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

March 17, 2025

Abstract

I propose a new semi-parametric estimation approach that recovers heterogeneous output elasticities, markups, and revenue total factor productivity from company accounts data. The method requires no assumption on demand, the law of motion of productivity, or the invertibility of input demand. It is also immune to the issues with the estimation of production functions when there is market power in output markets. My approach relies on input cost shares as proxies of output elasticities. I characterize the sign and magnitude of the bias in cost-share-based output elasticity estimates from the most plausible forms of measurement error in input expenditures (error in the cost of capital). I show analytically that this measurement error can always be purged from noisy data on the cost of inputs. I then propose a novel two-stage production function estimation approach that recovers heterogeneous output elasticities, markups, and revenue productivity. I use Monte Carlo simulations to show that this approach performs extremely well on multiple dimensions in a very general setting. Finally, I apply the new estimator to Compustat, and show that the biggest driver of the variance of revenue productivity and its rise is markup dispersion.

*I thank Filippo Biondi, Lukas Boehnert, Andrea Chiavari, Howard Smith, Ahmed Tohamy, Leonard Treuren, and especially Steve Bond for the helpful comments and discussions.

1 Introduction

Knowledge of production parameters, markups, and productivity is central to answering several economic questions, e.g. on the secular trends of market power (De Loecker et al., 2020), 'the rise of superstar firms' (Autor et al., 2020; Peters, 2020), the evolution of the labor share (Karabarbounis and Neiman, 2013), how intangible capital accumulation affects the macro-economy and competition (Chiavari, 2021; De Ridder, 2024), etc. Production function estimation is a powerful tool in this realm because it can also be used to estimate price-cost markups, following the approach by De Loecker and Warzynski (2012), and total factor productivity. However, estimating production functions in a setting with market power and data from company accounts is remarkably challenging. The most popular methods are designed for perfectly competitive settings (Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020), and they are either not usable or require strong assumptions about demand when firms have pricing power (Bond et al., 2021; Doraszelski and Jaumandreu, 2023; Biondi, 2024; Ackerberg and De Loecker, 2024).

This paper introduces a novel semi-parametric estimation method that recovers heterogeneous output elasticities, markups, and revenue total factor productivity. This new approach can be applied to standard company accounts data on firm revenue and input expenditures alone, and requires competitive input markets, fixing the returns to scale of the production function, and at least one flexible input (chosen by the firm with full information of the current state and without frictions). Albeit restrictive, these assumptions are not new to the literature (De Loecker and Warzynski, 2012; Flynn et al., 2019), and I show that they are sufficient to avoid all issues raised by the recent literature on the estimation of production functions when there is heterogeneous and unobserved market power. In addition, and unlike most current production function estimation methods, my approach does not require any assumption on the law of motion of unobserved productivity, nor the invsertibility of investment or input demand.

My method relies on input cost shares (expenditure on inputs divided by total cost) as 'proxies' of true output elasticities. It is well known that, given an assumption on the returns to scale, cost shares could be used directly as output elasticity estimates absent measurement error in input expenditures, or under a Cobb Douglas assumption (De Loecker and Syverson, 2021; De Loecker

et al., 2024; Raval, 2023*a*).[1] However, especially when it comes to dynamic inputs such as capital, the true economic cost of which is rarely observed, noisy data on input expenditures turns cost shares into biased proxies. In this paper, I characterize and quantify this bias analytically and with Monte Carlo simulations in a very general setting, and propose a two-stage econometric approach to eliminate it. My proposed method involves purging measurement error from the cost of inputs and cost shares, and then estimating output elasticities non-parametrically from the first order condition of cost minimization. These output elasticity estimates can be used to calculate markups and revenue total factor productivity.

I test my estimator with Monte Carlo simulations in a setting with Translog production and linear demand, and show that it correctly recovers the level and distribution of both output elasticities and markups, even with extremely noisy data. It also substantially overperforms the naive cost-share approach that does not deal with measurement error in input expenditures. I check the sensitivity of my new estimator to different sources and forms of measurement error (especially in revenue and in the capital stock), and show that it performs well for plausible amounts of unaccounted-for noise. I also show it can detect a Cobb Douglas production function (constant output elasticity) when this assumption is not imposed *ex ante*.

Finally, I apply the new estimator to data on company accounts of US listed firms and find average output elasticities and markups that are in line with the past literature, but different distributions. This seems to largely come from the use of a Cobb Douglas assumption of past papers (Traina, 2018; De Loecker et al., 2020; Chiavari, 2021). I then decompose the variance of revenue productivity into a markup and an 'production efficiency' term, and show that the markup dispersion accounts for most productivity dispersion. The same is true for the its increase over time. My results also point to non-trivial interactions between market power and quantity productivity, even though I cannot measure the latter.

The paper is structured as follows: Section 2 derives all the first-order conditions that underpin my approach in the most general setting, Section 3 provides an overview of the theoretical and practical limitations of existing approaches; Section 4 characterizes some conditions for the identification of

[1] If any input faces adjustment frictions, then the cost share is still a valid proxy of the true output elasticity as long as expenditure on the input with frictions is calculated using its 'true economic cost' instead of its nominal price (more below).

output elasticities through cost shares, characterizes the bias from the most plausible form of noise in input expenditures, and shows analytically that measurement error can be be purged from noisy data on the cost of inputs; Section 5 outlines the new approach using the results of the previous Section; Section 6 presents the results of Monte Carlo simulations in a setting with heterogeneous and endogenous output elasticities and markups, and increasingly noisy data; and Section 7 is the Compustat application.

2 Theory

In this Section I characterize a simple but general model of a profit-maximizing firm. 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.

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 $\phi_i(I_{it})$. 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:

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

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 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}. For ease of notation, denote current period profits as $\Pi_{it} = P_{it} \cdot Q_{it} - X_{it} - \phi_i(I_{it})$. The envelope condition can be rearranged as follows:

$$\frac{\partial \Pi_{it}}{\partial K_{it}} = \frac{\partial V_{it}}{\partial K_{it}} - \beta_i \cdot (1 - \delta_i) \cdot \mathrm{E}_t \left[\frac{\partial V_{it+1}}{\partial K_{it+1}} \right] \tag{3}$$

This equation pins down the return that the marginal unit of capital must yield in the current period, i.e. the 'rental rate' or the cost of capital. If I differentiate the profit function and multiply

both sides by $\frac{K_{it}}{Q_{it}}$, I obtain:[2]

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

But since $\frac{\partial \Pi_{it}}{\partial K_{it}}$ is just the equilibrium rental rate, I can express this 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. In addition, identical results 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 (expected) 'marginal cost',[3] which implies that all 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 $\rho_{it} := \frac{\partial \Pi_{it}}{\partial K_{it}}$ is the user cost of capital (the rental rate from (4)). Notice that current period capital K_{it} is not a choice variable because 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:

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

[2] 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).

[3] The marginal cost of one unit of a dynamic input depends on expectations on its future opportunity-cost, i.e. the second term of the right hand side of (4).

[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 $\rho_{it} = \frac{\partial \Pi_{it}}{\partial K_{it}}$ is that which equates capital's output elasticity and revenue share when there is no market power ($\mu_{it} = 1$).

This definition of the cost of capital 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 accounting markup, i.e. the ratio of revenue and true total cost, scaled by the returns to scale. 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 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 cost share of the flexible input, and $\alpha_{it}^{C,X}$ is the 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. Sections 4 and 5 deal with this by leveraging (4), which expresses the true economic cost of an input as a function of current variables.

3 Identification Issues

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, which will then motivate the method I propose in Section 4.

In all its applications I am aware of, the De Loecker and Warzynski (2012) approach is used to estimate firm-level markups *after* the output elasticity of a flexible input is estimated with standard production function techniques (Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003; Ackerberg et al., 2015), or with a cost share approach under a Cobb Douglas assumption (De Loecker and Fleitas, 2021; Raval, 2023*a*; De Loecker et al., 2024). Production function methods are likely to fail at estimating the required output elasticity correctly 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 et al. (2021); Doraszelski and Jaumandreu (2023); Biondi (2024). Existing cost share approaches, albeit closer to my own proposal of Section 4, appear to be limited to a Cobb Douglas production function, or to competitive settings (Gandhi et al., 2020).

3.1 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):

$$x_{it} = x(\omega_{it}^{Q}, k_{it}) \tag{11}$$

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:

$$x_{it} = x(\omega_{it}^{Q}, k_{it}, \Upsilon_{it}) \tag{12}$$

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 4) 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.[5] 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.

3.2 Quantities Are Not Observed

In many - if not most data applications physical output is not observed at the firm-level. Notice that, if output prices differ across firms, deflating the value of output (or revenue, or sales) with economy or industry-wide price deflators does not yield a valid proxy for physical output. Since common deflators cannot capture price heterogeneity, the deflated variable will still contain price variation, and all the following arguments apply.

The mapping from inputs to quantity (the production function) differs from that of inputs to

[5] $x(\Omega_{it}^{Q}, k_{it}, \Upsilon_{it})$ 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.

revenue (the revenue function) when output markets are not competitive. In logs, revenue is:

$$r_{it} = p_{it} + q_{it} \tag{13}$$

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

$$\theta_{it}^{R,X} := \frac{\partial r_{it}}{\partial x_{it}} = \frac{\theta_{it}^{Q,X}}{\mu_{it}} \tag{14}$$

i.e. the revenue elasticity is the output elasticity divided by the markup.[6]

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:

$$r_{it} = \tilde{\omega}_{it}^{R} + \frac{\theta^{Q,X}}{\mu_{it}} x_{it} + \frac{\theta^{Q,K}}{\mu_{it}} k_{it} \tag{15}$$

where $\tilde{a}_{it}^{R}$ is the residual of the revenue function.[7] It follows that the revenue function has heterogeneous coefficients in the presence of heterogeneous markups.[8] 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 (15) 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:

$$r_{it} = q_{it} + p_{it} = \omega_{it}^{Q} + f(x_{it}, k_{it}) + p(\mathbf{z}_{it}^{P}) \tag{16}$$

[6] 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).

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

[8] The only revenue function that is straightforward to estimate is the one covered in Klette and Griliches (1996), which is the same as (15) but with common markups, i.e. $\mu_{it} = \mu$ for every i and t.

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 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:

$$r_{it} = f(x_{it}, k_{it}) + \omega_{it}^R = f(x_{it}, k_{it}) + g(r_{it-1} - f(x_{it-1}, k_{it-1})) + \xi_{it}^R \tag{17}$$

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.

3.3 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, instead of the true cost of capital ρ_{it}, some noisy proxy $\tilde{\rho}_{it}$ s.t. $\mathrm{E}[\tilde{\rho}_{it}] = \rho_{it}$ is observed. Call the resulting noisy cost share of the flexible input $\tilde{\alpha}_{it}^{C,X}$. Clearly $\tilde{\alpha}_{it}^{C,X} \neq \theta_{it}^{Q,X}$. In addition, since measurement error enters the denominator of the cost share, because of Jensen's inequality:

$$\mathrm{E}\left[\tilde{\alpha}_{it}^{C,X}\right] \neq \mathrm{E}\left[\theta_{it}^{Q,X}\right] \tag{18}$$

Currently available approaches successfully address this problem in a Cobb Douglas setting (Raval, 2023*a*; 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*a*), 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 $\frac{\mathrm{E}[X_{it}]}{\mathrm{E}[X_{it}+\tilde{\rho}_{it}K_{it}]} \neq \mathrm{E}\left[\frac{X_{it}}{X_{it}+\rho_{it}K_{it}}\right] = \mathrm{E}\left[\theta_{it}^{Q,X}\right]$, and $\mathrm{median}\left[\frac{X_{it}}{X_{it}+\tilde{\rho}_{it}K_{it}}\right] \neq \mathrm{E}\left[\frac{X_{it}}{X_{it}+\rho_{it}K_{it}}\right] = \mathrm{E}\left[\theta_{it}^{Q,X}\right]$ unless $\theta_{it}^{Q,X} = \theta^{Q,X}$ for every i and t.

In the next Sections I will characterize this type of bias in noisy cost shares more generally, and

extend the cost share approach beyond Cobb Douglas in light of the results.

4 Semi-parametric 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 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. I also make no assumption on the functional form of the production function. My arguments require two assumptions: (i) fixed returns to scale in the production function $\mathcal{T}$, and (ii) there is no market power in input markets. 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).

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. E.g. the available data on input prices differs from the true economic cost per unit of input faced by the firm. In this Section, I show what conditions measurement error in input expenditures must satisfy for the semi-parametric identification of output elasticities and provide some examples that satisfy the conditions. In addition, I explicitly characterize the bias under the most plausible type of error: additive noise in the rental rate, i.e. when the data records $\tilde{\rho}_{it}$, a noisy proxy of the true economic cost of capital instead of ρ_{it} (which would require knowledge of the profit function as shown by (4)).

$$\tilde{\rho}_{it} = \rho_{it} + \varepsilon_{it}^{K} \tag{19}$$

where ε_{it}^K is white noise, i.e. $\mathrm{E}[\varepsilon_{it}^K|x_{it}, k_{it}] = 0$. Notice that this type of measurement error contaminates observed capital expenditure. See Appendix C for log additive measurement error in ρ_{it}.

I do not need to take a specific stance on measurement error in revenue for the estimation of output elasticities because my estimator only relies on input expenditures. However, to estimate markups and TFPR revenue data is also required. The estimation of markups also warrants a third assumption: (iii) at least one input is flexible. In line with the literature, I assume measurement error in revenue is log-additive:

$$\tilde{r}_{it} = p_{it} + q_{it} + \varepsilon_{it}^R \tag{20}$$

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

4.1 Non-parametric Identification of the Output Elasticity with Market Power

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:

$$\log \alpha_{it}^{R,X} = \log \theta^{Q,X}(x_{it}, k_{it}) - \varepsilon_{it}^R \tag{21}$$

The log of the elasticity is identified because ε_{it}^R is white noise and thus $\mathrm{E}[\log \alpha_{it}^{R,X}|x_{it}, k_{it}] = \log \theta^{Q,X}(x_{it}, k_{it})$.[9] This identification argument fails in the presence of market power without additional markup information. (21) becomes:

$$\log \alpha_{it}^{R,X} = \log \theta^{Q,X}(x_{it}, k_{it}) - \log \mu_{it} - \varepsilon_{it}^R \tag{22}$$

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 and demand parameters. Since output is a function of input levels and TFPQ, I can write $\log \mu_{it} =$

[9] The only difference between this argument and that of Gandhi et al. (2020) is that they model ε_{it}^R 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 $\mathrm{E}[\exp(\varepsilon_{it}^R)]$, because my shock does not affect true revenue, but just its observed counterpart.

$\log \mu(x_{it}, k_{it}, \omega_{it}^{Q}, \Upsilon_{it})$. 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 TFPQ 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, (10) still holds. Under the assumption that the returns to scale parameter is $\mathcal{T}$ and that *input expenditures are perfectly observed*, I can write:

$$\alpha_{it}^{C,X} \cdot \mathcal{T} = \theta^{Q,X}(x_{it}, k_{it}) \tag{23}$$

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[10] identification argument does not extend to additive or log-additive measurement error in the cost of capital, which is the most pressing concern with the use of cost shares.

4.2 Bias in Output Elasticities from Noisy Data

As already mentioned, the true economic cost of non-flexible inputs is unlikely to be perfectly observed, which in turn implies that the semi-parametric identification argument of (23) is generally invalid. Say that the cost of capital is noisily observed, with additive error as in (19), i.e. $\tilde{\rho}_{it} = \rho_{it} + \varepsilon_{it}^{K}$ s.t. $\mathrm{E}[\varepsilon_{it}^{K} | x_{it}, k_{it}] = 0$. With some algebraic manipulation (23) becomes:

$$\begin{aligned} \tilde{\alpha}_{it}^{C,X} \cdot \mathcal{T} &= \theta^{Q,X}(x_{it}, k_{it}) + u_{it} \\ \text{s.t.} \quad u_{it} &= -\frac{\varepsilon_{it}^{K} X_{it} K_{it}}{(X_{it} + (\rho_{it} + \varepsilon_{it}^{K}) K_{it})(X_{it} + \rho_{it} K_{it})} \end{aligned} \tag{24}$$

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

where ε_{it}^K is the additive error in the cost of capital. As the residual u_{it} depends on ε_{it}^K and input levels in a non-linear fashion, $\mathrm{E}[u_{it}|x_{it}, k_{it}] \neq 0$. In words, measurement error in the cost of an(y) input introduces bias. Proposition 1 establishes that this bias is positive.

Proposition 1. *If* $E[\varepsilon_{it}^K|x_{it}, k_{it}] = 0$ *and* ε_{it}^K *is additive, then*

$$E\left[\tilde{\alpha}_{it}^{C,X} \cdot \mathcal{T}|x_{it}, k_{it}\right] \geq \theta^{Q,X}(x_{it}, k_{it}) \tag{25}$$

Proof. From (24), $\mathrm{E}\left[\tilde{\alpha}_{it}^{C,X} \cdot \mathcal{T}|x_{it}, k_{it}\right] = \log \theta^{Q,X}(x_{it}, k_{it}) + \mathrm{E}\left[u_{it}|x_{it}, k_{it}\right]$. The definition of u_{it} from (24) is convex in ε_{it}^K, thus by Jensen's inequality $\mathrm{E}\left[u_{it}|x_{it}, k_{it}\right] \geq -\frac{\mathrm{E}[\varepsilon_{it}^K|x_{it},k_{it}]X_{it}K_{it}}{(X_{it}+(\rho_{it}+\mathrm{E}[\varepsilon_{it}^K|x_{it},k_{it}])K_{it})(X_{it}+\rho_{it}K_{it})} = 0$. □

I can also quantify the Jensen's Gap of Proposition 1 up to a second-order approximation, to show that it increases with the variance of noise in the cost of the dynamic input.

Proposition 2. *If* $E[\varepsilon_{it}^K|x_{it}, k_{it}] = 0$, $Var[\varepsilon_{it}^K|x_{it}, k_{it}] = \sigma_\varepsilon^2$, *and* ε_{it}^K *is additive, then*

$$\begin{aligned} E\left[u_{it}|x_{it}, k_{it}\right] &\approx \frac{1}{2}u''(0) \cdot \sigma_\varepsilon^2 \\ \textit{and} \quad u''(0) &> 0 \end{aligned} \tag{26}$$

Proof. Take a second-order Taylor expansion around $\varepsilon_{it}^K = 0$ of the definition of u_{it} from equation (24). I can write $u_{it}(\varepsilon_{it}^K) \approx u_{it}(0) + u'_{it}(0) \cdot \varepsilon_{it}^K + \frac{1}{2} \cdot u''_{it}(0) \cdot \left(\varepsilon_{it}^K\right)^2$. Taking the conditional expectation of both sides and evaluating u''_{it} at $\varepsilon_{it}^K = 0$ yields the result. □

An intuitive interpretation of Proposition 2 is that measurement error in input expenditures is not a concern when its variance is small. The convexity of u_{it} in ε_{it}^K may depend on other variables for different forms of measurement error (e.g. log-additive ε^K), which prevents the re-statement of these Propositions in equally general terms. However, in Appendix C I perform Monte Carlo simulations with log-additive measurement error in the cost of the dynamic input, and show that (i) it introduces less severe problems for cost shares than the additive error analyzed here, (ii) my proposed estimator of Section 5 still performs very well despite the assumption violation.

4.3 Purging Measurement Error from the Cost of Inputs

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:

$$\rho_{it} = \frac{\theta_{it}^{Q,K} X_{it}}{\theta_{it}^{Q,X} K_{it}} \tag{27}$$

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.[11]

The RHS of (27) 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 (27) into (19) suggests the following non-parametric regression equation:

$$\tilde{\rho}_{it} = \rho(x_{it}, k_{it}) + \varepsilon_{it}^{K} \tag{28}$$

where $\rho(\cdot)$ is a non-linear function. Notice that $\mathrm{E}[\varepsilon_{it}^{K} | x_{it}, k_{it}] = 0$ by the definition of ε_{it}^{K}. 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 (28) 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 $\mathrm{E}[\tilde{\rho}_{it}] > \mathrm{E}[\rho_{it}]$. 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}^{K} is log-additive instead of simply additive, this non-parametric regression can still recover the true

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

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 Appendix C I show that for plausible values of the variance of log-additive ε_{it}^{K}, this is a very minor issue.

5 A General Two-Stage Cost Share Estimator

In this Section I propose a two-stage estimator for (all) output elasticities, markups, and the distribution of TFPR, based on the remarks of Section 4. The first stage involves purging noise from the cost of the dynamic input by operationalizing (28) and reconstructing noise-purged input cost shares. In the second stage, these cost shares are regressed on non-parametric functions of input levels to obtain estimates of output elasticities. The required assumptions are the same as Section 4: (i) the returns to scale of the production function are known and equal $\mathcal{T}$; and (ii) there is no market power in input markets. To estimate markups, I also require that (iii) there is at least one flexible input (X_{it} in this case). (i) cannot be tested unlike in Syverson (2004), because I am working under the assumption that the researcher lacks data on physical output Q_{it}. But robustness checks can be performed by changing the assumption on $\mathcal{T}$ and recalculating all second-stage estimates, as the first stage is unaffected by the choice of $\mathcal{T}$. (ii) cannot be tested either, but if external markdown estimates are available (as in Rubens (2021) for example), they can be incorporated in cost shares to relax this assumption. (iii) can be tested in the presence of multiple (alleged) flexible inputs by performing the test proposed by Raval (2023*b*), where markup estimates based on different inputs are compared.

5.1 First Stage

Both stages only require data on $[X_{it}, K_{it}, \tilde{\rho}_{it}]$ (input levels and expenditures). However, in real data applications $\tilde{\rho}_{it}$ is usually calculated or estimated in advance. For listed companies, $\tilde{\rho}_{it}$ can be the weighted average cost of capital (WACC), or just the cost of debt/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 the estimation of a capital asset pricing model (CAPM) from accounts/stock market data (Fama and French, 1993). 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. However, any proxy of choice for the cost of capital should satisfy the following: (i) it is correct on average, i.e. $\mathrm{E}[\tilde{\rho}_{it}] = \mathrm{E}[\rho_{it}]$ (in a CAPM setting, this largely boils down to the correct choice of the risk-free yield); (ii) it contains sufficient variation across observations. (ii) rules out using annual and/or sector averages, unless the assumption that the cost of capital is the same within year and/or sector is made explicitly. This is needed because (28) requires data $\tilde{\rho}_{it}$ to capture *all* the heterogeneity in the underlying true value ρ_{it}, so that input levels can be used to remove measurement error.

Irrespective of the origin of $\tilde{\rho}_{it}$, data on $[X_{it}, K_{it}, \tilde{\rho}_{it}]$ can be used to estimate (28) 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 (28) 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 $\hat{\alpha}_{it}^{C,X}$ and $\hat{\alpha}_{it}^{C,K}$.

5.2 Second Stage

For the second stage, purged cost shares can be used to estimate (23) semi-parametrically:

$$\begin{aligned} \hat{\alpha}_{it}^{C,X} \cdot \mathcal{T} &= \theta^{Q,X}(x_{it}, k_{it}) + u_{it}^{SS,X} \\ \hat{\alpha}_{it}^{C,K} \cdot \mathcal{T} &= \theta^{Q,K}(x_{it}, k_{it}) + u_{it}^{SS,K} \end{aligned} \tag{29}$$

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 (29) delivers a perfect fit, i.e. $u_{it}^{SS,X} = u_{it}^{SS,K} = 0$ for every i and t. However, in light of the fact that both stages rely on non-parametric approximations, $u_{it}^{SS,X}$ and $u_{it}^{SS,K}$ are allowed to differ and take non-zero values. For example, if the first stage somehow fails at fully removing measurement error from $\tilde{\rho}_{it}$,[12] the fit will not be perfect. In this

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

case, Proposition 2 still applies. By virtue of having removed at least *some* noise from the cost shares, Proposition 2 implies that the bias in the output elasticity must be smaller up to a second order approximation. The Monte Carlo simulations of Section 6 and of the Appendices show that performing the first stage usually achieves a considerable improvement over using naive cost shares, even when the purging of ε_{it}^K 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:

$$\hat{\mu}_{it} = \frac{\hat{\theta}_{it}^{Q,X}}{\tilde{\alpha}_{it}^{R,X}} \tag{30}$$

Notice that the revenue share $\tilde{\alpha}_{it}^{R,X} := \frac{X_{it}}{R_{it}}$ 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 3, 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:

$$\hat{\omega}_{it}^R = (\omega_{it}^Q \widehat{+ p_{it}} + \varepsilon_{it}^R) = \tilde{r}_{it} - \int_0^{x_{it}} \hat{\theta}^{Q,X}(x, k_{it})\,\mathrm{d}x - \int_0^{k_{it}} \hat{\theta}^{Q,K}(0, k)\,\mathrm{d}k_{it} \tag{31}$$

6 Monte Carlo Simulations

I test the proposed estimator 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_{it}^Q, 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:

$$\begin{aligned} V(K_{it}, \Omega^Q_{it}, \Upsilon_{it}) = \max_{X_{it}, I_{it}} & P(Q_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \frac{1}{2}I^2 + \beta \cdot \mathrm{E}_t[V(K_{it+1}, A^Q_{it+1}, \Upsilon_{it+1})] \\ \text{s.t.} \quad & Q_{it} \leq \Omega^Q_{it} \cdot F(X_{it}, K_{it}) \\ & K_{it+1} = (1-\delta) \cdot K_{it} + I_{it} \end{aligned} \tag{32}$$

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

$$q_{it} = \omega^Q_{it} + \beta^X x_{it} + \beta^K k_{it} + \beta^{XK} x_{it} k_{it} + \beta^{XX} x^2_{it} + \beta^{KK} k^2_{it} \tag{33}$$

βs are chosen so that the production function is concave and constant returns to scale regardless of input choice.[13] $P_{it}(\cdot)$ 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:

$$P_{it}(Q_{it}, \Upsilon_{it}) = \Upsilon^0_{it} - \Upsilon^1_{it} Q_{it} \tag{34}$$

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 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 B I report the results of the same exercise when the production function is Cobb Douglas.

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

6.1 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 (19) and (20). 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.

$$\begin{aligned} \hat{\theta}_{it}^{Q,X} &= \tilde{\alpha}_{it}^{C,X} \\ \hat{\theta}_{it}^{Q,K} &= \tilde{\alpha}_{it}^{C,K} \\ \hat{\mu}_{it} &= \frac{\hat{\theta}_{it}^{Q,X}}{\tilde{\alpha}_{it}^{R}} \end{aligned} \tag{35}$$

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 5: 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 (31).

6.2 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}^{K} 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 B I perform the same exercise when production is Cobb Douglas, and obtain equally good results.

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

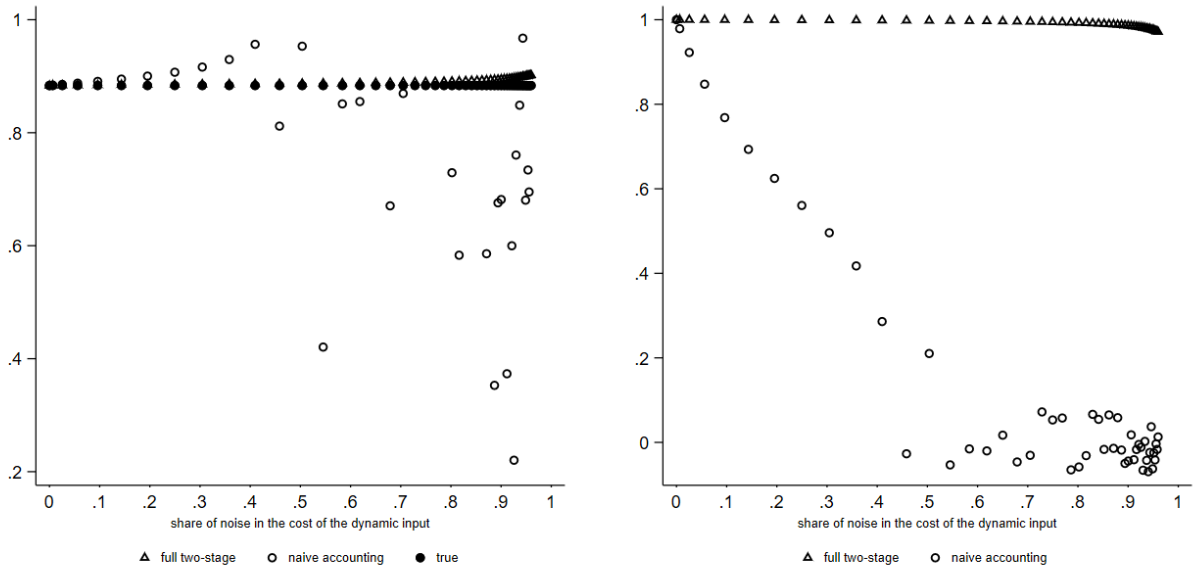
	flexible input elasticity ($\theta_{it}^{Q,X}$)		
	true value	two-stage	naive accounting
mean	0.8813	0.8820	0.9566
	(0.0042)	(0.0078)	(0.2261)
std. dev.	0.0878	0.0860	2.8563
	(0.0019)	(0.0093)	(6.3184)
corr. (with true)	–	0.9962	0.1564
	–	(0.0048)	(0.1269)
	dynamic input elasticity ($\theta_{it}^{Q,K}$)		
	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}^{K}) accounts for roughly (50%) of the variance of ($\tilde{\rho}_{it}$).

6.3 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}^{K} 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.

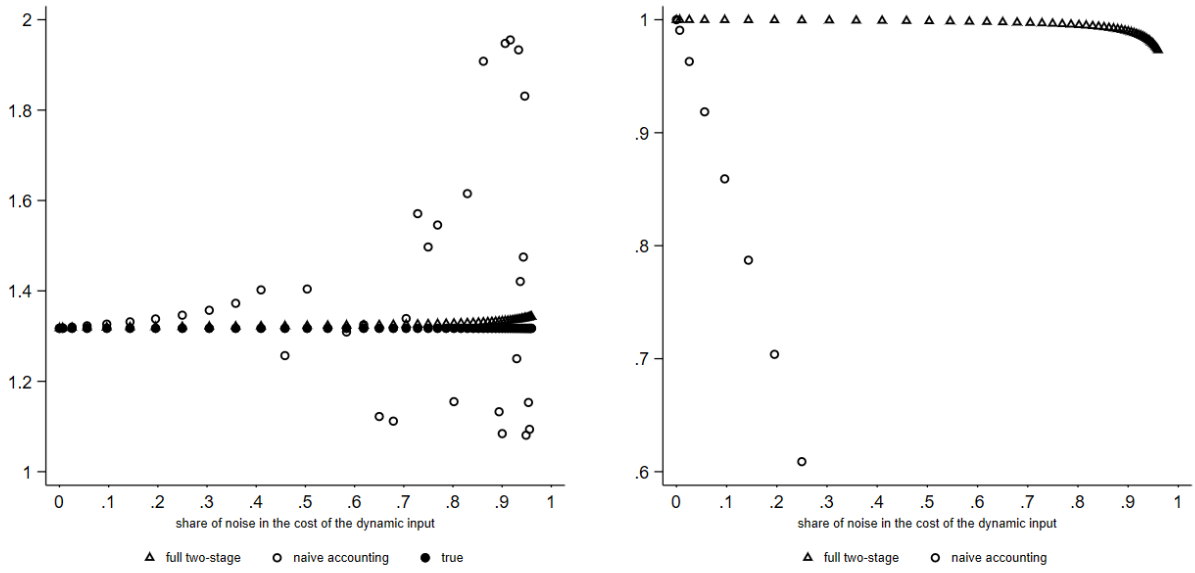
Figure 1: Sensitivity to measurement error in the cost of the dynamic input: average $\hat{\theta}_{it}^{Q,X}$ (left panel), and corr$(\hat{\theta}_{it}^{Q,X}, \theta_{it}^{Q,X})$ (right panel)



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

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. 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)

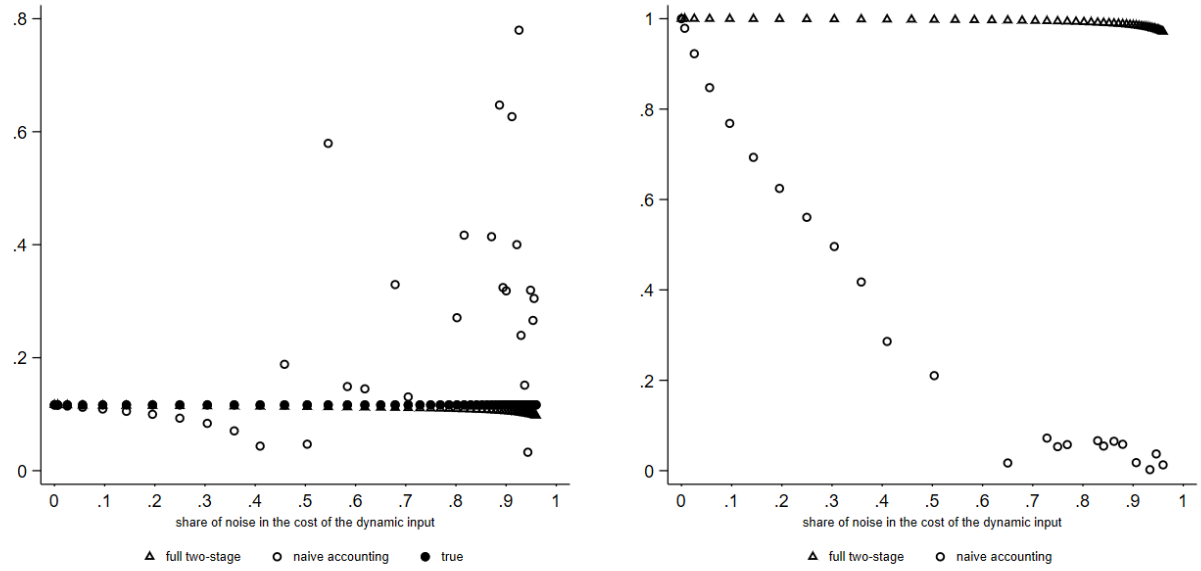


Note: extreme results from the naive accounting approach are omitted.

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 $\hat{\theta}_{it}^{Q,X}$.

Figure 3 shows how the three estimators of the output elasticity of the dynamic input $\theta_{it}^{Q,K}$ perform as the share of noise in $\tilde{\rho}_{it}$ increases. Note that noise in the cost of capital enters both the denominator and the numerator of the cost share of this input, unlike with the flexible one, whose expenditure is perfectly measured.

Figure 3: Sensitivity to measurement error in the cost of the dynamic input: average $\hat{\theta}_{it}^{Q,K}$ (left panel), and corr$(\hat{\theta}_{it}^{Q,K}, \theta_{it}^{Q,K})$ (right panel)

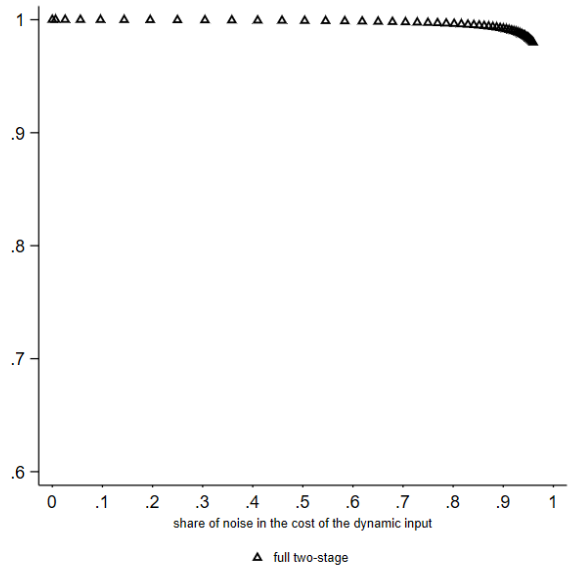


Note: extreme results from the naive accounting approach are omitted.

In short, the conclusions are the same as those of Figure 1: 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_ε^2 grows.

Finally, Figure 4 shows how the two-stage estimator of log TFPR ω_{it}^R performs as the share of noise in $\tilde{\rho}_{it}$ increases, with similar implications as the previous results.

Figure 4: Sensitivity to measurement error in the cost of the dynamic input: $\text{corr}(\hat{\omega}_{it}^R, \omega_{it}^R)$

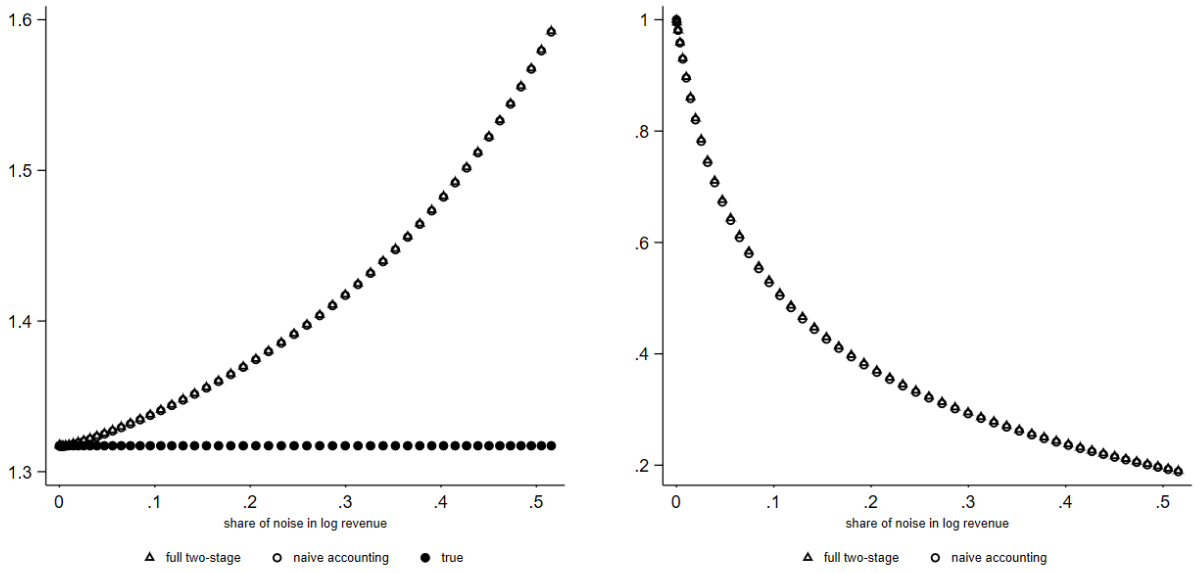


6.4 Sensitivity to Measurement Error in Revenue and Input Levels

I carry out two equivalent exercises to investigate sensitivity to measurement error in revenue (log additive, as in (20)) 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.

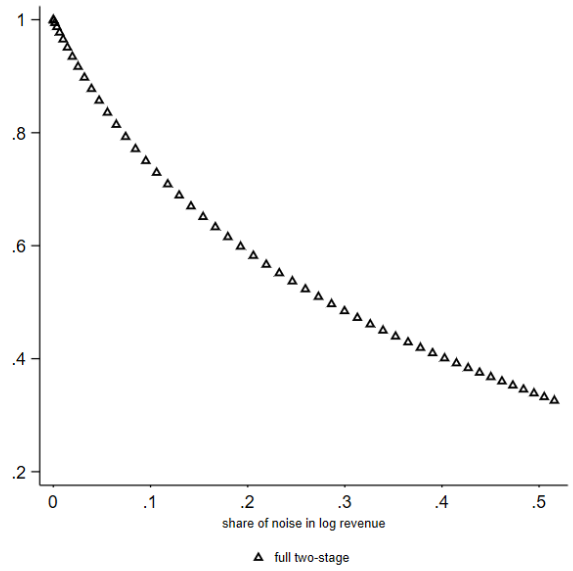
Measurement error in revenue does not affect output elasticities, because input cost shares do not rely on revenue data. Therefore, only markups and TFPR are affected, both in a log-linear fashion (i.e. error is 'added' to log markups and log TFPR estimates). Figure 5 presents the performance of the accounting benchmark and the two-stage approach in terms of markup estimates.

Figure 5: Sensitivity to measurement error in revenue: average $\hat{\mu}_{it}$ (left panel), and corr$(\hat{\mu}_{it}, \mu_{it})$ (right panel)



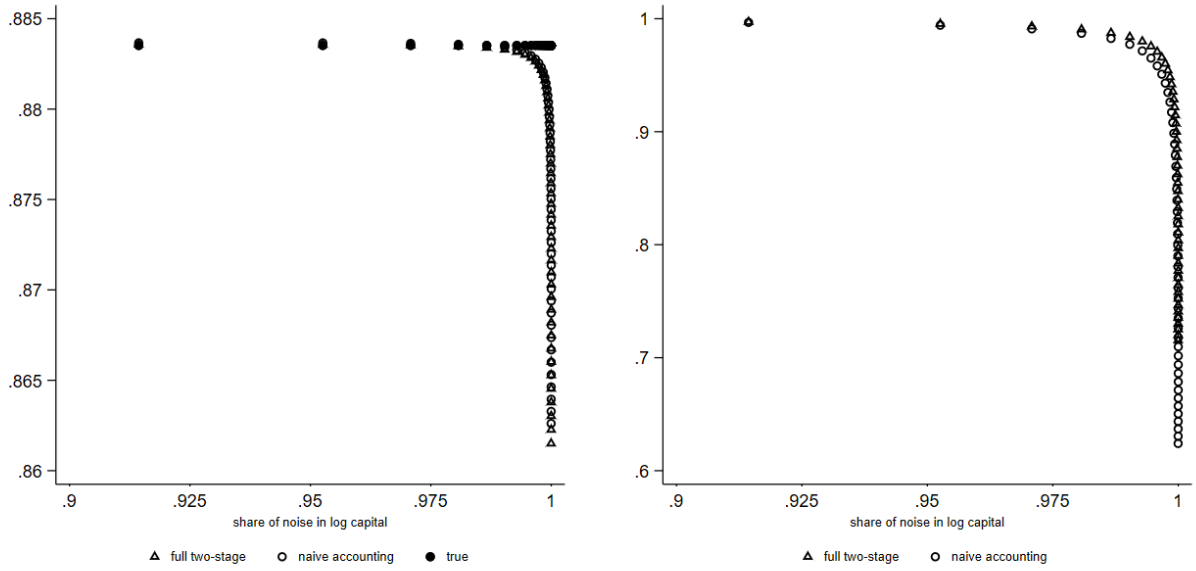
Measurement error in revenue biases the level and distribution of two-stage estimates as expected, i.e. by inflating revenue on average and adding noise across the board. Figure 6 shows the effect of log-additive measurement error in revenue on log TFPR estimates $\hat{\omega}_{it}^{R}$, which is similar to markup estimates.

Figure 6: 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. 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. 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 7 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 7: 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)

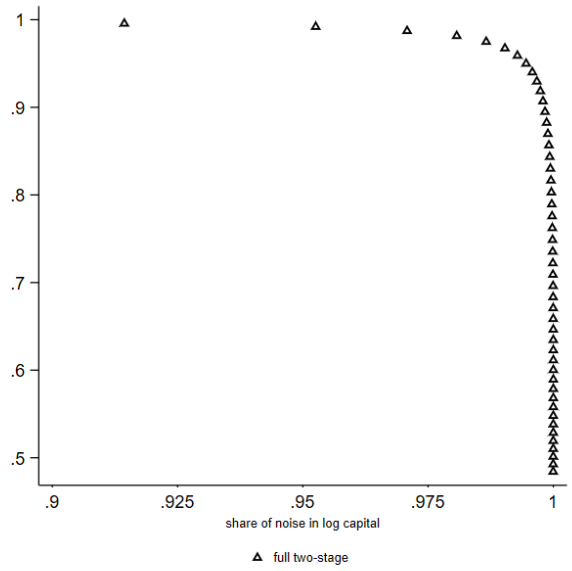


Note: the horizontal axis covers the $0.9 - 1$ interval. For lower levels of noise, estimates and true values are essentially the same.

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.[14] 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 8 shows that the effect of this type of measurement error on the TFPR distribution is also similar.

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

Figure 8: Sensitivity to measurement error in the level of the dynamic input: $\text{corr}(\hat{\omega}_{it}^R, \omega_{it}^R)$



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

7 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 decomposing revenue productivity dispersion. The TFPR distribution has been the focus of many papers since Hsieh and Klenow (2009) pointed out that productivity dispersion is informative about input misallocation. In this paper, I do not take a stance on the underlying model of the macro-economy nor on the appropriate measure of misallocation, but just show that the increase in TFPR dispersion of US listed firms is largely due to increased markup dispersion. To do so, I employ the following decomposition of log TFPR:

$$\omega_{it}^R := p_{it} + \omega_{it}^Q = \log \mu_{it} + (\log \lambda_{it} + \omega_{it}^Q) := \log \mu_{it} + \psi_{it} \tag{36}$$

where p_{it} is the log of the price of output, ω_{it}^{Q} is log TFPQ, and λ_{it} is the marginal cost of output. $\psi_{it} := (\log \lambda_{it} + \omega_{it}^{Q})$ 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, (36) can be thought of as a decomposition of TFPR between output market distortions ($\log \mu_{it}$) and efficiency (ψ_{it}). With data on revenue and input expenditures alone, ω_{it}^{R} and $\log \mu_{it}$ can be estimated with my two-stage approach, and I can thus calculate $\hat{\psi}_{it} = \hat{\omega}_{it}^{R} - \log \hat{\mu}_{it}$. These estimates can be used to perform the following variance decompositions:

$$\begin{aligned} \text{Var}(\hat{\omega}_{it}^{R}) &= \text{Var}(\log \hat{\mu}_{it}) + \text{Var}(\hat{\psi}_{it}) + 2\text{Cov}(\log \hat{\mu}_{it}, \hat{\psi}_{it}) \\ \frac{\Delta \text{Var}(\hat{\omega}_{it}^{R})}{\text{Var}(\hat{\omega}_{it-1}^{R})} &= \alpha_{it-1}^{\mu} \frac{\Delta \text{Var}(\log \hat{\mu}_{it})}{\text{Var}(\log \hat{\mu}_{it-1})} + \alpha_{it-1}^{\psi} \frac{\Delta \text{Var}(\hat{\psi}_{it})}{\text{Var}(\hat{\psi}_{it-1})} + 2\alpha_{it-1}^{\mu,\psi} \frac{\Delta \text{Cov}(\log \hat{\mu}_{it}, \hat{\psi}_{it})}{\text{Cov}(\log \hat{\mu}_{it-1}, \hat{\psi}_{it-1})} \end{aligned} \tag{37}$$

where $\alpha_{it-1}^{\mu} = \frac{\text{Var}(\log \hat{\mu}_{it-1})}{\text{Var}(\hat{\omega}_{it-1}^{R})}$. 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.

7.1 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 1960-2022 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 D and Bond and Gottardo (2024) for further details).

My cleaning routine is the same as De Loecker et al. (2020)): I drop firms which lack information on any variable of interest, or report negative sales and expenditures, and drop the 1^{st} and 99^{th} percentiles of the revenue share of COGS.

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 three-stage approach. Here I present results when SGA is treated as a sunk cost and thus omitted from total cost. In Appendix E I present the same results when SGA is treated as expenditure on a non-flexible production input.

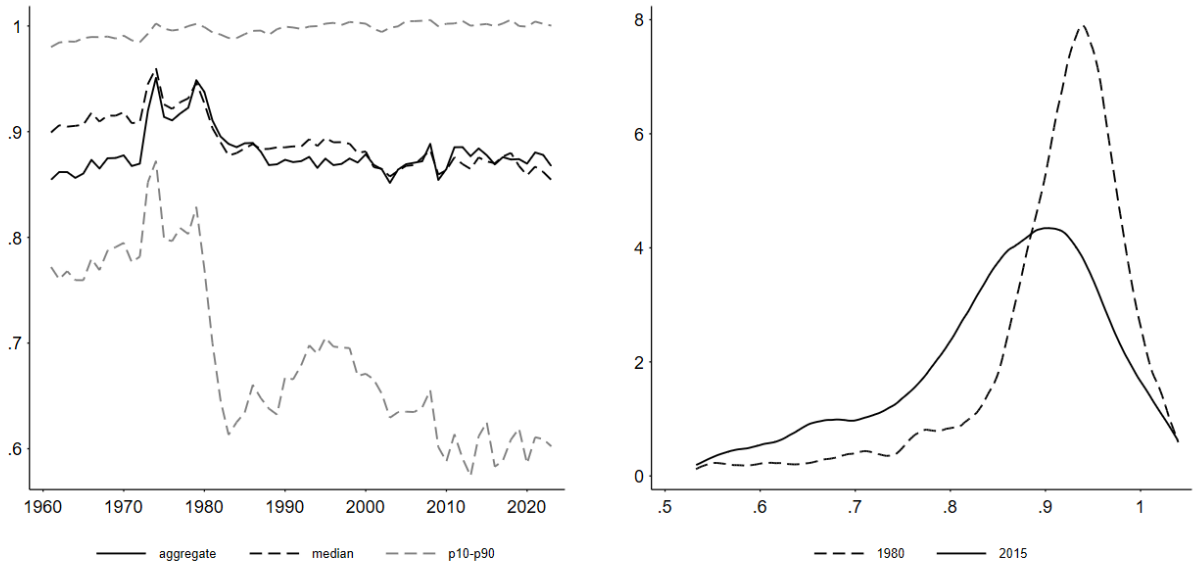
7.2 Estimation and Production Function Estimation Results

Total 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. I assume constant returns to scale throughout the analysis, i.e. $\mathcal{T} = 1$.

Figure 9 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 E total cost also includes SGA.

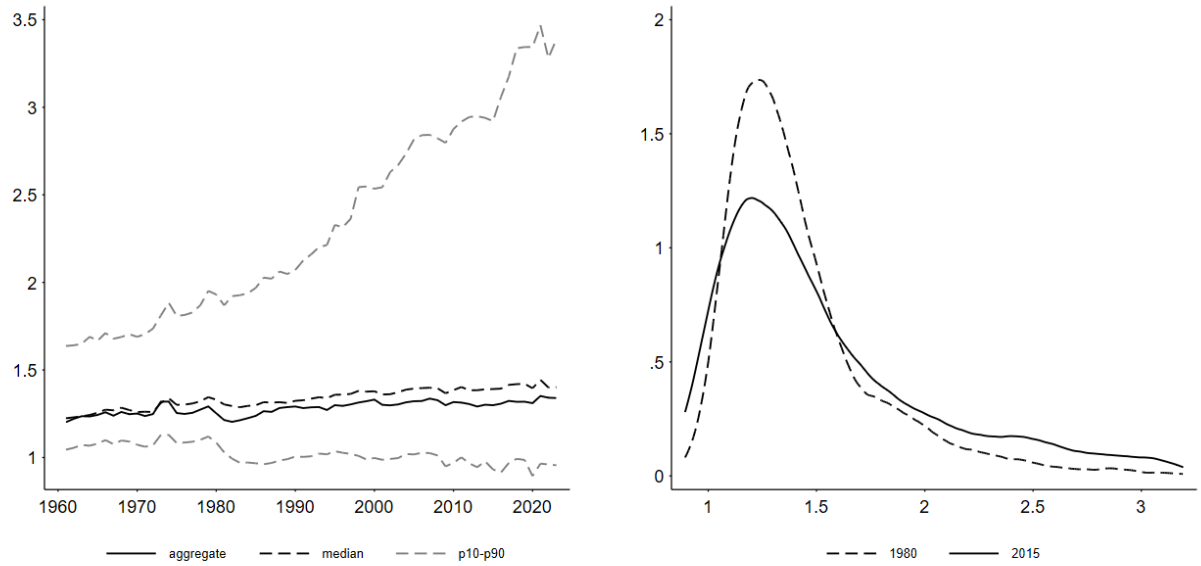
Figure 9: COGS output elasticity estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)



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'.

Figure 10 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).

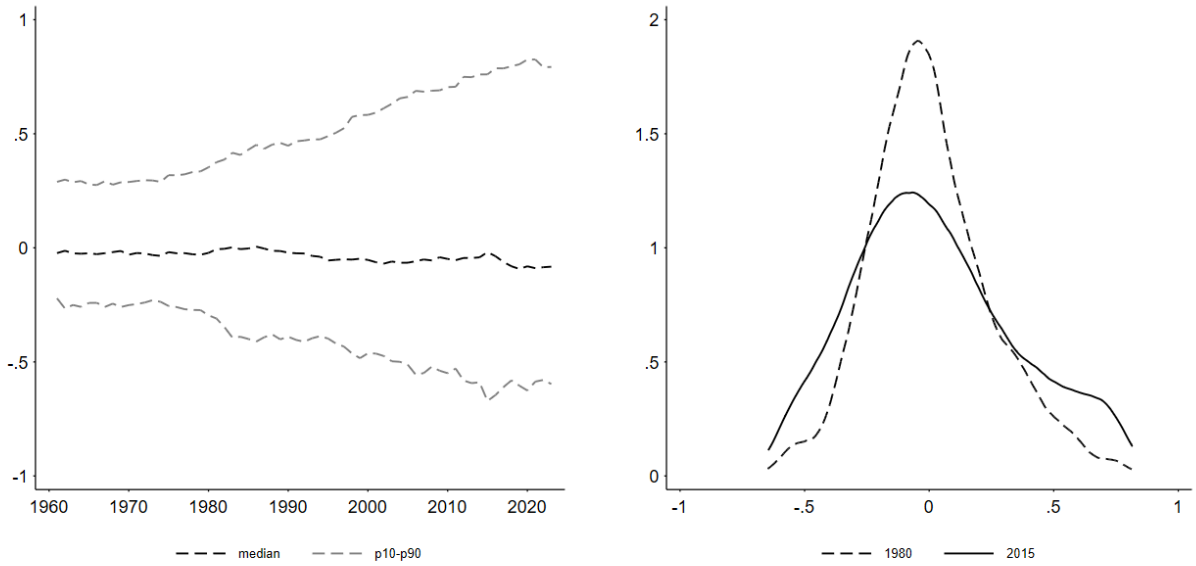
Figure 10: Markup estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)



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 90^{th} percentile grows from around 1.6 in the 1960s to 3.3 around 2020.

Finally, Figure 11 presents the evolution of the TFPR distribution for the whole sample, and compares its pdf in 1980 and 2015.

Figure 11: Log TFPR estimates (SGA omitted): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

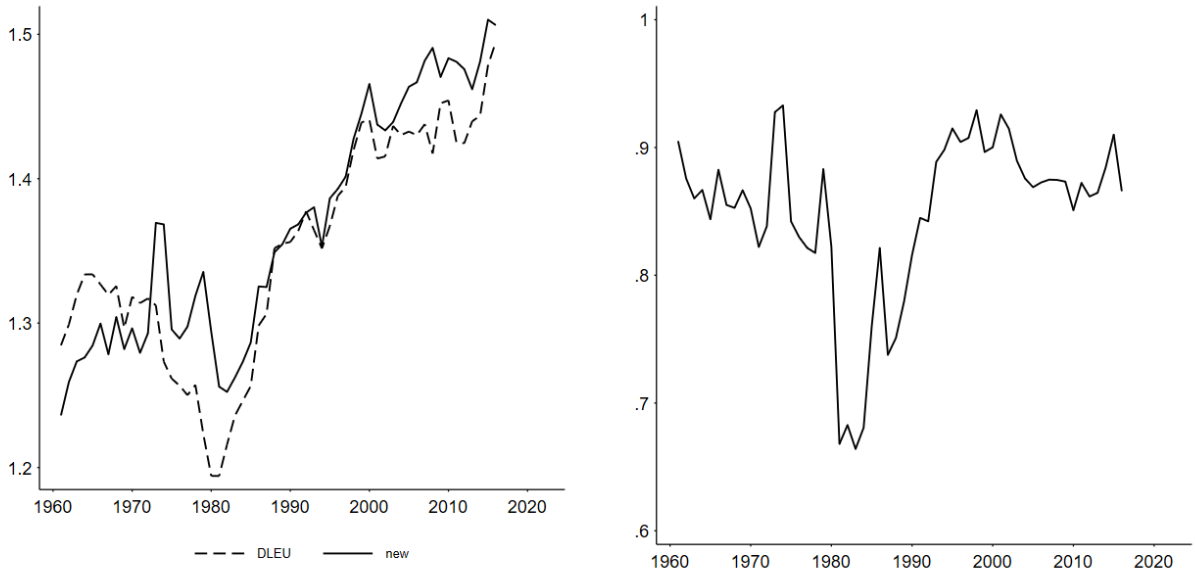


The increase in the log TFPR distribution is driven both by the right and left tails becoming fatter. To

7.3 Comparison with the Literature

Compared to De Loecker et al. (2020) (DLEU), my markup estimates are rather similar on average, but somewhat different in distribution (Figure 12), under the assumption that SGA is a sunk cost.

Figure 12: Comparison with De Loecker et al. (2020) markups: sales-weighted arithmetic average (left panel) and correlation (right panel)



Note: DLEU markups are calculated from the same data, using the output elasticities reported by the authors of the paper.

Note that the only difference from DLEU in how I calculate markups is the COGS output elasticity estimate. In their main specification, they employ a Cobb Douglas output elasticity estimate obtained with a control function method with price controls applied on revenue data (Ackerberg and De Loecker, 2024). 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, i.e. the Cobb Douglas assumption is rejected. This in turn leads to the differences in the markup distribution. On the other hand, the lack of meaningful differences in levels could 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:

$$\tilde{r}_{it} = \frac{\theta^{Q,X}}{\mu_{it}} x_{it} + \frac{\theta^{Q,K}}{\mu_{it}} k_{it} + \tilde{\omega}^R_{it} + \varepsilon^R_{it} \tag{38}$$

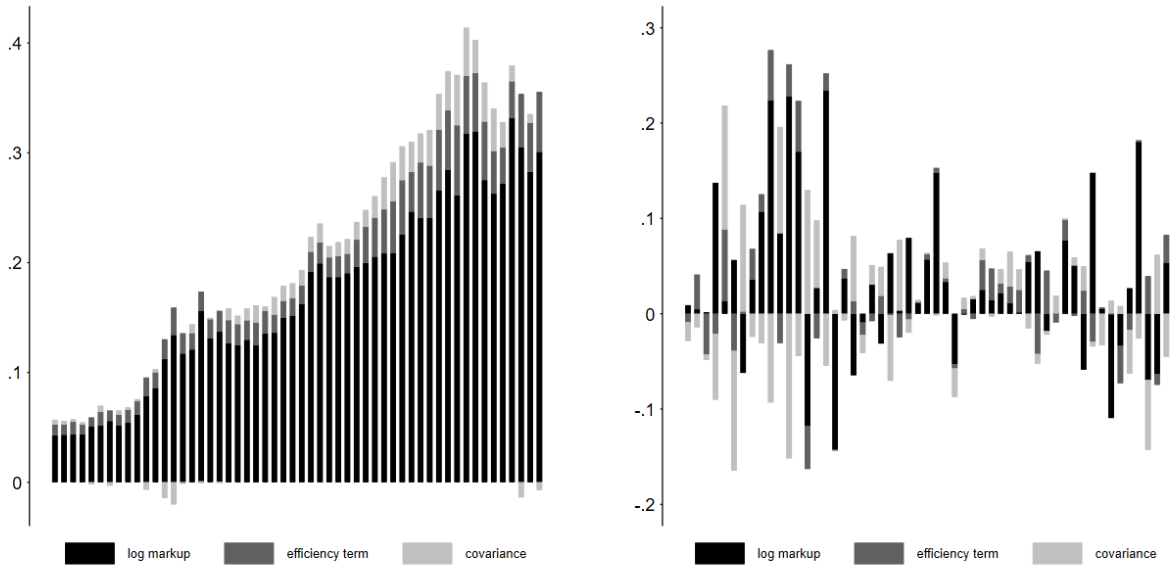
where $\tilde{\omega}^R_{it}$ is a function of TFPQ and prices. Absent any markup heterogeneity and provided that $\tilde{\omega}^R_{it}$ is controlled for successfully, the estimated coefficients suffer from negative bias because $\mu_{it} \geq 1$. At the same time, if the controls for $\tilde{\omega}^R_{it}$ fail, then positive transmission bias is also present. Since

$\tilde{\omega}_{it}^R$ contains both TFPQ and price variation - as argued in Section 3 - valid controls are likely to be much more demanding than in competitive settings in terms of required structure on residual demand and productivity. As extensively argued in De Ridder et al. (2022), the overall bias in this estimate can take any sign, so it may very well be quite close to the true value.

7.4 TFPR Variance Decomposition

Figure 13 presents the decomposition of the variance of log TFPR estimates of expression (37), both in levels and changes from 1970-2023.

Figure 13: Log TFPR variance decomposition 1970-2023 (SGA omitted): level (left panel) and changes (right panel)



Note: each column represent one year.

In short, log markup dispersion represents the biggest driver of log TFPR dispersion both in levels and in changes. The share of revenue productivity variance accounted for by markup variance has fluctuated between 50 and 90% without a trend over time, but its high level ensured that it remained the biggest driver of TFPR dispersion. 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.[16] The relevance of this covariance

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

term is robust to the choice on the treatment of SGA, but its sign flips (see Appendix E). This may hint to the fact - under the assumption that it is a sunk cost - SGA may be an efficiency *and* a market power driver, more or less directly. This should not be too surprising, given that SGA includes both R&D and advertisement expenditures.

8 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 competitive input markets, fixing the returns to scale of the production function, and at least one flexible input (the latter only to estimate markups). It requires no assumption on demand, the law of motion of productivity, or the invertibility of input demand. I show that cost shares can be 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 economic cost of inputs can be purged, and use this finding to develop a two-stage estimator for output elasticities, markups, and revenue productivity which only requires data on revenue and input expenditures.

I 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, compare results to the existing literature, and decompose the dispersion of revenue productivity into a markup term and a 'technical efficiency' one. My findings indicate that Cobb Douglas-based markups can mistake differences in output elasticities for differences in markups to a quantitatively relevant extent, but are also generally correct in terms of average level. My revenue productivity variance decomposition suggests that its main component and the driver of its increase is markup dispersion, and that market power and technical efficiency interact in non-trivial and quantitatively relevant ways.

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_{it}^{Q} as given. X_{it} is expenditure on the flexible input, and output is governed by the same production function $Q_{it} = \Omega_{it}^{Q} F(X_{it}, K_{it})$. Let expenditure on the sunk cost be $P_{it}^{S} S_{it}$. I assume P_{it}^{S} is exogenous and that the sunk cost has no dynamic effects for ease of exposition.[17]

First, I focus on the exogenous sunk cost case, e.g. $P_{it}^{S} S_{it}$ is an entry cost that the firms 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:

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

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 accounting markup, because it is not part of the cost of output. It follows that the accounting markup is unaffected 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 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:

$$\max_{X_{it}, S_{it}} P(Q_{it}, S_{it}, \Upsilon_{it}) \cdot Q_{it} - X_{it} - \rho_{it} K_{it} - P_{it}^{S} S_{it} \tag{40}$$

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:

$$\alpha_{it}^{R,S} = \frac{\partial r_{it}}{\partial s_{it}} \tag{41}$$

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.[18] 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.

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

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

B 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:

$$\hat{\alpha}_{it}^{C,X} \cdot \mathcal{T} = \theta^{Q,X}(x_{it}, k_{it}) + u_{it}^{SS,X} \tag{42}$$

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 $u_{it}^{SS,X}$ may create concerns. Note that this is addressed by the first stage and Proposition 2, which ensure that in the worst case scenario the bias from this kind of correlation should be small (and ideally zero, with a perfect first stage).

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}^{K}) 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.

C Different Form of Measurement Error

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

$$\log \tilde{\rho}_{it} = \log \rho_{it} + \varepsilon_{it}^{K} \tag{43}$$

with $E[\varepsilon_{it}^{K} | x_{it}, k_{it}] = 0$. 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 $\sigma_{\varepsilon}^{2} > 0$. In addition, the first stage is misspecified if it takes the form of (28), which could lead to additional bias.

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}^{K}) 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.

D 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:

$$\rho_{it} = i_t^{10\mathrm{Y}} - \pi_t + 0.07\beta_{it} + 0.05 \tag{44}$$

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 1^{st} or above the 99^{th} percentile of the COGS expenditure-sales ratio. For the exercise where I consider SGA a production input, I clean its sales-share similarly.

Figure 14 reports the sample size (number of companies), and the evolution of total deflated sales for every year.

Figure 14: Sample size and evolution of deflated sales

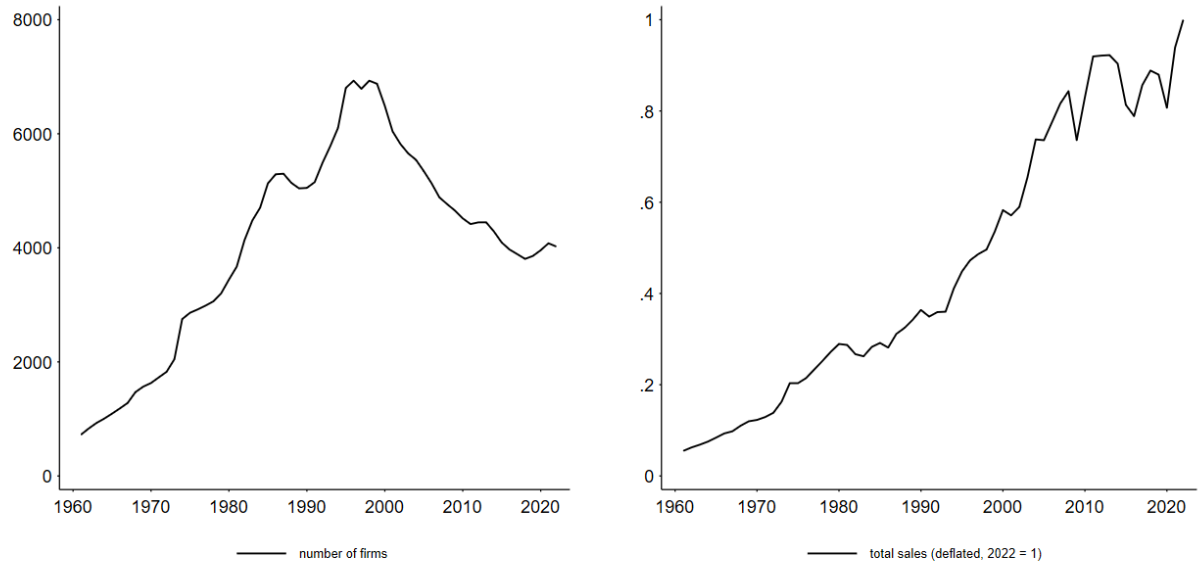
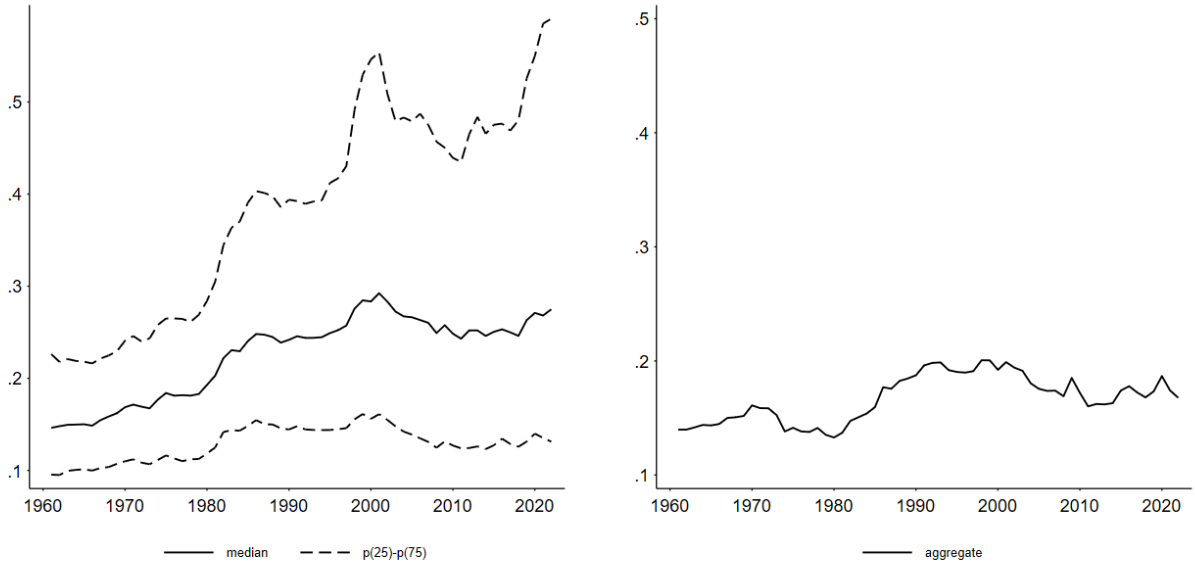


Figure 15 presents the evolution of the share of SGA in sales. The aggregate is obtained by taking the sales-weighted arithmetic mean of firm-level shares, which equals the ratio of total SGA expenditure to total sales.

Figure 15: SGA share of sales



E 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 1^{st} and 99^{th} 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 16, 17, 18, and Figure 19 are the equivalents of Figures 9, 10, 11, and 13 respectively.

Figure 16: COGS output elasticity estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

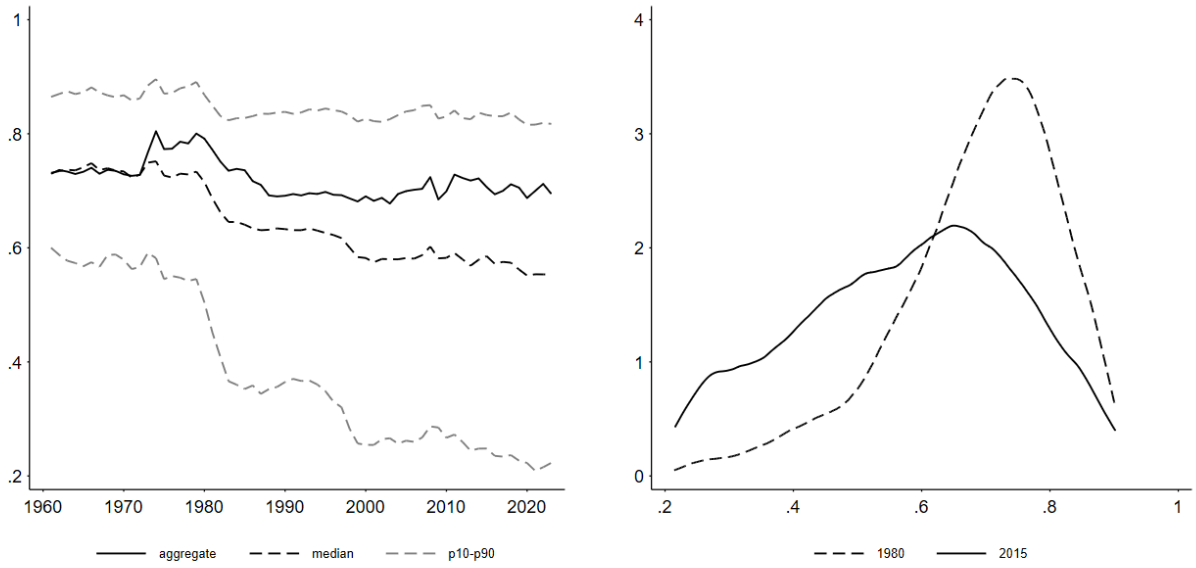


Figure 17: Markup estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

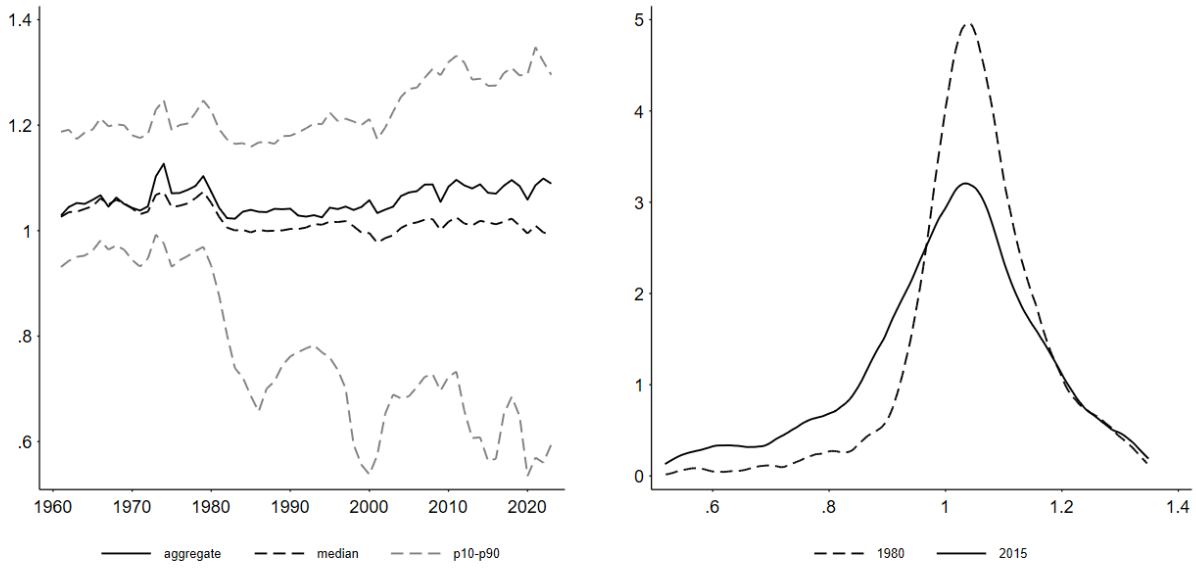


Figure 18: Log TFPR estimates (SGA included): evolution over time (left panel) and pdf in 1980 and 2015 (right panel)

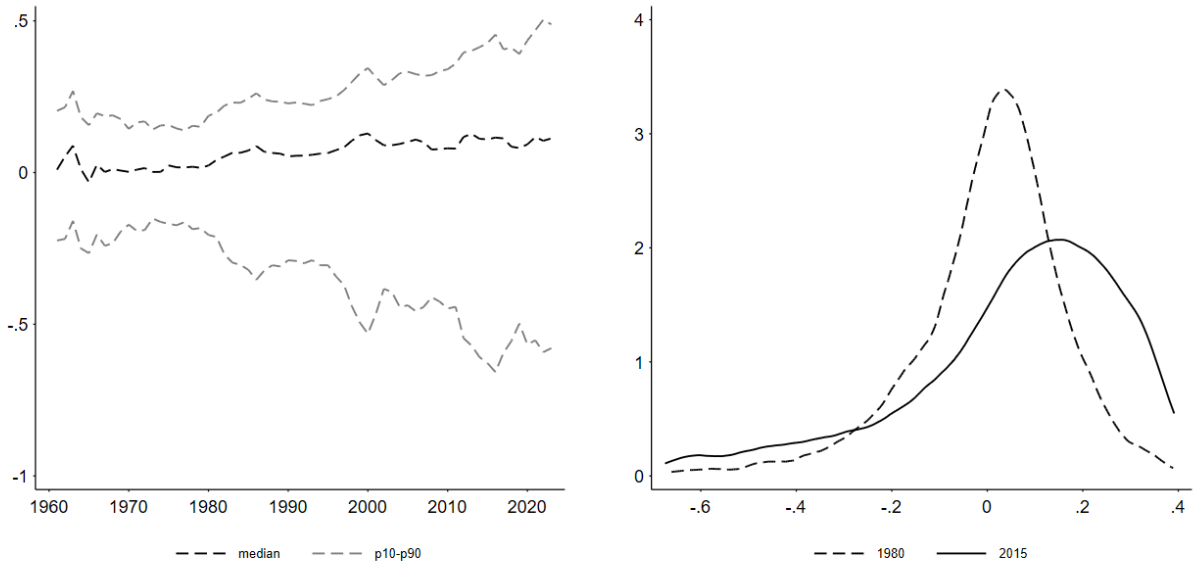
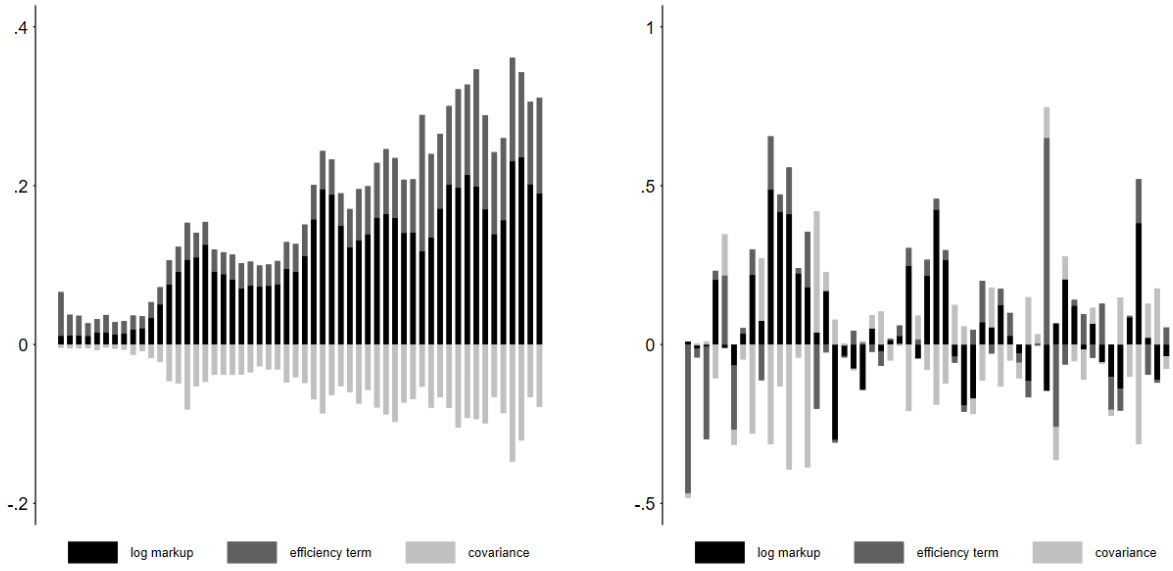


Figure 19: Log TFPR variance decomposition 1970-2023 (SGA included): level (left panel) and changes (right panel)



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 15). E.g. the number of firms with SGA expenditure greater than revenue increases from 2% in 1980 to almost 10% in 2015. 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 13 and 19 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.

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 identiÖcation under imperfect competition', *Working paper* .

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* .

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. (2024), 'Market power and innovation in the intangible economy', *American Economic Review* **114**(1), 199–251.

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.

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).

Hsieh, C.-T. and Klenow, P. J. (2009), 'Misallocation and manufacturing tfp in china and india*', *The Quarterly Journal of Economics* **124**(4), 1403–1448.

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

Karabarbounis, L. and Neiman, B. (2013), 'The global decline of the labor share*', *The Quarterly Journal of Economics* **129**(1), 61–103.

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.

Peters, M. (2020), 'Heterogeneous markups, growth, and endogenous misallocation', *Econometrica* **88**(5), 2037–2073.

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

Raval, D. (2023*b*), 'Testing the production approach to markup estimation', *The Review of Economic Studies* **90**(5), 2592–2611.

Rubens, M. (2021), 'Market structure, oligopsony power, and productivity', *Working Paper* .

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* .