The Cost Share Approach to Production Functions

A New Semi-parametric Method for Estimating Production Functions with Noisy Data on Input Expenditures and Revenue[*]

Giulio Gottardo

University of Oxford and Pembroke College

July 9, 2025

Abstract

I introduce a new method to estimate heterogeneous output elasticities and markups using only standard firm-level revenue and expenditure data, which avoids assumptions on demand, the evolution of productivity, and output measurement. My approach circumvents common biases that affect existing methods in non-competitive settings by exploiting firms' cost-minimization behavior. To correct for bias from measurement error in input costs (e.g., capital), I develop a two-stage procedure that purges this error non-parametrically, and then uses corrected cost shares to estimate firm-time-specific elasticities, markups, and revenue productivity. Monte Carlo simulations confirm the estimator's accuracy, even with substantial measurement error from multiple sources. Applying my method to Compustat, I find markup dispersion increased six-fold and drives around 80% of revenue productivity dispersion. Among Compustat firms, 'superstar' outcomes in markups and productivity are associated not with the largest firms, but with smaller, more sunk-cost-intensive companies.

---

[*] I declare that I have no relevant or material financial interests that relate to the research described in this paper. I thank David Abraham, Filippo Biondi, Lukas Boehnert, Andrea Chiavari, Howard Smith, Ahmed Tohamy, Leonard Treuren, and especially Steve Bond for the helpful comments and discussions.

1 Introduction

A large literature seeks to measure firm-level productivity and price-cost markups to understand secular trends in resource allocation, growth, and market power (e.g., De Loecker et al. (2020); Autor et al. (2020)). However, a fundamental identification problem arises when using the standard production function framework with firm-level revenue data. Without observing physical quantities or output prices, estimates of output elasticities — which are necessary for calculating productivity and markups, and for performing counterfactuals — are generally confounded by unobserved firm-specific price variation (Klette and Griliches, 1996). Recent work has shown that this issue renders production function estimates from common methods biased unless strong, often untestable, assumptions are imposed on the structure of demand and/or production (Bond et al., 2021; Doraszelski and Jaumandreu, 2023; Biondi, 2024).

This paper develops and implements an alternative identification strategy that circumvents these problems by exploiting information from firms' cost-minimization decisions. I show that under the assumption of fixed returns to scale, a profit-maximizing firm's output elasticities are proportional to its input cost shares. This relationship is derived from first-order conditions and holds for any form of demand or production technology, thereby severing the link between estimation and the problematic unobserved output price. The main impediment to implementing this cost-share approach is that the true economic cost of some inputs is not observed. In particular, standard data on the cost of dynamic inputs like capital contains substantial measurement error, which induces a severe bias in naive cost-share calculations.

My central methodological contribution is a two-stage semi-parametric procedure that removes this type of measurement error and estimates unbiased output elasticities without imposing functional form assumptions. The first stage addresses the noise in input costs: I show that the equilibrium condition equating the ratio of marginal products to the ratio of true input costs implies that the true economic cost of capital is a non-parametric function of a firm's input choices. I leverage this to obtain a consistent estimate of this function by regressing the noisy proxy on a flexible function of input levels, thereby purging measurement error from the data. The second stage uses the resulting purged costs to construct corrected cost shares. These, in turn, identify the non-parametric output elasticity functions. Finally, output elasticity functions can be used to estimate the markup and revenue productivity. This method requires profit-maximizing firms, Hicks-neutral technology, at least one flexible input, competitive input markets, and to fix a returns to scale parameter (I also discuss how to relax the last two assumptions). This new approach does not rely on specifying a law of motion for productivity, inverting any input demand function, and does not require any data on physical output or output price proxy.

I validate the estimator's performance through extensive Monte Carlo simulations. The experiments, which feature a translog production technology and heterogeneous linear demand, confirm that the two-stage procedure recovers the true firm-level distributions of output elasticities and markups with high precision. This holds even when measurement error in the cost of capital accounts for over 95% of the observed variance in the proxy—a level of noise at which naive cost-share estimators completely break down. I also investigate the effect of other sources of measurement error (in revenue and input levels), and find that my estimator is fairly robust despite not accounting for them explicitly.

Armed with this estimator, I apply it to a comprehensive panel of U.S. listed firms from Compustat (1962-2024) to investigate the rise of 'superstar' firms. I find that markup dispersion increased six-fold over the last six decades, and that this dispersion accounts for around 80% of the variance of revenue productivity. More strikingly, my analysis reveals a new fact about the nature of superstar firms within the population of public companies. Somewhat in contrast to the narrative that market power is concentrated in the largest firms, I find a robust negative relationship between firm size (measured by sales) and both markups and revenue productivity. The highest markups are instead found in smaller listed firms, which are characterized by high sunk-cost intensity (i.e., high ratios of SGA or intangible capital to sales). This finding qualifies the current understanding of market power, suggesting that among the economy's top-performing firms, superstar outcomes are a feature of nimble, high-overhead firms rather than of the largest companies.

The paper proceeds as follows. Section 2 lays out the theoretical framework. Section 3 details the identification strategy and the two-stage estimation procedure. Section 4 discusses the existing literature and its limitations. Section 5 presents the Monte Carlo evidence. Section 6 contains the empirical application and results. Section 7 concludes.

2 Theory

In this Section I characterize a simple but general model of a profit-maximizing firm in the same spirit as Jorgenson (1963). A firm chooses two inputs to maximize its net present value taking current productivity and market conditions (residual demand) as given. One input is 'flexible„ i.e. frictionless and chosen optimally in every period. The other input is 'dynamic', i.e. accumulated over time through investment and subject to depreciation. I use this model to characterize relationships between observables (input levels and shares) and objects of interest (output elasticities and markups) in a general and common setting. These relationships will be used in the remainder of the paper to outline the issues of current methods and to construct my proposed estimator. See Appendix A for an extension to exogenous and endogenous sunk costs and Appendix B for an extension to non-competitive input markets.

2.1 Setting

Firm i at time t maximizes its net present value by choosing input X_{it} without frictions, and investment I_{it}, conditional on its capital stock K_{it}, productivity Ω_{it}^{Q} (TFPQ), and a vector of residual demand conditions Υ_{it}. Capital depreciates at rate δ_i, future profits are discounted at rate β_i, and the cost of investment is governed by the function $\phi_i(I_{it}, K_{it})$.[1] Depreciation, the discount factor, and the cost of investment are allowed differ across firms. The Bellman equation and the constraints associated with this problem are:

$$\begin{aligned} V(K_{it}, \Omega_{it}^{Q}, \Upsilon_{it}) = \max_{X_{it}, I_{it}} \; & P(Q_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \phi_i(I_{it}, K_{it}) + \beta_i \cdot \mathrm{E}_t[V(K_{it+1}, \Omega_{it+1}^{Q}, \Upsilon_{it+1})] \\ \text{s.t. } & \Omega_{it}^{Q} \cdot F(X_{it}, K_{it}) \geq Q_{it} \\ & K_{it+1} = (1 - \delta_i) \cdot K_{it} + I_{it} \end{aligned} \tag{1}$$

where $P(\cdot)$ is the residual demand function, and $F(\cdot)$ is the Hicks-neutral production function. Υ_{it} captures all the relevant differences across firms and time in terms of residual demand conditions, including the mode of competition, which is thus allowed to differ by i and t. X_{it} is measured in currency units. This is without loss of generality as long as it is purchased by the firm from a competitive market without frictions, and the differences in its price only and fully capture

---

[1] Note that, besides investment frictions $\phi_i(I_{it}, K_{it})$ can also capture any direct payment to capital.

differences in quality. I call X_{it} the flexible input, and K_{it} the dynamic, or non-flexible input interchangeably. Notice that the choice of X_{it} is static because it does not affect future profits. A somewhat implicit assumption of this model is that the firm can use the flexible input to set current output Q_{it} to the NPV-maximizing value without unanticipated shocks. This is a restriction compared to other papers, which allow for white noise shocks to TFPQ after the choice of the flexible input (Gandhi et al., 2020). I need to impose this because, in a setting with market power, unanticipated TFPQ shocks also affect the equilibrium price and markup (through quantity), which adds considerable complication to the arguments below.

2.2 FOCs and Markups

The first order condition of (1) with respect to the flexible input X_{it} is the standard static profit-maximizing solution:

$$\alpha_{it}^{R,X} = \theta_{it}^{Q,X} \cdot \left(\frac{\partial P_{it}}{\partial Q_{it}} \frac{Q_{it}}{P_{it}} + 1 \right) \tag{2}$$

where $\alpha_{it}^{R,X} = \frac{X_{it}}{P_{it} \cdot Q_{it}}$ is the revenue share of the flexible input, and $\theta_{it}^{Q,X} = \frac{\partial Q_{it}}{\partial X_{it}} \frac{X_{it}}{Q_{it}}$ is its output elasticity.

The optimal choice of investment I_{it} implies a similar equilibrium condition for capital K_{it}. The envelope condition of (1) wrt K_{it} can be rearranged as follows:

$$\begin{aligned} \frac{\partial (P_{it} Q_{it})}{\partial K_{it}} &= \frac{\partial \phi_{it}}{\partial K_{it}} + \frac{\partial V_{it}}{\partial K_{it}} - \beta_i \cdot (1 - \delta_i) \cdot \mathrm{E}_t \left[\frac{\partial V_{it+1}}{\partial K_{it+1}} \right] \\ \frac{\partial (P_{it} Q_{it})}{\partial K_{it}} &= \rho_{it} \end{aligned} \tag{3}$$

The LHS of this equation pins down the return that the marginal unit of capital yields in the current period. It follows that the RHS, ρ_{it} is the endogenous 'rental rate', or the cost of capital.[2] The expression for the cost of capital of equation (3) can be broken down into a current component $\frac{\partial \phi_{it}}{\partial K_{it}}$ (the cost part that affects current profits), plus a friction one (the other terms). In fact, if K_{it} too were flexible, all the friction terms would be zero. The results that follow do not depend on how the frictions are modeled, and can be derived similarly for alternative models with non-flexible inputs.

---

[2] For a flexible input the RHS equals the exogenous price of the input.

By differentiating revenue, $P_{it}Q_{it}$ wrt K_{it} and rearranging I obtain the revenue elasticity of capital:[3]

$$\frac{\partial(P_{it}Q_{it})}{\partial K_{it}} \cdot \frac{K_{it}}{P_{it} \cdot Q_{it}} = \theta_{it}^{Q,K} \cdot \left(\frac{\partial P_{it}}{\partial Q_{it}} \frac{Q_{it}}{P_{it}} + 1 \right) \tag{4}$$

By plugging (4) into (3), I can express the latter in the same notation as (2):

$$\alpha_{it}^{R,K} = \theta_{it}^{Q,K} \cdot \left(\frac{\partial P_{it}}{\partial Q_{it}} \frac{Q_{it}}{P_{it}} + 1 \right) \tag{5}$$

(2) and (5) do not require any restriction on the production function, demand, or the shape of investment frictions. As just mentioned, identical result can be derived for simpler static models and models with any number of inputs. In general, the profit maximizing solution requires the marginal revenue of any input to equal its 'effective marginal cost', which implies that all the appropriately defined revenue shares must equal the corresponding revenue elasticities in equilibrium.

To characterize the price-marginal cost markup, I need to first characterize marginal cost. I can solve the firm's cost minimization problem and express marginal cost as its associated Lagrange multiplier, as in De Loecker and Warzynski (2012). In the current period the firm only chooses X_{it} to minimize the cost of producing the optimal level of output:[4]

$$\begin{aligned} &\min_{X_{it}} X_{it} + \rho_{it} \cdot K_{it} \\ &\text{s.t. } \Omega_{it}^{Q} \cdot F(X_{it}, K_{it}) \geq Q_{it} \end{aligned} \tag{6}$$

where ρ_{it} is the user cost of capital, defined by (3), which the firm takes as given when choosing X_{it}. This happens because current period capital K_{it} is not a choice variable, as it cannot be adjusted by the firm in the current period. Manipulating the first order condition with respect to X_{it} yields the well known expression that relates the price-marginal cost markup to the output elasticity of the flexible input and its share of revenue:

$$\mu_{it} := \frac{P_{it}}{\lambda_{it}} = \frac{\theta_{it}^{Q,X}}{\alpha_{it}^{R,X}} \tag{7}$$

---

[3] K_{it} also has a second-order effect on profits through the choice of X_{it}. However (2) ensures that this is zero in equilibrium (Envelope Theorem).

[4] Since there is only one flexible input in this setting, the solution to the cost minimization problem is trivial. However, it is still useful to characterize marginal cost.

where λ_{it} is the Lagrange multiplier of the cost minimization problem, which equals the marginal cost of output. Comparing (2) to (7) allows me to conclude that in the profit maximizing equilibrium $\mu_{it} = \left(\frac{\partial P_{it}}{\partial Q_{it}} \frac{Q_{it}}{P_{it}} + 1\right)^{-1}$. This result can be plugged into (5) to also write:

$$\mu_{it} = \frac{\theta_{it}^{Q,K}}{\alpha_{it}^{R,K}} \tag{8}$$

Notice that I can treat capital *as if* it were a flexible input despite its dynamic nature because of how I defined the rental rate. To see this, notice that the rental rate ρ_{it} is the value which equates capital's output elasticity and revenue share when there is no market power ($\mu_{it} = 1$).

The definition of the cost of capital provided by (3) allows me to characterize the *cost* shares of inputs. Let the returns to scale of the production function be $\mathcal{T} := \theta_{it}^{Q,X} + \theta_{it}^{Q,K}$, I can sum (7) and (8) and rearrange:

$$\mu_{it} = \mathcal{T} \cdot \frac{P_{it} Q_{it}}{X_{it} + \rho_{it} K_{it}} \tag{9}$$

This expression is the same as the accounting markup when capital is flexible too, i.e. the ratio of revenue and total cost, scaled by the returns to scale (this is used as a markup measure in Antràs et al. (2017); Autor et al. (2020)). Like (8), (9) holds as long as the economic cost of the non-flexible input(s) is appropriately defined. Now I can plug (9) back into (7) and (8) to express output elasticities as 'true' cost shares:

$$\begin{aligned} \alpha_{it}^{C,X} \cdot \mathcal{T} &= \theta_{it}^{Q,X} \\ \alpha_{it}^{C,K} \cdot \mathcal{T} &= \theta_{it}^{Q,K} \end{aligned} \tag{10}$$

where $\alpha_{it}^{C,X} := \frac{X_{it}}{X_{it} + \rho_{it} K_{it}}$ is the true cost share of the flexible input, and $\alpha_{it}^{C,X}$ is the true cost share of the dynamic input. (9) and (10) only require the firm to be profit maximizing, and no further structure on demand or production beyond the returns to scale $\mathcal{T}$.

The obstacle to using cost shares and accounting markups in data applications is that the economic cost of non-flexible inputs is rarely observed, and thus cost shares and accounting markups are contaminated by what is essentially measurement error in the price of inputs. In the following Section I deal with this issue by leveraging (4), which expresses the true economic cost of an input as a function of current variables.

3 Identification of Output Elasticities

In this Section I assume that a researcher observes noisy revenue $\tilde{R}_{it}$, expenditure on the flexible input X_{it}, the level of the dynamic input K_{it}, and a noisy proxy for its economic cost $\tilde{\rho}_{it}$. Data on $\left[\tilde{R}_{it}, X_{it}, K_{it}, \tilde{\rho}_{it}\right]$ is available for a large N, small T sample of firms ($T \geq 1$ since I do not need time variation for identification). These firms have the same production function $F(\cdot)$, but face different residual demand schedules and productivities. I do not specify the data generating process of TFPQ Ω_{it}^{Q} nor that of the residual demand vector Υ_{it}. I also make no assumption on the functional form of the production function, besides the returns to scale.

I characterize the conditions under which output elasticities are identified in this setting with unmodeled and heterogeneous market power, and noisy data on revenue and input expenditures. My arguments require the following assumptions: (i) fixed returns to scale in the production function $\theta_{it}^{Q,X} + \theta_{it}^{Q,K} = \mathcal{T}$ for every i, t; (ii) at least one input is flexible (X_{it} in this case); (iii) the flexible input is chosen to maximize current profits, as in (1);[5] (iv) quantity productivity Ω_{it}^{Q} is Hicks-neutral; (v) there is no market power in input markets. See Appendix B for a discussion on how to relax and/or test (i) and (v). Note that I only require the returns to scale to be fixed at some level $\mathcal{T}$, but constant returns to scale $\mathcal{T} = 1$ is by far the most common in the recent literature (Gandhi et al., 2020; Flynn et al., 2019).

3.1 Noisy Data

As anticipated in Section 2, with perfect data on input expenditures it would be possible to directly measure output elasticities by calculating the corresponding cost shares from the data, and this paper would not be interesting to write or read. The main practical obstacle to the cost share approach is that input expenditures are not perfectly observed. This is generally most concerning for inputs that face adjustment frictions, such that their economic cost differs from payments to the input. In the first half of this Section I explicitly characterize the bias under the most plausible type of error: additive noise in the rental rate of capital, i.e. when the data records $\tilde{\rho}_{it}$, a noisy

---

[5] This is weaker than simple profit maximization, because it allows for dynamic considerations to affect the choice of K_{it}.

proxy of the true economic cost of capital instead of ρ_{it}:

$$\tilde{\rho}_{it} = \rho_{it} + \varepsilon^{\rho}_{it} \tag{11}$$

where ε^{ρ}_{it} is white noise, i.e. $\mathrm{E}[\varepsilon^{\rho}_{it}|x_{it}, k_{it}] = 0$. This form of measurement error in the cost of capital is present in the data if the firm-level rental rate ρ_{it} is estimated with the Capital Asset Pricing Model (CAPM) (see Fama and MacBeth (1973) for a seminal application). In Appendix C (11) is derived from the empirical CAPM.

Unlike most of the existing literature, I do not need to take a specific stance on measurement error in revenue or untransmitted TFPQ shocks for the estimation of output elasticities, because my estimator relies on input expenditures alone. However, to estimate markups and TFPR revenue data is also required. In line with the literature, I assume measurement error in revenue is log-additive:

$$\tilde{r}_{it} = p_{it} + q_{it} + \varepsilon^{R}_{it} \tag{12}$$

where ε^{R}_{it} is white noise, i.e. $\mathrm{E}[\varepsilon^{R}_{it}|x_{it}, k_{it}] = 0$.

3.2 Identification of the Output Elasticity with Market Power and Noisy Data

Gandhi et al. (2020) show that in a competitive setting with noisy revenue the output elasticity of the flexible input is non-parametrically identified from the log of its first order condition, equation (7) with $\mu_{it} = 1$ for every i and t:

$$\log \alpha^{R,X}_{it} = \log \theta^{Q,X}(x_{it}, k_{it}) - \varepsilon^{R}_{it} \tag{13}$$

The log of the elasticity is identified because ε^{R}_{it} is white noise and thus $\mathrm{E}[\log \alpha^{R,X}_{it}|x_{it}, k_{it}] = \log \theta^{Q,X}(x_{it}, k_{it})$.[6] This identification argument fails in the presence of market power without

---

[6] The only difference between this argument and that of Gandhi et al. (2020) is that they model ε^{R}_{it} as a unanticipated TFPQ shock, while I model it as measurement error in revenue. This difference implies that I do not need make identification arguments about $\mathrm{E}[\exp(\varepsilon^{R}_{it})]$, because my shock does not affect true revenue, but just its observed counterpart.

additional markup information. (13) becomes:

$$\log \alpha_{it}^{R,X} = \log \theta^{Q,X}(x_{it}, k_{it}) - \log \mu_{it} - \varepsilon_{it}^{R} \tag{14}$$

It immediately follows that $\mathrm{E}[\log \alpha_{it}^{R,X} | x_{it}, k_{it}] \neq \log \theta^{Q,X}(x_{it}, k_{it})$ because the log of the markup is not zero as in the competitive setting. In addition, the markup may depend on output q_{it} and demand conditions Υ_{it}. Since output is in turn a function of input levels and TFPQ, I can write $\log \mu_{it} = \log \mu(x_{it}, k_{it}, \omega_{it}^{Q}, \Upsilon_{it})$. This is useful to show that even the *revenue* elasticity is not non-parametrically identified with information on input levels alone: although it is true that $\mathrm{E}[\log \alpha_{it}^{R,X}] = \mathrm{E}[\log \theta^{Q,X}(x_{it}, k_{it}) - \log \mu_{it}]$ (unconditional expectation), I should *also* condition on unobserved ω_{it}^{Q} and Υ_{it} to write $\mathrm{E}[\log \alpha_{it}^{R,X} | x_{it}, k_{it}, \omega_{it}^{Q}, \Upsilon_{it}] = \log \theta^{Q,X}(x_{it}, k_{it}) - \log \mu(x_{it}, k_{it}, \omega_{it}^{Q}, \Upsilon_{it})$. In words, unless output and all relevant demand information are observed alongside input levels, neither the output nor the revenue elasticity are identified.

The Gandhi et al. (2020) identification argument relies on a competitive output market ($\mu_{it} = 1$ for every i and t), which in turn implies that the output elasticity and the revenue share coincide. In addition, any concern relating to measurement error in the true economic cost of non-flexible inputs can be ignored because there is no need for total cost data to calculate the revenue share. Departing from the competitive output market assumption, the relationship between *cost* shares and output elasticities, equation (10) still holds. If the returns to scale parameter is $\mathcal{T}$ and *input expenditures are perfectly observed*, I can write:

$$\alpha_{it}^{C,X} \cdot \mathcal{T} = \theta^{Q,X}(x_{it}, k_{it}) \tag{15}$$

It immediately follows that $\mathrm{E}\left[\alpha_{it}^{C,K} \cdot \mathcal{T} | x_{it}, k_{it}\right] = \theta^{Q,X}(x_{it}, k_{it})$. However, this semi-parametric[7] identification argument does not apply when there is measurement error in the cost of capital, which is bound to be a concern when $\tilde{\rho}_{it}$ is an estimate itself (see Appendices C and H).

Say that the cost of capital is noisily observed, with additive error as in (11), i.e. $\tilde{\rho}_{it} = \rho_{it} + \varepsilon_{it}^{\rho}$ s.t.

---

[7] (15) is not fully non-parametric because the returns to scale are fixed at $\mathcal{T}$.

$\mathrm{E}[\varepsilon^{\rho}_{it}|x_{it}, k_{it}] = 0$. With some algebraic manipulation (15) becomes:

$$\begin{aligned} &\tilde{\alpha}^{C,X}_{it} \cdot \mathcal{T} = \theta^{Q,X}(x_{it}, k_{it}) + u_{it} \\ &\text{s.t.} \quad u_{it} = -\frac{\varepsilon^{\rho}_{it} X_{it} K_{it}}{(X_{it} + (\rho_{it} + \varepsilon^{\rho}_{it})K_{it})(X_{it} + \rho_{it} K_{it})} \end{aligned} \tag{16}$$

where ε^{ρ}_{it} is the additive error in the cost of capital. As the residual u_{it} depends on ε^{ρ}_{it} and input levels in a non-linear fashion, $\mathrm{E}[u_{it}|x_{it}, k_{it}] \neq 0$. It immediately follows that $\mathrm{E}\left[\tilde{\alpha}^{C,K}_{it} \cdot \mathcal{T}|x_{it}, k_{it}\right] \neq \theta^{Q,X}(x_{it}, k_{it})$. In words, measurement error in the cost of an(y) input introduces bias. As shown in Appendix H with Monte Carlo simulations, this is also the case when ρ_{it} is measured with log-additive measurement error.

When measurement error in the cost of the dynamic input is additive, it is not even possible to sign the bias in $\mathrm{E}\left[\tilde{\alpha}^{C,K}_{it} \cdot \mathcal{T}|x_{it}, k_{it}\right]$, as established by Proposition 1.

Proposition 1. *If* $\tilde{\rho}_{it} = \rho_{it} + \varepsilon^{\rho}_{it}$ *with* $\varepsilon^{\rho}_{it} \in \mathbb{R}$ *and* $E[\varepsilon^{\rho}_{it}|x_{it}, k_{it}] = 0$*, then* $E\left[\tilde{\alpha}^{C,K}_{it} \cdot \mathcal{T}|x_{it}, k_{it}\right]$ *does not exist.*

Proof. See Appendix D. ☐

The intuition behind Proposition 1 and its proof is that negative noise shocks (when $\varepsilon^{\rho}_{it} \sim -\frac{X_{it}}{K_{it}} - \rho_{it}$) can cause u_{it} to diverge to both plus and minus infinity. In the data, these materialize as extreme cost shares (both implausibly high and low), which prevent inference on the output elasticity. I find evidence of this in the Monte Carlo simulations of Section 5, where a naive cost share estimator is implemented, whose reliability steeply falls as the variance of ε^{ρ}_{it} is increased, while its bias cannot be signed.

3.3 Purging Measurement Error from the Cost of Inputs (First Stage)

By definition, the economic cost of an input must equal the value it generates in the current period for a profit-maximizing firm. For dynamic inputs that face adjustment costs, this is captured by equation (3), which pins down the current rental rate. But regardless of how the nature of an input is modeled (static-dynamic, with-without adjustment frictions), its true economic cost must be consistent with an equation of the form of (2) or (5), i.e. its expenditure's share of revenue must equal its revenue elasticity. If I plug (7) into (3) and rearrange, I obtain:

$$\rho_{it} = \frac{\theta_{it}^{Q,K} X_{it}}{\theta_{it}^{Q,X} K_{it}} \tag{17}$$

Note that this is just a general version of the standard static condition for cost-minimizing firms, i.e. 'the ratio of expenditures equals the ratio of marginal products'. The reason why I can restate it for dynamic inputs too is that their economic cost is appropriately defined to capture all frictions.[8]

The RHS of (17) is a non-linear function of input levels alone, since both output elasticities are at most functions of input levels due to the assumption of Hicks-neutrality. Plugging (17) into (11) suggests the following non-parametric regression equation:

$$\tilde{\rho}_{it} = \rho(x_{it}, k_{it}) + \varepsilon_{it}^{\rho} \tag{18}$$

where $\rho(\cdot)$ is a non-linear function, common across firms that share the same production function. Notice that $\mathrm{E}[\varepsilon_{it}^{\rho}|x_{it}, k_{it}] = 0$ by the definition of ε_{it}^{ρ}. It follows that I can obtain consistent firm-year estimates of the true economic cost of the dynamic input $\hat{\rho}_{it}$ by running this regression of noisy data on input levels.

Whereas (18) can purge virtually any amount of white noise measurement error from an input's price, it cannot remove bias in the average level, i.e. it does not deal with $\mathrm{E}[\tilde{\rho}_{it}] \gtrless \mathrm{E}[\rho_{it}]$. In fact, it will always be the case that $\mathrm{E}[\hat{\rho}_{it}] = \mathrm{E}[\tilde{\rho}_{it}]$, which implies that any bias in level will produce an identical bias in the level of the noise-purged cost of the dynamic input. For example, if ε_{it}^{ρ} is log-additive instead of simply additive, this non-parametric regression can still recover the true distribution of ρ_{it}, but Jensen's Inequality ensures that $\mathrm{E}[\hat{\rho}_{it}] = \mathrm{E}[\tilde{\rho}_{it}] \geq \mathrm{E}[\rho_{it}]$.

In practice, for listed companies $\tilde{\rho}_{it}$ can be a CAPM-based estimate of the weighted average cost of capital (WACC), or just the cost of equity depending on assumptions and data availability. Both tend to involve the choice of a safe asset return (e.g. the real yield of 10-year Treasuries) and a depreciation rate (which can be sector-time specific), and the estimation of an empirical CAPM

---

[8] If X_{it} were not the numeraire, I could write an equivalent (albeit trivial) representation of its price, i.e. $P_{it}^{X} = \frac{\theta_{it}^{Q,X} \rho_{it} K_{it}}{\theta_{it}^{Q,K} X_{it}}$.

from accounts/stock market data (Fama and MacBeth, 1973; Fama and French, 1993, 2015), and will introduce mean-zero, additive error in $\tilde{\rho}_{it}$. Alternatively, data on firms' perceptions about their cost of capital can also be used (Gormsen and Huber, 2024). For data on non-listed firms, interest expenditures and dividends can be employed to construct proxies of the cost of debt and equity, respectively. I do not need to take a specific stance on the different merits of these approaches and their alternatives, given that my method is specifically designed to purge measurement error from any of these. Clearly, if $\tilde{\rho}_{it}$ is a CAPM or augmented-CAPM estimate, purging should be performed in levels. In general, depending on the data at hand the researcher must take a stance on whether $\tilde{\rho}_{it}$ has to be purged in levels or in logs, since this will affect average cost shares. However, misspecifying (18) in levels when the true error is log-additive only introduces a minor amount of bias, as shown in Appendix H.

Irrespective of the origin of $\tilde{\rho}_{it}$, data on $[X_{it}, K_{it}, \tilde{\rho}_{it}]$ can be used to estimate (18) non-parametrically, e.g. by approximating $\rho(\cdot)$ with a polynomial and using OLS. Note that the function $\rho(\cdot)$ is time and firm-invariant by definition if firms share the same production function, but considerations on how $\tilde{\rho}_{it}$ is measured may warrant splitting the sample for estimation. For example, if a year-specific risk-free return is used within a CAPM, then (18) can be estimated separately for every year to avoid level issues. The objects of interest of the first stage are the fitted values, $\hat{\rho}_{it}$. These can be used to construct purged cost shares for all inputs, in this case $\hat{\alpha}^{C,X}_{it}$ and $\hat{\alpha}^{C,K}_{it}$.

3.4 Estimation of Output Elasticities with Purged Cost Shares (Second Stage)

For the second stage, purged cost shares $\hat{\alpha}^{C,X}_{it}$ and $\hat{\alpha}^{C,K}_{it}$ can be used to estimate (15) semi-parametrically:

$$\begin{aligned}\hat{\alpha}^{C,X}_{it} \cdot \mathcal{T} &= \theta^{Q,X}(x_{it}, k_{it}) + u^{SS,X}_{it} \\ \hat{\alpha}^{C,K}_{it} \cdot \mathcal{T} &= \theta^{Q,K}(x_{it}, k_{it}) + u^{SS,K}_{it}\end{aligned} \tag{19}$$

where $\theta^{Q,X}(\cdot)$ and $\theta^{Q,K}(\cdot)$ can be approximated with polynomials. If the first stage perfectly purges measurement error from the cost of the dynamic input, and if the true functional forms for the output elasticities are specified, then (19) delivers a perfect fit, i.e. $u^{SS,X}_{it} = u^{SS,K}_{it} = 0$ for every i and t. However, in light of the fact that both stages rely on non-parametric approximations, $u^{SS,X}_{it}$ and $u^{SS,K}_{it}$ are allowed to differ and take non-zero values. This can also happen if the first stage fails at fully removing measurement error from $\tilde{\rho}_{it}$,[9]. The Monte Carlo simulations of Section 5 and in the Appendices show that performing the first stage always achieves a considerable improvement over using naive cost shares, even when the purging of ε^{ρ}_{it} is not perfect, or when the capital stock K_{it} is noisily measured itself.

Second stage estimates of the output elasticity of the flexible input can be used to calculate markups with the De Loecker and Warzynski (2012) production approach:

$$\hat{\mu}_{it} = \frac{\hat{\theta}^{Q,X}_{it}}{\tilde{\alpha}^{R,X}_{it}} \tag{20}$$

Notice that the revenue share $\tilde{\alpha}^{R,X}_{it} := \frac{X_{it}}{\tilde{R}_{it}}$ is unaffected by measurement error in the cost of the dynamic input, but can be noisy because of measurement error in revenue. If revenue is measured with error, then the markup can only be recovered up to this form of noise, which cannot be purged with standard methods in the presence of heterogeneous market power and revenue data (see Section 4, Doraszelski and Jaumandreu (2023), and Biondi (2024)). Finally, I can also integrate the output elasticities over the corresponding inputs to estimate the TFPR distribution up to measurement error in revenue:

$$\hat{\omega}^{R}_{it} = (\widehat{\omega^{Q}_{it} + p_{it}} + \varepsilon^{R}_{it}) = \tilde{r}_{it} - \int_0^{x_{it}} \hat{\theta}^{Q,X}(x, k_{it})\, \mathrm{d}x - \int_0^{k_{it}} \hat{\theta}^{Q,K}(0, k)\, \mathrm{d}k \tag{21}$$

4 Identification Issues in the Previous Literature

In this section, I use the model of Section 2 to characterize the identification issues that researchers face when working with production data outside of a competitive setting, in order to outline some advantages of the new proposed approach.

Standard production function methods (Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003; Ackerberg et al., 2015) can fail at the estimation of output elasticities, markups and TFPR in non-competitive settings for several reasons. I separate these depending on the data available to the researcher, i.e. whether physical output quantities Q_{it} are observed. For more extensive treatment of some of the issues I am about to discuss, see, among others, Bond

---

[9] As hinted to above, this may happen if $\tilde{\rho}_{it}$ contains insufficient variation to capture the heterogeneity of ρ_{it}, or because ε^{ρ}_{it} is not additive.

et al. (2021); Doraszelski and Jaumandreu (2023); Biondi (2024). Existing share-based approaches (De Loecker and Fleitas, 2021; Raval, 2023; De Loecker et al., 2024), albeit closer to my own proposal of Section 3, appear to be limited to a Cobb Douglas production function, or to competitive settings (Gandhi et al., 2020).

4.1 Quantities Are Observed

With the exception of the panel approach of Blundell and Bond (2000) - where an AR(1) process for log TFPQ is assumed - production function estimation methods rely on proxying for TFPQ with observables (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015). The proxy approach is justified with the inversion of the demand for the flexible input. In a perfectly competitive setting I can write (lowercases denote logs):

$$x_{it} = x(\omega_{it}^{Q}, k_{it}) \tag{22}$$

i.e. x_{it} is a function of k_{it} and log TFPQ ω_{it}^{Q} only (the 'scalar unobservable' assumption). Perfect competition also implies - for reasonably well behaved production functions - that $x(\cdot)$ is increasing in ω_{it}^{Q}, which in turn means that it is invertible. It immediately follows that I can proxy TFPQ with observables as $\omega_{it}^{Q} = x^{-1}(x_{it}, k_{it})$. However, as shown by Doraszelski and Jaumandreu (2023), in a setting with market power such as the one of Section 2, the demand for the flexible input also depends on market conditions:

$$x_{it} = x(\omega_{it}^{Q}, k_{it}, \Upsilon_{it}) \tag{23}$$

Unless the full vector of residual demand conditions Υ_{it} is observed, the scalar unobservable assumption is not met because differences in the level of x_{it} are not only caused by TFPQ differences. Finding a valid proxy for Υ_{it} to avoid this problem is hard for two reasons: (i) the validity of any sufficient statistic corresponds to some assumption on demand and conduct (e.g. those discussed in Ackerberg and De Loecker (2024)); (ii) a price or markup proxy (such as the one I will discuss in Section 3) may not suffice in this context because the researcher must capture all demand differences across firms for the successful inversion of $x(\cdot)$. Two firms may happen to charge the same price at the same markup, but face different demand schedules. In this case a valid price or markup proxy is a bad demand proxy.

Now suppose Υ_{it} is observed. Biondi (2024) shows that the relationship between TFPQ and input demand is not necessarily monotonic. This prevents the researcher from inverting $x(\omega_{it}^{Q}, k_{it}, \Upsilon_{it})$ for TFPQ because the function is not monotonic.[10] In light of this, both Doraszelski and Jaumandreu (2023) and Biondi (2024) suggest using proxy methods with considerable extra care, or using the panel approach by Blundell and Bond (2000), which imposes additional structure on the TFPQ process. My approach avoids all these issues by not relying on any inversion of input demand, nor on specifying the law of motion of TFPQ/TFPR. By not requiring data on physical output to be implemented, my method is also avoids any concerns relating to the measurement of Q_{it}, whether across firms (e.g. quality) or within firm (multiproduct firms).

4.2 Quantities Are Not Observed

In many - if not most data applications physical output is not observed at the firm-level. What follows also applies if output prices differ across firms, but firm revenue is deflated with economy or industry-wide price deflators, which cannot yield a valid proxy for physical output. In fact, common deflators cannot capture price heterogeneity, so the deflated variable will still contain price variation.

The mapping from inputs to quantity (the production function) differs from that of inputs to revenue (the revenue function) when output markets are not competitive. In logs, revenue is:

$$r_{it} = p_{it} + q_{it} \tag{24}$$

In a non-competitive setting like the one of Section 2, (2) and (7) hold in equilibrium:

$$\theta_{it}^{R,X} := \frac{\partial r_{it}}{\partial x_{it}} = \frac{\theta_{it}^{Q,X}}{\mu_{it}} \tag{25}$$

i.e. the revenue elasticity is the output elasticity divided by the markup.[11]

---

[10] $x(\Omega_{it}^{Q}, k_{it}, \Upsilon_{it})$ is monotonic in TFPQ if the residual demand slope faced by every firm is CES (Biondi, 2024). It follows that under the assumption of CES demand it is possible to use proxy methods when Q_{it} is observed.

[11] This is a generalization of the same result on industries with common CES demand of Klette and Griliches (1996), and is also present in Bond et al. (2021); Bond and Gottardo (2024).

It may be tempting to try and estimate the *revenue* function akin to the *production* function. In the simple Cobb Douglas production and CES demand case the revenue function reduces to:

$$r_{it} = \tilde{\omega}_{it}^{R} + \frac{\theta^{Q,X}}{\mu_{it}} x_{it} + \frac{\theta^{Q,K}}{\mu_{it}} k_{it} \tag{26}$$

where $\tilde{a}_{it}^{R}$ is the residual of the revenue function.[12] It follows that the revenue function has heterogeneous coefficients in the presence of heterogeneous markups.[13] Without the restrictive CES demand assumption, markups are also equilibrium objects, i.e. they will depend on input levels and unobserved primitives (Ω_{it}^{Q} and Υ_{it}). Therefore, unless demand is explicitly modeled in a way that allows the researcher to rewrite the revenue function as a function of of input levels and primitives only, attempting to directly estimate (26) will result in bias from model misspecification.

Some papers have attempted to solve the issues with the revenue function by using price proxies (De Loecker et al., 2020) to control for output prices and estimate the production function parameters without price bias (De Loecker et al., 2020; Ackerberg and De Loecker, 2024). The price proxy approach consists in using a vector of price proxies $\mathbf{z}_{it}^{P}$ to separate log output from log price:

$$r_{it} = q_{it} + p_{it} = \omega_{it}^{Q} + f(x_{it}, k_{it}) + p(\mathbf{z}_{it}^{P}) \tag{27}$$

In order to successfully separate prices from output, the price proxies must capture all price differences across i and t. If this is not the case, the residual will contain price variation and the argument above about price bias applies. Ackerberg and De Loecker (2024) show that market shares can proxy for output prices when demand takes a specific logit form. Beyond the issues raised in Doraszelski and Jaumandreu (2023) on the correct measurement of these market shares, it is worth noting that any violation of the assumption about the functional form of residual demand results in price bias. It follows that any application of De Loecker and Warzynski (2012) where an output elasticity is estimated with price proxies is not agnostic about demand.

Flynn et al. (2019) and Kirov et al. (2023) follow a different route, by proposing to control for revenue TFP, i.e. $\omega_{it}^{R} = \omega_{it}^{Q} + p_{it}$ by using its law of motion. This is in the same fashion as most

---

[12] $\tilde{\omega}_{it}^{R}$ differs from revenue TFP, which the literature defines as the sum of quantity TFP and price (in logs), i.e. $\omega_{it}^{R} = \omega_{it}^{Q} + p_{it}$.

[13] The only revenue function that is straightforward to estimate is the one covered in Klette and Griliches (1996), which is the same as (26) but with common markups, i.e. $\mu_{it} = \mu$ for every i and t.

of the production function literature in competitive settings, where a common first order Markov process is assumed for log TFPQ, i.e. $\omega_{it}^{Q} = g(\omega_{it-1}^{Q}) + \xi_{it}^{Q}$. Their strategy consists in doing the same for TFPR:

$$r_{it} = f(x_{it}, k_{it}) + \omega_{it}^{R} = f(x_{it}, k_{it}) + g(r_{it-1} - f(x_{it-1}, k_{it-1})) + \xi_{it}^{R} \tag{28}$$

In a sense, the law of motion of TFPR plays the role of price proxy because it captures both unobserved TFPQ *and* price differences. As such, it is subject to the same limitations of any price proxy approach, i.e. it is underpinned by assumptions about the functional form (and dynamics in this case) of residual demand. First, for the TFPR shock ξ_{it}^{R} to be a shock *strictu sensu*, price cannot be a function of current output levels. If the current price depends on current output levels, then ξ_{it}^{R} is a function of input levels x_{it} and k_{it}, which makes instrumenting inputs impossible because any informative instrumental variable will have to be endogenous. A similar argument applies to the law of motion $g(\cdot)$: for it to be common across firms, price changes over time cannot depend on changes in output levels, since it would imply a heterogeneous law of motion. These conditions for ξ_{it}^{R} and $g(\cdot)$ are not satisfied in the common CES demand with monopolistic competition case, where price depends on own output - and hence input levels. They are instead trivially satisfied in the competitive setting, where the price is exogenous by definition. In applications, controlling for TFPR with past revenue and input levels is likely to leave endogenous price variation in the residual ξ_{it}^{R}, so the argument on price bias applies.

All the concerns outlined here stream from the fact that - in a general setting - prices and markups are equilibrium objects, unlike TFPQ. This makes proxying for them harder, and appears to require strong assumptions on residual demand. By essentially avoiding a regression of log revenue on log input levels, my proposed approach does not require to take a stance on prices (and hence demand). This is achieved by using first order conditions (equation (10)), and at the ‘expense’ of having to fix the returns to scale of the underlying production function.

4.3 Cost Share Approaches

Inferring output elasticities from the cost shares of inputs is quite common, and generally provides the advantage of not requiring complex econometric techniques, if any at all (De Loecker and

Syverson, 2021). Instead of regressing output or revenue on inputs levels with appropriate controls, these methods recover the elasticity from first order conditions, i.e. equation (10). As anticipated in Section 2, the main concern is the measurement of the true economic cost of non-flexible inputs such as capital. This can be easily avoided when output markets are competitive, which ensures that total cost and revenue coincide. This in turn allows the researcher to work with revenue shares, from which the output elasticity of the flexible input is non-parametrically identified (Gandhi et al., 2020). On the contrary, in the presence of market power the researcher is forced to work with cost shares, which are contaminated by measurement error in input expenditures. For example, suppose that the noisy proxy $\tilde{\rho}_{it}$ s.t. $\mathrm{E}[\tilde{\rho}_{it}] = \rho_{it}$ is observed. As already shown, the noisy cost share of the flexible input differs from the output elasticity, i.e. $\tilde{\alpha}_{it}^{C,X} \cdot \mathcal{T} \neq \theta_{it}^{Q,X}$ and:

$$\mathrm{E}\left[\tilde{\alpha}_{it}^{C,X} \cdot \mathcal{T}\right] \neq \mathrm{E}\left[\theta_{it}^{Q,X}\right] \tag{29}$$

Currently available approaches successfully address this problem in a Cobb Douglas setting (Raval, 2023; De Loecker et al., 2024). If the elasticity is common across firms $\theta_{it}^{Q,X} = \theta^{Q,X}$ for every i and t, it is possible to either (i) average expenditure on the flexible input and total cost *separately*, and then take their ratio (Raval, 2023), or (ii) take the median of observed cost shares (Collard-Wexler and De Loecker, 2024). Both these solutions rely on the Cobb Douglas assumption, which implies that as all the observed cost share dispersion across firms has to come from measurement error in input expenditures. As such, if the (common) Cobb Douglas production function assumption is violated, these approaches will not account for any of the elasticity heterogeneity and run into Jensen's inequality issues when it comes to the average elasticity. To see the latter, even when $\tilde{\rho}_{it} = \rho_{it}$, simply notice that $\frac{\mathrm{E}[X_{it}]}{\mathrm{E}[X_{it}+\tilde{\rho}_{it}K_{it}]} \cdot \mathcal{T} \neq \mathrm{E}\left[\frac{X_{it}}{X_{it}+\rho_{it}K_{it}} \cdot \mathcal{T}\right] = \mathrm{E}\left[\theta_{it}^{Q,X}\right]$, and median $\left[\frac{X_{it}}{X_{it}+\tilde{\rho}_{it}K_{it}} \cdot \mathcal{T}\right] \neq \mathrm{E}\left[\frac{X_{it}}{X_{it}+\rho_{it}K_{it}} \cdot \mathcal{T}\right] = \mathrm{E}\left[\theta_{it}^{Q,X}\right]$ unless $\theta_{it}^{Q,X} = \theta^{Q,X}$ for every i and t. Therefore, my proposed estimator can be considered a generalization of the existing cost share approaches, as it removes the Cobb Douglas requirement.

5 Monte Carlo Simulations

I test the estimator proposed in Section 3 by simulating data from 100 samples of $N = 100$ firms over $T = 10$ time periods. The setting is the same as the one described in Section 2. Firm i at time t maximizes its net present value by choosing input X_{it} without frictions, and investment I_{it}, conditional on its capital stock K_{it}, productivity Ω^Q_{it}, and a vector of residual demand conditions Υ_{it}. Capital depreciates at rate δ, future profits are discounted at rate β, and the cost of investment is quadratic, i.e. $\frac{1}{2}I^2$. The Bellman equation and the constraints associated to this problem are:

$$
\begin{aligned}
V(K_{it}, \Omega^Q_{it}, \Upsilon_{it}) = \max_{X_{it}, I_{it}} & \; P(Q_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \frac{1}{2}I^2 + \beta \cdot \mathrm{E}_t[V(K_{it+1}, A^Q_{it+1}, \Upsilon_{it+1})] \\
\text{s.t.} & \;\; Q_{it} \leq \Omega^Q_{it} \cdot F(X_{it}, K_{it}) \\
& K_{it+1} = (1-\delta) \cdot K_{it} + I_{it}
\end{aligned}
\tag{30}
$$

I assume that production is Translog, so that output elasticities are endogenous:

$$
q_{it} = \omega^Q_{it} + \beta^X x_{it} + \beta^K k_{it} + \beta^{XK} x_{it} k_{it} + \beta^{XX} x^2_{it} + \beta^{KK} k^2_{it} \tag{31}
$$

βs are chosen so that the production function is concave and constant returns to scale regardless of input choice.[14] $P_{it}(\cdot)$ is a linear inverse demand schedule with heterogeneous parameters $\Upsilon_{it} = [\Upsilon^0_{it}, \Upsilon^1_{it}]$, so that both prices and markups are endogenous:

$$
P_{it}(Q_{it}, \Upsilon_{it}) = \Upsilon^0_{it} - \Upsilon^1_{it} Q_{it} \tag{32}
$$

The dynamic problem is solved with value function iteration (Judd, 1998) on three fine grids of 50 values each for K_{it}, I_{it}, and X_{it}, and three coarse grids for Ω^Q_{it}, Υ^0_{it}, and Υ^1_{it} of 5, 2, and 2 values respectively. The transition probabilities of these parameters are fixed and common across firms. Therefore, firms choose investment while facing uncertain future productivity and demand. Note that, like in Section 2 the uncertainty in Υ_{it} can also capture uncertainty about future competitive pressure. I simulate the data for 100 periods, and discard the first 90 to minimize the impact of the choice of the initial values of parameters and of the capital stock. After finding optimal input

---

[14] I choose $\beta^X = 0.8$, $\beta^K = 0.2$, $\beta^{XK} = 0.2$, $\beta^{XX} = \beta^{KK} = -0.1$.

levels, and the resulting output and revenue, I calculate the true markup using equation (7), and the true cost of capital using (8).

In Appendix G I report the results of the same exercise when the production function is Cobb Douglas.

5.1 Data and Benchmarks

I assume that the researcher observes noisy revenue, input levels, and a noisy economic cost of the dynamic input, $\{\tilde{R}_{it}, X_{it}, K_{it}, \tilde{\rho}_{it}\}$. Noise is log additive for revenue and additive for the cost of capital, as in (11) and (12). I show sensitivity to the variance of both sources of noise. The researcher also knows the true returns to scale parameter, $\mathcal{T} = 1$.

I present output elasticity and markup estimates - $\hat{\theta}_{it}^{Q,X}$, $\hat{\theta}_{it}^{Q,K}$, and $\hat{\mu}_{it}$ - from two different approaches. The first is the 'naive accounting approach', taken as a benchmark: the researcher disregards measurement error and simply equates cost shares to the corresponding elasticities.

$$\begin{aligned} \hat{\theta}_{it}^{Q,X} &= \tilde{\alpha}_{it}^{C,X} \\ \hat{\theta}_{it}^{Q,K} &= \tilde{\alpha}_{it}^{C,K} \\ \hat{\mu}_{it} &= \frac{\hat{\theta}_{it}^{Q,X}}{\tilde{\alpha}_{it}^{R}} \end{aligned} \tag{33}$$

This approach is by far the most sensitive to measurement error, but it is also computationally straightforward and quite used in the literature (e.g. Autor et al. (2020)).

The second approach is my two-stage estimator outlined in Section 3: measurement error is purged from the cost of capital with a non-parametric regression, and then output elasticities and markups are estimated from purged cost shares. I use this approach to also estimate TFPR as in equation (21).

5.2 Baseline Results

Table 1 presents my baseline results, which are obtained in a setting with no measurement error in revenue, and where measurement error in the cost of capital ε_{it}^{ρ} accounts for roughly 50% of the variance of $\tilde{\rho}_{it}$. The 'naive accounting' benchmark suffers from some bias in levels, and considerable bias in distribution, with the average estimate of the elasticity of the flexible input being around +8.5% above the true value, and the individual values being only weakly correlated with true ones (+0.1564). On the contrary, two-stage estimates of all output elasticities fit the true distributions close to perfectly (the averages match, and individual values are correlated above +0.99). The quality of output elasticity estimates is reflected in markup and TFPR estimates, which are of similar quality. In Appendix G I perform the same exercise when production is Cobb Douglas, and obtain equally good results.

Table 1: Monte Carlo Simulations: Translog + Linear Demand Baseline

	flexible input elasticity ($\theta_{it}^{Q,X}$)		
	true value	two-stage	naive accounting
mean	0.8813	0.8820	0.9566
	(0.0042)	(0.0078)	(0.2261)
std. dev.	0.0878	0.0860	2.8563
	(0.0019)	(0.0093)	(6.3184)
corr. (with true)	–	0.9962	0.1564
	–	(0.0048)	(0.1269)

	dynamic input elasticity ($\theta_{it}^{Q,K}$)		
	true value	two-stage	naive accounting
mean	0.1187	0.1180	0.0434
	(0.0042)	(0.0078)	(0.2261)
std. dev.	0.0878	0.0860	2.8563
	(0.0019)	(0.0093)	(6.3184)
corr. (with true)	–	0.9962	0.1564
	–	(0.0048)	(0.1269)

	markup (μ_{it})		
	true value	two-stage	naive accounting
mean	1.3255	1.3270	1.4187
	(0.0097)	(0.0122)	(0.2312)
std. dev.	0.1705	0.1727	2.9819
	(0.0048)	(0.0073)	(6.3853)
corr. (with true)	–	0.9956	0.1302
	–	(0.0048)	(0.0941)

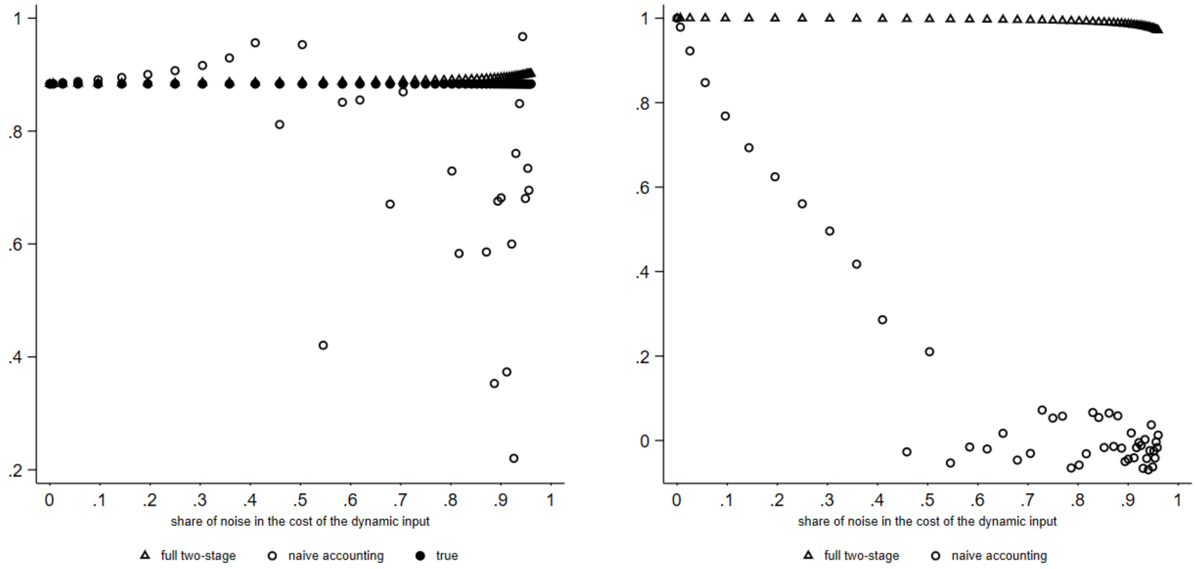
	TFPR (ω_{it}^{R})		
	true value	two-stage	naive accounting
mean	-	-	-
	-	-	-
std. dev.	0.2307	0.2311	-
	(0.0049)	(0.0069)	-
corr. (with true)	–	0.9983	-
	–	(0.0021)	-

Note: standard deviations in parentheses. For each estimator, I report the mean estimate across 100 Monte Carlo samples of 100 firms and 10 time periods each. Standard deviations are calculated across these 100 samples. In these samples there is no noise in revenue, and noise in the cost of the dynamic input (ε_{it}^{ρ}) accounts for roughly (50%) of the variance of ($\tilde{\rho}_{it}$).

5.3 Sensitivity to Measurement Error in the Cost of the Dynamic Input

To investigate the sensitivity of the three estimators to measurement error in the cost of capital, I repeat the simulation exercise 50 times on the same data, while increasing the standard deviation of said error from 0 to 4 times that of the true value of the economic cost of the dynamic input ρ_{it}. When the standard deviation of ε_{it}^{ρ} is 4 times that of ρ_{it}, around 96% of the variance of $\tilde{\rho}_{it}$ is coming from the noise term. Figure 1 shows how the accounting and the two-stage estimators of the output elasticity of the flexible input $\theta_{it}^{Q,X}$ perform in terms of average (left panel) and correlation with true values (right panel), as the 'share of noise' in $\tilde{\rho}_{it}$ increases.

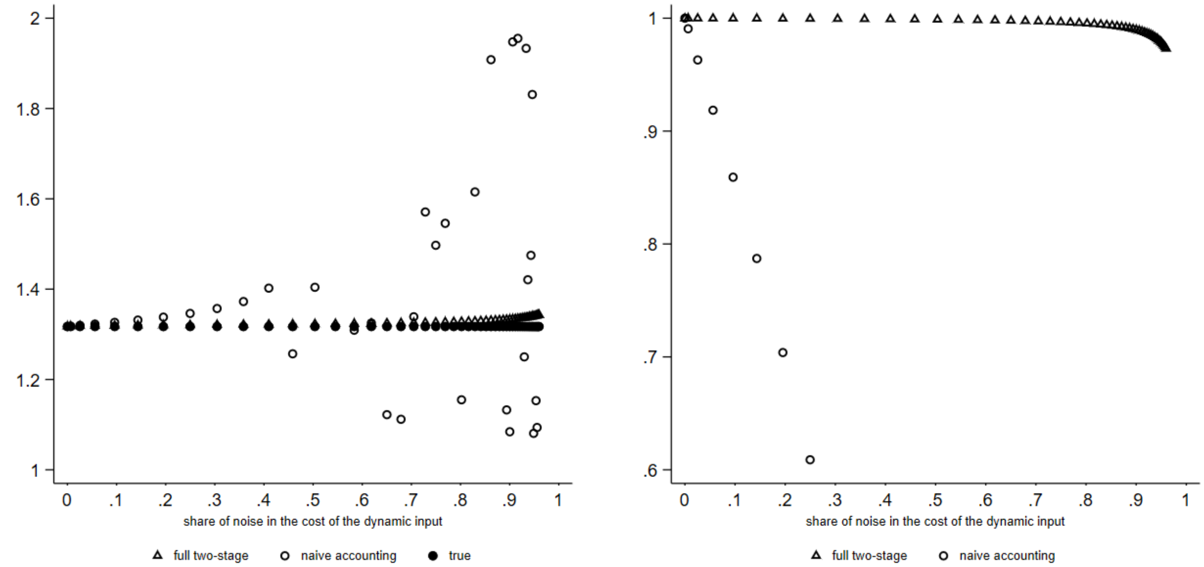
Figure 1: Sensitivity to measurement error in the cost of the dynamic input: average $\hat{\theta}_{it}^{Q,X}$ (left panel), and corr$(\hat{\theta}_{it}^{Q,X}, \theta_{it}^{Q,X})$ (right panel)



Note: extreme results from the naive accounting approach are omitted (see Proposition 1 for a discussion).

First, the accounting approach tends to 'break down' when noise in the cost of capital is above 50%, and for lower shares of noise still displays substantial (and growing) bias of both signs, as indicated by Proposition 1. It follows that the benefit of purging measurement error from the cost of capital increases with σ_{ε}^2, as expected. The bias in the average output elasticity from the 'full two-stage' estimator when noise accounts for 96% of the variance of $\tilde{\rho}_{it}$ is around just 2% of the true counterpart. Similarly, the correlation between two-stage estimates and true value does not drop below +0.97. Figure 2 shows the effect on markups of increasing the same type of noise.

Figure 2: Sensitivity to measurement error in the cost of the dynamic input: average $\hat{\mu}_{it}$ (left panel), and corr$(\hat{\mu}_{it}, \mu_{it})$ (right panel)



Note: extreme results from the naive accounting approach are omitted.

Predictably, markup estimates are affected in the same way as the output elasticity of the flexible input because - absent measurement error in revenue - the only source of bias in markups can be $\hat{\theta}^{Q,X}_{it}$.

In Appendix E I show that the same results hold for the capital elasticity $\theta^{Q,K}_{it}$ and revenue productivity ω^{R}_{it}: the two-stage approach performs extremely well at any level of noise, while the naive accounting approach is strictly worse, and increasingly so as σ^2_{ε} grows. In Appendix F I show how the two estimators perform when there is log additive measurement error in revenue and in the level of the dynamic input K_{it}. Measurement error in revenue does not affect output elasticities, because input cost shares do not rely on revenue data. Therefore, only markups and TFPR are affected. Instead, measurement error in the capital stock enters both the first stage and all cost shares. This leads to two concerns: attenuation bias (Klepper and Leamer, 1984) in the first stage, and division bias (Borjas, 1980) in the second stage for the elasticity of the flexible input. These do not appear to be quantitatively relevant concerns for plausible values of $\theta^{Q,K}_{it}$ and variance of this type of measurement error.

6 Compustat Application

I apply the proposed estimator to data on the accounts of North American listed companies (Compustat). This is essentially the same data as Traina (2018), De Loecker et al. (2020), and Bond and Gottardo (2024), among many others. Besides showing how the new estimator performs in a real data setting and comparing the results to the previous literature, I am interested in identifying the characteristics of the so-called 'superstar firms', i.e. the high markup, high productivity companies that appear to be driving the increase in the dispersion of both (Autor et al., 2020).

6.1 Data and Cleaning

Compustat covers the company accounts of all firms listed on North American exchanges between the mid 50s until the present. I focus on the 1962-2024 period and use annual data. The variables of interest for the analysis are the same as in the above cited papers: sales, cost of goods sold (COGS), property, plant, and equipment (PPE, a measure of tangible capital), and selling, general, and administrative expenses (SGA). In line with the previous literature, I define sectors at the 2-digit NAICS level. In addition to these variables, I construct the user cost of capital as in Bond and Gottardo (2024): I use the yield on 10-year treasuries as the safe asset, and company-year specific betas from WRDS Beta Suite to calculate the risk premium (see Appendix I and Bond and Gottardo (2024) for further details). I also construct the stock of intangible capital following Chiavari and Goraya (2025) closely. I first add amortization and remove goodwill from the book value of intangibles. I then capitalize SGA spending with the perpetual inventory method, and add this stock to the previous one.

My cleaning routine is similar to De Loecker et al. (2020) and Bond and Gottardo (2024): I drop firms which lack information on any variable of interest, or report negative sales and expenditures, and drop the 1st and 99th percentiles of the revenue share of COGS, PPE, and SGA.

In line with Traina (2018); De Loecker et al. (2020); Bond and Gottardo (2024) I treat COGS as the flexible input, and PPE as non-flexible. The three papers differ in their treatment of SGA: De Loecker et al. (2020) treat it as a sunk cost, while the others treat it as a production input. This distinction affects results substantially regardless of the estimation approach of choice because SGA accounts for an increasing share of revenue and total cost. As argued in Appendix A, any sunk cost that does not enter the production function directly does not enter the marginal cost of output. As such, sunk costs must be omitted from the total cost of output and from all non-parametric regressions of my two-stage approach. Here I present results when SGA/intangible capital is treated as a sunk cost and thus omitted from total cost. In Appendix J I present the same results when SGA is treated as a production input.

6.2 Estimation

In what follows, total variable cost is defined as the sum of COGS and PPE times the firm-level cost of capital.[15] The two stages are performed separately for every 2-digit NAICS sector-year sub-sample, as my approach does not require variation across time periods. I rely on a second-order polynomial approximation for the first stage, and on a first-order approximation for the second stage.[16] I assume constant returns to scale throughout the analysis, i.e. $\mathcal{T} = 1$. In Appendix K I compare and discuss my results on output elasticities and markups to those in De Loecker et al. (2020).

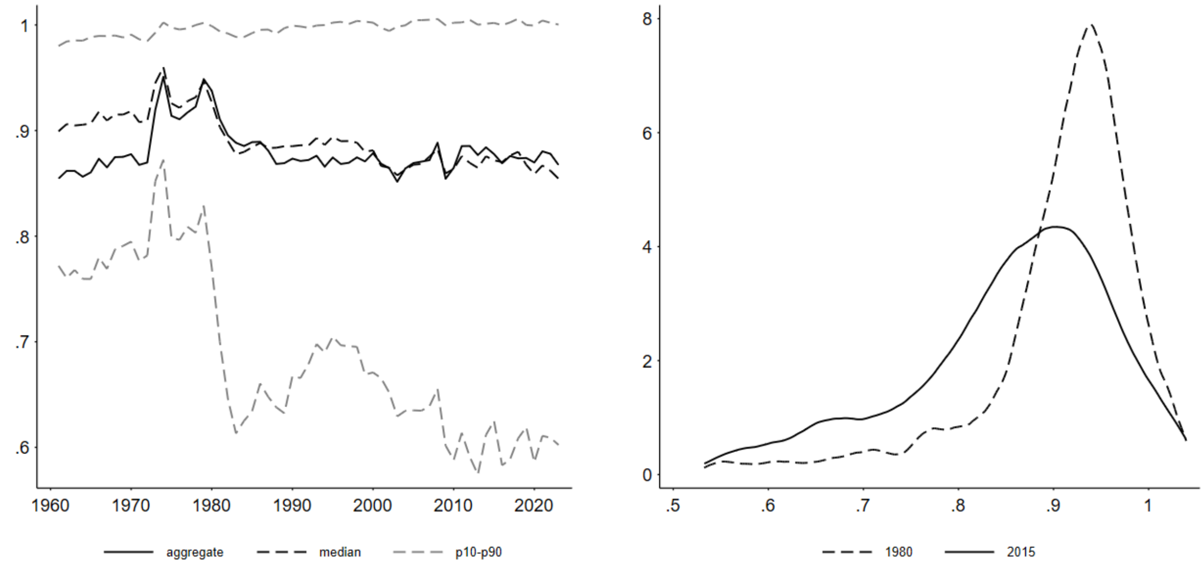
Figure 3 presents the evolution of the output elasticity of COGS for the whole sample, and compares its pdf in 1980 and 2015. The aggregate elasticity is obtained by taking the sales-weighted arithmetic mean of firm-level output elasticities.

---

[15] In Appendix J total cost also includes the cost of intangible capital.

[16] A first-order approximation in the second stage is exact if the underlying production function is Translog.

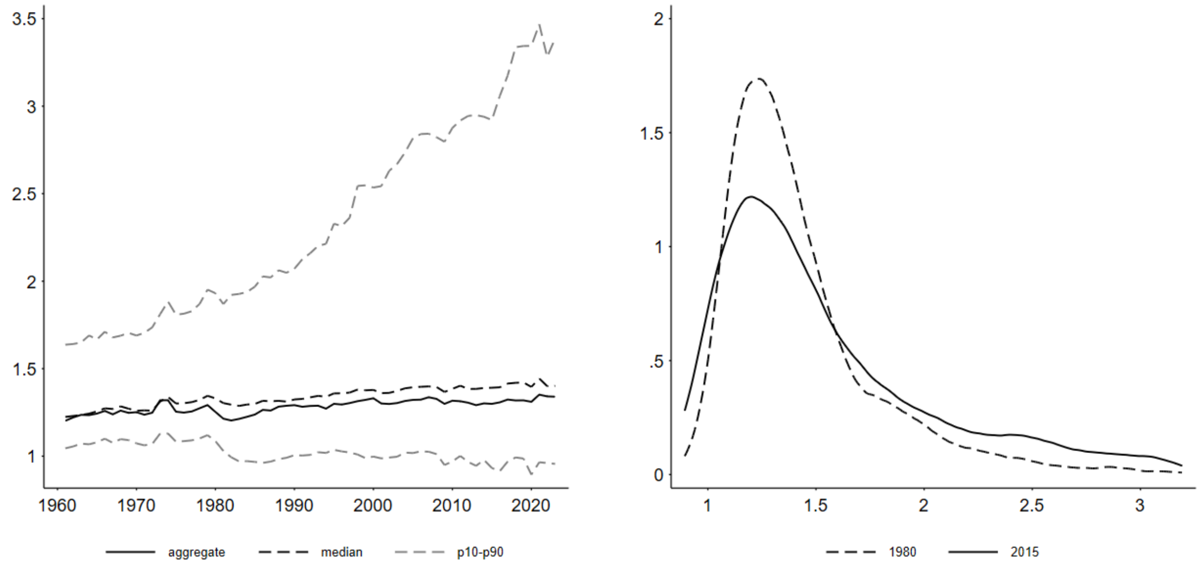
Figure 3: COGS output elasticity estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)



The aggregate and the median output elasticity of COGS are quite stable over time, never outside the $0.85-0.95$ range, and slightly falling between 1980 and 1990. However, the underlying distribution evolves visibly, with the 10^{th} percentile of $\hat{\theta}_{it}^{Q,X}$ dropping from 0.80 in the 1970s to 0.60 in the early 2010s. These trends are driven by the cost share of COGS, which remained almost flat on average but whose variance increased substantially because of its 'left tail' (see the right panel).

Figure 4 presents the evolution of the markup for the whole sample, and compares its pdf in 1980 and 2015. The aggregate markup is obtained by taking the sales-weighted harmonic mean of firm-level markups (Edmond et al., 2023).

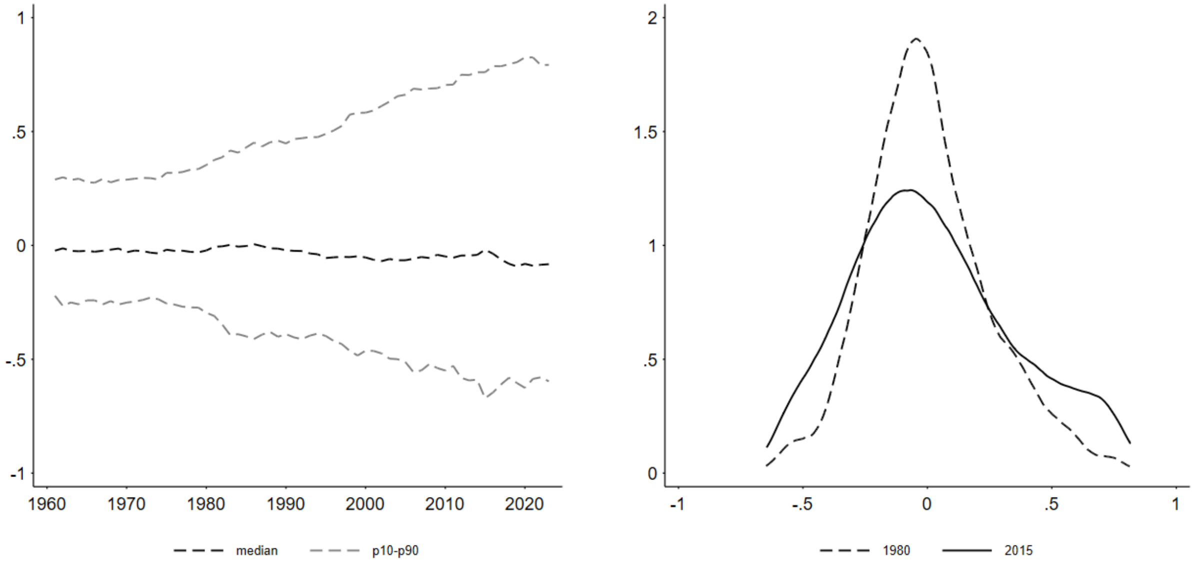
Figure 4: Markup estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)



The increase in the aggregate markup from around 1.25 in the 1960s to 1.35 today is dwarfed by the increase in dispersion, which is driven largely by the right tail of the markup distribution, as the 90th percentile grows from around 1.6 in the 1960s to 3.3 around 2020.

Finally, Figure 5 presents the evolution of the TFPR distribution for the whole sample, and compares its pdf in 1980 and 2015.

Figure 5: Log TFPR estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)



The increase in the log TFPR distribution is driven both by the right and left tails becoming fatter.

6.3 TFPR Variance Decomposition

The similarity between the evolution of the distribution of markups and that of TFPR can be explained with the following decomposition of the latter:

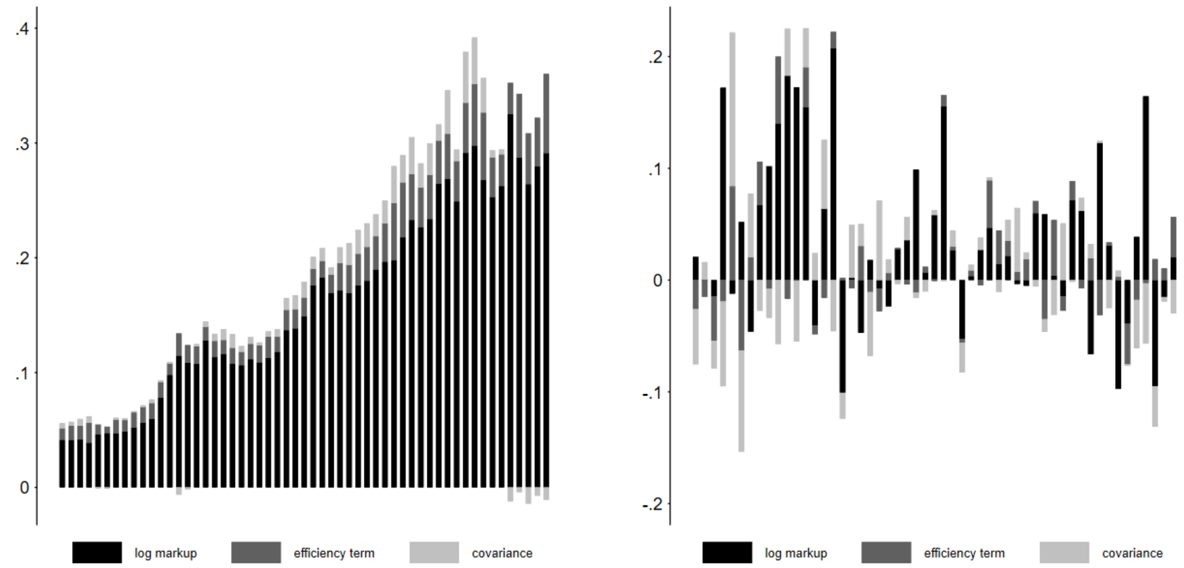
$$\omega^R_{it} := p_{it} + \omega^Q_{it} = \log \mu_{it} + (\log \lambda_{it} + \omega^Q_{it}) := \log \mu_{it} + \psi_{it} \tag{34}$$

where p_{it} is the log of the price of output, ω^Q_{it} is log TFPQ, and λ_{it} is the marginal cost of output. $\psi_{it} := (\log \lambda_{it} + \omega^Q_{it})$ is not productivity and cannot be interpreted as such, because it is the sum of actual TFPQ and marginal cost, and the latter is most likely decreasing in TFPQ. However, this term does not directly depend on output market outcomes and distortions, but just on the productive efficiency of the firm. Therefore, (34) can be thought of as a decomposition of TFPR between output market distortions ($\log \mu_{it}$) and an efficiency term (ψ_{it}). With data on revenue and input expenditures alone, ω^R_{it} and $\log \mu_{it}$ can be estimated with my two-stage approach, and I can thus calculate $\hat{\psi}_{it} = \hat{\omega}^R_{it} - \log \hat{\mu}_{it}$. These estimates can be used to perform the following variance decompositions:

$$\begin{aligned}
\mathrm{Var}(\hat{\omega}^R_{it}) &= \mathrm{Var}(\log \hat{\mu}_{it}) + \mathrm{Var}(\hat{\psi}_{it}) + 2\mathrm{Cov}(\log \hat{\mu}_{it}, \hat{\psi}_{it}) \\
\frac{\Delta \mathrm{Var}(\hat{\omega}^R_{it})}{\mathrm{Var}(\hat{\omega}^R_{it-1})} &= \alpha^{\mu}_{it-1}\frac{\Delta \mathrm{Var}(\log \hat{\mu}_{it})}{\mathrm{Var}(\log \hat{\mu}_{it-1})} + \alpha^{\psi}_{it-1}\frac{\Delta \mathrm{Var}(\hat{\psi}_{it})}{\mathrm{Var}(\hat{\psi}_{it-1})} + 2\alpha^{\mu,\psi}_{it-1}\frac{\Delta \mathrm{Cov}(\log \hat{\mu}_{it}, \hat{\psi}_{it})}{\mathrm{Cov}(\log \hat{\mu}_{it-1}, \hat{\psi}_{it-1})}
\end{aligned} \tag{35}$$

where $\alpha^{\mu}_{it-1} = \frac{\mathrm{Var}(\log \hat{\mu}_{it-1})}{\mathrm{Var}(\hat{\omega}^R_{it-1})}$. The first decomposition breaks down the level of the variance in log TFPR into markup variance, efficiency term variance, and their covariance. The second expression does the same for changes over time in the variance of log TFPR. Figure 6 presents the decomposition of the variance of log TFPR estimates of expression (35), both in levels and changes from 1962 until 2024.

Figure 6: Log TFPR variance decomposition 1962-2023 (SGA omitted): level (left panel) and changes (right panel)



Note: each column represent one year.

In short, log markup dispersion is the main component of log TFPR dispersion. The share of revenue productivity variance accounted for by markup variance has fluctuated between 62% and 95% (81% on average), without a clear trend over time. The markup-efficiency covariance term is almost always positive and plays a role similar to the variance of the efficiency term, meaning that there is a non-trivial interplay between productivity and market power.[17]

---

[17] I cannot take a stance on the correlation between TFPQ and markups because data limitations prevent me from

6.4 Who Are the Superstars?

The methodological approach introduced in this paper is particularly assumption-light compared to the existing ones in terms of the functional forms of the production function, residual demand, and the law of motion of productivity. This makes it suited for investigating patterns and differences across firms, especially in sector-year cross sections, where Cobb-Douglas and/or demand assumptions would be more limiting. I therefore focus on the relationship between revenue productivity, markups, overhead/intangibles, and firm size. Several papers report the same finding of Figure 6, i.e. two closely related secular increases in markup and productivity dispersion, driven by a small number of "superstar" firms (Autor et al., 2020; De Loecker et al., 2020). In particular, Autor et al. (2020), using US Economic Census data, argue that *larger* firms have higher markups. I show that this is not the case on Compustat. This is not a falsification of their finding because the underlying data is different: in order to be on Compustat, a firm needs to be listed (at some point) on some stock exchange, a very rare occurrence for the typical firm. The US Census Bureau reports around 6.4 million firms and 8.3 million establishments in the US in 2022, whereas Compustat only contains 11,415 observations in the same year, before I perform any cleaning (see Appendix I). It follows that Compustat covers less than 0.2% of US firms - even though their share of output, employment, investment etc. is much larger because of their size. Thus, the results below require careful interpretation: they should not be used to draw inference about the US macro-economy; rather, they concern the drivers of markup and revenue productivity differences "at the top" of the distribution of firms. In other words, these results concern that $\sim 0.2\%$ of companies that are successful enough to be on Compustat.

To investigate the relationships between revenue productivity, market power, overhead, intangibles, and size, I employ regressions of the following form:

$$Y_{it} = \beta_0 + \sum_{p=2}^{100} \beta_p \cdot D_{it}^{p,V} + \tau_{est} + \varepsilon_{it}^{V} \tag{36}$$

where $D_{it}^{p,V} = 1$ if firm i at time t belongs to the pth percentile of the distribution of variable V, and τ_{est} are exchange-by-sector-by-year fixed effects. I employ this specification instead of a simple

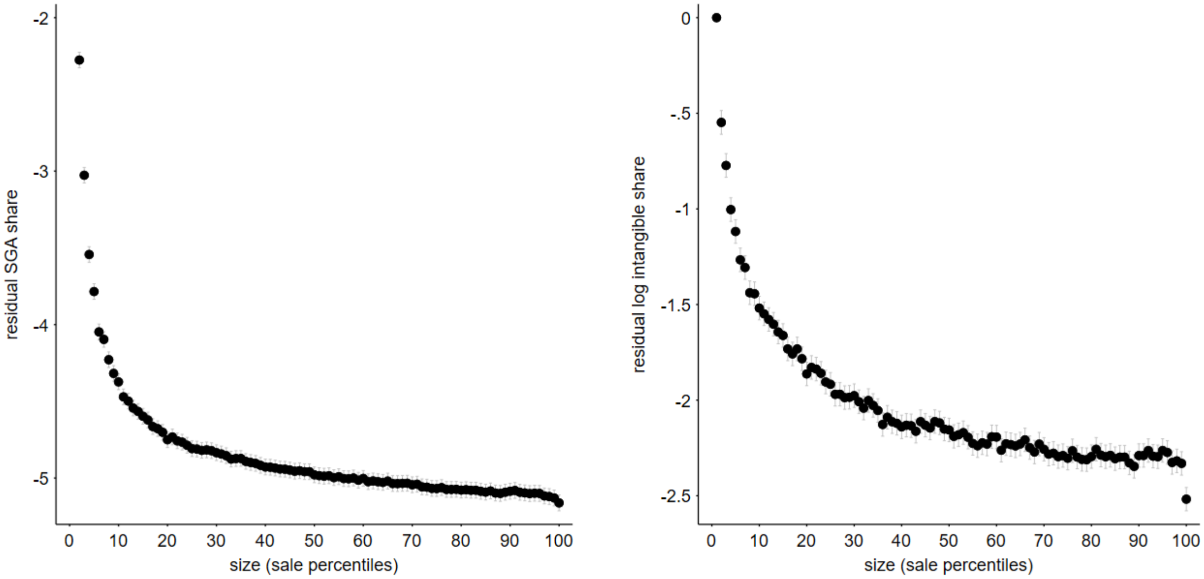
---

measuring the former.

linear one to avoid taking a strong stance on the functional form of the relationship between any Y and V of interest. Y_{it} is always a log, since log additive measurement error in revenue enters markups, revenue TFP, and revenue shares log-additively.

I define firm size as sales. All results are robust to defining it as total cost. Figure 7 displays the relationship between overhead (SGA)/intangibles and size.

Figure 7: Relationship between size (sales percentiles) and the log of the sales share of SGA (left panel) and the log of the sales share of the intangible capital stock (right panel)



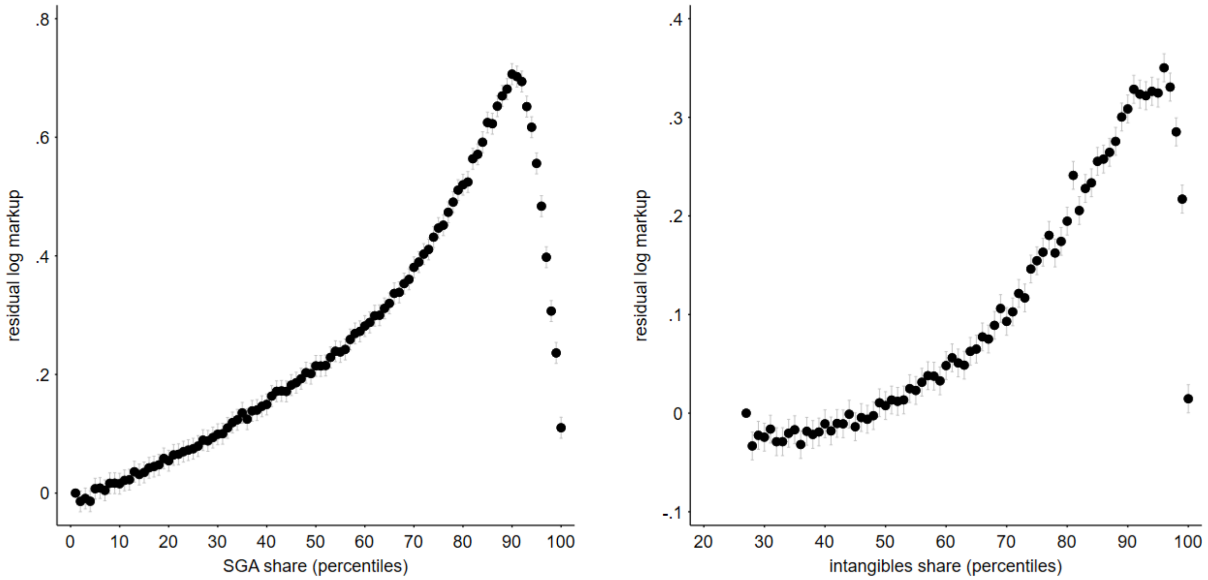
Note: OLS estimates of (36). 95% confidence intervals are in grey.

The takeaway of Figure 7 is that smaller firms are more SGA/intangible intensive. This is consistent with a sunk cost interpretation of overhead and intangible capital.[18] Figures 8 and 9 display the relationship between markups and revenue productivity, and sunk costs.

---

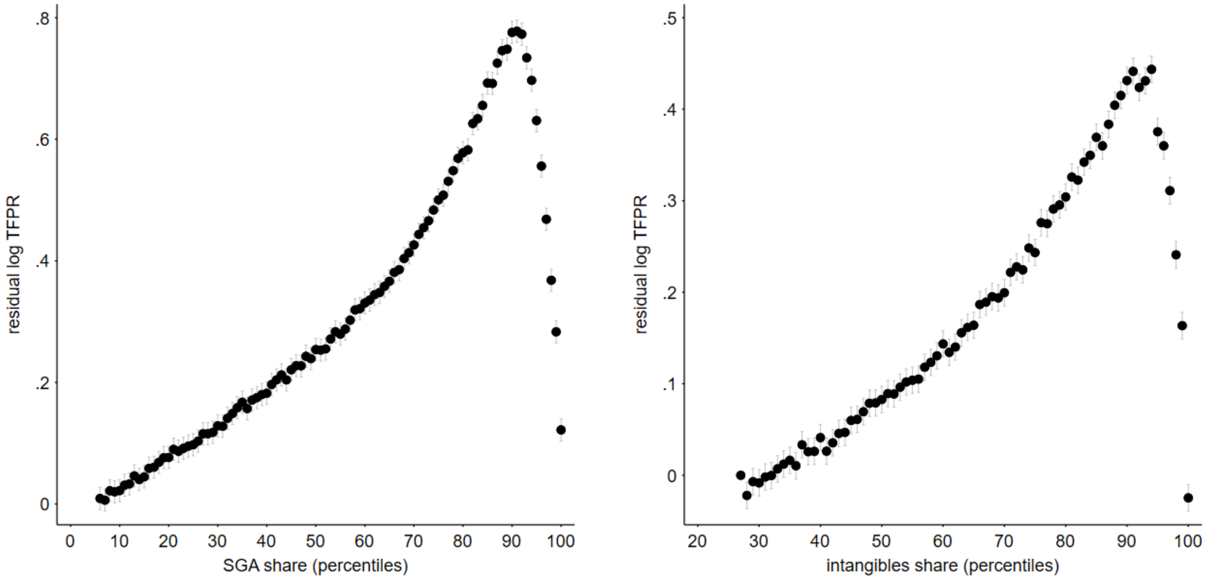
[18] Note, however, that larger firms have larger SGA expenditures and intangible capital stocks. This is consistent with and endogenous sunk cost interpretation, but not an exogenous sunk cost one, which would predict similar levels of expenditure on overhead for different size classes.

Figure 8: Relationship between markups and sunk costs (SGA share of sales percentiles, left panel, intangible share of sales percentiles, right panel).



Note: OLS estimates of (36). 95% confidence intervals are in grey. The first 26 percentiles of the sales share of intangibles are zero, and thus omitted.

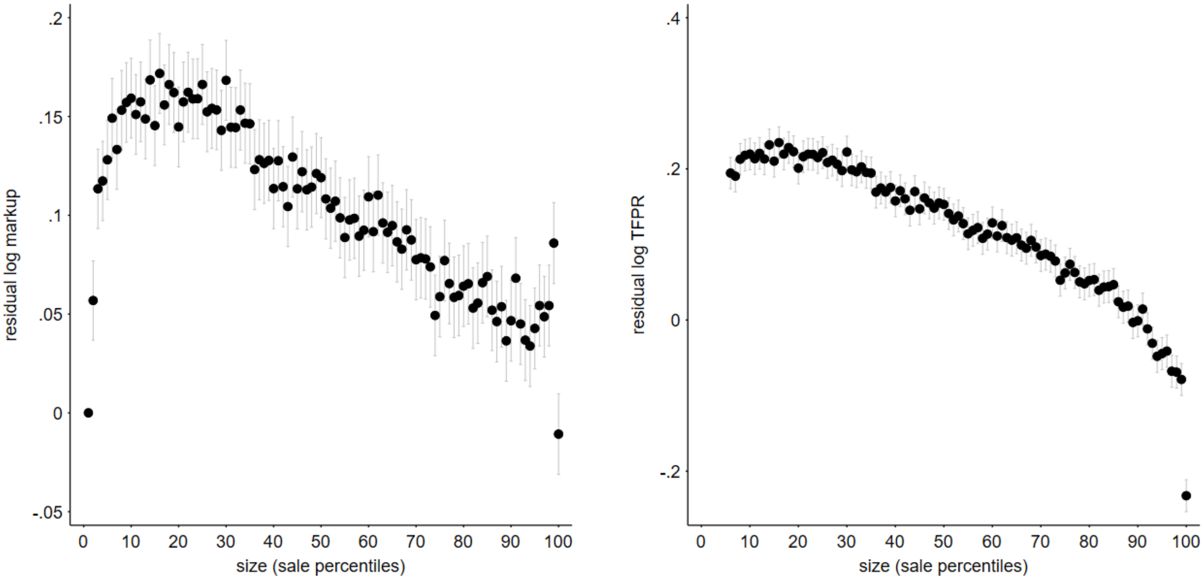
Figure 9: Relationship between revenue productivity and sunk costs (SGA share of sales percentiles, left panel, intangible share of sales percentiles, right panel).



Note: OLS estimates of (36). 95% confidence intervals are in grey. The first 26 percentiles of the sales share of intangibles are zero, and thus omitted.

Albeit at a much more granular level, the evidence provided in Figure 8 is in line with De Loecker et al. (2020), who find that firms that spend relatively more on overhead also record higher markups. I find the same for revenue productivity, which should be unsurprising, as TFPR differences are largely driven by markup differences (see Figure 6). Once again, the results are consistent with a sunk cost story: firms that spend relatively more on sunk costs may *need* higher markups to support their spending on non-production items, which would mechanically translate into higher revenue productivity. The negative association between sunk costs and size, together with the positive one between sunk costs and markups/TFPR hint to the relationship between markups and revenue productivity, and firm size, reported in Figure 10.

Figure 10: Relationship between markups (left panel) and revenue productivity (right panel) and firm size (sales percentiles).



Note: OLS estimates of (36). 95% confidence intervals are in grey.

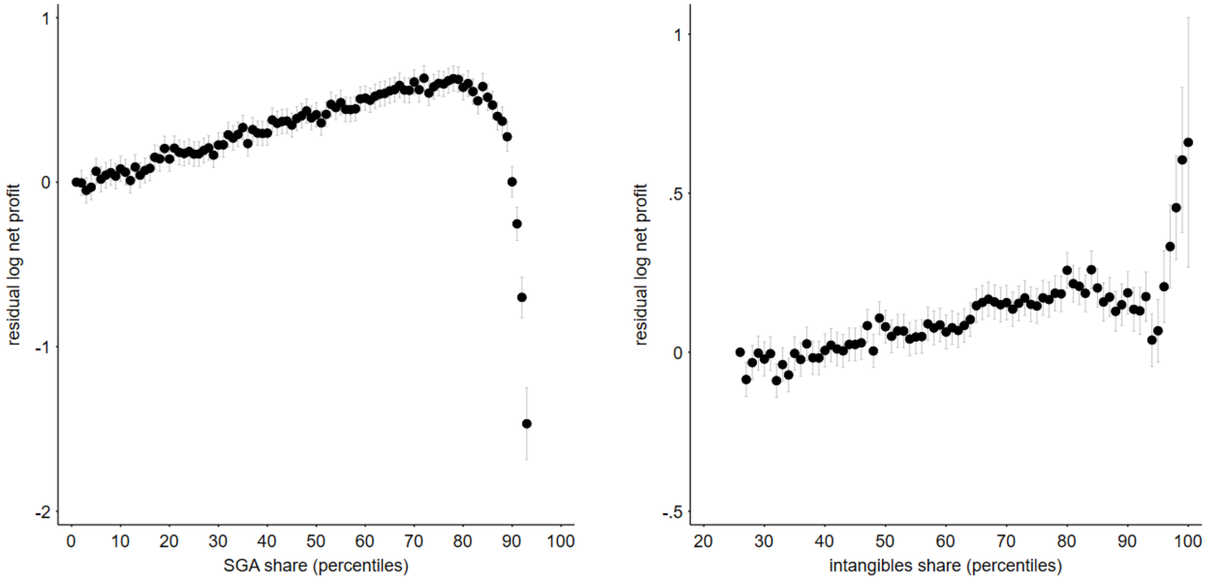
This negative relationship should not be surprising, as it is entirely due to the previous results. Small firms spend relatively more on intangibles, and thus record higher markups and revenue productivity, the latter being mostly driven by the former. It follows that on Compustat the "average" superstar (in a markup/productivity sense) is not a large firm, but rather a small, recently listed firm that spends a lot on SGA/intangible capital, which in turn requires charging a higher markup.

The presence of sunk costs requires a distinction between the firm markup (ratio of price to marginal cost) and firm net profitability (the flow of profits *net* of sunk costs). I define the net profit rate as follows:

$$\pi_{it} := \frac{\text{revenue}_{it} - \text{cogs}_{it} - \hat{\rho}_{it} \cdot \text{ppegt}_{it} - \text{sga}_{it}}{\text{revenue}_{it}} = 1 - \frac{1}{\hat{\mu}_{it}} - \alpha_{it}^{R,\text{SGA}} \tag{37}$$

This definition is very close to a measure of accounting profits, since the only estimate required to compute it is the purged cost of capital, $\hat{\rho}_{it}$. All results that follow are robust to using the CAPM-based noisy $\tilde{\rho}_{it}$ instead of its purged counterpart. Figure 11 shows the relationship between sunk costs and profitability. Note that the highest percentiles of the SGA share of sales tend to be associated with lower net profits on average (as these are firms that spend more on overhead than their total sales), but the highest percentiles of the intangible share are not, because intangible capital is approximately a capitalization of past SGA expenditure.

Figure 11: Relationship between log net profits and sunk costs (SGA share of sales percentiles, left panel, intangible share of sales percentiles, right panel).

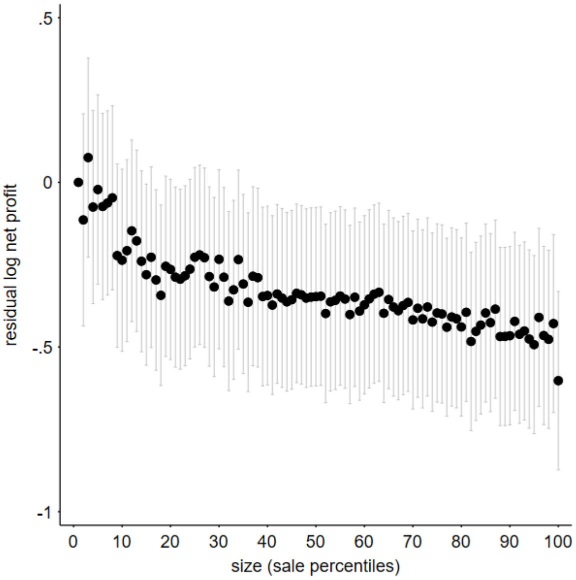


Note: OLS estimates of (36). 95% confidence intervals are in grey. The first 26 percentiles of the sales share of intangibles are zero, and thus omitted.

Sunk costs appear to boost net profitability too, albeit more weakly than they map into higher markups. Moreover, the top 95% of firms in terms of intangible intensity are substantially clear net profitability superstars, as their accumulated intangible stock does not weigh on their profits the same way as current SGA spending does. Are these intangible superstars larger than other firms?

Figure 12 shows that this is not the case.

Figure 12: Relationship between log net profits and firm size (sales percentiles).



Note: OLS estimates of (36). 95% confidence intervals are in grey.

Unlike overhead and intangible capital, size is a very poor predictor of net profitability. This implies that the most profitable firms that benefit from larger intangible stocks are not larger nor smaller than average.

7 Conclusion

In this paper, I propose a novel approach to estimate output elasticities, markups, and revenue total factor productivity with standard data on company accounts. This new method is not subject to the shortcomings pointed out in the current literature on production function estimation, and it is robust to realistic sources and amounts of measurement error in the data. It relies on noisy cost shares and requires optimizing firms, Hicks-neutral productivity, a flexible input, competitive input markets, and fixing the returns to scale of the production function. It requires no assumption on demand, the law of motion of productivity, or the invertibility of input demand, and it does not require data on physical output at all. I show that cost shares are biased estimators of output elasticities because of measurement error in input expenditures, e.g. from noisy data on the cost of dynamic inputs such as capital. I then show that noise in the cost of inputs can always be purged, and use this finding to develop my two-stage estimator.

I then test my estimator with Monte Carlo simulations in a setting with heterogeneous output elasticities and linear residual demand, and show it is robust to extremely noisy data on input expenditures/total cost.

Finally, I apply the new estimator to Compustat over the 1962=2024 time period. In line with the previous literature, I find that markup and revenue productivity dispersion increased around 6-fold over the whole period. I also find that the dispersion in output elasticities has increased, and decompose the increase in TFPR dispersion to show that markup dispersion accounts for around 81% of it, meaning that it is also the main driver of the former. I then investigate the relationship between productivity, markups, profits, sunk costs, and firm size. I find that: (i) smaller firms are more sunk cost intensive; (ii) sunk cost intensive firms record higher markups and profits; (iii) smaller firms have higher markups; (iv) profitability does not vary significantly with size; (v) revenue productivity - being largely driven by markups - is also higher in sunk cost intensive, and thus smaller firms. These results are not directly comparable to those of Autor et al. (2020), because of the different underlying data. However, they imply that, among listed firms, markup and productivity superstars are not necessarily the largest firms, but rather sunk cost intensive recently-listed companies.

References

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

Ackerberg, D. A. and De Loecker, J. (2024), ‘Production function identification under imperfect competition’, *Working paper* .

Antràs, P., Fort, T. C. and Tintelnot, F. (2017), ‘The margins of global sourcing: Theory and evidence from us firms’, *American Economic Review* **107**(9), 2514–64.

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.

Biondi, F. (2024), ‘Firm productivity and derived factor demand under variable markups’, *Working paper* .

Blundell, R. and Bond, S. R. (2000), ‘Gmm estimation with persistent panel data: An application to production functions’, *Econometric Reviews* **19**, 321–340.

Bond, S. and Gottardo, G. (2024), ‘Non-parametric bounds for average values of revenue elasticities, revenue returns to scale, and price-cost markups’, *Working Paper* .

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

Borjas, G. J. (1980), ‘The relationship between wages and weekly hours of work: The role of division bias’, *The Journal of Human Resources* **15**(3), 409–423.

Chiavari, A. (2021), ‘The macroeconomics of rising returns to scale: Customer acquisition, markups, and dynamism’, *Working Paper* .

Chiavari, A. and Goraya, S. (2025), ‘The rise of intangible capital and the macroeconomic implications’, *Working paper* .

Collard-Wexler, A. and De Loecker, J. (2024), 'Productivity and capital measurement error', *NBER 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 Fleitas, S. (2021), 'Markups and mergers in the us hospital industry', *working paper* .

De Loecker, J., Fuss, C., Quiller-Doust, N. and Treuren, L. (2024), 'The anatomy of costs and firm performance: Evidence from belgium', *CEPR Discussion Paper DP19641* .

De Loecker, J. and Syverson, C. (2021), An industrial organization perspective on productivity, Vol. 4 of *Handbook of Industrial Organization*, Elsevier, pp. 141–223.

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

De Ridder, M., Grassi, B. and Morzenti, G. (2022), 'The hitchhiker's guide to markup estimation', *Working Paper* .

Doraszelski, U. and Jaumandreu, J. (2023), 'Reexamining the de loecker warzynski (2012) method for estimating markups', *CEPR Discussion Paper No. DP16027* .

Edmond, C., Midrigan, V. and Xu, D. Y. (2023), 'How costly are markups?', *Journal of Political Economy* **131**(7), 1619–1675.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. F. and French, K. R. (2015), 'A five-factor asset pricing model', *Journal of Financial Economics* **116**(1), 1–22.

Fama, E. F. and MacBeth, J. D. (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Flynn, Z., Gandhi, A. and Traina, J. (2019), 'Measuring markups with production data', *SSRN Working Paper* .

Gandhi, A., Navarro, S. and Rivers, D. A. (2020), ‘On the identification of gross output production functions’, *Journal of Political Economy* **128**(8), 2973 – 3016.

Gormsen, N. J. and Huber, K. (2024), ‘Firms’ perceived cost of capital’, *NBER Working Paper Series* (32611).

Jorgenson, D. W. (1963), ‘Capital theory and investment behavior’, *The American Economic Review* **53**(2), 247–259.

Judd, K. L. (1998), *Numerical Methods in Economics*, Vol. 1 of *MIT Press Books*, The MIT Press.

Kirov, I., Traina, J. and Mengano, P. (2023), ‘Measuring markups with revenue data’, *SSRN Working Paper* .

Klepper, S. and Leamer, E. E. (1984), ‘Consistent sets of estimates for regressions with errors in all variables’, *Econometrica* **52**(1), 163–183.

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.

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

Raval, D. (2023), ‘A flexible cost share approach to markup estimation’, *Economics Letters* **230**, 111262.

Sutton, J. (1991), *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*, MIT Press.

Syverson, C. (2004), ‘Market structure and productivity: A concrete example’, *Journal of Political Economy* **112**(6), 1181–1222.

Traina, J. (2018), ‘Is aggregate market power increasing? production trends using financial statements’, *SSRN working paper* .

A Exogenous and Endogenous Sunk Costs

I extend the simple model of Section 2 to include expenditures on sunk costs that do not directly contribute to production in the spirit of Sutton (1991). The setup is the same: a firm maximizes its current NPV by taking market conditions Υ_{it}, current capital K_{it}, and TFPQ Ω^Q_{it} as given. X_{it} is expenditure on the flexible input, and output is governed by the same production function $Q_{it} = \Omega^Q_{it} F(X_{it}, K_{it})$. Let expenditure on the sunk cost be $P^S_{it} S_{it}$. I assume P^S_{it} is exogenous and that the sunk cost has no dynamic effects for ease of exposition.[19]

First, I focus on the exogenous sunk cost case, e.g. $P^S_{it} S_{it}$ is an entry cost that the firm pays to be active. The timing of the payment of said entry cost is irrelevant for the purpose of this paper: either the firm enters with full information about payoff-relevant state variables, and thus it will enter only if it can generate a non-negative NPV; or the sunk cost is paid in advance, and the firm may realize an ex-post negative NPV. In either case, the firm's choice of X_{it} solves:

$$\max_{X_{it}} P(Q_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \rho_{it} K_{it} - P^S_{it} S_{it} \tag{38}$$

The exogenous sunk cost is taken as given, and does not affect output or price. Therefore, the first order condition wrt the flexible input is still (2). The same can be argued for the equivalent cost-minimization problem: the sunk cost does not enter the variable cost of output and the marginal cost of output, so (7) holds too. Note that expenditure on the sunk cost must be omitted from the denominator of the markup, because it is not part of the cost of output. It follows that the markup is not directly by the presence or absence of exogenous sunk costs. Notice that any returns to scale assumption (parameter $\mathcal{T}$) concerns the production function alone $F(\cdot)$, which does not take sunk costs as arguments by definition. As a consequence, in the presence of constant returns to scale in output, an exogenous sunk cost implies increasing returns in revenue (total cost need not be double to double revenue, because the sunk cost is held fixed).

Sutton (1991) also models endogenous sunk costs, which are essentially expenditures that can contribute to revenue and profits without directly affecting output, e.g. advertising. Now the sunk cost does not enter the production function, but can be modeled as a price shifter, i.e. it can affect

---

[19] If the sunk cost is 'accumulated' in a similar fashion as a dynamic input, the relevant implications are the same.

the willingness to pay for the same product. Without loss of generality, I assume that S_{it} is chosen with full knowledge of state variables and without frictions or dynamic implications, like the flexible input. The firm solves:

$$\max_{X_{it},S_{it}} P(Q_{it}, S_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \rho_{it} K_{it} - P^S_{it} S_{it} \tag{39}$$

The first order condition wrt the flexible input is still (2), although now S_{it} enters price and the inverse demand elasticity. As before, the dual equivalent cost minimization problem is unaffected because S_{it} does not enter the production function. Therefore, (7) holds too again. The first order condition wrt the sunk cost reads:

$$\alpha^{R,S}_{it} = \frac{\partial r_{it}}{\partial s_{it}} \tag{40}$$

i.e. the revenue share equals the revenue elasticity. Note that S_{it} is not a production input, so its output elasticity is zero, and it still does not enter the variable cost of output and its marginal cost.[20] As with an exogenous sunk cost, expenditure on S_{it} must be omitted from the denominator of the accounting markup, because it is not part of the cost of output.

B Relaxing Assumptions

Non-Competitive Input Markets

Add a third input to the model of Section 2, L_{it} (labour). Like the flexible input X_{it}, labour faces no adjustment frictions and is chosen to maximize current profits. However, unlike the flexible input, the price of L_{it} is endogenous, i.e. $W_{it} = W(L_{it}, \Upsilon_{it})$, where Υ_{it} also captures input market conditions. Let $\nu_{it} := 1 + \frac{\partial w_{it}}{\partial l_{it}}$ be the markdown on L_{it}, which in equilibrium must equal the ratio

---

[20] S_{it} now enters the marginal cost of revenue because it is an argument of the revenue function.

between the marginal revenue of L_{it} and W_{it}. Solving for the cost shares and the markup yields:

$$\begin{aligned}
\mu_{it} &= \mathcal{T} \cdot \frac{P_{it}Q_{it}}{X_{it} + \rho_{it}K_{it} + \nu_{it}W_{it}L_{it}} \\
\theta_{it}^{Q,X} &= \mathcal{T} \cdot \frac{X_{it}}{X_{it} + \rho_{it}K_{it} + \nu_{it}W_{it}L_{it}} \\
\theta_{it}^{Q,K} &= \mathcal{T} \cdot \frac{\rho_{it}K_{it}}{X_{it} + \rho_{it}K_{it} + \nu_{it}W_{it}L_{it}} \\
\theta_{it}^{Q,L} &= \mathcal{T} \cdot \frac{\nu_{it}W_{it}L_{it}}{X_{it} + \rho_{it}K_{it} + \nu_{it}W_{it}L_{it}}
\end{aligned} \tag{41}$$

With input market power, all cost shares should be ‘corrected’ with the appropriate markdown. Note that with data on revenue and input expenditures alone $[R_{it}, X_{it}, K_{it}, L_{it}, \rho_{it}, W_{it}]$ it is not possible to separately identify the markup and the markdown because there are two unknowns in the first line of equation (41). The intuition behind this fact is that the (observed) gross profit margin of the firm can be split arbitrarily between the (unobserved) markup and the markdown. The same holds for output elasticities: without additional information or assumptions on the markdown, they are not separately identified, as (41) is a system of four equations with five unknowns $\left(\mu_{it}, \nu_{it}, \theta_{it}^{Q,X}, \theta_{it}^{Q,K}, \theta_{it}^{Q,L}\right)$.

To relax the competitive input market assumption, an alternative assumption or external data/estimates of markdowns are required. Corrected cost shares can in fact be calculated with any external ν_{it} before performing the second stage, whereas the first stage is unaffected as it is a non-parametric function of input levels alone.

Returns to Scale

As mentioned in Section 3, the returns to scale parameter $\mathcal{T}$ can be fixed arbitrarily at any level, which means that robustness checks can be performed without external data. In practice, these are likely to deliver very similar results for plausible values of $\mathcal{T}$, as most papers either cannot reject constant returns to scale or find minimal deviations from it (e.g. $\mathcal{T} = 1.075$ in Chiavari (2021)). Since heterogeneity across observations in cost shares, input intensities, and profit margins is considerably larger than these deviations from constant returns, inference on the level of output elasticities and markups is barely affected.

For the same reason that markups and markdowns are not separately identified with data on revenue and input expenditures alone, so are markups and the returns to scale, as equations (9) and (10) would gain one degree of freedom. However, with data on physical output Q_{it}, it is possible to estimate the returns to scale beforehand in the same fashion as Syverson (2004). Suppose output elasticities are estimated under the wrong assumption on the returns to scale $\mathcal{T}^{\text{wrong}}$, such that $\tilde{\mathcal{T}}\hat{\theta}_{it}^{Q,X} = \theta_{it}^{Q,X}$ and $\hat{\theta}_{it}^{Q,K} = \tilde{\mathcal{T}}\theta_{it}^{Q,K}$, where $\tilde{\mathcal{T}} = \frac{\mathcal{T}^{\text{true}}}{\mathcal{T}^{\text{wrong}}}$. I can write:

$$q_{it} = \int_0^{x_{it}} \theta^{Q,X}(x, k_{it})\,\mathrm{d}x + \int_0^{k_{it}} \theta^{Q,K}(0, k)\,\mathrm{d}k + \omega_{it}^Q = \tilde{\mathcal{T}}\left(\int_0^{x_{it}} \hat{\theta}^{Q,X}(x, k_{it})\,\mathrm{d}x + \int_0^{k_{it}} \hat{\theta}^{Q,K}(0, k)\,\mathrm{d}k\right) + \omega_{it}^Q \tag{42}$$

$\tilde{\mathcal{T}}$ can be estimated from equation (42) provided that valid instruments $\mathbf{z}_{it}$ (and controls) such that $\mathrm{E}[\mathbf{z}_{it}\omega_{it}^Q] = 0$ are available. For example, Syverson (2004) proposes input price shifters alongside fixed effects. Once $\tilde{\mathcal{T}}$ is estimated, all output elasticities, markups, and productivities (both TFPR and TFPQ in this case) can be appropriately scaled.

C Noise in the Cost of Capital from the CAPM

Say that the standard empirical CAPM is estimated with data on firm i's stock returns. The econometric specification is:

$$\rho_{it}^{\text{stock}} - \rho_t^{\text{safe}} = \alpha_i + \beta_i\left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right) + e_{it} \tag{43}$$

Where ρ_{it}^{stock} is i's stock return at time t, ρ_t^{safe} is the return on the safe asset, and ρ_t^{mkt} is the return of the market portfolio. In this model, β_i represents the risk premium associated with firm i. An OLS estimate $\hat{\beta}_i$ is consistent if the assumptions underlying the CAPM are satisfied, and it includes a mean-zero disturbance in a finite sample:

$$\hat{\beta}_i = \beta_i + \frac{\sum_1^T\left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right)e_{it}}{\sum_1^T\left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right)^2} = \beta_i + e_i^\beta \tag{44}$$

With $\mathrm{E}[e_{it}^\beta] = 0$. In equilibrium the cost of capital (gross of depreciation) faced by firm i is

$$\rho_{it} = \rho_t^{\text{safe}} + \beta_i\left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right) + \delta_i \tag{45}$$

i.e. the return of the safe asset, plus the equity premium scaled by the firm-specific risk factor, plus depreciation. Substituting (44) into (45):

$$\tilde{\rho}_{it} = \rho_t^{\text{safe}} + \hat{\beta}_i \left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right) + \delta_i = \rho_{it} + e_i^{\beta} \left(\rho_t^{\text{mkt}} - \rho_t^{\text{safe}}\right) = \rho_{it} + \varepsilon_{it}^{\rho} \tag{46}$$

With $\mathrm{E}[\varepsilon_{it}^{\rho}] = 0$.

Last, I need to establish that $\mathrm{E}[\varepsilon_{it}^{\rho} | x_{it}, k_{it}] = 0$, i.e. that ε_{it}^{ρ} is also mean-independent of (x_{it}, k_{it}). From (46) and (44), this boils down to whether e_{it}, the shock to firm i's time t returns is mean-independent of (x_{it}, k_{it}). Under the assumption that the simple CAPM captures all non-idiosyncratic differences in stock returns (i.e. under the assumptions that underpin the micro foundation of the CAPM), this is satisfied. However, as shown in Fama and French (1993, 2015) (and many other), the CAPM can be augmented with additional factors to better explain stock returns. With the addition of other factors to (43), the correlation between e_{it} and other firm-level variables should be mechanically reduced, especially given that the augmenting factors are functions of (x_{it}, k_{it}) (e.g. size). In short, concerns about the mean-independence of e_{it} and thus ε_{it}^{ρ} can be addressed by opting for augmented versions of the CAPM.

See Appendix H for log additive measurement error in ρ_{it}.

D Proof of Proposition 1

For ease of notation, I drop all firm and time subscripts. I establish that $\mathrm{E}[\tilde{\alpha}^{C,X} | x, k]$ does not exist by showing that $\mathrm{E}[u | x, k]$ from equation (16) diverges. From equation (16), the residual is:

$$u = -\frac{\varepsilon^{\rho} X K}{(X + (\rho + \varepsilon^{\rho})K)(X + \rho K)} \tag{47}$$

Let $c = -\frac{X}{K} - \rho$, and take the following limits:

$$\begin{aligned} &\lim_{\varepsilon^{\rho} \to \infty} u(\varepsilon^{\rho}) = -\frac{X}{X + \rho K} < \infty \\ &\lim_{\varepsilon^{\rho} \to c^{+}} u(\varepsilon^{\rho}) = +\infty \\ &\lim_{\varepsilon^{\rho} \to c^{-}} u(\varepsilon^{\rho}) = -\infty \end{aligned} \tag{48}$$

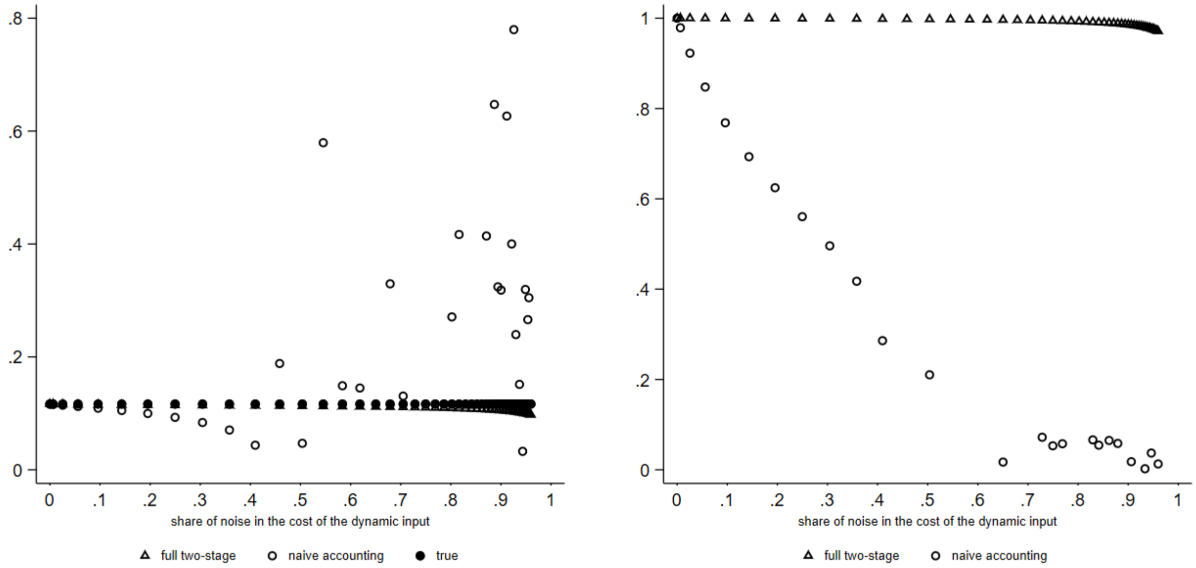
Now let $f_\varepsilon : \mathbb{R} \to \mathbb{R}^+$ be the probability density function of ε^ρ, which is independent of (x, k). Hence, the conditional density of ε^ρ is also $f_\varepsilon(\varepsilon^\rho)$. By the definition of conditional expectation:

$$\mathrm{E}[u|x,k] = \int_{-\infty}^{+\infty} u(\varepsilon^\rho) f_\varepsilon(\varepsilon^\rho)\,\mathrm{d}\varepsilon^\rho = \int_{-\infty}^{c} u(\varepsilon^\rho) f_\varepsilon(\varepsilon^\rho)\,\mathrm{d}\varepsilon^\rho + \int_{c}^{+\infty} u(\varepsilon^\rho) f_\varepsilon(\varepsilon^\rho)\,\mathrm{d}\varepsilon^\rho = -\infty + \infty \quad (49)$$

Where the last equality comes from (48) and the fact that $f_\varepsilon(\varepsilon^\rho) \geq 0$ for every ε^ρ. (49) establishes that $\mathrm{E}[u|x,k]$ does not exist. It immediately follows that $\mathrm{E}[\tilde{\alpha}^{C,X}|x,k] = \theta^{Q,X}(x,k) + \mathrm{E}[u|x,k]$ does not exist either.

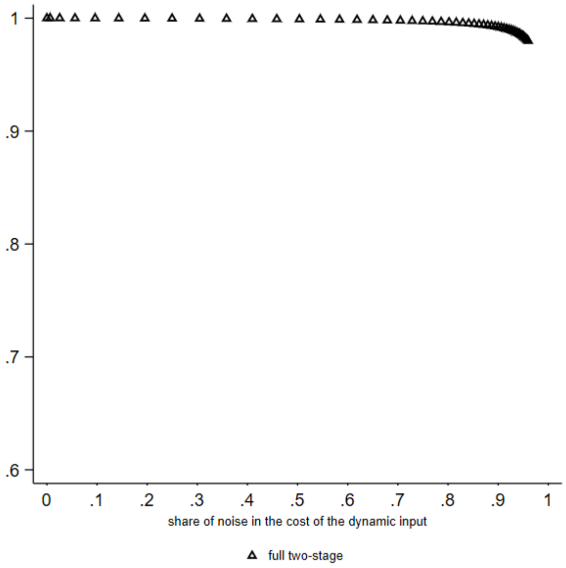
E Additional Simulations Figures

Figure 13: Sensitivity to measurement error in the cost of the dynamic input: average $\hat{\theta}_{it}^{Q,K}$ (left panel), and $\mathrm{corr}(\hat{\theta}_{it}^{Q,K}, \theta_{it}^{Q,K})$ (right panel)



Note: extreme results from the naive accounting approach are omitted.

Figure 14: Sensitivity to measurement error in the cost of the dynamic input: corr$(\hat{\omega}_{it}^{R}, \omega_{it}^{R})$

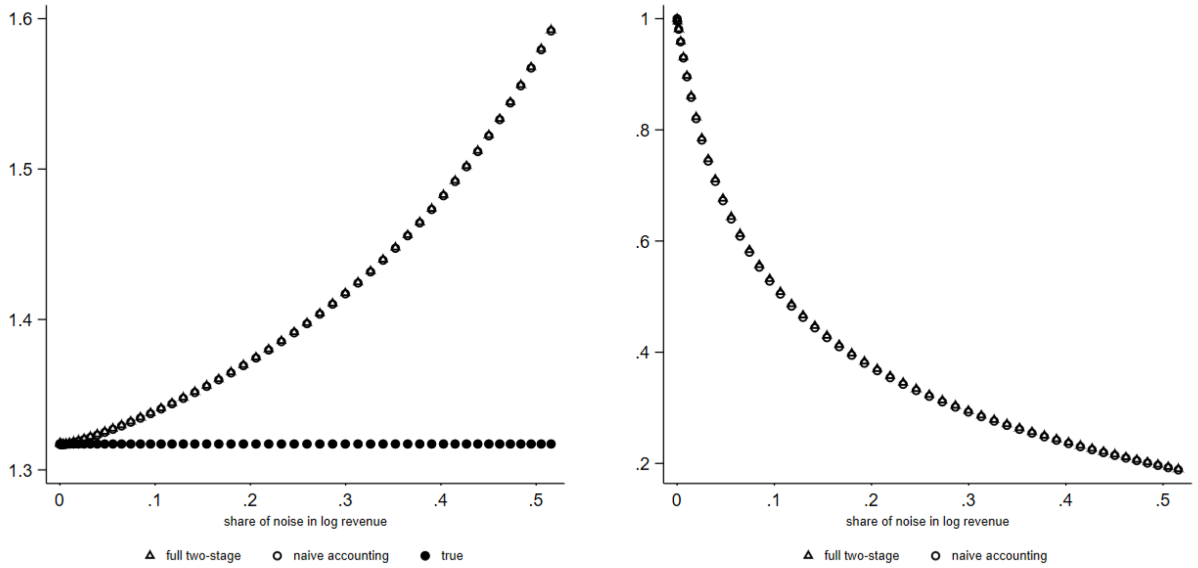


F Sensitivity to Measurement Error in Revenue and Input Levels

I carry out two equivalent exercises to investigate sensitivity to measurement error in revenue (log additive, as in (12)) and in the capital stock (again, log additive). To fully capture the direct effect of these forms of error, I remove noise in the cost of capital, and increase the variance of measurement error in revenue or the log of the capital stock separately. In practical terms, the share of the variance of observed log revenue $\tilde{r}_{it}$ and observed log capital $\tilde{k}_{it}$ that is accounted for by noise is increased from zero to around 50%. Note that when there is no measurement error in any variable, the accounting estimators are all correct by construction, and so is the two-stage approach.

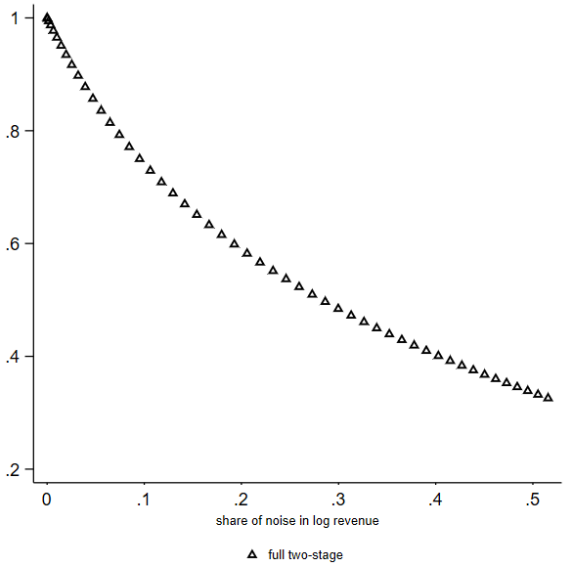
Measurement error in revenue does not affect output elasticities, because input cost shares do not rely on revenue data. Therefore, only markups and TFPR are affected, both in a log-linear fashion (i.e. error is ‘added’ to log markups and log TFPR estimates). Figure 15 presents the performance of the accounting benchmark and the two-stage approach in terms of markup estimates.

Figure 15: Sensitivity to measurement error in revenue: average $\hat{\mu}_{it}$ (left panel), and corr$(\hat{\mu}_{it}, \mu_{it})$ (right panel)



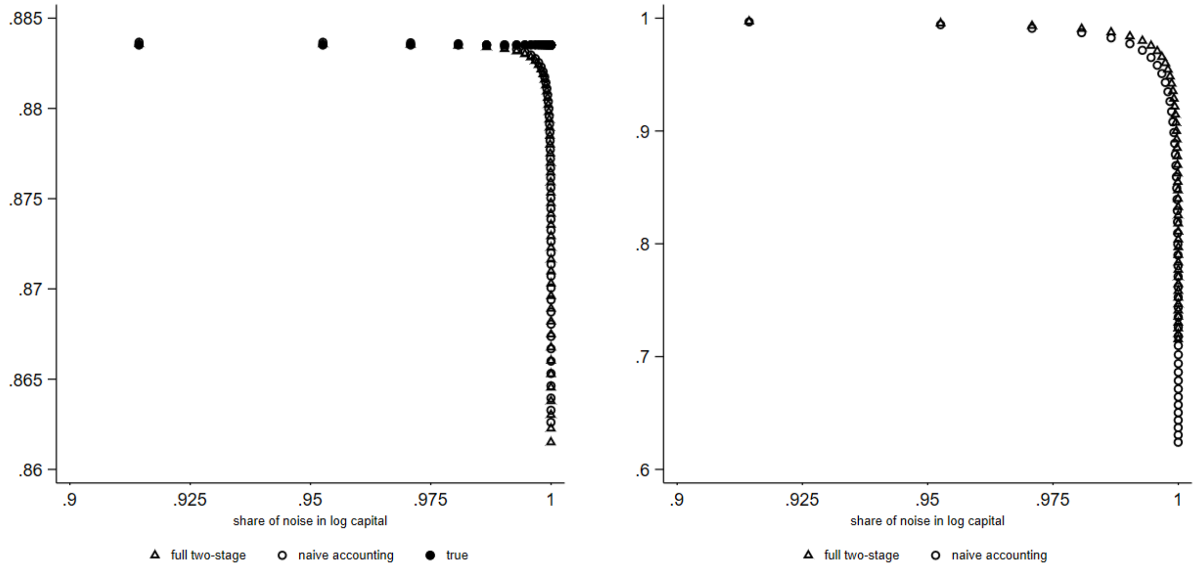
Measurement error in revenue biases the level and distribution of two-stage estimates as expected, i.e. by inflating revenue on average and adding noise across the board. Figure 16 shows the effect of log-additive measurement error in revenue on log TFPR estimates $\hat{\omega}^R_{it}$, which is similar to markup estimates.

Figure 16: Sensitivity to measurement error in revenue: corr($\hat{\omega}_{it}^{R}, \omega_{it}^{R}$) (right panel)



Unlike measurement error in revenue, error in the capital stock enters both the first stage and all cost shares, leading to attenuation bias (Klepper and Leamer, 1984) in the first stage, and division bias (Borjas, 1980) in the second stage for the elasticity of the flexible input. In general, these will depend on (i) how noisy the measurement of the capital stock is, and (ii) how 'important' capital is in production (i.e. its true cost share). In all my Translog simulations the average true capital elasticity (and thus the true cost share) is around 0.12, and the average true economic cost of capital is around 0.14. These values are similar to those used or estimated by most papers (e.g. De Loecker et al. (2020); Gandhi et al. (2020)), and greatly reduce the relevance of capital measurement error because of (ii), as capital is not a large component of total cost in practice. Figure 17 shows how the accounting and the two-stage estimators of the output elasticity of the flexible input $\theta_{it}^{Q,X}$ perform in terms of average (left panel) and correlation with true values (right panel), as the 'share of noise' in $\tilde{k}_{it}$ increases.

Figure 17: Sensitivity to measurement error in the level of the dynamic input: average $\hat{\theta}_{it}^{Q,X}$ (left panel), and corr$(\hat{\theta}_{it}^{Q,X}, \theta_{it}^{Q,X})$ (right panel)



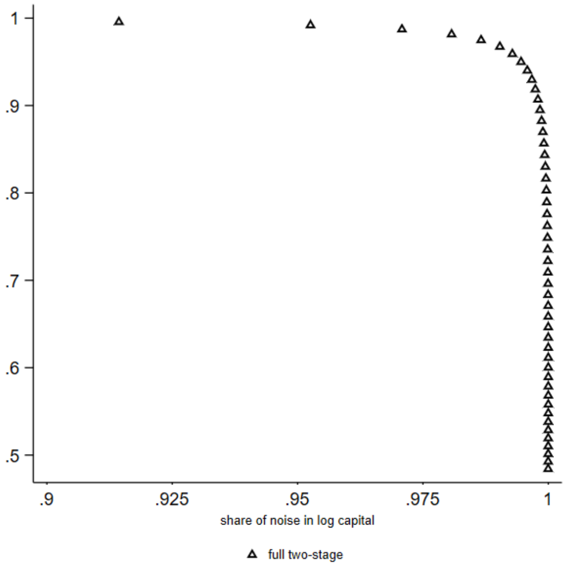
Note: the horizontal axis covers the $0.9 - 1$ interval. For lower levels of noise, estimates and true values are essentially the same.

For this realistically low levels of the cost of capital and the capital elasticity, attenuation and division bias concerns become quantitatively relevant for extremely high noise shares of the observed capital stock ($> 98\%$), and the two-stage approach still overperforms the naive accounting one.[21] This result holds for markups too, and it is reversed for the average capital elasticity because measurement error enters the numerator of the capital cost share. Figure 18 shows that the effect of this type of measurement error on the TFPR distribution is also similar.

---

[21] The overperformance of the two-stage approach is more marked if the user cost of capital is also measured with noise.

Figure 18: Sensitivity to measurement error in the level of the dynamic input: $\text{corr}(\hat{\omega}^R_{it}, \omega^R_{it})$



Note: the horizontal axis covers the $0.9 - 1$ interval. For lower levels of noise, estimates and true values are essentially the same.

G Cobb Douglas Special Case

The proposed estimator does not require the researcher to impose any structure on the functional form of output elasticities. However, there may be a concern that if the true production function is Cobb Douglas, the estimator may pick up non-existing variance in the output elasticity. More formally, the second stage estimation equation for the flexible input elasticity is:

$$\hat{\alpha}^{C,X}_{it} \cdot \mathcal{T} = \theta^{Q,X}(x_{it}, k_{it}) + u^{SS,X}_{it} \tag{50}$$

If the researcher knows in advance that the production function is Cobb Douglas, i.e. $\theta^{Q,X}(x_{it}, k_{it}) = \theta^{Q,X}$ for every i and t, then this restriction can be imposed and the second stage reduces to regressing the (scaled) cost share on a constant. In the more realistic case where the researcher does not know functional forms in advance, the potential correlation between input levels used to approximate $\theta^{Q,X}(\cdot)$ and $u^{SS,X}_{it}$ may create concerns.

Table 2 shows the same Monte Carlo exercise as Table 1 when the production function is Cobb Douglas instead of Translog, and the researcher still regresses purged cost shares on input levels in the second stage, i.e. $\theta^{Q,X}(x_{it}, k_{it}) = \theta^{Q,X}$ is not imposed ex ante.

Table 2: Monte Carlo Simulations: Cobb Douglas + Linear Demand

	flexible input elasticity ($\theta_{it}^{Q,X}$)		
	true value	two-stage	naive accounting
mean	0.8000	0.8003	0.8259
	(0.0000)	(0.0037)	(0.0314)
std. dev.	0.0000	0.0060	0.5149
	(0.0000)	(0.0036)	(0.8282)
	dynamic input elasticity ($\theta_{it}^{Q,K}$)		
	true value	two-stage	naive accounting
mean	0.2000	0.1997	0.1741
	(0.0000)	(0.0037)	(0.0314)
std. dev.	0.0000	0.0060	0.5149
	(0.0000)	(0.0036)	(0.8282)
	markup (μ_{it})		
	true value	two-stage	naive accounting
mean	1.3229	1.3234	1.3615
	(0.0100)	(0.0112)	(0.0442)
std. dev.	0.1801	0.1810	0.7331
	(0.0049)	(0.0073)	(1.0969)
corr. (with true)	–	0.9985	0.3544
	–	(0.0018)	(0.1655)
	TFPR (ω_{it}^{R})		
	true value	two-stage	naive accounting
mean	-	-	-
	-	-	-
std. dev.	0.2214	0.2216	-
	(0.0050)	(0.0057)	-
corr. (with true)	–	0.9994	-
	–	(0.0009)	-

Note: standard deviations in parentheses. For each estimator, I report the mean estimate across 100 Monte Carlo samples of 100 firms and 10 time periods each. Standard deviations are calculated across these 100 samples. In these samples there is no noise in revenue, and noise in the cost of the dynamic input (ε_{it}^{ρ}) accounts for roughly (50%) of the variance of ($\tilde{\rho}_{it}$).

In summary, the two-stage approach correctly detects little to no variance in the output elasticities, which are also correct in levels. The quality of markup and TFPR estimates follows accordingly. This contrasts with the accounting approach, which wrongly assigns the variance in cost shares coming from measurement to the output elasticity estimates, resulting in an estimated elasticity standard deviation around 86 times larger than the two-stage estimator.

H Different Form of Measurement Error in ρ_{it}

Suppose measurement error in the cost of the dynamic input takes the following log additive form:

$$\log \tilde{\rho}_{it} = \log \rho_{it} + \varepsilon^{\rho}_{it} \tag{51}$$

with $E[\varepsilon^{\rho}_{it}|x_{it}, k_{it}] = 0$. First, this type of measurement error leads to bias in the *level* of the cost of capital. In fact, by Jensen's Inequality: $\mathrm{E}[\tilde{\rho}_{it}] \geq \mathrm{E}[\rho_{it}]$, with strict inequality if $\sigma^2_{\varepsilon} > 0$. In addition, the first stage is misspecified if it takes the form of (18), which could lead to additional bias.

First, note that this type of measurement error is almost guaranteed to be present when $\tilde{\rho}$ is calculated by taking a ratio of noisy items from company accounts. Say that a company only finances investment through debt, on which it pays interest. A natural estimator from the cost of capital is:

$$\tilde{\rho}_{it} = \frac{\tilde{\mathcal{I}}_{it}}{\tilde{\mathcal{D}}_{it}} \tag{52}$$

Where $\mathcal{I}_{it}$ is interest payments and $\mathcal{D}_{it}$ is outstanding debt. If either (or both) is measured with log-additive measurement error - similarly to what the literature assumes for revenue - then:

$$\tilde{\rho}_{it} = \frac{\tilde{\mathcal{I}}_{it}}{\tilde{\mathcal{D}}_{it}} = \frac{\mathcal{I}_{it}}{\mathcal{D}_{it}} \cdot \exp\left(\varepsilon^{\mathcal{I}}_{it} - \varepsilon^{\mathcal{D}}_{it}\right) \tag{53}$$

Where $\mathrm{E}[\varepsilon^{\mathcal{I}}_{it}|x_{it}, k_{it}] = \mathrm{E}[\varepsilon^{\mathcal{D}}_{it}|x_{it}, k_{it}] = 0$. (51) follows by taking logs and defining $\varepsilon^{\rho}_{it} := \varepsilon^{\mathcal{I}}_{it} - \varepsilon^{\mathcal{D}}_{it}$.

The practical implication of (51) and (53) is that, when the cost of capital is directly from company accounts, the first stage should be performed in logs to avid bias in the level of the purged rental rate, $\hat{\rho}_{it}$. However, it is worth investigating what happens when the first stage is misspecificed in levels, despite measurement error being log-additive. Table 3 shows that this type of measurement error affects naive accounting estimates substantially less than the additive one investigated in the main text, and leaves the quality of two-stage estimates mostly unchanged.

Table 3: Monte Carlo Simulations: Translog + Linear Demand with log Additive Error in ρ_{it}

	flexible input elasticity ($\theta_{it}^{Q,X}$)		
	true value	two-stage	naive accounting
mean	0.8813	0.8676	0.8734
	(0.0042)	(0.0051)	(0.0047)
std. dev.	0.0878	0.0946	0.1135
	(0.0019)	(0.0030)	(0.0033)
corr. (with true)	–	0.9997	0.8185
	–	(0.0004)	(0.0132)

	dynamic input elasticity ($\theta_{it}^{Q,K}$)		
	true value	two-stage	naive accounting
mean	0.1187	0.1324	0.1266
	(0.0042)	(0.0051)	(0.0047)
std. dev.	0.0878	0.0946	0.1135
	(0.0019)	(0.0030)	(0.0033)
corr. (with true)	–	0.9997	0.8185
	–	(0.0004)	(0.0132)

	markup (μ_{it})		
	true value	two-stage	naive accounting
mean	1.3255	1.3030	1.3122
	(0.0097)	(0.0100)	(0.0101)
std. dev.	0.1705	0.1628	0.1984
	(0.0048)	(0.0052)	(0.0057)
corr. (with true)	–	0.9975	0.8287
	–	(0.0014)	(0.0181)

	TFPR (ω_{it}^{R})		
	true value	two-stage	naive accounting
mean	-	-	-
	-	-	-
std. dev.	0.2307	0.2358	-
	(0.0049)	(0.0053)	-
corr. (with true)	–	0.9998	-
	–	(0.0002)	-

Note: standard deviations in parentheses. For each estimator, I report the mean estimate across 100 Monte Carlo samples of 100 firms and 10 time periods each. Standard deviations are calculated across these 100 samples. In these samples there is no noise in revenue, and log additive noise in the cost of the dynamic input (ε_{it}^{ρ}) accounts for roughly (50%) of the variance of ($\tilde{\rho}_{it}$), which also implies $\mathrm{E}[\tilde{\rho}_{it}] \approx 1.16 \times \mathrm{E}[\rho_{it}]$.

Besides the remarkably better performance of the naive accounting approach, the two-stage approach is comparably good in levels, and still substantially better in distribution for output elasticities and markups. Similarly, it still performs well on log TFPR estimates. Note that the bias in the average level of the data on the cost of capital grows very fast with σ_{ε}^2, and results in implausibly large values of $\mathrm{E}[\tilde{\rho}_{it}]$ for variances that are substantially bigger than the one investigated here.

I Data, Cleaning and Summaries

My data cleaning procedure follows Bond and Gottardo (2024) and De Loecker et al. (2020) quite closely to ensure comparability.

First, I construct the same firm-year user cost of capital as in the former. The 10-year US Treasury is used as the safe asset. For company risk premia, I use firm-year betas from WRDS Beta Suite, which are matched to Compustat observations by ticker symbol. For companies without a matched beta, I use the average beta among observations belonging to the same year, sector, sales decile, and age (since appearing on Compustat) decile. Then, the cost of capital is:

$$\rho_{it} = i_t^{10\mathrm{Y}} - \pi_t + 0.07\beta_{it} + 0.05 \tag{54}$$

where $i_t^{10\mathrm{Y}}$ is the average annual yield of the 10-year US Treasury, π_t is CPI inflation, β_{it} is the firm-year-specific beta, and 0.07 is the assumed market risk premium, and 0.05 is depreciation.

I drop all observations with missing, negative, or null sales, COGS, SGA, or PPE, and then those observations which are below the 1st or above the 99th percentile of the COGS expenditure-sales ratio. For the exercise where I consider SGA a production input, I clean its sales-share similarly.

Figure 19 reports the sample size (number of companies), and the evolution of total deflated sales for every year.

Figure 19: Sample size and evolution of deflated sales

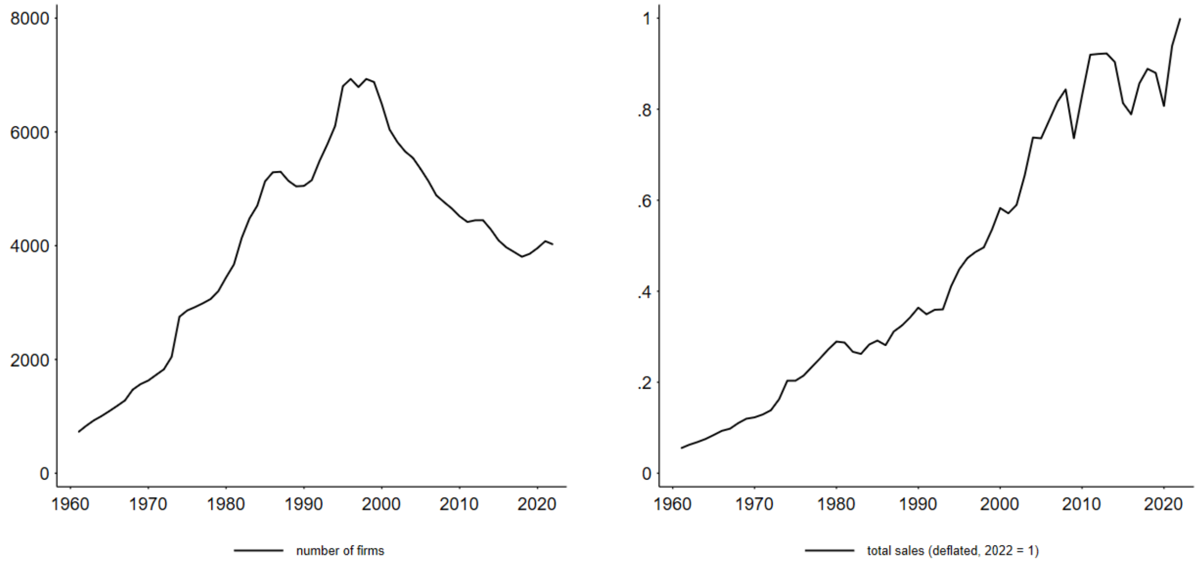
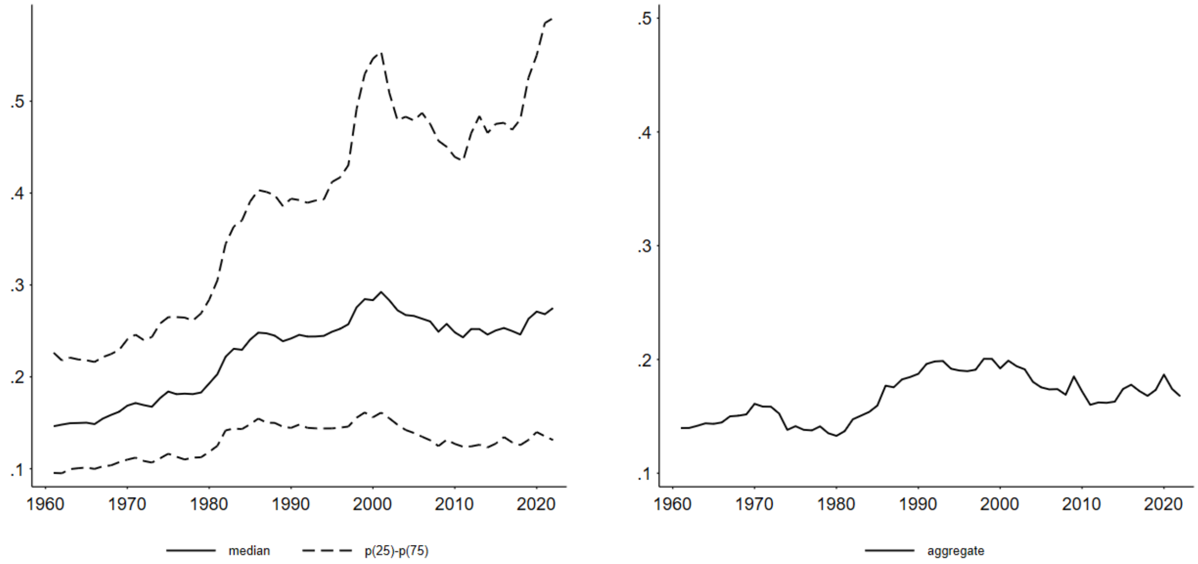


Figure 20 presents the evolution of the share of SGA in sales. The aggregate is obtained by taking the sales-weighted arithmetic mean of firm-level shares, which equals the ratio of total SGA expenditure to total sales.

Figure 20: SGA share of sales



J Compustat Application: SGA Included in Total Cost

I repeat the same estimation procedure while adding SGA to total cost, and treating it as expenditure on a dynamic input. The sample differs marginally from that of the main text only because I also clean the 1st and 99th percentiles of the SGA revenue share, for consistency. Unlike with capital, I cannot perform the second stage on SGA to remove measurement error from its price, because it is measured in currency units. Figures 21, 22, 23, and Figure 24 are the equivalents of Figures 3, 4, 5, and 6 respectively.

Figure 21: COGS output elasticity estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

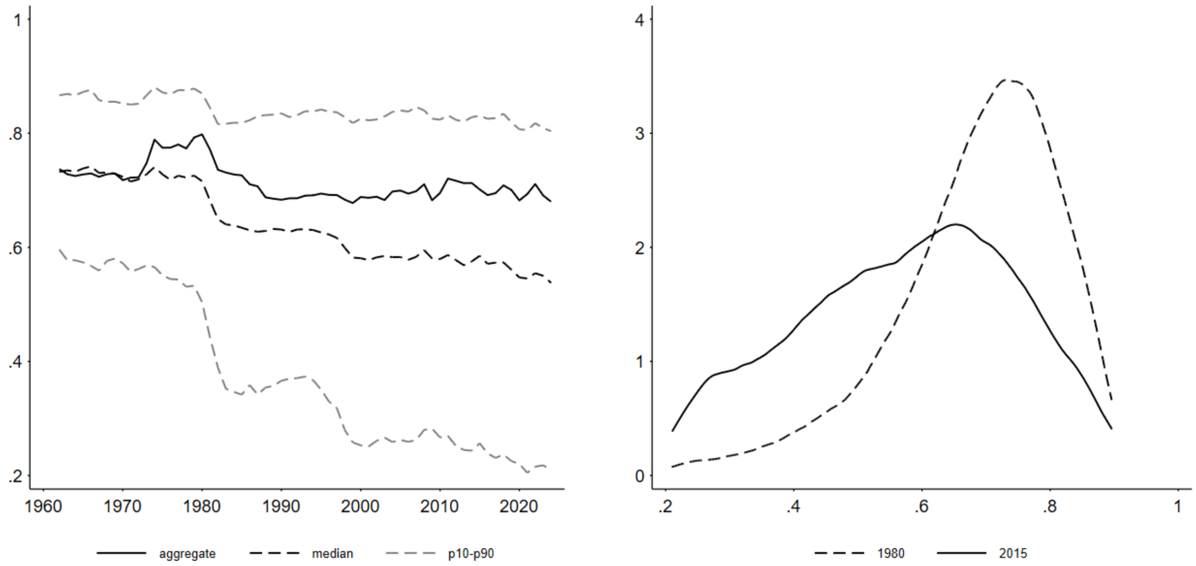


Figure 22: Markup estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

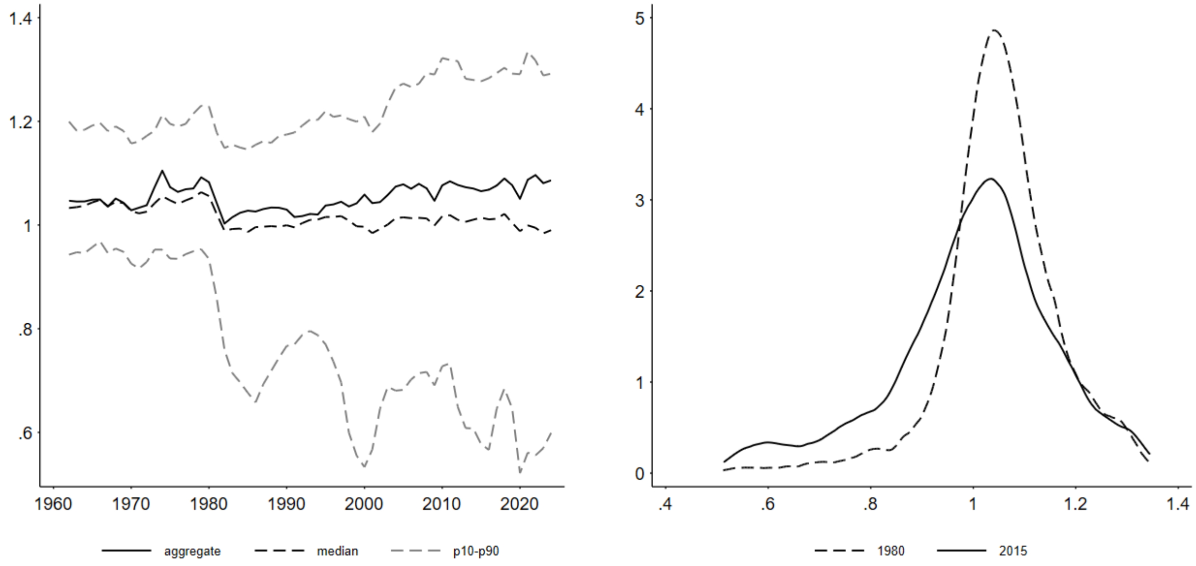


Figure 23: Log TFPR estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

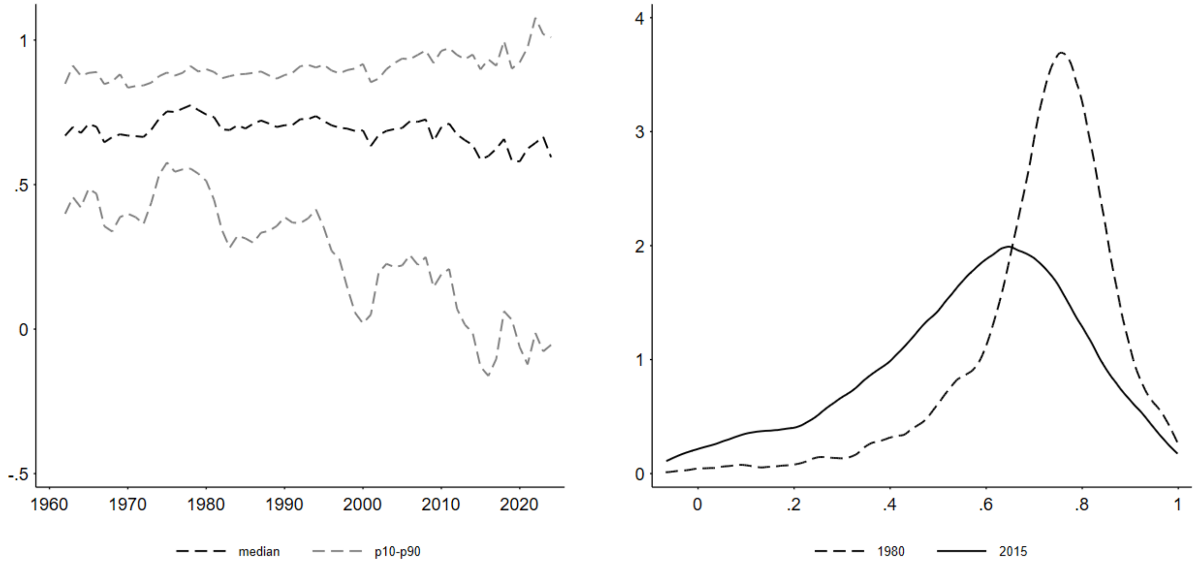
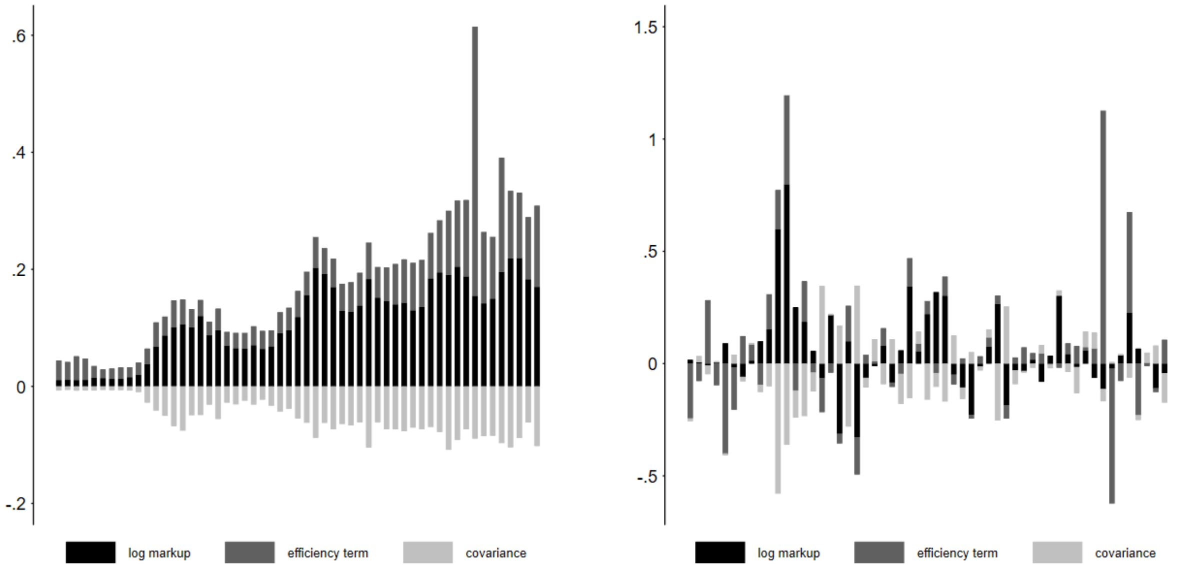


Figure 24: Log TFPR variance decomposition 1970-2023 (SGA included): level (left panel) and changes (right panel)



Note: each column represents one year.

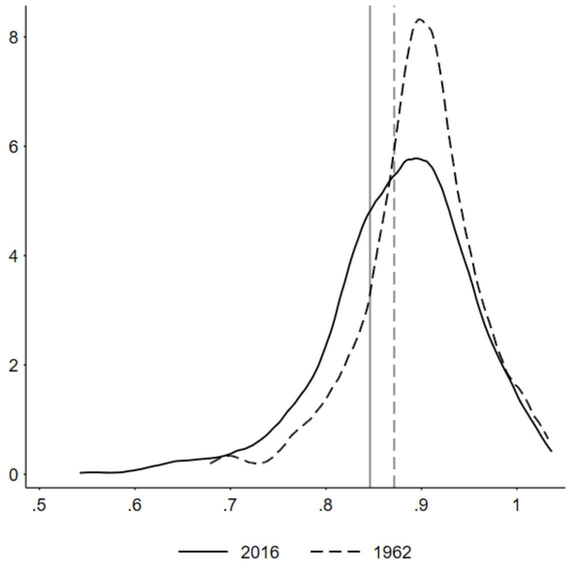
The main common trend between estimates with and without SGA is the general increase in dispersion over time, especially since the 1980s. Including SGA in total cost introduces a negative trend in the COGS elasticity and in the left tails of the markup and TFPR distributions. This is driven by the increase in the average (sales and cost) share of SGA and its dispersion over time (see Figure 20). E.g. the number of firms with SGA expenditure greater than revenue increases from 1% in 1980 to 10% in 2024. It follows that the number of firms with negative accounting profits and price below marginal cost ($\hat{\mu}_{it} < 1$) grows mechanically too. The inclusion of an additional input that grows over time also skews the TFPR distribution and its evolution to the left. However, it also leaves log markup dispersion as the main driver of log TFPR dispersion in levels and over time. the main difference between Figures 6 and 24 is the role of the markup-efficiency covariance term. If SGA is a production input, this covariance is always negative and thus substantially reduces overall log TFPR dispersion.

K Comparison with De Loecker et al. (2020)

Here I compare my estimates to De Loecker et al. (2020), which provides a reference point for many other subsequent papers that analyze the same data under similar assumptions. For their main specification, they assume a year-sector specific Cobb Douglas production function, treat SGA/intangible capital as a sunk cost, and COGS as flexible. There are two main sources of differences between this paper and De Loecker et al. (2020): (i) by treating output elasticities non-parametrically, I allow for departures from Cobb Douglas; and (ii) they employ a proxy approach along the lines of Ackerberg and De Loecker (2024), where price bias is addressed with a price control function, and the demand for the flexible input is inverted (see Section 4 for details).

Departing from Cobb Douglas provides the most quantitatively relevant effect in terms of output elasticity estimates within year and sector, where my approach allows for an arbitrary amount of heterogeneity. Figure 25 shows the estimates of the output elasticity of COGS in 1962 (start of the sample) and 2016 (end of their sample) for sector NAICS 33 (the largest in the data).

Figure 25: Comparison with De Loecker et al. (2020), output elasticity of COGS of sector NAICS 33: distribution of two-stage estimates (black) and DLEU Cobb Douglas estimate (grey). 1962 (dashed) and 2016 (solid).

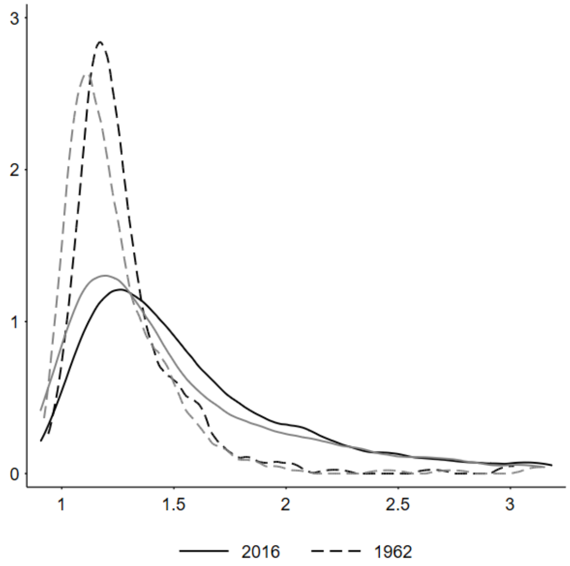


Note: DLEU estimates are those reported by the authors of the paper.

On average, the elasticities of De Loecker et al. (2020) are similar to mine, but the increase in the dispersion of $\theta_{it}^{Q,X}$ over time implies that the Cobb Douglas assumption artificially removes an increasing amount of heterogeneity.

The Cobb Douglas assumption is less problematic for markups, whose variance is largely driven by differences in the revenue share of COGS, rather than differences in output elasticities. Therefore, my markup estimates are rather similar on average and in distribution (Figure 26).

Figure 26: Comparison with De Loecker et al. (2020), distribution of markups in sector NAICS 33: two-stage estimates (black) and DLEU Cobb Douglas estimate (grey). 1962 (dashed) and 2016 (solid).



The Cobb Douglas assumption inevitably ‘assigns’ all dispersion in observed revenue shares to markups. When the common elasticity assumption is relaxed with my approach, some of the dispersion in revenue shares is accounted for by (purged) cost share differences. This leads to the differences in the markup distribution.

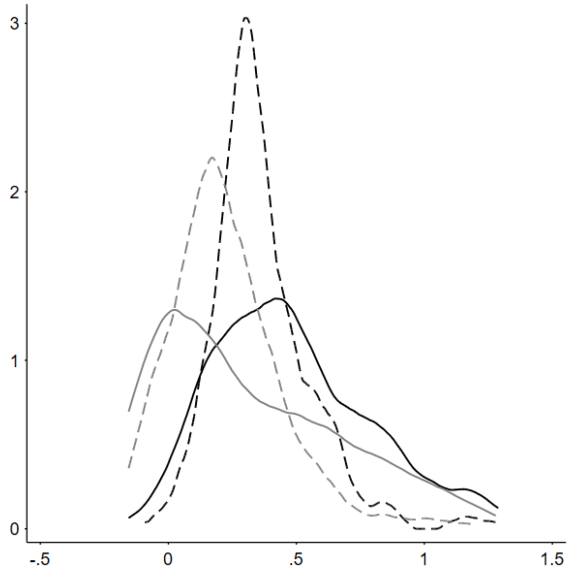
The lack of meaningful differences in levels (both for $\theta_{it}^{Q,X}$ and μ_{it}) may be the result of two countervailing sources of bias in econometric estimates of the output elasticity. If production is Cobb Douglas, when log revenue is regressed on the logs of input levels the true model is:

$$\tilde{r}_{it} = \frac{\theta^{Q,X}}{\mu_{it}} x_{it} + \frac{\theta^{Q,K}}{\mu_{it}} k_{it} + \tilde{\omega}_{it}^{R} + \varepsilon_{it}^{R} \tag{55}$$

where $\tilde{\omega}_{it}^{R}$ is a function of TFPQ and prices. Absent any markup heterogeneity and provided that $\tilde{\omega}_{it}^{R}$ is controlled for successfully, the estimated coefficients suffer from negative bias because $\mu_{it} \geq 1$ (price bias). At the same time, if the controls for $\tilde{\omega}_{it}^{R}$ fail, then positive transmission bias is also present. Since $\tilde{\omega}_{it}^{R}$ contains both TFPQ and price variation - as argued in Section 4 - valid controls are likely to be much more demanding than in competitive settings in terms of required structure on residual demand and productivity. As argued in De Ridder et al. (2022), the overall bias in level can take any sign, so it may very well be quite close to the true value.

Finally, even though De Loecker et al. (2020) do not use their output elasticity estimates to estimate TFPR, this can be calculated from equation (21). Figure 27 shows the estimates of revenue productivity in 1962 and 2016 for sector NAICS 33 (the largest in the data).

Figure 27: Comparison with De Loecker et al. (2020), distribution of log TFPR in sector NAICS 33: two-stage estimates (black) and DLEU Cobb Douglas estimate (grey). 1962 (dashed) and 2016 (solid).



Note: log TFPR is calculated using equation (21).

Log TFPR estimates differ more than markup estimates (correlation in sector NAICS 33 and year 2016 +0.83 and +0.97, respectively),[22] as differences in revenue productivity beyond the Cobb Douglas case are also driven by the higher order terms of the production function.

---

[22] Over the whole sample, the correlation between my log TFPR estimates and De Loecker et al. (2020)'s is +0.66, and that of markups is +0.88.