + \underbrace{\sum_{i\in\mathcal{C}_t}\Delta\omega_{it}\left(\overline{\mu_i^{-1}} - \overline{\mu^{-1}}^{*}\right)}_{between} + \underbrace{\sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1} - \overline{\mu^{-1}}^{*}\right) - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\left(\mu_{i,t-1}^{-1} - \overline{\mu^{-1}}^{*}\right)}_{net\ entry}\Bigg], \quad (12)$$

where i and t index producers and time, respectively, $\mathcal{C}_t := \{i \in \mathcal{I}_t \cap \mathcal{I}_{t-1}\}$ is the set of incumbents, $\mathcal{E}_t := \{i \in \mathcal{I}_t \setminus \mathcal{I}_{t-1}\}$ is the set of entrants, $\mathcal{X}_{t-1} := \{i \in \mathcal{I}_{t-1} \setminus \mathcal{I}_t\}$ is the set of exiting firms, $\mathcal{I}_t = \mathcal{C}_t \cup \mathcal{E}_t$, $\mathcal{I}_{t-1} = \mathcal{C}_t \cup \mathcal{X}_{t-1}$, $\overline{\mu}_t^{hsw}$ is the harmonic sales-weighted markup in period t, $\Delta X_{it} := X_{it} - X_{i,t-1}$, $\overline{X}_i := 0.5(X_{it} + X_{i,t-1})$, and $\overline{\mu^{-1}}^{*} := 0.5\left(\overline{\mu}_{hsw,t}^{-1} + \overline{\mu}_{hsw,t-1}^{-1}\right)$.

Proof. See Appendix B.3. □

The results of this decomposition are shown in Figure 7, which decomposes the cumulative change in the aggregate markup into the three components of the proposition: a within component capturing changes in markups for continuing firms, a between component capturing the reallocation of sales toward firms with higher-than-average markups, and a net-entry component capturing the contribution of entrants relative to exiters. Each line plots the cumulative contribution of one component over time, starting from the initial level of the aggregate markup. The solid dark blue line is the actual aggregate markup, which by construction equals the sum of the three components plus the initial level. This decomposition is the harmonic-mean analogue of the well-known Foster et al. (2001) decomposition of aggregate productivity, with between and net-entry terms centered around the period-average inverse markup so that each term has a clean economic interpretation: the between term is positive only if reallocation favors above-average-markup firms, and the net-entry term is positive only if entrants have above-average markups or exiters had below-average markups.

Figure 7: Decomposition of the Aggregate U.S. Markup.



The rise in the aggregate markup is almost entirely a reallocation phenomenon. Of the 0.16 cumulative increase in the aggregate markup between 1958 and 2022, the within component contributes −5 percent (continuing firms' markups have, if anything, declined slightly on average), while reallocation toward higher-markup firms contributes 70 percent and net entry contributes 35 percent. In other words, the aggregate markup has risen not because existing firms are charging more, but because high-markup firms have grown relative to low-markup firms, and because entrants tend to have higher markups than exiters. This finding sharpens the "superstar firms" interpretation of rising market power emphasized by Autor et al. (2020) and others: in our data, virtually none of the increase in the aggregate markup reflects within-firm markup growth.

Taken together, Figure 6 and the decomposition in Figure 7 show that the rise in the aggregate markup is not the result of a uniform increase in firm-level markups. Rather, it reflects a profound change in the cross-sectional distribution of markups and in the allocation of economic activity across firms. Markups have become substantially more dispersed, with the upper tail pulling away from the rest of the distribution, and the aggregate markup has risen primarily because high-markup firms expand relative to low-markup firms and because entrants tend to have higher markups than exiters, especially since the 1990s.

Industry-level heterogeneity. Finally, we examine heterogeneity in markups, returns to scale, and profitability across US industries. This exercise serves two purposes. First, it documents how market power, cost structure, and profitability vary across sectors. Second, it provides a useful plausibility check for our empirical results. If higher markups are associated with larger fixed costs and stronger scale economies, industries with greater returns to scale should also tend to exhibit higher markups.

Figure 8 summarizes the cross-industry evidence for 2022. The horizontal axis reports industry-level markups, the vertical axis reports industry-level returns to scale, and dot size reflects the industry profit share. The figure reveals a clear positive association between returns to scale and markups across industries. This pattern is consistent with the view that sectors with larger overall fixed-cost requirements must charge higher markups to cover those costs. Information technology stands out in the upper-right region of the figure, combining the highest markup with high returns to scale, while sectors such as utilities and education also feature high returns to scale. By contrast, sectors such as wholesale, transportation, and hospitality tend to display both lower markups and returns to scale closer to one. The figure also shows that industries with higher markups and higher returns to scale tend to have larger profit shares. This pattern is consistent with the idea that sectors facing larger fixed costs tend to sustain both higher markups and higher profitability. At the same time, the figure should be interpreted with caution. A high markup in a given year need not imply that firms are pricing far above the level required to cover overall fixed costs, since current markups may also reflect the need to recoup sunk costs or previous investments. Appendices F.2–F.3 report the time series of returns to scale and markups for each industry, and Appendix F.4 provides a snapshot of industry markups and returns to scale for 2022.

Figure 8: Industry-level Markups, Returns to Scale, and Profitability in the U.S., 2022.

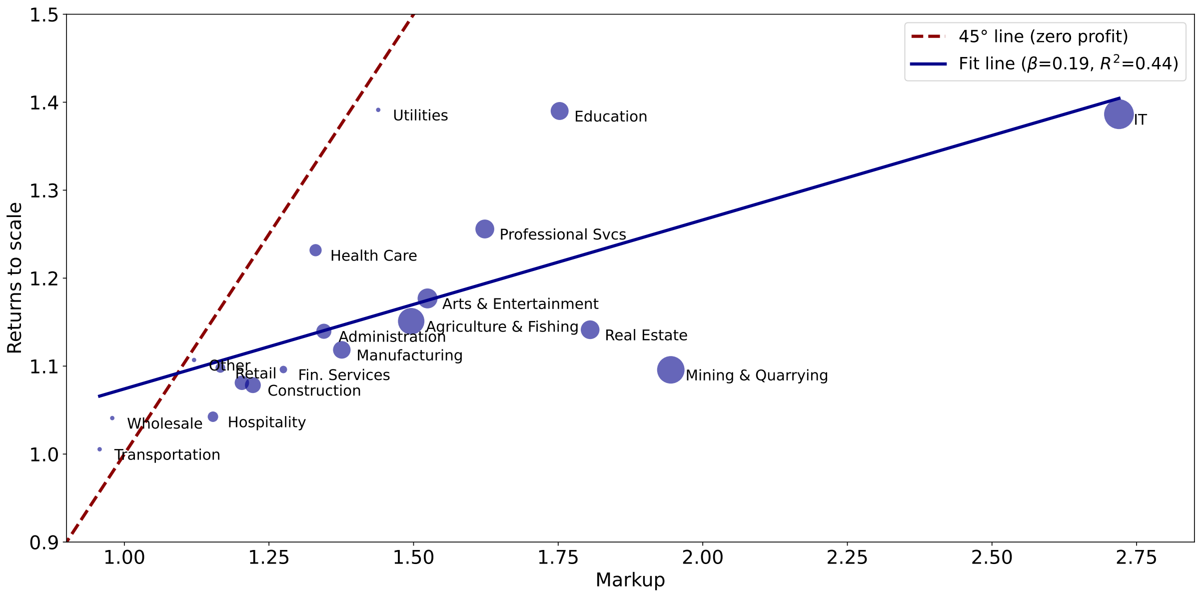


Figure Notes. Markups are harmonic sales-weighted markups. Returns to scale are sales-weighted scale elasticities adjusted for fixed costs. Dot size is proportional to the industry profit share.

4 Manufacturing Evidence from the U.S. Census Bureau

We complement the Compustat analysis with establishment-level data from the U.S. Census Bureau. Specifically, we use data from the Census of Manufactures (CMF), conducted every five years, and the Annual Survey of Manufactures (ASM), which surveys a rotating sample of manufacturing establishments in non-census years.

The Census data offer two main advantages. First, they cover the universe of manufacturing establishments rather than only publicly traded firms, thereby mitigating the representativeness concern inherent in Compustat. Second, they provide a detailed breakdown of variable costs, which allows us to separately identify the elasticities of materials and labor rather than relying on a single measure of COGS or operating expenses. This makes it possible to also estimate markdowns and to construct labor shares directly from the micro data.

Sample construction. Our sample covers the period 1977-–2021. Following Yeh, Macaluso and Hershbein (2022), in census years we restrict the CMF to establishments that also appear in an adjacent ASM year. This avoids artificial spikes in aggregates arising from the fact that the CMF covers the full universe of manufacturing establishments, whereas the ASM is based on a substantially smaller rotating sample. We classify establishments into 18 BEA manufacturing industry groups based on three-digit NAICS codes. We also trim observations with input levels below the 1st percentile or above the 99th percentile within each year and drop establishments whose materials cost share falls below 30 percent of the year-specific mean.

For each establishment, we observe output, materials, energy costs, labor costs, and capital stocks. Output is measured as the total value of shipments and deflated using industry-specific output price deflators. Labor costs are measured as salaries and wages, and capital is measured as the sum of equipment and structures. Relative to Compustat, the Census data provide a much richer decomposition of variable costs, allowing us to estimate a production function with separate measures of materials, energy, labor, and capital. Table X provides summary statistics.

[Insert Table here]

Estimation. We estimate a Cobb–Douglas production function in which output depends on materials, labor, energy, and capital. Materials, labor, and energy are treated as flexible inputs, while capital is treated as a state variable. We implement the Levinsohn and Petrin (2003) estimator with the ACF correction using five-year rolling windows by industry group, with materials serving as the proxy variable in the control function.

Identification. We identify markups and markdowns using a two-step procedure following Yeh, Macaluso and Hershbein. In the first step, we compute the markup as the ratio of the output elasticity of materials to the materials cost share, under the assumption that firms do not exercise monopsony power in input markets for materials. We then use the estimated markups to recover labor-market markdowns as the ratio of the labor output elasticity to the product of the markup and the labor cost share.

4.1 Aggregate Facts for U.S. Manufacturing

[To be added]

4.2 Heterogeneity and Industry-Level Results

[To be added]

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. Among these, the profit share has attracted sustained attention from both economists and policymakers.

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*. This result is purely-accounting based and does not rely on modeling assumptions. This procedure links micro data and macro analysis more closely, allowing researchers to leverage interesting features of micro-level datasets to inform aggregate outcomes.

Since profits 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. Under 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 fixed costs and relaxes the assumption of price-taking behavior in factor markets. Beyond standard regularity conditions on production functions, the only requirement is cost minimization.

The main theoretical contribution of the 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, that is, how profits propagate through input-output linkages. 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; and (iv) calibrating 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 1960 to 2020 using firm-level data from US Compustat. We have four main findings. First, we find that an economically meaningful notion of aggregate markup—the harmonic, sales-weighted markup—has increased from roughly 10% of price over marginal cost in 1960 to 25% in 2020. Second, we find that the rise in the aggregate markup is mostly explained by the reallocation of economic activity towards firms with higher markups. Third, we find that aggregate returns to scale have risen from 1.03 in 1960 to 1.15 in 2020, reflecting both rising fixed costs and changes in technology. Fourth, despite the rise in market power, we find that the aggregate profit share in the United States has been roughly constant at 16% of GDP over the past six decades. We reconcile these facts by decomposing the profit share into monopoly rents and fixed costs and changing technologies, and showing that the increase in the former has been offset by the latter.

These findings have important policy implications. In particular, a rise in market power does not necessarily imply a higher aggregate profit share. If firms use higher markups to cover rising fixed costs, aggregate profits as a share of GDP may remain stable. Policies that compress markups without accounting for underlying cost structures may therefore have unintended effects on firm entry, exit, and market structure. More broadly, the substantial rise in firm-level heterogeneity in markups and returns to scale that we document further complicates competition policy. In a relatively homogeneous economy, such as that of the 1960s and 1970s, uniform regulatory rules are likely to have broadly similar effects across firms. In today's more dispersed economy, however, a uniform policy that compresses markups may have sharply asymmetric effects: some firms may absorb the reduction through lower rents, while others may no longer be able to cover fixed costs and may exit. The rise in firm-level heterogeneity therefore suggests that effective competition policy may need to be more granular and more tailored to firm- or sector-specific conditions than in the past.

References

Abreu, D. (1986). Extremal equilibria of oligopolistic supergames, *Journal of Economic Theory* **39**(1): 191–225.

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.

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.

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

Baqaee, D. R. and Farhi, E. (2019). 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.

Basu, S. and Fernald, J. G. (1997). Returns to scale in us production: Estimates and implications, *Journal of political economy* **105**(2): 249–283.

Basu, S. and Fernald, J. G. (2002). Aggregate productivity and aggregate technology, *European Economic Review* **46**(6): 963–991.

Benkard, C. L., Miller, N. H. and Yurukoglu, A. (2025). The rise of market power and the macroeconomic implications: Comment, *Technical report*, National Bureau of Economic Research.

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.

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.

Brand, J. (2021). Differences in differentiation: Rising variety and markups in retail food stores, *Available at SSRN 3712513* .

Bresnahan, T. (1987). Competition and collusion in the american auto& mobile market: The 1955 price war, *Journal of Industrial Economics* **35**(4).

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.

Collard-Wexler, A. and De Loecker, J. (2015). Reallocation and technology: Evidence from the us steel industry, *American Economic Review* **105**(1): 131–171.

Corrado, C., Hulten, C. and Sichel, D. (2009). Intangible capital and us economic growth, *Review of income and wealth* **55**(3): 661–685.

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 Loecker, J. and Scott, P. (2022). Markup estimation using production and demand data. an application to the us brewing industry, *Technical report*, Working paper.

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

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. A., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2016). Declining business dynamism: What we know and the way forward, *American Economic Review* **106**(5): 203–207.

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* .

Doraszelski, U. and Jaumandreu, J. (2020). The inconsistency of de loecker and warzynski's (2012) method to estimate markups and some robust alternatives, *Technical report*, Working Paper.

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.

Farhi, E. and François, G. (2018). Accounting for macro-finance trends: Market power, intangibles, and risk premia, *Brookings Papers on Economic Activity* p. 147.

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.

Foster, L., Haltiwanger, J. C. and Krizan, C. J. (2001). Aggregate productivity growth: Lessons from microeconomic evidence, *New developments in productivity analysis*, University of Chicago Press, pp. 303–372.

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.

Ganapati, S. (2021). Growing oligopolies, prices, output, and productivity, *American Economic Journal: Microeconomics* **13**(3): 309–327.

Gandhi, A., Navarro, S. and Rivers, D. (2014). On the identification of production functions: How heterogeneous is productivity?, *Technical report*, University of Wisconsin-Madison.

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.

Grassi, B. (2017). Io in i-o: Size, industrial organization, and the input-output network make a firm structurally important, *Working Paper, Bocconi University* .

Green, E. J. and Porter, R. H. (1984). Noncooperative collusion under imperfect price information, *Econometrica: Journal of the Econometric Society* pp. 87–100.

Grieco, P. L., Murry, C. and Yurukoglu, A. (2023). The evolution of market power in the us automobile industry, *The Quarterly Journal of Economics* p. qjad047.

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.

Hsieh, C.-T. and Rossi-Hansberg, E. (2023). The industrial revolution in services, *Journal of Political Economy Macroeconomics* **1**(1): 3–42.

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.

Kehrig, M. and Vincent, N. (2021). The micro-level anatomy of the labor share decline, *The Quarterly Journal of Economics* **136**(2): 1031–1087.

Klette, T. J. and Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous, *Journal of applied econometrics* **11**(4): 343–361.

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

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

Miller, N., Osborne, M., Sheu, G. and Sileo, G. (2022). Technology and market power: The united states cement industry, 1974-2019, *Georgetown McDonough School of Business Research Paper* (4041168).

Morlacco, M. (2019). Market power in input markets: Theory and evidence from french manufacturing, *Unpublished, March* **20**: 2019.

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.

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

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 — FRB of Richmond
Luis Pérez — SMU

April 26, 2026

List of Appendices

A Dynamic Foundations for the User Cost

This appendix provides a dynamic foundation for the user cost of predetermined inputs. The relevant object is the within-period shadow rental of installed input services recovered from a restricted cost-minimization problem conditional on the installed stock. We show that this object remains the correct within-period opportunity cost for economic operating profits when the installed inputs are owned by the producer and chosen dynamically in the presence of adjustment costs, discounting, and risk premia.

Dynamic valuation of predetermined inputs. Time is discrete. At the beginning of period t, the producer observes the state vector $(\boldsymbol{z}_t, \boldsymbol{s}_t)$, where $\boldsymbol{z}_t = \{z_{kt}\}_{k\in\mathcal{Z}}$ denotes installed inputs and $\boldsymbol{s}_t$ collects exogenous state variables, including productivity and demand shifters. Within the period, the producer chooses variable inputs $\boldsymbol{x}_t = \{x_{jt}\}_{j\in\mathcal{N}}$ and investment $\boldsymbol{i}_t = \{i_{kt}\}_{k\in\mathcal{Z}}$. Production occurs according to

$$y_t = F(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t),$$

where $A_t \subset \boldsymbol{s}_t$, and installed inputs evolve according to

$$z_{kt+1} = (1-\delta_k)z_{kt} + i_{kt}. \tag{A.1}$$

The producer maximizes firm value. Let $V(\boldsymbol{z}_t, \boldsymbol{s}_t)$ denote the value function. The recursive problem is

$$V(\boldsymbol{z}_t, \boldsymbol{s}_t) = \max_{\boldsymbol{x}_t, \boldsymbol{i}_t \geq 0} \Bigg\{ \underbrace{p_t y_t - \sum_{j\in\mathcal{N}} w_{jt}x_{jt} - \sum_{k\in\mathcal{Z}} \Big(i_{kt} + \Phi_k(i_{kt}, z_{kt})\Big) - \mathrm{FC}_t}_{\text{cash-flow profits}} + \underbrace{\mathbb{E}_t\Big[\Xi_{t,t+1} V(\boldsymbol{z}_{t+1}, \boldsymbol{s}_{t+1})\Big]}_{\text{continuation value}} \Bigg\}, \tag{A.2}$$

where $\Xi_{t,t+1}$ is the stochastic discount factor that prices the continuation value.

Assumptions. Throughout, we maintain the following assumptions:

A1. *Differentiability*. The production function F and the adjustment-cost functions Φ_k are continuously differentiable, and Φ_k is twice continuously differentiable in i_k.

A2. *Convex adjustment costs*. Each Φ_k is convex in investment i_k.

A3. *Interior optimum and transversality*. An interior optimum exists, the value function is differentiable in $\boldsymbol{z}$, and the transversality condition holds.

A4. *Pricing*. There exists a stochastic discount factor $\Xi_{t,t+1} > 0$ satisfying standard no-arbitrage restrictions, and it is used to discount the continuation value.

Define the shadow value of installed input z_{kt} as

$$q_{kt} := \frac{\partial V(\boldsymbol{z}_t, \boldsymbol{s}_t)}{\partial z_{kt}}. \tag{A.3}$$

The next lemma gives its standard recursive characterization.

Lemma 2 (Shadow value recursion)**.** *Under assumptions A1–A4, the shadow value satisfies*

$$q_{kt} = p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) - \Phi_{k,z}(i_{kt}, z_{kt}) + \mathbb{E}_t\Big[\Xi_{t,t+1}(1-\delta_k) q_{k,t+1}\Big]. \tag{A.4}$$

Proof. Under assumptions A1–A4, the value function is differentiable and the optimum is interior, so the envelope theorem applies. Differentiating the Bellman equation with respect to z_{kt}, holding the optimal choices fixed, yields

$$\frac{\partial V(\boldsymbol{z}_t, \boldsymbol{s}_t)}{\partial z_{kt}} = p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) - \Phi_{k,z}(i_{kt}, z_{kt}) + \frac{\partial}{\partial z_{kt}} \mathbb{E}_t\Big[\Xi_{t,t+1} V(\boldsymbol{z}_{t+1}, \boldsymbol{s}_{t+1})\Big].$$

Since the continuation term depends on z_{kt} through $z_{k,t+1}$, the chain rule gives

$$\frac{\partial}{\partial z_{kt}} \mathbb{E}_t\Big[\Xi_{t,t+1} V(\boldsymbol{z}_{t+1}, \boldsymbol{s}_{t+1})\Big] = \mathbb{E}_t\left[\Xi_{t,t+1} \frac{\partial V(\boldsymbol{z}_{t+1}, \boldsymbol{s}_{t+1})}{\partial z_{k,t+1}} \frac{\partial z_{k,t+1}}{\partial z_{k,t}}\right]$$

Using the law of motion,

$$\frac{\partial z_{k,t+1}}{\partial z_{k,t}} = 1 - \delta_k.$$

Substituting and using the definition of q_{kt} gives

$$q_{kt} = p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) - \Phi_{k,z}(i_{kt}, z_{kt}) + \mathbb{E}_t\Big[\Xi_{t,t+1}(1-\delta_k) q_{k,t+1}\Big],$$

which proves the result. □

This recursion has the familiar asset-pricing structure: the shadow value of the installed input equals its marginal contribution to firm value, net of adjustment-cost effects, plus its discounted continuation value. This motivates the following definition.

Definition 1 (Dynamic user cost)**.** *The dynamic user cost of predetermined input z_{kt} is*

$$r^D_{kt} := q_{kt} - \mathbb{E}_t\Big[\Xi_{t,t+1}(1-\delta_k) q_{k,t+1}\Big]. \tag{A.5}$$

Equivalently, by Lemma 2,

$$r^D_{kt} = p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) - \Phi_{k,z}(i_{kt}, z_{kt}). \tag{A.6}$$

Thus, the dynamic user cost is the current marginal contribution of installed input k to firm value, net of the effect of the installed stock on adjustment costs. The first expression provides its intertemporal valuation representation, while the second gives its current-flow representation. In stochastic environments, the dynamic user cost reflects the required return on the input, including compensation for risk. In deterministic environments, it reduces to the familiar required-return-plus-depreciation logic.

To characterize how the installed stock is chosen, we next turn to the first-order condition for investment. This condition does not appear in the main text's profit decomposition, but it is useful for showing how adjustment costs, discounting, and risk premia shape the equilibrium evolution of installed inputs.

Proposition 3 (Investment Euler equation)**.** *For each $k \in \mathcal{Z}$, optimal investment satisfies*

$$1 + \Phi_{k,i}(i_{kt}, z_{kt}) = \mathbb{E}_t[\Xi_{t,t+1}\, q_{k,t+1}]. \tag{A.7}$$

Proof. Differentiate the Bellman equation with respect to $i_{k,t}$. Investment affects the current payoff through

$$-i_{kt} - \Phi_k(i_{kt}, z_{kt}),$$

and it affects the continuation value through the law of motion

$$z_{k,t+1} = (1 - \delta_k) z_{kt} + i_{kt}.$$

Hence,

$$\begin{aligned}
0 &= \frac{\partial}{\partial i_{k,t}}\Big(- i_{kt} - \Phi_k(i_{kt}, z_{kt})\Big) + \frac{\partial}{\partial i_{kt}} \mathbb{E}_t[\Xi_{t,t+1} V(s_{t+1})] \\
&= -(1 + \Phi_{k,i}(i_{kt}, z_{kt})) + \mathbb{E}_t\left[\Xi_{t,t+1} \frac{\partial V(s_{t+1})}{\partial z_{k,t+1}} \cdot \frac{\partial z_{k,t+1}}{\partial i_{kt}}\right] \\
&= -(1 + \Phi_{k,i}(i_{kt}, z_{kt})) + \mathbb{E}_t[\Xi_{t,t+1} q_{k,t+1}].
\end{aligned}$$

□

Having characterized the producer's intertemporal problem, we now connect it to the within-period shadow rental recovered from a static constrained cost-minimization problem. The key point is that, conditional on the realized installed stock, the producer's within-period input choice solves a static problem, even though the stock itself was chosen dynamically. This distinction allows us to recover the relevant opportunity cost of predetermined input services from static optimality conditions while preserving the role of adjustment costs, discounting, and risk in shaping the equilibrium stock.

Restricted cost minimization and the within-period user cost. Fix period t and treat installed inputs $\boldsymbol{z}_t$ as predetermined. Define the restricted variable cost function

$$C_t(\overline{y};\boldsymbol{z}_t,A_t)=\min_{\boldsymbol{x}_t\geq 0}\sum_{j\in\mathcal{N}} w_{jt}(x_{jt})x_{jt} \qquad \text{s.t.} \quad F(\boldsymbol{x}_t,\boldsymbol{z}_t;A_t)\geq \overline{y}. \tag{A.8}$$

The next two lemmas recover the shadow value of relaxing the output requirement and the shadow value of increasing the installed stock.

Lemma 3 (Envelope for output)**.** *Assume $C_t(\cdot)$ is differentiable in $\overline{y}$. Then*

$$\frac{\partial C_t(\overline{y};\boldsymbol{z}_t,A_t)}{\partial \overline{y}}=\lambda_t,$$

where λ_t denotes the multiplier on the constraint in the restricted cost minimization problem.

Proof. Let the Lagrangian for the restricted cost minimization problem be

$$\mathcal{L}(\boldsymbol{x}_t,\lambda_t)=\sum_{j\in\mathcal{N}} w_{jt}(x_{jt})x_{jt}+\lambda_t\Big[\overline{y}-F(\boldsymbol{x}_t,\boldsymbol{z}_t;A_t)\Big].$$

By the envelope theorem,

$$\frac{\partial C_t(\overline{y};\boldsymbol{z}_t,A_t)}{\partial \overline{y}}=\frac{\partial \mathcal{L}(\boldsymbol{x}_t^*,\lambda_t^*)}{\partial \overline{y}}=\lambda_t^*,$$

which proves the claim. □

Definition 2 (Marginal cost)**.** *Define marginal cost by*

$$MC_t=\left.\frac{\partial C_t(\overline{y};\boldsymbol{z}_t,A_t)}{\partial \overline{y}}\right|_{\overline{y}=y_t}.$$

By Lemma 3, $MC_t=\lambda_t$.

Lemma 4 (Envelope for installed inputs)**.** *Assume $C_t(\cdot)$ is differentiable in z_{kt}. Then*

$$\frac{\partial C_t(\overline{y};\boldsymbol{z}_t,A_t)}{\partial z_{kt}}=-\lambda_t F_{z_k}(\boldsymbol{x}_t,\boldsymbol{z}_t;A_t).$$

Proof. Using the same Lagrangian, the envelope theorem gives

$$\frac{\partial C_t(\overline{y};\boldsymbol{z}_t,A_t)}{\partial z_{kt}}=\frac{\partial \mathcal{L}(\boldsymbol{x}_t^*,\lambda_t^*)}{\partial z_{kt}}=-\lambda_t F_{z_k}(\boldsymbol{x}_t,\boldsymbol{z}_t;A_t),$$

which establishes the result. □

Proposition 4 (Restricted-cost characterization of the within-period user cost)**.** *Define the within-period user cost by*

$$r_{kt} = -\frac{\partial C_t(\overline{y}; \boldsymbol{z}_t, A_t)}{\partial z_{kt}}\bigg|_{\overline{y}=y_t}.$$

Then

$$r_{kt} = MC_t \times F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t).$$

Equivalently, if

$$\mu_t = \frac{p_t}{MC_t} \qquad \text{and} \qquad \theta_{kt} = \frac{F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) z_{kt}}{y_t},$$

then

$$r_{kt} = \frac{\theta_{kt}}{\mu_t} \frac{p_t y_t}{z_{kt}}.$$

This equality holds irrespective of whether the installed input is owned or rented and does not require the existence of a rental market.

Proof. By Lemma 4,

$$\frac{\partial C_t(\overline{y}; \boldsymbol{z}_t, A_t)}{\partial z_{kt}} = -\lambda_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t).$$

Therefore,

$$r_{kt} = -\frac{\partial C_t(\overline{y}; \boldsymbol{z}_t, A_t)}{\partial z_{kt}}\bigg|_{\overline{y}=y_t} = \lambda_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t)$$

Using Lemma 3 and the definition of marginal cost gives $\lambda_t = \mathrm{MC}_t$, so

$$r_{kt} = \mathrm{MC}_t \times F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t).$$

Finally, since $F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) = \theta_{kt} y_t / z_{kt}$ and $\mathrm{mc}_t = p_t/\mu_t$, we obtain

$$r_{kt} = \frac{\theta_{kt}}{\mu_t} \frac{p_t y_t}{z_{kt}}.$$

□

Proposition 4 shows that the user cost recovered from the static problem is the relevant within-period opportunity cost for economic operating profits. It is the reduction in minimum variable expenditure generated by an additional unit of the installed input, holding output fixed. As such, it is the correct shadow rental to subtract when measuring profits net of the opportunity cost of all inputs.

Proposition 5 (Dependence of the within-period user cost on the dynamic stock choice)**.** *Suppose the producer solves the dynamic problem* (A.2)*. Let* $\{\boldsymbol{z}_t\}$ *denote the equilibrium path of installed inputs induced by the Euler equation* (A.7) *and the law of motion for installed inputs* (A.1)*. Then* $r_{kt} = MC_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t)$ *varies with adjustment costs, discounting, and risk premia through their impact on the equilibrium evolution of* $\boldsymbol{z}_t$*.*

Proof. The Euler equation for investment (A.7) shows that the stochastic discount factor $\Xi_{t,t+1}$ and the adjustment-cost function Φ_k jointly determine optimal investment and therefore the evolution of the installed stock. Any change in adjustment costs, discounting, or risk premia changes the equilibrium path $\{\boldsymbol{z}_t\}$. Since the within-period user cost is computed as $r_{kt} = \mathrm{MC}_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t)$ and F_{z_k} is evaluated at the realized stock $\boldsymbol{z}_t$, these dynamic forces affect the empirically recovered user cost through the induced variation in the marginal product. Thus, the user cost recovered empirically from the production elasticity

$$r_{kt} = \frac{\theta_{kt}}{\mu_t} \frac{p_t y_t}{z_{kt}},$$

is a within-period shadow rental rate, but it is evaluated at the observed installed stock, which is itself the outcome of intertemporal optimization. □

Taken together, Propositions 4 and 5 make clear that the user cost entering the profit-rate formula is the within-period shadow rental $r_{kt} = \mathrm{MC}_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t)$, because this is the opportunity cost of installed services in period t. Although this object is characterized from a static restricted-cost problem, the user cost recovered in the data reflects adjustment costs, discounting, and risk premia through the observed (intertemporally chosen) equilibrium stock at which the marginal product is evaluated.

Relation between the dynamic and the static user costs. Finally, it is useful to relate the dynamic valuation object to the within-period user cost recovered in the data. By Definition 1 and Lemma 2,

$$r_{kt}^D = p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) - \Phi_{k,z}(i_{kt}, z_{kt}).$$

Hence,

$$p_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t) = r^D_{kt} + \Phi_{k,z}(i_{kt}, z_{kt}).$$

Using $r_{kt} = \text{mc}_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t)$ and $\mu_t = p_t/\text{mc}_t$, it follows that

$$r_{kt} = \frac{1}{\mu_t}\left[r^D_{kt} + \Phi_{k,z}(i_{kt}, z_{kt})\right].$$

This expression clarifies the distinction between the two objects. The object that enters the profit-rate decomposition is the within-period user cost

$$r_{kt} = \text{mc}_t F_{z_k}(\boldsymbol{x}_t, \boldsymbol{z}_t; A_t),$$

because it is the relevant opportunity cost of installed services in period t. Although the within-period user cost is characterized as a static shadow rental from a restricted cost-minimization problem, it is recovered in the data using an equivalent expression based on the markup and output elasticity. The object recovered empirically is nevertheless shaped by dynamic optimization, because the marginal product F_{z_k} is evaluated at the installed stock chosen by the producer's intertemporal problem. In this sense, the empirically recovered within-period user cost is a dynamic equilibrium object, even though its characterization is static.

Thus, the dynamic and static user costs are closely related but conceptually distinct. The object entering the profit-rate decomposition is the static within-period shadow rental r_{kt}, not the dynamic user cost r^D_{kt} itself. But because r_{kt} is evaluated at the stock chosen by the producer's dynamic problem, the user cost recovered in the data reflects the intertemporal forces that shape the optimal stock.

B Omitted Proofs

B.1 Proof of Lemma 5

Lemma 5 (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

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

□

B.2 Proof of Proposition 1

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

$$s_{\pi_t} = 1 - \frac{\text{RS}_t}{\mu_t} = 1 - \frac{\text{SE}_t^{\text{adj}}}{\mu_t} + \frac{\mathcal{M}_t}{\mu_t}, \tag{3}$$

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

$$\text{RS}_t = \text{SE}_t^{\text{adj}} - \mathcal{M}_t.$$

The adjusted scale elasticity is given by

$$\text{SE}_t^{\text{adj}} := \text{SE}_t \times \left(\frac{\text{TC}_t}{\text{TC}_t - \text{FC}_t}\right), \tag{4}$$

where $SE = \sum_{j \in \mathcal{N}} \theta_j$ is the scale elasticity of the production function, TC denotes total costs, and FC are fixed operating costs.

The monopsony term is given by

$$\mathcal{M}_t := \left(\frac{\mathrm{TC}_t}{\mathrm{TC}_t - \mathrm{FC}_t} \right) \sum_{j \in \mathcal{F}} \theta_{jt}(1 - \nu_{jt}), \tag{5}$$

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 $j \in \mathcal{F}$, 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. Fix installed inputs $\boldsymbol{z}$ and productivity A. The restricted cost-minimization problem for a generic producer may be written as

$$C(\overline{y}; \boldsymbol{z}, A) = \min_{\boldsymbol{x} \geq 0} \sum_{j \in \mathcal{N}} w_j(x_j) x_j \qquad \text{s.t.} \quad F(\boldsymbol{x}, \boldsymbol{z}; A) \geq \overline{y}.$$

Let λ denote the multiplier on the output constraint. By the envelope theorem,

$$\lambda = \left. \frac{\partial C(\overline{y}; \boldsymbol{z}, A)}{\partial \overline{y}} \right|_{\overline{y} = y} = \mathrm{MC},$$

where MC denotes marginal cost.

For any interior variable input $j \in \mathcal{N}$, the first-order condition is

$$w_j(x_j) + w_j'(x_j) x_j = \lambda F_{x_j},$$

which implies

$$\left(1 + \frac{w_j'(x_j) x_j}{w_j(x_j)} \right) w_j(x_j) x_j = \lambda F_{x_j} \frac{x_j}{y} y.$$

Defining the producer's (perceived) inverse elasticity of supply of input j as

$$\varepsilon_{Sj}^{-1} := \frac{w_j'(x_j) x_j}{w_j(x_j)},$$

we can write

$$\left(1 + \varepsilon_{Sj}^{-1} \right) w_j(x_j) x_j = \lambda y \theta_j,$$

where $\theta_j := F_{x_j} x_j / y$ is the output elasticity of input j.

Using the equivalence between the restricted cost-minimization problem and the associated conditional profit-maximization problem, $1+\varepsilon_{Sj}^{-1}$ is the ratio of the marginal revenue product of input j to its rental rate. Let $\nu_j \in (0,1]$ denote the markdown in input j. Then,

$$\nu_j = \left(1+\varepsilon_{Sj}^{-1}\right)^{-1}.$$

Using $\lambda = \text{MC}$, it follows that

$$w_j(x_j)x_j = \text{MC}\, y\, \theta_j \nu_j.$$

Now consider a predetermined input $k \in \mathcal{Z}$. By the envelope result for installed inputs, its within period user cost is

$$r_k = -\left.\frac{\partial C(\overline{y}; \boldsymbol{z}, A)}{\partial z_k}\right|_{\overline{y}=y} = \text{MC}\, F_{z_k}.$$

Multiplying by z_k,

$$r_k z_k = \text{MC}\, F_{z_k} z_k = \text{MC}\, y \theta_k,$$

where $\theta_k = F_{z_k} z_k / y$ is the output elasticity of predetermined input k.

Define total economic cost by

$$\text{TC} := \sum_{j\in\mathcal{N}} w_j(x_j)x_j + \sum_{k\in\mathcal{Z}} r_k z_k + \text{FC}.$$

Hence,

$$\text{TC} - \text{FC} = \sum_{j\in\mathcal{N}} w_j(x_j)x_j + \sum_{k\in\mathcal{Z}} r_k z_k = \text{MC}\, y \left(\sum_{j\in\mathcal{N}} \theta_j \nu_j + \sum_{k\in\mathcal{Z}} \theta_k\right).$$

Since monopsony distortions are present only for factors $j \in \mathcal{F} \subseteq \mathcal{N}$, we have $\nu_j = 1$ for $j \notin \mathcal{F}$. Therefore,

$$\sum_{j\in\mathcal{N}} \theta_j \nu_j = \sum_{j\in\mathcal{N}} \theta_j + \sum_{j\in\mathcal{F}} \theta_j (1-\nu_j).$$

Using the definition of scale elasticity,

$$\text{SE} = \sum_{j\in\mathcal{N}} \theta_j + \sum_{k\in\mathcal{Z}} \theta_k,$$

it follows that

$$\mathrm{TC} - \mathrm{FC} = \mathrm{MC}\, y \left(\mathrm{SE} - \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j) \right).$$

Rearranging,

$$\frac{\mathrm{TC} - \mathrm{FC}}{\mathrm{MC}\, y} = \mathrm{SE} - \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j).$$

Multiplying both sides by TC/(TC − FC), we obtain

$$\frac{\mathrm{TC}}{\mathrm{MC}\, y} = \mathrm{SE}\left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right) - \left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j).$$

Since average cost is $\mathrm{AC} := \mathrm{TC}/y$, this becomes

$$\frac{\mathrm{AC}}{\mathrm{MC}} = \mathrm{SE}\left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right) - \left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j).$$

Define

$$\mathrm{RS} := \frac{\mathrm{AC}}{\mathrm{MC}}, \qquad \mathrm{SE}^{\mathrm{adj}} := \mathrm{SE}\left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right), \qquad \mathcal{M} := \left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}} \right) \sum_{j \in \mathcal{F}} \theta_j (1 - \nu_j).$$

Then,

$$\mathrm{RS} = \mathrm{SE}^{\mathrm{adj}} - \mathcal{M}.$$

Finally, economic profits are $\pi := py - \mathrm{TC}$, so the profit rate in terms of sales is

$$s_\pi := \frac{\pi}{py} = 1 - \frac{\mathrm{TC}}{py} = 1 - \frac{\mathrm{AC}}{p} = 1 - \frac{\mathrm{AC}/\mathrm{MC}}{p/\mathrm{MC}} = 1 - \frac{\mathrm{RS}}{\mu}.$$

Hence,

$$s_\pi = 1 - \frac{\mathrm{RS}}{\mu} = 1 - \frac{\mathrm{SE}^{\mathrm{adj}}}{\mu} + \frac{\mathcal{M}}{\mu}.$$

□

B.3 Proof of Proposition 2

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

$$\Delta\overline{\mu}_t^{\text{hsw}} = -\overline{\mu}_t^{\text{hsw}}\,\overline{\mu}_{t-1}^{\text{hsw}}\Bigg[\underbrace{\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\,\Delta\mu_{it}^{-1}}_{\text{within}} + \underbrace{\sum_{i\in\mathcal{C}_t}\Delta\omega_{it}\left(\overline{\mu_i^{-1}} - \overline{\mu^{-1}}^{*}\right)}_{\text{between}}$$
$$+\underbrace{\sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1} - \overline{\mu^{-1}}^{*}\right) - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\left(\mu_{i,t-1}^{-1} - \overline{\mu^{-1}}^{*}\right)}_{\text{net entry}}\Bigg], \quad (12)$$

where i and t index producers and time, respectively, $\mathcal{C}_t := \{i \in \mathcal{I}_t \cap \mathcal{I}_{t-1}\}$ is the set of incumbents, $\mathcal{E}_t := \{i \in \mathcal{I}_t \setminus \mathcal{I}_{t-1}\}$ is the set of entrants, $\mathcal{X}_{t-1} := \{i \in \mathcal{I}_{t-1} \setminus \mathcal{I}_t\}$ is the set of exiting firms, $\mathcal{I}_t = \mathcal{C}_t \cup \mathcal{E}_t$, $\mathcal{I}_{t-1} = \mathcal{C}_t \cup \mathcal{X}_{t-1}$, $\overline{\mu}_t^{hsw}$ is the harmonic sales-weighted markup in period t, $\Delta X_{it} := X_{it} - X_{i,t-1}$, $\overline{X}_i := 0.5(X_{it} + X_{i,t-1})$, and $\overline{\mu^{-1}}^{*} := 0.5\left((\overline{\mu}_t^{hsw})^{-1} + (\overline{\mu}_{t-1}^{hsw})^{-1}\right)$.

Proof. By definition,

$$\begin{aligned}\Delta\overline{\mu}_t^{\text{hsw}} &= \overline{\mu}_t^{\text{hsw}} - \overline{\mu}_{t-1}^{\text{hsw}} \\ &= \left(\sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1}\right)^{-1} - \left(\sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right)^{-1}.\end{aligned}$$

Let

$$A_t := \sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1}.$$

Then

$$\begin{aligned}\Delta\overline{\mu}_t^{\text{hsw}} &= A_t^{-1} - A_{t-1}^{-1} \\ &= \frac{A_{t-1} - A_t}{A_t A_{t-1}} \\ &= -\overline{\mu}_t^{\text{hsw}}\,\overline{\mu}_{t-1}^{\text{hsw}}\left[\sum_{i\in\mathcal{I}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right].\end{aligned}$$

Splitting each sum into continuing firms and entrants/exiters,

$$\Delta\overline{\mu}_t^{\text{hsw}} = -\overline{\mu}_t^{\text{hsw}}\,\overline{\mu}_{t-1}^{\text{hsw}}\left[\sum_{i\in\mathcal{C}_t}\left(\omega_{it}\mu_{it}^{-1} - \omega_{i,t-1}\mu_{i,t-1}^{-1}\right) + \sum_{i\in\mathcal{E}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right].$$

For each continuing firm,

$$\omega_{it}\mu_{it}^{-1} - \omega_{i,t-1}\mu_{i,t-1}^{-1} = \overline{\omega}_i\,\Delta\mu_{it}^{-1} + \Delta\omega_{it}\,\overline{\mu_i^{-1}},$$

where

$$\overline{\omega}_i := \frac{1}{2}(\omega_{it} + \omega_{i,t-1}), \qquad \overline{\mu_i^{-1}} := \frac{1}{2}\left(\mu_{it}^{-1} + \mu_{i,t-1}^{-1}\right).$$

Substituting this identity into the previous expression yields

$$\Delta\overline{\mu}_t^{\text{hsw}} = -\overline{\mu}_t^{\text{hsw}}\,\overline{\mu}_{t-1}^{\text{hsw}}\left[\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\,\Delta\mu_{it}^{-1} + \sum_{i\in\mathcal{C}_t}\Delta\omega_{it}\,\overline{\mu_i^{-1}} + \sum_{i\in\mathcal{E}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1}\right].$$

Now define

$$\overline{\mu^{-1}}^{*} := \frac{1}{2}\left[(\overline{\mu}_t^{\text{hsw}})^{-1} + (\overline{\mu}_{t-1}^{\text{hsw}})^{-1}\right].$$

Adding and subtracting $\overline{\mu^{-1}}^{*}$ inside the entry and exit terms gives

$$\begin{aligned}\sum_{i\in\mathcal{E}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1} &= \sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1} - \overline{\mu^{-1}}^{*}\right) - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\left(\mu_{i,t-1}^{-1} - \overline{\mu^{-1}}^{*}\right)\\ &\quad + \overline{\mu^{-1}}^{*}\left(\sum_{i\in\mathcal{E}_t}\omega_{it} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\right).\end{aligned}$$

Using $\sum_{i\in\mathcal{I}_t}\omega_{it} = 1$ and $\sum_{i\in\mathcal{I}_{t-1}}\omega_{i,t-1} = 1$, together with $\mathcal{I}_t = \mathcal{C}_t \cup \mathcal{E}_t$ and $\mathcal{I}_{t-1} = \mathcal{C}_t \cup \mathcal{X}_{t-1}$,

$$\sum_{i\in\mathcal{E}_t}\omega_{it} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1} = \left(1 - \sum_{i\in\mathcal{C}_t}\omega_{it}\right) - \left(1 - \sum_{i\in\mathcal{C}_t}\omega_{i,t-1}\right) - \sum_{i\in\mathcal{C}_t}\Delta\omega_{it}.$$

Therefore,

$$\sum_{i\in\mathcal{E}_t}\omega_{it}\mu_{it}^{-1} - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\mu_{i,t-1}^{-1} = \sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1} - \overline{\mu^{-1}}^{*}\right) - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\left(\mu_{i,t-1}^{-1} - \overline{\mu^{-1}}^{*}\right) - \overline{\mu^{-1}}^{*}\sum_{i\in\mathcal{C}_t}\Delta\omega_{it}.$$

Substituting this expression into the previous equation and collecting terms yields

$$\begin{aligned}\Delta\overline{\mu}_t^{\text{hsw}} = -\overline{\mu}_t^{\text{hsw}}\,\overline{\mu}_{t-1}^{\text{hsw}}\Bigg[&\sum_{i\in\mathcal{C}_t}\overline{\omega}_i\,\Delta\mu_{it}^{-1} + \sum_{i\in\mathcal{C}_t}\Delta\omega_{it}\left(\overline{\mu_i^{-1}} - \overline{\mu^{-1}}^{*}\right)\\ &+ \sum_{i\in\mathcal{E}_t}\omega_{it}\left(\mu_{it}^{-1} - \overline{\mu^{-1}}^{*}\right) - \sum_{i\in\mathcal{X}_{t-1}}\omega_{i,t-1}\left(\mu_{i,t-1}^{-1} - \overline{\mu^{-1}}^{*}\right)\Bigg],\end{aligned}$$

which is the desired decomposition. □

B.4 Proof of Theorem 1

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*

$$\Lambda_{\Pi t} = \underbrace{\chi_t}_{\text{IO multiplier}} \times \underbrace{\left(1 - \frac{\overline{\mathrm{SE}}_t^{\mathrm{adj}}}{\overline{\mu}_t^{\mathrm{hsw}}} + \frac{\overline{\mathcal{M}}_t}{\overline{\mu}_t^{\mathrm{hsw}}} - \mathrm{Cov}_\omega\left[\mathrm{SE}_t^{\mathrm{adj}}, \frac{1}{\mu_t}\right] + \mathrm{Cov}_\omega\left[\mathcal{M}_t, \frac{1}{\mu_t}\right]\right)}_{\text{sales-weighted profit rate}}, \tag{7}$$

where $\chi = \sum_{k \in \mathcal{I}} \frac{p_k y_k}{GDP}$ denotes the input-output multiplier, SE^{adj} the scale elasticity adjusted for fixed costs given by (4), μ the markup, and $\mathcal{M}$ the monopsony term given by (5). The term $\overline{X}$ is the sales-weighted average of X, $\overline{\mu}^{hsw}$ is the harmonic sales-weighted markup, and $\mathrm{Cov}_\omega(X, Y)$ is the sales-weighted covariance of variables X and Y.

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

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

Proposition 1 can be used to establish

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

Hence,

$$\begin{aligned}
\Lambda_\Pi &= \sum_{i \in \mathcal{I}} \frac{p_i y_i}{\mathrm{GDP}} \left(1 - \frac{\mathrm{SE}_i^{\mathrm{adj}}}{\mu_i} + \frac{\mathcal{M}_i}{\mu_i}\right) \\
&= \underbrace{\left(\sum_{k \in \mathcal{I}} \frac{p_k y_k}{\mathrm{GDP}}\right)}_{\equiv \chi} \sum_{i \in \mathcal{I}} \underbrace{\frac{p_i y_i}{\sum_k p_k y_k}}_{\equiv \omega_i} \left(1 - \frac{\mathrm{SE}_i^{\mathrm{adj}}}{\mu_i} + \frac{\mathcal{M}_i}{\mu_i}\right) \\
&= \chi \left(1 - \mathbb{E}_\omega\left[\frac{\mathrm{SE}_i^{\mathrm{adj}}}{\mu_i}\right] + \mathbb{E}_\omega\left[\frac{\mathcal{M}_i}{\mu_i}\right]\right) \\
&= \chi \left(1 - \frac{\overline{\mathrm{SE}}^{\mathrm{adj}}}{\overline{\mu}^{\mathrm{hsw}}} + \frac{\overline{\mathcal{M}}}{\overline{\mu}^{\mathrm{hsw}}} - \mathrm{Cov}_\omega\left[\mathrm{SE}^{\mathrm{adj}}, \frac{1}{\mu}\right] + \mathrm{Cov}_\omega\left[\mathcal{M}, \frac{1}{\mu}\right]\right)
\end{aligned}$$

□

C Fixed Costs and the Measurement of Scale Elasticities

In this appendix, we provide a simple example showing that the object one would estimate and refer to as a scale elasticity may, in fact, reflect both technological curvature and fixed-cost requirements when fixed and variable inputs are not separately observed.

Suppose there is a firm with production technology

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

where y is output, $A \in \mathbb{R}_{++}$ is a productivity shifter, $\ell \in \mathbb{R}_{+}$ is labor, $\overline{\ell} \in \mathbb{R}_{++}$ is a minimum requirement on labor to produce positive output, and $\alpha \in \mathbb{R}_{++}$ governs the curvature of the production technology with respect to the net labor input, $\ell - \overline{\ell}$.

Assuming the production function is continuously differentiable and that there is no monopsony power, the derivative of output with respect to labor ℓ is

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

Multiplying both sides of (C.2) by $(\ell - \overline{\ell})$, we obtain

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

Rearranging yields

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

which implies

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

Thus, the elasticity of output with respect to observed labor exceeds the underlying technological parameter α whenever $\overline{\ell} > 0$. In particular, even if the underlying technology is constant returns in net labor, that is, $\alpha = 1$, the measured elasticity with respect to observed labor is greater than one.

In this environment, this elasticity also coincides with returns to scale. Since there is a single input and the firm minimizes costs, average cost and marginal cost are given by

$$\mathrm{AC} = \frac{w\ell}{y}, \qquad \mathrm{MC} = \frac{w}{\mathrm{MP}_\ell},$$

where $\mathrm{MP}_\ell = \mathrm{d}y/\mathrm{d}\ell$ denotes the marginal product of labor. Hence,

$$\frac{\mathrm{AC}}{\mathrm{MC}} = \frac{w\ell/y}{w/\mathrm{MP}_\ell} = \mathrm{MP}_\ell \frac{\ell}{y} = \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y}.$$

Therefore,

$$\mathrm{RS} = \frac{\mathrm{AC}}{\mathrm{MC}} = \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y} = \alpha \frac{\ell}{\ell - \overline{\ell}}.$$

To connect this object to empirical practice, suppose we observe cross-sectional data on $\{y, \ell\}_i$, where y is output, ℓ is total labor input, and i indexes firms. Importantly, total labor includes both the variable component and the fixed labor requirement. Consider the regression

$$\log y_i = \beta \log \ell_i + \tilde{\omega}_i,$$

where β is the coefficient of interest. Note that

$$\beta = \frac{\mathrm{d}\log y}{\mathrm{d}\log \ell} = \frac{\mathrm{d}y}{\mathrm{d}\ell}\frac{\ell}{y}.$$

Given the production function above, it follows that

$$\hat{\beta} = \mathrm{RS} = \alpha \frac{\ell}{\ell - \overline{\ell}}.$$

That is, the object one would estimate and interpret as scale elasticity corresponds to returns to scale, which conflate the underlying technological curvature α with the fixed-cost adjustment factor $\ell/(\ell - \overline{\ell})$.

This simple example illustrates the identification problem relevant for our empirical analysis. When fixed and variable input requirements are not separately observed, estimated scale elasticities do not admit a purely technological interpretation. Measured increasing returns may arise either because the underlying technology is non-constant returns or because production requires a fixed input component. Without separate information on fixed inputs, these two forces cannot be disentangled. We use this logic to interpret the estimated scale elasticities in our empirical analysis.

D Biases Associated with the Use of Revenue Elasticities

In this appendix, we discuss how using revenue, rather than physical output, affects the estimation of elasticities and, in turn, measures of markups, returns to scale, and profitability. More specifically, we show that although revenue-based elasticities may bias markup estimates downward, our estimates of profit rates and profit shares are unaffected by this source of bias.

D.1 Markups

As first noted by Klette and Griliches (1996), using deflated revenue as a proxy for real output introduces an omitted-variable problem when the firms within an industry charge different prices. The source of the problem is that firm-level prices may be correlated with input choices, so that revenue-based regressions need not recover output elasticities. In general, this causes revenue-based elasticities, and hence markup estimates, to differ from their output-based counterpart. More recently, Bond et al. (2021) and De Ridder et al. (2022) have emphasized additional concerns with using revenue data to estimate output elasticities.

A central criticism in Bond et al. (2021) is that, if one replaces output elasticities with revenue elasticities in the standard markup formula for a firm that maximizes current-period profits, the resulting markup estimate is equal to one and therefore uninformative about true output-market power. The key to this argument is that firms maximize static profits. In more general environments in which firms maximize the discounted sum of profits, static profit maximization need not apply, although firms may still minimize current-period costs, as we show in Appendix A.[25] Imposing only cost-minimizing behavior, the next proposition shows—similarly to Klette and Griliches and Bond et al.—that revenue-based markups understate true markups when firms face downward-sloping demand.

Proposition 6. *Revenue-based markups μ^R understate true markups μ under monopolistic competition; that is,*

$$\mu^R := \frac{\theta_\ell^R}{\alpha_\ell} \leq \frac{\theta_\ell}{\alpha_\ell} := \mu,$$

where θ_ℓ^R denotes the revenue elasticity of a flexible input, θ_ℓ denotes the corresponding output elasticity, and α_ℓ is the revenue share of that input.

[25] Several papers in the repeated-games literature show that, with sufficiently low discount factors and strategic interactions, equilibrium outcomes can differ sharply from those implied by static profit maximization. See, for example, Green and Porter (1984) and Abreu (1986).

Proof. Cost minimization implies that the true markup can be written as

$$\mu = \frac{\theta_\ell}{\alpha_\ell},$$

where θ_ℓ is the output elasticity of variable input ℓ, and $\alpha_\ell := p_\ell \ell/(py)$ its revenue share.

The corresponding revenue-based markup is

$$\mu^R = \frac{\theta_\ell^R}{\alpha_\ell},$$

where $\theta_\ell^R := (\partial R/\partial \ell)(\ell/R)$ is the revenue elasticity of input ℓ, and $R = py$ is revenue.

Applying the chain rule,

$$\theta_\ell^R = \frac{\partial R}{\partial y}\frac{y}{R} \cdot \frac{\partial y}{\partial \ell}\frac{\ell}{y} = \theta_y^R \theta_\ell,$$

where $\theta_y^R := (\mathrm{d}R/\mathrm{d}y)(y/R)$ is the revenue elasticity of output. It follows that

$$\mu^R = \theta_y^R \mu.$$

Under monopolistic competition, the firm faces a downward-sloping demand curve with perceived elasticity $\epsilon > 1$. Since revenue is given by $R(y) = p(y)y$,

$$\theta^R = \frac{\mathrm{d}R}{\mathrm{d}y}\frac{y}{R} = [p'(y)y + p(y)]\frac{1}{p(y)} = 1 + \frac{p'(y)y}{p(y)}.$$

Using the definition of perceived elasticity of demand,

$$\epsilon := -\frac{\mathrm{d}y}{\mathrm{d}p}\frac{p}{y},$$

we have

$$\frac{p'(y)y}{p(y)} = -\frac{1}{\epsilon},$$

so that

$$\theta_y^R = \frac{\epsilon - 1}{\epsilon} \in (0, 1).$$

Hence, $\mu^R = \theta_y^R \mu \leq \mu$. □

Proposition 6 shows that revenue-based markups understate true markups when firms face downward-sloping demand. The magnitude of this bias depends on the elasticity of demand and vanishes in the limiting case of perfectly-elastic demand.

D.2 Profit Rates and the Profit Share

Our next proposition shows that even when revenue-based markups understate true markups, the corresponding revenue-based profit rates are unaffected by this bias.

Proposition 7. *Revenue-based and output-based profit rates are equal, that is,* $s_\pi^R = s_\pi$.

Proof. By proposition 1, in the absence of factor-market power the output-based profit rate of a monopolistic producer is

$$s_\pi = 1 - \frac{\mathrm{RS}}{\mu},$$

where RS denotes returns to scale and μ is the output-based markup. Recall that

$$\mathrm{RS} = \mathrm{SE}\left(\frac{\mathrm{TC}}{\mathrm{TC} - \mathrm{FC}}\right), \qquad \text{where } \mathrm{SE} = \sum_j \theta_j.$$

Under revenue-based estimation, the elasticity of each input is scaled by a common factor, $\theta_j^R = \theta_y^R \theta_j$. Therefore, the revenue-based elasticity is $\mathrm{SE}^R = \theta_y^R \mathrm{SE}$, and the corresponding revenue-based returns to scale satisfy $\mathrm{RS}^R = \theta_y^R \mathrm{RS}$, since the fixed-cost adjustment factor is unchanged. From Proposition 6, $\mu^R = \theta_y^R \mu$. Hence,

$$\frac{\mathrm{RS}^R}{\mu^R} = \frac{\mathrm{RS}}{\mu},$$

which establishes the result. □

Propositions 6 and 7 imply that although revenue-based elasticities may bias markup estimates downward, this bias does not affect the profit rates or profit shares that are central to our analysis.

D.3 Other Sources of Biases

Recent literature has identified additional sources of bias in markup estimation that are distinct from the revenue-versus-output issue discussed above.

First, De Ridder et al. (2022) show that estimating revenue-based elasticities under a Cobb-Douglas production function, as we do, can overstate the variance of the implied markup distribution. Second, Bond et al. (2021) emphasize two concerns that arise even when one is able to recover output elasticities. One is that, if the flexible input used in the markup formula is itself distorted, the estimated markup will capture not only output-market power but also other frictions. A particularly relevant example is monopsony power. If the producer has monopsony power in the variable input used to construct the markup, then the resulting object captures both output- and input-market power, as discussed by Morlacco (2019), Brooks et al. (2021), and Yeh et al. (2022). A second concern is that the estimated markup may be biased downward if the flexible input affects not only production but also demand, as would be the case for expenditures on marketing or similar activities. Finally, Doraszelski and Jaumandreu (2020) argue that the first stage of the Ackerberg, Caves and Frazer (2015)'s procedure conceptually requires knowledge of markups even though recovering markups is itself the goal of the exercise. De Ridder et al. (2022), however, argue that failing to control for markups in the first stage appears to matter little in practice.

E 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 obtained from NIPA tables, was around 62% of GDP. Barkai (2020) made Basu's argument more formally by explicitly using the expression

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

where Λ_{Π} is the profit share, χ is the ratio of sales to aggregate value added, RS denotes 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 sales-weighted markup of De Loecker, Eeckhout and Unger cannot be used to infer the aggregate profit share.

Assuming no monopsony power, we can characterize the aggregate profit share, as a corollary to our Theorem 1, as

$$\Lambda_\Pi = \chi\left(1 - \frac{\overline{\mathrm{RS}}}{\overline{\mu}^{\mathrm{hsw}}} - \mathrm{Cov}_\omega\left[\mathrm{RS}, \frac{1}{\mu}\right]\right), \tag{E.2}$$

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

Inspection of this expression reveals several problems with the Basu–Barkai inference. First, as noted by De Loecker, Eeckhout and Unger, if there is heterogeneity in markups and returns to scale across producers, aggregation is nonlinear. This is why the covariance term appears: the aggregate profit share depends not only on the aggregate markup and aggregate returns to scale, but also on how the two are jointly distributed across firms. Second, our aggregation result demands the harmonic sales-weighted markup, 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 depends on the variance of the cross-sectional distribution. Thus, using the sales-weighted markup would generally lead to overstating the aggregate profit share. Third, our characterization calls for sales-weighted returns to scale, not the sales-weighted scale elasticity estimated by De Loecker et al. (2020).

This last distinction is important because, in the presence of fixed costs, returns to scale are not equal to the technological scale elasticity. Rather, returns to scale equal the scale elasticity multiplied by a fixed-cost adjustment factor. This adjustment is quantitatively relevant because De Loecker et al. treat SG&A as a fixed cost. Once we aggregate correctly and account for fixed costs, we obtain an expression that is well suited to assess the consistency of micro estimates of markups and returns to scale with aggregate profit shares.[26] In this sense, the Basu–Barkai calculation is not incorrect in spirit, but it does not use the properly-aggregated micro estimates.

In Figure 9, we map micro-level estimates of markups and returns to scale to the aggregate profit share using equation (E.2). 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 roughly constant at around 15% of GDP until 1995

[26] Equation (E.2) nests Basu's and Barkai's expression (E.1) as the special case in which there are no fixed costs and no heterogeneity in markups.

and then increases until 20% by 2020. By contrast, the alternative series map firm-level estimates of markups to the aggregate profit share using flawed aggregation schemes and ignoring relevant features of the data such as fixed costs.

Figure 9: Mapping Micro-Level Estimates of Markups to the US Profit Share.

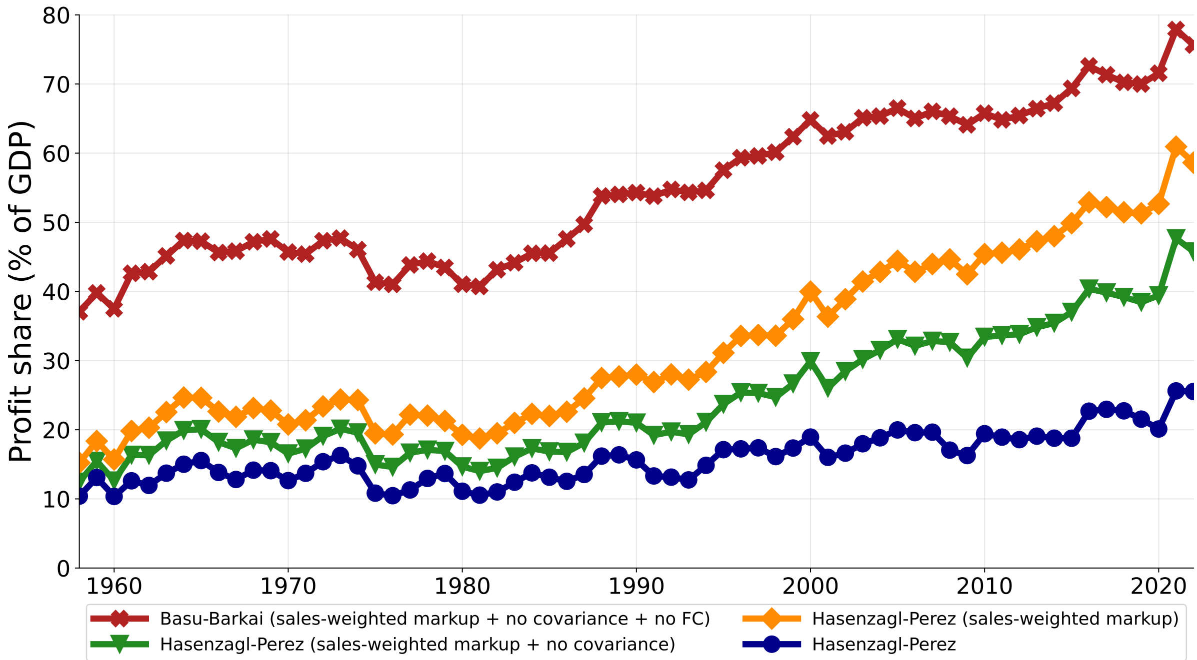
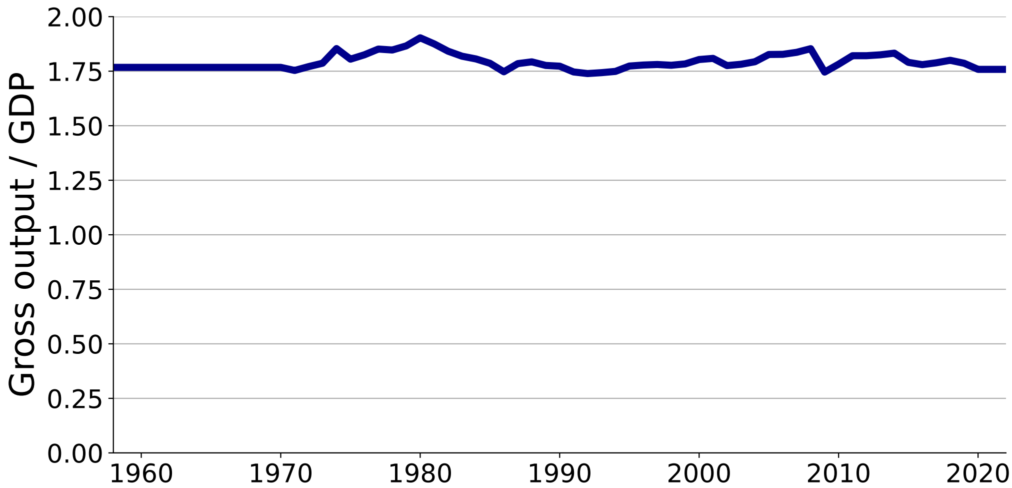


Figure Notes. Production function parameters are estimated using the cost of goods sold (COGS) as variable input and a measure of capital which includes physical capital, as in De Loecker et al. (2020)'s procedure. To calculate the fixed-cost adjustment factor we use selling, general, and administrative expenses (SG&A) as a measure of fixed costs in line with De Loecker et al..

F Additional Results

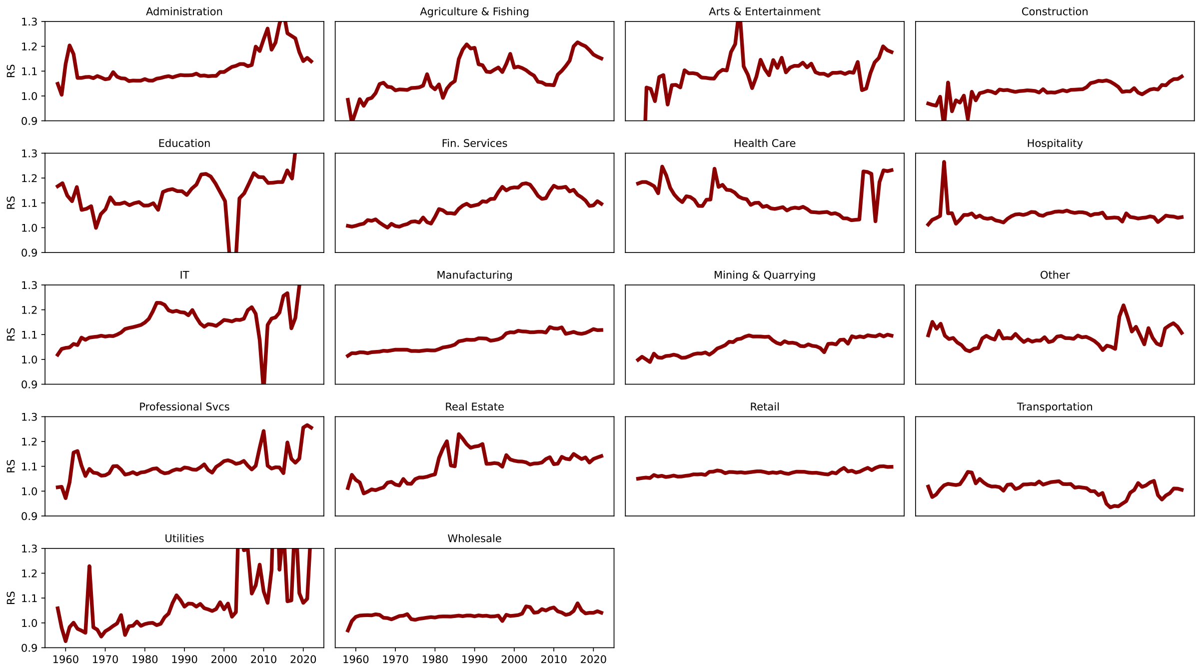
F.1 The Input–Output Multiplier in the United States

Figure 10: The U.S. Input–Output Multiplier.



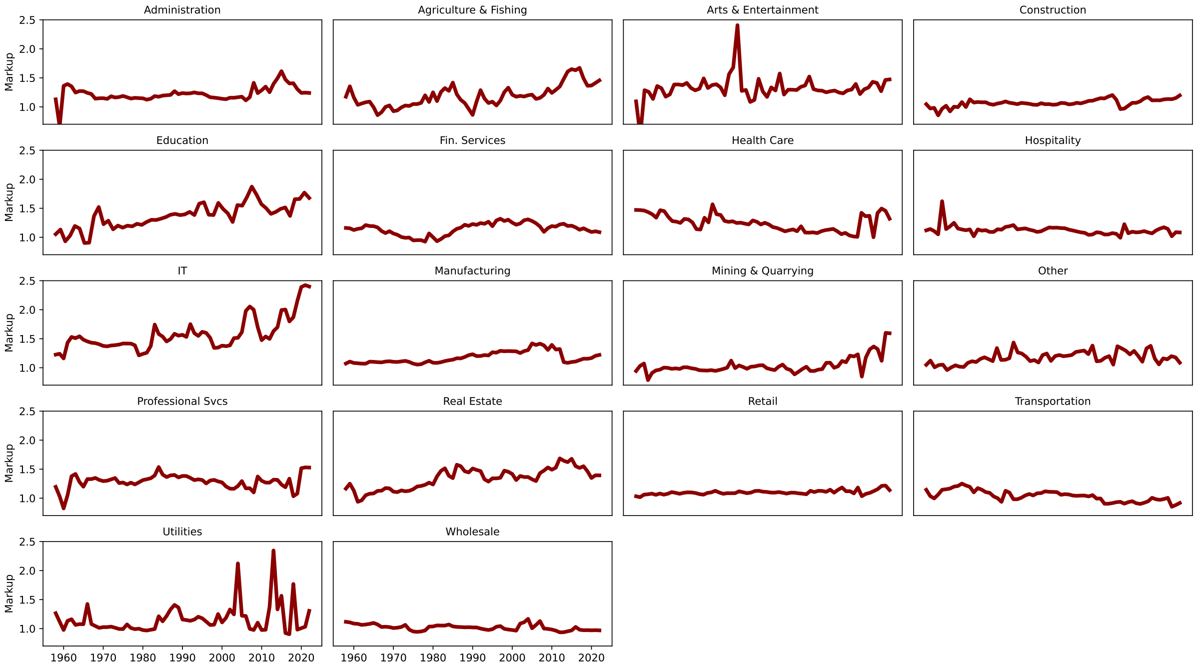
F.2 Heterogeneity in Returns to Scale

Figure 11: Returns to Scale Across Industries, 1970–2019.



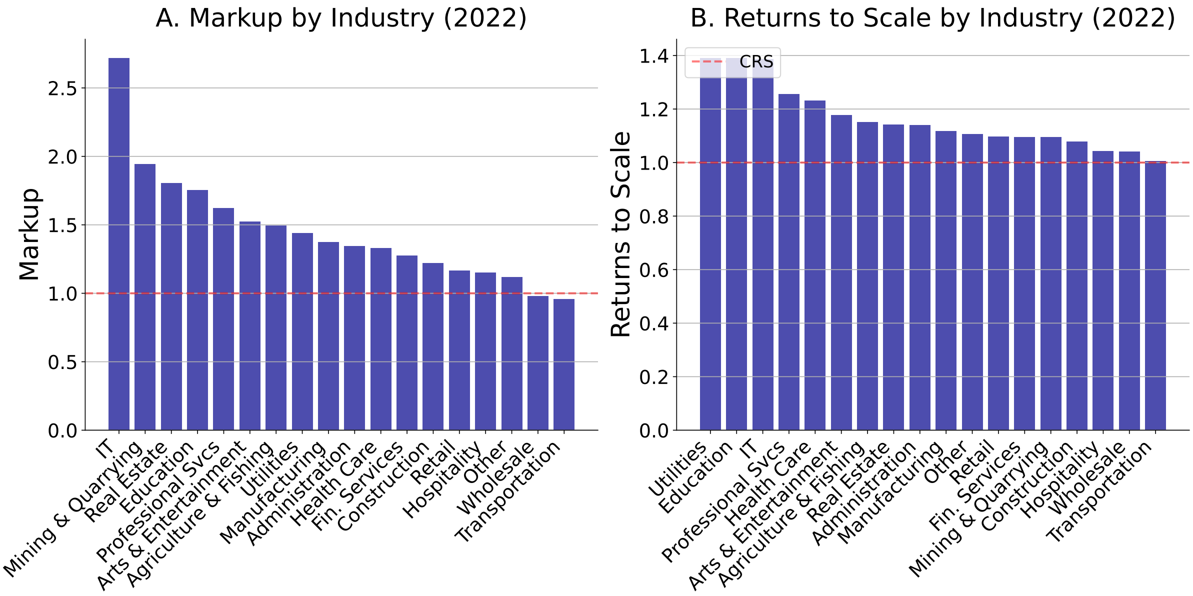
F.3 Heterogeneity in Markups

Figure 12: Industry-level Markups, 1970–2020.



F.4 Industry-Level Markups and Returns to Scale in 2022

Figure 13: Industry-level Markups and Returns to Scale.



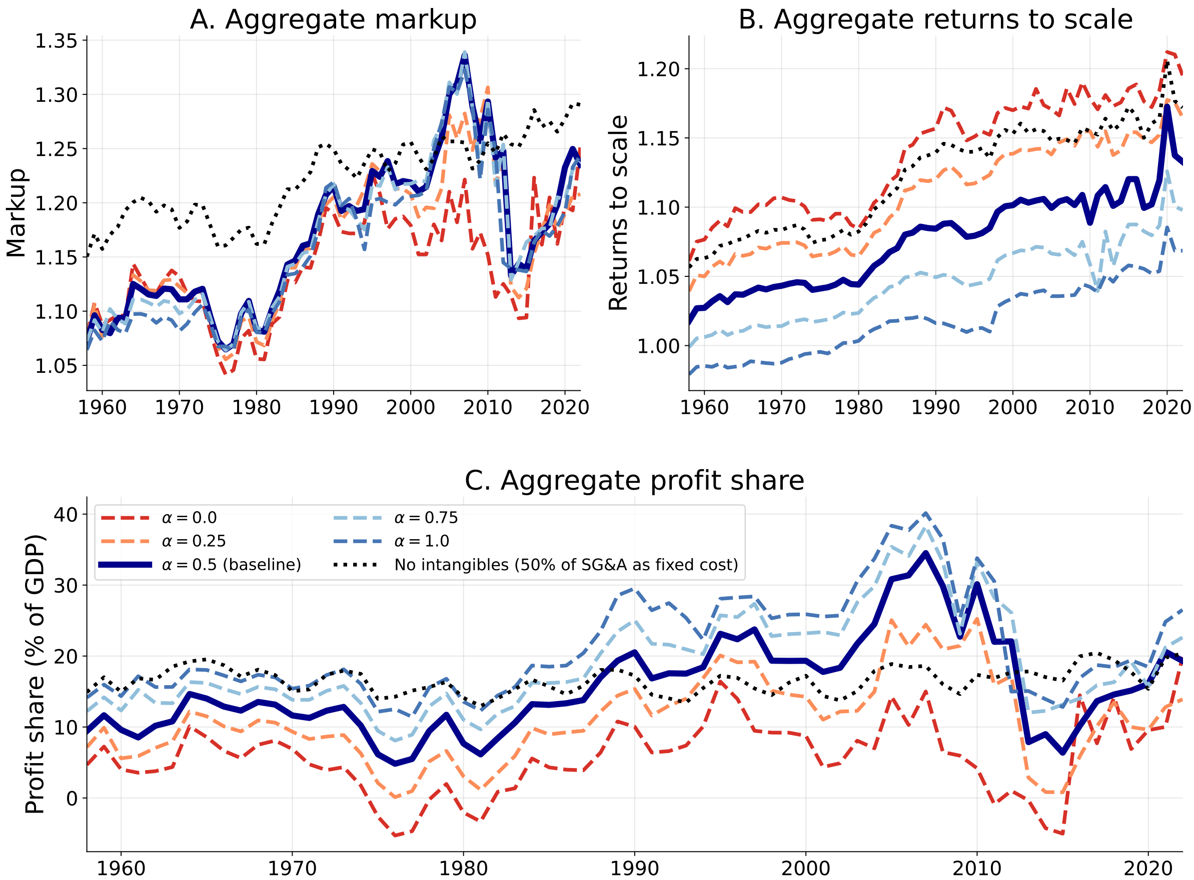
F.5 Robustness: Intangible Capital and the Split of SG&A

Figure 14 reports aggregate markups (Panel A), aggregate returns to scale (Panel B), and aggregate profit shares (Panel C) under two sets of robustness exercises. In the first, we abstract from intangible capital and treat 50% of gross SG&A (including R&D) as a fixed cost, using physical capital as the only capital input and COGS + 50% of SG&A as the variable input (dotted black line). In the second exercise, we retain our baseline specification where the capital stock is the sum of both physical and intangible capital and vary the parameter $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ that governs the share of non-capitalized SG&A that is treated as a variable input, with the complementary share $1 - \alpha$ treated as fixed cost. For each value of α, we re-estimate the production function using variable input

$$\text{OPEX}_{\alpha} = \text{COGS} + \alpha \cdot 0.70 \cdot \text{SG\&A},$$

and recompute the aggregate markups, returns to scale, and profit shares. Our baseline corresponds to $\alpha = 0.5$.

Figure 14: Robustness: Aggregate Markups, Returns to Scale, and the Profit Share.



Three features of Figure 14 are worth emphasizing. First, in Panel A, the aggregate markup series rises by roughly 15% between 1958 and 2022 in every specification. The differences across specifications are differences in the level and the volatility of the series, not its trend. For all series with intangibles, the level is approximately the same, and the only difference between those specifications are concentrated in the period of the Great Recession, 2007-2013.

Second, in Panel B, returns to scale similarly rise by 9–13% for every specification. The level of returns to scale depends on α by construction: as a larger share of SG&A is treated as fixed cost, the fixed-cost adjustment factor increases and the measured returns to scale rise mechanically. The comparison across specifications thus identifies only how much of the level is attributable to the SG&A split, not how much of the trend is. The trend itself—common to all six variants—reflects secular changes in the production structure captured by the estimated elasticities.

Third, in Panel C, the aggregate profit share is more sensitive to modeling choices than the aggregate markup or returns to scale, but its level remains economically reasonable across the empirically plausible specifications. In our baseline specification, the profit share averages about 16% of GDP. The no-intangibles variant delivers a series of very similar magnitude—roughly 17% of GDP—although it is notably less volatile. Both specifications therefore support the same broad conclusion: the U.S. aggregate profit share is on the order of 15-17% of GDP, consistent with an average profit rate of roughly 7-9%. Alternative values of α primarily shift the level of the series without altering its profile. The two extremes, $\alpha = 0$ and $\alpha = 1$, put lower and upper bounds for the full-sample average at about 6% and 21%, respectively, but these cases are best interpreted as sensitivity bounds rather than empirically plausible alternatives: $\alpha = 0$ attributes all non-capitalized SG&A to fixed costs, whereas $\alpha = 1$ treats it entirely as a variable cost and therefore eliminates fixed costs.

Taken together, the robustness exercises support two conclusions. First, the trends in aggregate markups and returns to scale are robust to the treatment of SG&A and to the inclusion of intangible capital: both rise over the sample under every specification, and their differences across specifications are differences in level rather than in trend. Second, the trend in the aggregate profit share is robust across all specifications, with our baseline in the middle serving as a reasonable specification given the role of intangibles in production and the neutral stance on the split of non-capitalized SG&A between variable and fixed costs.

G Replication and Comparison: De Loecker et al. (2020)

In this appendix, we replicate the main empirical results of De Loecker, Eeckhout and Unger (2020) (henceforth, DEU) using their original sample and specification, and compare them with our estimates. The exercise serves three purposes. First, it documents the differences in sample construction between DEU's exercise and ours, in light of the recent comment by Benkard et al. (2025) (henceforth, BMY) showing that two unstated restrictions in DEU's code drive a substantial portion of their headline result. Second, it allows us to decompose the differences between DEU's results and ours into the contributions of sample selection, the variable input, the user cost of capital, and the inclusion of intangible capital. Third, it shows that our main empirical findings are robust to alternative sample restrictions and specification choices.

G.1 Sample Comparison

DEU estimate sector-specific production functions on the U.S. nonfinancial corporate sector, using firm-level data from Compustat over 1955–2016. Their baseline sample applies several explicit restrictions (industrial format, standard data format, domestic population, consolidated accounts) and excludes finance and insurance (NAICS 52). BMY have recently shown that DEU's publicly available code applies two additional, unstated restrictions: it drops observations with missing SG&A and observations with non-positive capital. Together, these restrictions remove approximately 27 percent of the observations that satisfy DEU's stated criteria. BMY further show that re-running DEU's procedure on the unrestricted sample substantially attenuates the increase in the aggregate markup—the "rise of market power" that DEU document is, in their interpretation, partly an artifact of these unstated sample restrictions.

Our sample selection differs from both DEU and BMY in several respects. We retain DEU's format restrictions but make three changes:

1. *We require non-missing SG&A*. Our specification uses OPEX = COGS + 35% of SG&A net of R&D as the variable input. Treating firms that have positive but unreported SG&A as if their SG&A were zero would mechanically inflate our measured profit share, which is the central object in our paper. The SG&A restriction is not arbitrary and therefore dictated by the need to subtract all relevant costs when computing economic profits.

2. *We require positive capital*. Our baseline specification estimates a production function where the state variable is the total capital stock, which is the sum of both physical and intangible capital.

3. *We do not trim cost shares*. DEU's code applies an additional p1/p99 trimming on cost shares within each year. We do not impose this restriction, mainly because our user costs of capital are different as we will show below.

4. *We include the financial sector (NAICS 52)*. Our goal is to measure aggregate market power and profitability for the entire U.S. economy, not just the nonfinancial corporate sector. Excluding finance does not affect our headline results.

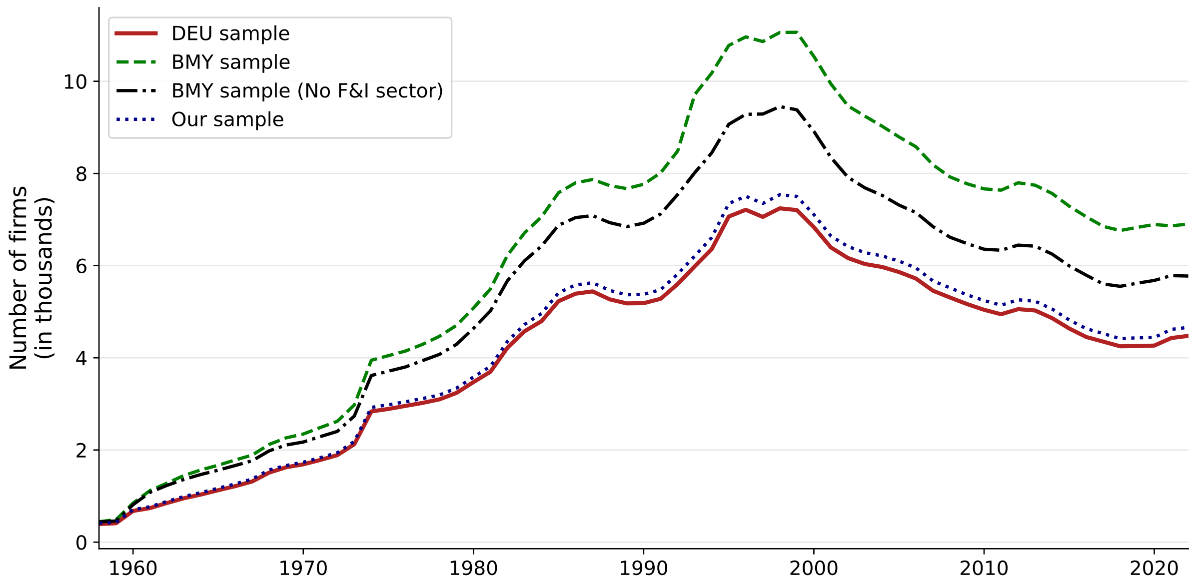
Table 3 summarizes the differences in sample selection.

Table 3: Sample selection.

Restriction	DEU 2020	BMY 2025	HP 2026 (this paper)
INDL/STD/D/C format	✓	✓	✓
Require non-missing SG&A	✓	✘	✓
Require positive capital	✓	✘	✓
Cost-share trimming (p1/p99)	✓	✘	✘
Exclude finance (NAICS 52)	✓	✓	✘
Variable input	COGS	COGS	OPEX = COGS + 35% of SG&A
User cost of capital	Assumed $(i - \pi + \delta)$	Assumed $(i - \pi + \delta)$	Recovered from data
Capital stock	PPEGT	PPEGT	PPEGT + K_INT

Figure 15 plots the number of firms in each sample over time. The DEU sample drops roughly 27 percent of firms relative to the unrestricted BMY sample. Our sample is slightly larger than the DEU sample because we do not trim cost shares, but as the figure shows the two samples are nearly identical.

Figure 15: Number of Firms in U.S. Compustat by Sample Selection Criteria.



G.2 DEU Replication

In Figure 16 , we replicate DEU's main results. In Panel A, we re-estimate their COGS-based markups on their sample and confirm that we recover their headline finding: the sales-weighted markup rises from approximately 1.20 in the early 1980s to 1.60 by 2016. We also reproduce their profit-share series, computed using their assumed user cost of capital ($r = i - \pi + 0.12$), where we use the federal funds rate as the short-term nominal interest rate and the GDP deflator to measure inflation. In both panels, we also report the corresponding series for our sample, applying their estimation procedure and user cost of capital to construct the profit rate. The aggregate markup is slightly lower in our sample throughout the period—reflecting the inclusion of firms with somewhat higher cost-to-sales ratios that DEU's cost-share trimming removes. The difference is more pronounced in the profit rate: our sample produces a profit rate that is roughly 2–3 percentage points lower than DEU's at every point in time. This gap reflects the same mechanism: by trimming the upper tail of cost shares, DEU's procedure removes firms with thin margins, which pushes the implied profit rates upward.

Figure 16: DEU Replication: Main Results.

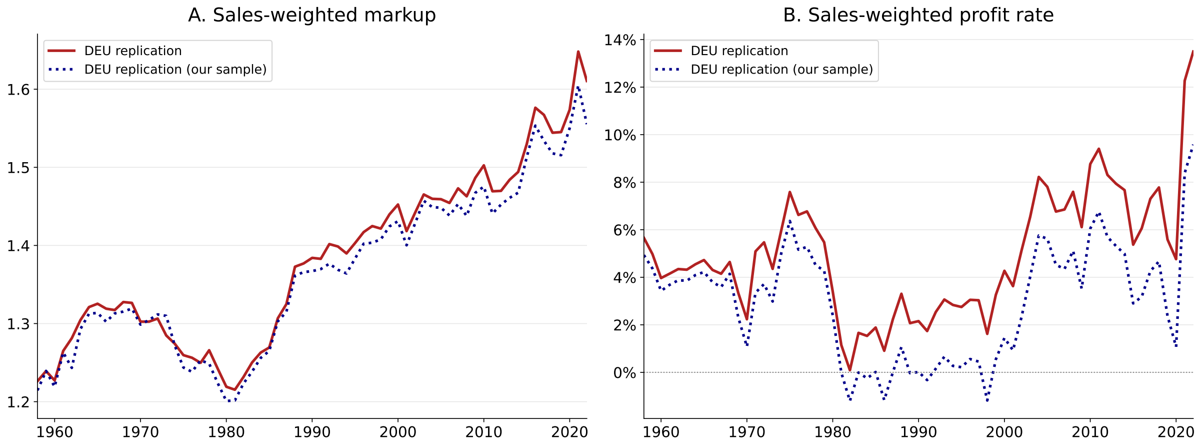


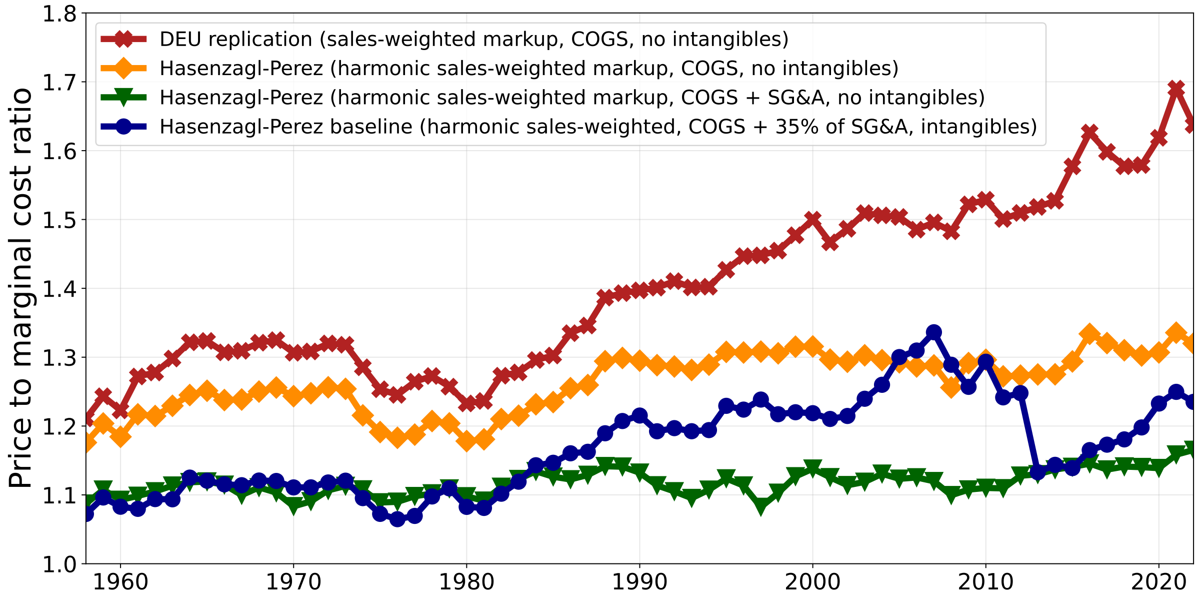
Figure Notes. Both series apply DEU's baseline specification (COGS-based markup, Cobb-Douglas with physical capital only, assumed user cost $r = i - \pi + 0.12$, where i is the federal funds rate and π is the GDP deflator). The firm-level profit rate is $1 - \text{cogs/sales} - rk/\text{sales} - \text{xsga/sales}$, treating SG&A (gross of R&D). The two series differ only in sample selection.

G.3 Benchmarking Our Estimates

Markups. Figure 17 shows four different markup series: the DEU replication on our sample and three other series that rely on the aggregation scheme implied by our theorem, but differ in the treatment of variable costs and the composition of the capital stock. In our baseline specification, the capital stock is the sum of physical and intangible capital, and variable costs are the sum of COGS and 35% of SG&A net of

R&D.[27] The two other intermediate specifications use physical capital only and differ only in whether SG&A (gross of R&D) is included in the variable input.

Figure 17: Markup Differences Between DEU 2020 and Hasenzagl–Perez 2026.



The Figure allows us to decompose the gap between DEU's markup estimate and ours into three components. First, the distance between the DEU replication (cross-marked) and the Hasenzagl–Perez COGS series (diamond-marked) isolates the effect of aggregation: holding fixed both the estimation procedure and the definition of variable input, switching from a simple sales-weighted mean to the harmonic sales-weighted mean—as implied by our theorem—accounts for about 60% of the level gap in recent years. Second, the distance between the DEU replication (cross-marked) and the Hasenzagl–Perez series (triangle-marked) captures the effect of adding SG&A to the variable input: when SG&A gross of R&D is treated as entirely variable, the rising share of SG&A in operating costs mechanically offsets the upward trend in markups, producing a much flatter series. This specification, however, is not reasonable: treating all of SG&A as variable expenses is problematic because SG&A includes fixed

[27]Since we include intangible capital in our measure of capital and follow Peters and Taylor (2017), we capitalize R&D and 30% of SG&A net of R&D as investments into intangible capital. This leaves us with 70% of SG&A net of R&D to allocate between variable and fixed costs. We adopt a neutral stance, and apply a 50-50 split, so that 35% of SG&A net of R&D enters variable costs and the remaining 35% is treated as a fixed cost. The motivation for including any part of SG&A in variable costs is that Traina (2018), Karabarbounis and Neiman (2019), and others argue that variable costs include portions of SG&A, since SG&A contains marketing expenditures, commissions, delivery expenses, and legal and accounting expenditures. While De Loecker et al. (2020) correctly argue that the categorization of variable costs should in theory not affect markup estimates—because all that is required is to identify a variable input—in practice this distinction matters quantitatively. We therefore parametrize the split as $\alpha \in [0, 1]$ of the non-capitalized SG&A going to variable costs and $(1 - \alpha)$ going to fixed costs. In Appendix F.5, we show robustness to this choice of α.

costs, R&D investments, and investments in organizational capital. Treating all these components as variable inputs biases markups downward and mutes their upward trend. Our baseline (circle-marked) series corrects this. By capitalizing R&D and 30% of SG&A net of R&D as investments in intangible capital, treating another 35% as fixed overhead, and including the remaining 35% in the variable input, our baseline restores the portion of SG&A that arguably varies with output. Because part of SG&A is treated as investment, we include intangible capital alongside physical capital in the production function. The resulting series lies above the triangle line since the mid-1980s and exhibits an upward trend like DEU's replication, but at a substantially lower level.

User costs and profit shares. DEU assume a common user cost of capital across firms, computed from aggregate data as

$$r_t = i_t - \pi_t + \delta, \tag{G.1}$$

where i_t is the effective federal funds rate (FF) and π_t is the growth rate of the GDP implicit price deflator (GDPDEF), which proxies for inflation. The depreciation rate δ is assumed to be constant and equal to 0.12 over the entire sample period.

Firm-level profit rates are then computed as

$$s_{\pi_{it}} = 1 - \frac{\theta_{jt}^{\text{COGS}}}{\mu_{it}} - \frac{r_t k_{it}^p}{p_{it} y_{it}} - \frac{\text{SG\&A}_{it}}{p_{it} y_{it}}, \tag{G.2}$$

where $\theta_{jt}^{\text{COGS}}$ is the elasticity of output with respect to COGS for the industry j in which firm i operates, μ is the markup calculated using the COGS elasticity, r is the aforementioned user cost, k^p is physical capital, py are sales, and SG&A is overhead.

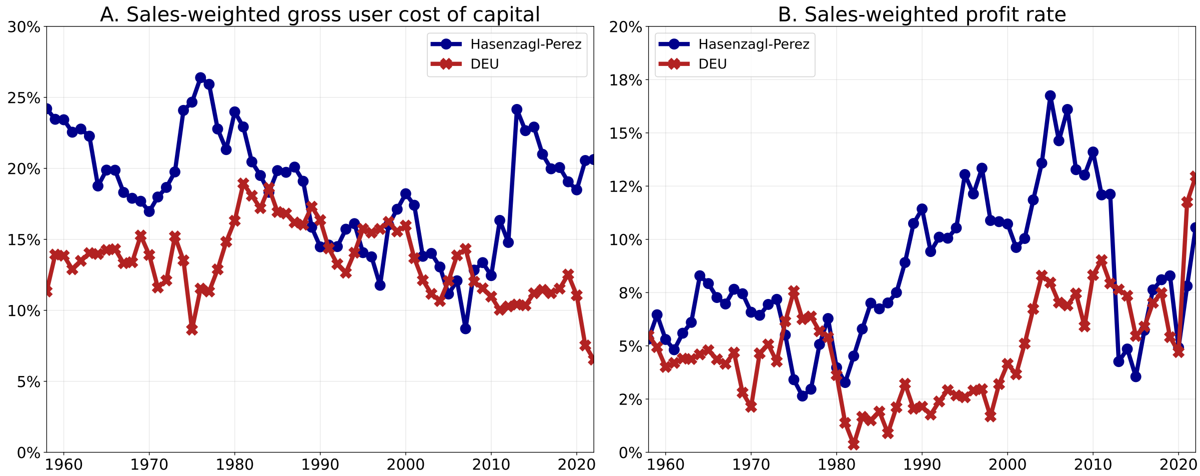
Our approach differs along three key margins. First, we compute a firm-specific user cost of capital using the shadow rental in equation (11). Second, we include both physical and intangible capital in the cost of capital term. Third, we classify fixed and variable costs differently. Since we include intangibles, we treat R&D and 30% of SG&A net of R&D as investment into intangible capital, following Peters and Taylor (2017). Accordingly, we do not include these in economic profits. Of the remaining 70% of SG&A we categorize 35% as variable inputs and 35% as fixed costs. Thus, the resulting firm-level profit rate is:

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

where $\theta_{jt}^{\text{OPEX}}$ is the elasticity of output with respect to OPEX (= COGS + 35% of SG&A) for the industry j in which firm i operates, μ_{it}^{HP} is the markup calculated using the OPEX elasticity, r^{HP} is our user cost of capital, k^{HO} is our measure of capital, which includes both physical and intangible capital, and FC (= 35% of SG&A) are fixed overhead costs.

Figure 18 compares the two approaches on our sample. Panel A reports the sales-weighted user cost of capital under the two constructions, and Panel B reports the corresponding sales-weighted profit rate.

Figure 18: User Costs and Profit Rates Comparison.



DEU's and our user-cost series measure conceptually different objects. DEU's is a market user cost derived from aggregate financial data: a no-arbitrage rental rate, common to all firms, computed independently of firm behavior. Ours is a firm-specific within-period shadow rental recovered from a restricted cost-minimization problem conditional on the installed capital stock. As shown in Appendix A, it is the reduction in minimum variable expenditure generated by a marginal unit of installed capital services, holding output fixed. Because this object depends on the firm's realized capital stock, technology, and variable-input usage, it varies across firms and over time with differences in capital intensity and production structure.

Economic profits are what remain after all productive factors have been compensated at the opportunity cost of their services. For variable inputs and observed fixed costs such as overhead, that opportunity cost is the price the firm actually pays. For installed inputs, it is the shadow rental: the value to the firm of a marginal unit of installed capital services, conditional on its realized stock, technology, and variable-input choices. Paying a firm's capital at $i_t - \pi_t + \delta$ subtracts a frictionless market benchmark rather than the relevant opportunity cost of installed services, and thus attributes to economic profits compensation that properly belongs to capital. For measuring economic profits in a heterogeneous cross-section of firms operating under adjustment costs and firm-specific technologies, the shadow rental is therefore the more appropriate object.

A second source of divergence between the two profit-rate formulas is the treatment of R&D and SG&A. If R&D and the capitalized share of SG&A build a stock of intangible capital that contributes to current production, then the relevant opportunity cost to subtract from sales is the shadow rental on that stock, not the investment flow that creates it. Equation (G.2) subtracts SG&A in full, which includes R&D, so it charges the firm for the investment flow while recording no rental compensation for the intangible stock it creates—effectively treating investment as current consumption. Our construction instead capitalizes R&D and 30% of SG&A net of R&D into intangible capital and prices this stock at its shadow rental, alongside physical capital. The gap between the two profit-rate series in Panel B therefore reflects the combined effect of two modeling choices—the concept of the user cost and the treatment of intangible investment—each of which raises our measure of economic profits relative to DEU's, especially for the period 1980–2010.