Production Function Estimation with Factor-Augmenting Technology: An Application to Markups

Mert Demirer

July 26, 2022

Abstract

Traditional production functions rely on factor-neutral technology and functional form assumptions. These assumptions impose strong economic restrictions and are often empirically rejected. This paper develops a new method for estimating production functions with labor-augmenting technology and applies it to markup estimation. The method does not impose parametric restrictions and generalizes prior approaches that rely on the CES production function. I first extend the canonical Olley-Pakes framework and then develop an identification strategy based on a novel control variable approach and first-order conditions. I use this method to estimate output elasticities and markups in manufacturing industries in the US and four developing countries. I find that neglecting labor-augmenting productivity underestimates variable input elasticity, and overestimates markups in all countries. These biases also affect markup growth: I estimate a more muted markup growth in the US manufacturing sector than recent estimates. My findings suggest that accommodating labor-augmenting productivity is crucial for markup estimation.

Keywords: Production Functions, Markups, Factor-Augmenting Technology, Control Variables, Productivity, Manufacturing

Assistant Professor, MIT Sloan School of Management, mdemirer@mit.edu. I am indebted to Nikhil Agarwal, Daron Acemoglu, Anna Mikusheva, and Victor Chernozhukov for invaluable guidance and support. This paper benefited from feedback from Allan Collard-Wexler, Glenn Ellison, Sara Fisher Ellison, Paul Joskow, Christopher Knittel, Jing Li, Whitney Newey, Alex Olssen, Nancy Rose, Tobias Salz, Karthik Sastry, John Van Reenen, Michael Whinston, Kamil Yilmaz, and Nathan Zorzi, as well as various seminar participants. I thank the State Institute of Statistics in Turkey for allowing me access to the data under the project "Growth and Technological Change in Turkey." I thank Devesh Raval for helping me to obtain the Colombian and Chilean data.

1 Introduction

Production functions are useful in many areas of economics. They are used to quantify productivity growth, misallocation of inputs, and market power. The typical exercise requires researchers to specify a model of production function and estimate it using firm-level data. However, a misspecified production function may produce biased estimates, which in turn generate incorrect answers to important economic questions. For example, a biased capital elasticity would imply misallocation in an economy with efficient allocation, and a biased flexible input elasticity would give incorrect markup estimates.

Much of the empirical literature relies on Hicks-neutral technology and functional form assumptions for production function estimation. These two elements of standard practice impose strong theoretical restrictions, which, as several papers have shown, are rejected by data. For example, the large firm-level heterogeneity in input ratios is not consistent with Hicks-neutral technology (Raval (2020)). The elasticity of substitution is often estimated to be less than one, contradicting the Cobb-Douglas functional form (Chirinko (2008)). This evidence suggests that firms' production functions do not take the form of commonly used specifications.

In this paper, I develop a method for estimating nonparametric production functions with factor-augmenting productivity and examine its implications empirically. The model has two key features. First, it has labor-augmenting productivity in addition to Hicks-neutral productivity. Second, it does not rely on parametric assumptions for identification; it only imposes a limited functional form structure, which nests the common parametric forms. Together, these features yield a flexible production model, with the ability to better explain the data.

This paper makes both methodological and empirical contributions. Methodologically, I extend the canonical Olley and Pakes (1996) framework to a model with multidimensional productivity, and then study the nonparametric identification of this model by building on the recent literature on factor-augmenting technical change (Doraszelski and Jaumandreu (2018), Raval (2019)). Empirically, I find that neglecting factor-augmenting technology mismeasures output elasticities and markups.

A major challenge in estimating production functions is the endogeneity of inputs. This problem generates additional complications in my model due to the multidimensional unobserved productivity and the absence of parametric restrictions. To address this challenge, I make three novel methodological contributions.

My first result establishes the invertibility of labor-augmenting productivity. In particular, I show how to express labor-augmenting productivity as a function of inputs by inverting input demand functions. This result is key to controlling for labor-augmenting productivity, and it generalizes the parametric inversion to a nonparametric model (Doraszelski and Jaumandreu (2018)). I establish invertibility under the assumption that the production function satisfies homothetic separability in labor and materials, a weak condition that nests the most commonly used production functions. Most importantly, it is an economic rather than a statistical restriction, with clear implications for firm behavior.

My second contribution is to develop a novel control variable approach for production function estimation, building on Imbens and Newey (2009). This result uses the timing assumption for capital and Markov property of productivity shocks, both of which are standard in the literature. The control variable approach differs from the standard "proxy variable approach" in that it does not directly condition on the observed inputs. Instead, it estimates control variables from data in the form of conditional quantiles and conditions on them to address endogeneity. The control variable approach generally cannot be applied to multidimensional unobserved heterogeneity due to the lack of invertibility (Kasy (2011)). I circumvent this problem by showing that the input demand functions form a triangular structure under the modeling assumptions, allowing for invertibility with two-dimensional unobserved heterogeneity.

The third methodological contribution is an identification strategy for output elasticities. After developing the control variable approach, I study which features of the production function can be identified. I first establish a negative result: the output elasticity of flexible inputs (labor and materials), the key input for markup estimation, is not identified due to a functional dependence problem. However, the sum of the flexible input elasticities is identified. To separately identify the flexible input elasticities, I use the first-order conditions of cost minimization, which imply that the ratio of two flexible inputs' elasticities is identified as the ratio of their expenditures. Importantly for markup estimation, this result does not rely on perfect competition in the output market.[1]

Using cost minimization to identify the ratio of output elasticities has an appealing feature: markup estimates from two different flexible inputs are identical. This addresses the well-documented problem that different flexible inputs often yield conflicting markup estimates (Raval (2020)). I show that allowing for labor-augmenting productivity provides a natural solution to this problem. Without labor-augmenting productivity, the model is not rich enough to explain the large firm-level heterogeneity in input ratios, leading to conflicting markup estimates. Labor-augmenting productivity introduces unobserved heterogeneity, makes the model internally valid, and generates identical markup estimates.

I use my method to estimate output elasticities in manufacturing industries in the US and four developing countries: Chile, Colombia, India, and Turkey. I compare my results with estimates from three production functions: (i) Cobb-Douglas, (ii) translog with Hicks-neutral productivity, and (iii) CES with labor-augmenting productivity. The results suggest that the Cobb-Douglas model underestimates the capital elasticity, on average, by 30 percent and overestimates the labor elasticity by 25 percent. Using a labor-augmenting CES or Hicks-neutral translog production function only partially corrects these biases, pointing to the importance of both relaxing parametric restrictions and introducing unobserved heterogeneity.

Estimates of output elasticities are typically used to measure important economic variables. A prime example is markups, which have recently been estimated using production functions (De Loecker et al. (2020)). After documenting biases in output elasticities, I study how these biases propagate into markup estimates.

Hicks-neutral production functions yield biased estimates for markups. First, the Cobb-Douglas model overestimates markups in all countries by 10 to 15 percentage points, an important magnitude when markups are interpreted as a measure of market power. This bias persists, to a smaller extent, when we allow for more flexible Hicks-neutral production functions, such as translog, or parametric labor-augmenting production functions, such as CES. Overall, my results suggest that one has to take into account labor-augmenting productivity and allow for a flexible functional form to reliably estimate markups.

[1]The paper also provides identification conditions and results for other objects of interest, such as elasticity of substitution and labor-augmenting productivity.

Next, I study how labor-augmenting productivity affects the estimates of markup changes. For this, I estimate the evolution of markups in US manufacturing using data from Compustat. Recently, De Loecker et al. (2020) found that the aggregate markup in the US has risen by 35 percentage points since the 1960s using Hicks-neutral production functions. Their finding has drawn significant attention, as it suggests a large increase in market power. Using the same dataset, I instead find that the aggregate markup in US manufacturing has increased by only 15 percentage points, from 1.3 in 1960 to 1.45 in 2012. This difference arises because the Hicks-neutral specification suggests a negligible change in production technology over the last fifty years, whereas I find that flexible input elasticity and its relationship with firm size have changed.

My paper contributes to the literature on production function estimation with proxy variable approach (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), Gandhi et al. (2020)). My approach builds on these papers but differs in three main respects. First, it includes factor-augmenting productivity in addition to Hicks-neutral productivity. Second, I use control variables (Imbens and Newey (2009)) rather than proxy variables to address endogeneity. Third, I use the first-order conditions of cost minimization for identification with respect to two inputs. Unlike the work of Gandhi et al. (2020), firms have market power in the output market, but I require two flexible inputs.

Three recent papers have studied factor-augmenting technology and its implications (Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019)). Common features of these papers are the CES production function and firm-level variation in input prices. They exploit the parameter restrictions between the production and input demand functions and parametrically invert the input demand functions to recover factor-augmenting productivity. I contribute to this literature by relaxing the CES assumption and analyzing identification with or without variation in input prices. With the nonparametric form, I relax some common restrictions the CES functional form imposes on the production technology, such as homogeneous returns to scale and elasticity of substitution across firms.[2]

This paper contributes to the literature on markup estimation from production

[2] Another strand of literature uses a random coefficient model to introduce firm-level unobserved heterogeneity (Kasahara et al. (2015), Li and Sasaki (2017), Balat et al. (2019)).

data (Hall (1988), De Loecker and Warzynski (2012), Doraszelski and Jaumandreu (2019), Raval (2020)). This literature demonstrates how to estimate markups from output elasticities under the cost minimization assumption and studies the properties of this approach. I complement this literature by providing an estimation method for labor-augmenting production functions and by showing that labor-augmenting technology is important for markup estimation. Lastly, a growing literature analyzes change in market using the Hicks-neutral productivity assumption and finds rising markups (Autor et al. (2020), De Loecker et al. (2020)). My results suggest that a labor-augmenting production function estimation points to a more muted rise in markups in the US manufacturing sector.

2 Model

I begin by introducing a production function model. The defining feature of my model is that it allows for both labor-augmenting and Hicks-neutral productivity without parametric restrictions.

2.1 Production Function with Labor-Augmenting Technology

Firm i produces output at time t by transforming three inputs—capital, K_{it}; labor, L_{it}; and materials, M_{it}—according to the following production function:

$$Y_{it} = F_t(K_{it}, \omega_{it}^L L_{it}, M_{it}) \exp(\omega_{it}^H) \exp(\epsilon_{it}), \tag{2.1}$$

where Y_{it} denotes the output produced by the firm. The production function includes two unobserved productivity terms. Labor-augmenting productivity, denoted by $\omega_{it}^L \in \mathbb{R}_+$, increases the effective units of the labor input. Hicks-neutral productivity, denoted by $\omega_{it}^H \in \mathbb{R}$, raises the quantity produced for any input composition. Finally, $\epsilon_{it} \in \mathbb{R}$ is a random shock to output.

The factors of production are either *flexible* or *predetermined*. I assume that labor and materials are *flexible inputs*, meaning that the firm optimizes them each period and their level does not affect future production.[3] In contrast, I assume that capital is a *predetermined input*; that is, the firm chooses the level of capital one period in advance. Each period, the firm chooses the level of flexible

[3]Whether labor is a flexible input depends on its measurement, the number of days vs. employees, specific industry, and country. In my empirical setting of manufacturing, it is more likely to hold for production workers than for white-collar workers.

inputs to minimize the production cost given its information set, denoted by $\mathcal{I}_{it}$, which includes productivity shocks, capital stock, past information sets, and other potential signals observed by the firm. The information set is orthogonal to the random shock, i.e., $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}] = 0$, so ϵ_{it} can be viewed as measurement error in output or an ex-post productivity shock not observed (or predicted) by the firm.

I assume that firms are price-takers in the input market. Input prices do not vary across firms, but they can vary over time. My model and identification strategy extend to the case with heterogeneous and observed input prices, which is provided in Online Appendix B.1. The model does not assume perfect competition in the output market but rules out market power in the input markets.

The form of the production function is *industry-specific* and *time-varying*. That is, all firms in the same industry produce according to the same functional form, which can change over time. The time-varying production function, unobserved firm-specific productivity terms, and the nonparametric form yield a flexible production model. Despite its flexibility, the production function has an important restriction: factor-augmenting productivity affects only labor, implying that capital and materials productivity are homogeneous across firms. More generally, my model can accommodate only one factor-augmenting productivity for a flexible input. This limitation comes from the fact that a non-flexible input has dynamic implications, requiring a different toolkit to model its productivity.

I choose to consider labor-augmenting technology for two reasons. First, heterogeneity in ω_{it}^L reflects firm-level differences in labor efficiency, which can be generated by different mechanisms, such as firms' management practices and human capital. Because these measures are typically not observed in the data, it is natural to model them as unobserved heterogeneity. Second, in most production datasets, labor's cost share has the most across-firm variation among all inputs, suggesting unobserved heterogeneity in labor productivity.

My model has two key features: (i) it contains factor-augmenting technology, and (ii) it does not impose a parametric structure. These features have important implications that are not captured by many other production functions. As an illustration, consider the Cobb-Douglas production function, $Y_{it} = K_{it}^{\beta_k} L_{it}^{\beta_l} M_{it}^{\beta_m} \exp(\omega_{it}^H) \exp(\epsilon_{it})$, which is nested in Equation (2.1). Cobb-Douglas has two key restrictions: (i) the production function is log-linear, and (ii) ω_{it}^H is the only source of unobserved heterogeneity in production technology. These are strong restrictions with strong implications. First, they imply that revenue shares of flexible inputs are the same across firms, a prediction rejected by the data (Raval (2019)). Second, the literature has documented large heterogeneity in factor intensities, which contradicts constant elasticity.[4] Finally, empirical evidence suggests that the elasticity of substitution is less than one (Chirinko (2008)).

A more flexible Hicks-neutral production function, such as translog, is still restrictive, as it does not have factor-specific productivity. The literature has documented a large heterogeneity (across firms) and a significant decline (over time) in labor share in advanced economies. Most importantly, these facts have been attributed to within-industry changes and reallocation across firms (Autor et al. (2020), Kehrig and Vincent (2021)), and heterogeneity in production technology has been proposed as a mechanism (Oberfield and Raval (2021)).

2.2 Assumptions

This section presents assumptions and discusses their implications. The first assumption imposes a homothetic separability restriction on the production function. This assumption allows me to invert the firm's input choices to express ω_{it}^L as an unknown function of inputs. Other assumptions concern firm behavior and productivity shocks, generalizing the standard Olley and Pakes (1996) production framework to a model with two productivity shocks. Throughout the paper, I assume that all functions are continuously differentiable as needed, and all random variables have a continuous and strictly increasing distribution function.

2.2.1 A Homothetic Separability Restriction

I first provide a set of conditions under which labor-augmenting productivity can be expressed as a function of the firm's inputs.

Assumption 2.1 (Homothetic Separability)**.** *Suppose that:*

(i) *The production function is of the form*

$$Y_{it} = F_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big)\exp(\omega_{it}^H)\exp(\epsilon_{it}). \tag{2.2}$$

(ii) $h_t(K_{it}, \cdot, \cdot)$ *is homogeneous of arbitrary degree for all* K_{it}.

[4] Large firms are more capital-intensive than small firms (Holmes and Schmitz (2010)), and exporting firms are more capital-intensive than domestic firms (Bernard et al. (2007)).

(iii) *The firm minimizes production cost with respect to* (L_{it}, M_{it}) *given* K_{it}*, productivity shocks* $(\omega^L_{it}, \omega^H_{it})$*, and input prices* (p^l_t, p^m_t)*.*
(iv) *Let* $\sigma_t(K, \omega^L L, M)$ *denote the elasticity of substitution between effective labor* $(\omega^L L)$ *and materials. We have* $\sigma_t(K, \omega^L L, M) < 1$ *or* $\sigma_t(K, \omega^L L, M) > 1$*.*

Assumption 2.1(i-ii) is called homothetic separability (Shephard (1953)), and it is the key assumption of the paper. It states that the production function is separable in K_{it} and a composite input h_t that is homogeneous of arbitrary degree in labor and materials. Homothetic separability is common in models of consumer preferences and production functions, and most parametric production functions satisfy this property. It has two key economic implications. First, the production is broken into stages, where h_t can be seen as an 'intermediate input' with its own production function, which is then combined with capital for production. Second, homotheticity of h_t implies that increasing the scale of labor and materials is equivalent to increasing the scale of h_t. This means the firm decides the optimal scale of h_t rather than the optimal scale of labor and materials separately. In other words, the optimal ratio of materials and labor is a sufficient statistic, allowing for dimension reduction in the firm's optimization. A production function that violates this assumption is the CES with nested capital and labor.

Assumption 2.1(iii) specifies that firms choose the level of flexible inputs to minimize production cost. The production cost does not involve capital, since it is a predetermined input. Moreover, cost minimization is a static problem, so it is agnostic about the firm's dynamic decisions. Assumption 2.1(iv) implies the effective labor and materials are either substitutes or complements. In a nonparametric production function, whether two inputs are substitutes or complements can change with the level of inputs. Assumption 2.1(iv) precludes this possibility.

Next, I provide two parametric forms that satisfy Assumption 2.1.

Example 1 (CES). The constant elasticity of substitution production function is:

$$Y_{it} = \left(\beta_k K^\sigma_{it} + \beta_l [\omega^L_{it} L_{it}]^\sigma + (1 - \beta_l - \beta_m) M^\sigma_{it}\right)^{v/\sigma} \exp(\omega^H_{it}) \exp(\epsilon_{it}).$$

My model nests the CES production function with $h_t(\cdot) = \beta_l \left[\omega^L_{it} L_{it}\right]^\sigma + (1 - \beta_l - \beta_m) M^\sigma_{it}$. This function is homogeneous of degree one with elasticity of substitution σ. The CES specification has been widely used in the literature to study factor-augmenting technology (Doraszelski and Jaumandreu (2018), Raval (2019)).

Example 2 (Nested CES). A more flexible parametric form is the nested CES, where the elasticity of substitution between effective labor and materials is σ_1:[5]

$$Y_{it} = \left(\beta_k K_{it}^{\sigma} + (1-\beta_k)\left(\beta_l \left[\omega_{it}^{L} L_{it}\right]^{\sigma_1} + (1-\beta_l) M_{it}^{\sigma_1}\right)^{\sigma/\sigma_1}\right)^{v/\sigma} \exp(\omega_{it}^{H}) \exp(\epsilon_{it}).$$

My production function differs from these examples in two important ways. First, in both examples, the elasticity of substitution between inputs is constant, which has strong theoretical implications (Nadiri (1982)). In contrast, I impose a mild restriction on the elasticity of substitution given by Assumption 2.1(iv), so it can vary freely subject to this restriction. Second, neither example allows for heterogeneity in returns to scale across firms, which equals v. In my model, returns to scale is heterogeneous across firms.

Proposition 2.1.

(i) *Under Assumptions 2.1(i-iii), the flexible input ratio, denoted by $\tilde{M}_{it} = M_{it}/L_{it}$, depends only on K_{it} and ω_{it}^{L} through an unknown function $r_t(K_{it}, \omega_{it}^{L})$:*

$$\tilde{M}_{it} \equiv r_t(K_{it}, \omega_{it}^{L}). \tag{2.3}$$

(ii) *Under Assumption 2.1(iv), $r_t(K_{it}, \omega_{it}^{L})$ is strictly monotone in ω_{it}^{L}.*

See Appendix B for the proof. The first part of the proposition states that the flexible input ratio is a function of ω_{it}^{L}, but not ω_{it}^{H}. The intuition is that the firm's labor and materials allocation depends the relative marginal products of these inputs, which in turn depends on the input ratio by homotheticity.

The second part of Proposition 2.1 establishes that $r_t(K_{it}, \omega_{it}^{L})$ is strictly monotone in ω_{it}^{L}. For strict monotonicity, the flexible input ratio should always move in the same direction as ω_{it}^{L}, which increases the ratio of marginal products of labor and materials. Because the relationship between the input ratio and the ratio of marginal products depends on the substitutability of inputs, Assumption 2.1(iv) restricts the elasticity of substitution. Together, these two results provide a function, $r_t(K_{it}, \omega_{it}^{L})$, that is strictly monotone in a scalar unobserved variable.[6]

To relate this result to parametric production functions, note that under the

[5]Note that in this example, as $\sigma_1 \to \infty$ the production function approaches Leontief in materials and labor, so my model can approximate the Leontief production function.

[6]Another common assumption is Leontief, where inputs are perfect complements. In this case, r_t becomes a linear function of ω_{it}^{L} conditional on K_{it}.

CES assumption, $\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^L)$ has a known functional form, which is log-linear in ω_{it}^L: $\log(\tilde{M}_{it}) = \sigma p_{it}^{l/m} + \log(\omega_{it}^L)$, where $p_{it}^{l/m}$ is the ratio of input prices. A common approach in the literature is to estimate this equation by instrumenting for input prices (Doraszelski and Jaumandreu (2018)).[7] However, this relies on the linear separability functional form obtained from CES, which might not hold in more flexible production functions. Therefore, one contribution of this paper is generalizing the CES production function to an arbitrary functional form (subject to Assumption 2.2) and showing invertibility under more general conditions.[8,9]

2.2.2 Other Assumptions

The rest of the assumptions generalize the standard production function assumptions to accommodate labor-augmenting technology.

Assumption 2.2 (First-Order Markov). *Productivity shocks follow an exogenous first-order Markov process:* $P(\omega_{it}^L, \omega_{it}^H \mid \mathcal{I}_{it-1}) = P(\omega_{it}^L, \omega_{it}^H \mid \omega_{it-1}^L, \omega_{it-1}^H)$.

This assumption is a natural generalization of the standard first-order Markov assumption from Olley and Pakes (1996) to accommodate two-dimensional productivity.[10] It does not restrict the joint distribution of productivity shocks, which can be arbitrarily correlated. Moreover, it allows for first-order dynamics in productivity shocks: higher ω_{it}^H today can be associated with higher ω_{it+1}^L tomorrow.

Assumption 2.3 (Timing). *Capital evolves according to* $K_{it} = \kappa(K_{it-1}, I_{it-1})$, *where* I_{it-1} *denotes investment made by firm* i *during period* $t-1$.

According to this assumption, firms choose capital one period in advance, implying that K_{it} belongs to the firm's information set at period $t-1$, that is, $K_{it} \in \mathcal{I}_{it-1}$.

Assumption 2.4 (Monotonicity). *Firms' materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^L, \omega_{it}^H), \tag{2.4}$$

where $s_t(K_{it}, \omega_{it}^L, \omega_{it}^H)$ *is an unknown function that is strictly increasing in* ω_{it}^H.

[7] Estimating this equation requires heterogeneous input prices at the firm level.

[8] CES may not be restrictive for some empirical problems, as it is a first-order approximation to any production functions with separability (Doraszelski and Jaumandreu (2018)).

[9] However, one drawback of this approach is that, contrary to the parametric setting, r_t depends on the derivatives of h_t in a complicated way (see Equation (B.4)). Therefore, it is more difficult to use the additional restrictions on r_t that could be obtained from a parametric specification, as in Doraszelski and Jaumandreu (2018).

[10] The model can accommodate a controlled Markov process, where observed variables, such as R&D and export, can affect productivity distribution (Doraszelski and Jaumandreu (2013)).

Introduced by Levinsohn and Petrin (2003), this assumption states that, holding all else constant, more productive firms have higher input demand. In my model, the materials demand function depends also on ω_{it}^L as it affects the marginal product of materials.[11] Since the firm's materials demand depends on its profit-maximizing level of output, verifying this assumption requires the primitives of the output market, such as the demand shifters and competition structure, which I do not model in this paper.[12] Implicit in this assumption is that there is no unobserved firm-specific heterogeneity in the firm's perceived residual demand curve in the output market; otherwise, the materials input demand function should include firm-specific unobserved demand shocks, violating two-dimensional unobserved heterogeneity required for identification. However, if there are observed demand shifters or proxies for demand shocks, one can include them, as in De Loecker (2011), to introduce firm-specific heterogeneity in demand.

Even though this assumption restricts the competition in the output market, it accommodates some commonly used demand models, such as monopolistic competition without unobserved demand shifters and Cournot competitions. Moreover, it allows for ex-post demand shocks after the firm chooses the planned output. Finally, this assumption does not imply constant markups. Ex-post demand shocks and some monopolistic competition models, such as variable elasticity of substitution, allow for heterogeneous markups. See also Jaumandreu (2018), Doraszelski and Jaumandreu (2019), and Bond et al. (2021) for more discussion on the importance of demand shocks in production function estimation.

2.3 Invertibility: Expressing Unobserved Productivity Using Inputs

Proposition 2.1 provides the necessary conditions, monotonicity and scalar unobserved heterogeneity, to invert out ω_{it}^L using the flexible input ratio:

$$\omega_{it}^L = r_t^{-1}(K_{it}, \tilde{M}_{it}) \equiv \bar{r}_t(K_{it}, \tilde{M}_{it}). \tag{2.5}$$

Similarly, Assumption 2.4 provides a monotonicity condition for ω_{it}^H using materials demand function in Equation (2.4). Inverting Equation (2.4) yields

[11] As discussed in Gandhi et al. (2020), this assumption imposes an implicit restriction on the distribution of ϵ_{it}, i.e., $\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}] = \mathbb{E}[\exp(\epsilon_{it}) \mid K_{it}, \omega_{it}^L, \omega_{it}^H]$.

[12] For derivation, see Equation (B.1), which includes the planned output. The implicit assumption is that no unobserved variable affects the firm's planned output.

$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \omega_{it}^L)$. Substituting for it from Equation (2.5):

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it})) \equiv \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}). \tag{2.6}$$

Equations (2.5) and (2.6) demonstrate that the modeling assumptions and optimal firm behavior allow me to write unobserved productivity shocks as unknown functions of inputs. Invertibility is a standard condition in the proxy variable approach, which uses observables, such as investments or materials, to control for unobserved productivity. In the next section, I use these invertibility results to develop a control variable approach to address endogeneity.

3 A Control Variable Approach to Production Functions

The control variable approach relies on constructing variables from data to control for endogenous variation. In this section, I show how to construct a control variable for each productivity shock using the Markov and timing assumptions.

My approach builds on Imbens and Newey (2009), who study the identification of non-separable models where a scalar unobservable has a strictly monotone relationship with the outcome and it is independent of an instrument. I make two innovations to apply the control variable approach to production function estimation. First, I show that the Markov and timing assumptions provide the necessary independence condition. Second, my model involves two-dimensional unobserved heterogeneity, for which the control variable approach generally does not apply due to the lack of invertibility (Kasy (2011)). To overcome this, I use the triangular structure of input demand functions in Equations (2.3) and (2.4).

3.1 Control Variable for Factor-Augmenting Technology

If productivity shocks are continuously distributed, we can relate labor-augmenting productivity to past productivity shocks in the following way:

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1), \qquad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0, 1). \tag{3.1}$$

This representation of ω_{it}^L is without loss of generality and follows from the Skorohod representation of random variables. Here, $g_1(\omega_{it-1}^L, \omega_{it-1}^H, \tau)$ corresponds to the τ-th conditional quantile of ω_{it}^L given $(\omega_{it-1}^L, \omega_{it-1}^H)$. So, we can view u_{it}^1 as the productivity rank of firm i relative to firms with the same past productivity levels.

Another interpretation of u_{it}^1 is unanticipated *innovation* to ω_{it}^L, which determines the current period's productivity given last period's productivity. Unlike the standard definition of 'innovation' to productivity, which is separable from and mean-independent of past productivity,[13] u_{it}^1 is non-separable and independent. These properties of u_{it}^1 are key for using the modeling assumptions to derive the control variables. The previous section showed that $\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^L)$. Substituting for ω_{it}^L from Equation (3.1) and using Equations (2.5-2.6), I obtain

$$\begin{aligned}
\tilde{M}_{it} &= r_t\big(K_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)\big), \\
&= r_t\big(K_{it}, g_1\big(\bar{r}_t(K_{it-1}, \tilde{M}_{it-1}), \bar{s}_t(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1\big)\big), \\
&\equiv \tilde{r}_t\left(K_{it}, W_{it-1}, u_{it}^1\right),
\end{aligned} \tag{3.2}$$

for some unknown function $\tilde{r}_t(\cdot)$ and $W_{it} := (K_{it}, M_{it}, L_{it})$. Note that $\tilde{M}_{it}$ is strictly monotone in u_{it}^1 because $r_t(\cdot)$ is strictly monotone in ω_{it}^L by Assumption 2.1, and $g_1(\cdot)$ is strictly monotone in u_{it}^1 by construction. Next, I establish the statistical independence of u_{it}^1 from other variables in Equation (3.2).

Lemma 3.1. *Under Assumptions 2.2 - 2.3, we have that* $u_{it}^1 \perp\!\!\!\perp (K_{it}, W_{it-1})$.

The proof is provided in Appendix B. The lemma shows that the modeling assumptions, Markov and timing, give the necessary independence condition to apply control variable approach. The intuition behind this result is as follows. Condition on $(\omega_{it-1}^L, \omega_{it-1}^H)$ throughout. By the timing assumption, $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$. Together with the Markov assumption, this implies that (K_{it}, W_{it-1}) is not informative about current productivity. Recall that u_{it}^1 contains all information related to current productivity. Since (K_{it}, W_{it-1}) does not contain information about current productivity, it is independent of u_{it}^1.

We now have the two conditions needed to derive a control variable: (i) $\tilde{r}_t\left(K_{it}, W_{it-1}, u_{it}^1\right)$ is strictly monotone in u_{it}^1 and (ii) u_{it}^1 is independent of (K_{it}, W_{it-1}). Since the distribution of u_{it}^1 is already normalized to a uniform distribution in Equation (3.1), we can identify u_{it}^1 from data as:

$$u_{it}^1 = F_{\tilde{M}_{it}|K_{it}, W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}), \tag{3.3}$$

where $F_{\tilde{M}_{it}|K_{it}, W_{it-1}}$ denotes the CDF of $\tilde{M}_{it}$ conditional on (K_{it}, W_{it-1}).[14] The

[13] In particular, the innovation ξ is defined as $\omega_{it}^L = g(\omega_{it-1}^L, \omega_{it-1}^H) + \xi_{it}$ with $\mathbb{E}[\xi_{it} \mid \mathcal{I}_{it}] = 0$.

[14] To simplify the exposition, I assume $\tilde{M}_{it}$ is strictly increasing in u_{it}^1. This is without loss of generality because I need to recover u_{it}^1 up to a monotone transformation.

main idea is that two firms, i and j, with the same capital and last period's inputs, but different materials-to-labor ratios, differ only in their innovations to labor-augmenting productivity. That is, conditional on $K_{it} = K_{jt}$ and $W_{it-1} = W_{jt-1}$, $\tilde{M}_{it} > \tilde{M}_{jt}$ if and only if $u_{it}^1 > u_{jt}^1$. As a result, I can recover u_{it}^1 from the firm's rank in the flexible input ratio. Using this result, I can express ω_{it}^L as a function of the control variable and past inputs:

$$\begin{aligned} \omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) &= g_1(\bar{r}_t(K_{it-1}, \tilde{M}_{it-1}), \bar{s}_t(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1), \\ &\equiv c_{1t}\left(W_{it-1}, u_{it}^1\right), \end{aligned} \tag{3.4}$$

where $c_{1t}(\cdot)$ is an unknown function.

3.2 Control Variable for Hicks-Neutral Technology

The derivation for the control variable for ω_{it}^H proceeds similarly:[15]

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \text{Uniform}(0,1). \tag{3.5}$$

Following the same steps as in Equation (3.2), I use the monotonicity of materials in ω_{it}^H given by Assumption 2.4 to write

$$M_{it} \equiv \tilde{s}_t\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right), \tag{3.6}$$

where $\tilde{s}_t(\cdot)$ is an unknown function. Note that $\tilde{s}_t\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right)$ is strictly increasing in u_{it}^2 because $s_t(K_{it}, \omega_{it}^L, \omega_{it}^H)$ is strictly increasing in ω_{it}^H by Assumption 2.4, and $g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2)$ is strictly increasing in u_{it}^2 by construction.

Lemma 3.2. *Under Assumptions 2.2 - 2.3, we have that* $u_{it}^2 \perp\!\!\!\perp (K_{it}, W_{it-1}, u_{it}^1)$.

See Appendix B for the proof. Now, we can use Equation (3.6) to identify u_{it}^2 as:

$$u_{it}^2 = F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1). \tag{3.7}$$

With this result, ω_{it}^H can be written as:

$$\omega_{it}^H \equiv c_{2t}\left(W_{it-1}, u_{it}^1, u_{it}^2\right) \tag{3.8}$$

for an unknown function $c_{2t}(\cdot)$ whose derivation is the same as Equation (3.4). This result and Equation (3.4) imply that conditional on last period's inputs and two control variables, there is no variation in productivity shocks. Hence, we can use these control variables to address endogeneity in production function estimation.

[15] u_{it}^1 is included in g_2 to account for the correlation between ω_{it}^L and ω_{it}^H. If one assumes productivity shocks are independent conditional on past productivity, u_{it}^1 need not be included.

Remark 3.1 (Comparison to the Proxy Variable Approach)**.** My approach differs from the proxy variable approach in that it relies on a different representation of productivity shocks. The proxy variable approach uses: $\omega_{it}^L = \bar{r}_t(K_{it}, \tilde{M}_{it})$, $\omega_{it}^H = \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it})$, whereas the control variable approach relies on: $\omega_{it}^L = c_{1t}\left(W_{it-1}, u_{it}^1\right)$, $\omega_{it}^H = c_{2t}\left(W_{it-1}, u_{it}^1, u_{it}^2\right)$ to control for endogeneity. The key difference is that control variables exploit the independence property given by the Markov assumption, whereas the proxy variable approach only uses its mean independence implication. Using the independence assumption fully might result in efficiency gains. However, if the mean independence holds but independence does not, then control variables would give inconsistent estimates, whereas the proxy variable estimator would remain consistent.

Remark 3.2 (Comparison to Ackerberg et al. (2022))**.** Another paper that allows for non-separable productivity using the control variable approach is Ackerberg et al. (2022). Their model allows for a single productivity shock in an unrestrictive way, whereas I allow for two productivity shocks under specific functional form restrictions. Another important difference is how the control function is constructed. Ackerberg et al. (2022) use the lagged inputs, whereas I use the input demand functions. These approaches complement each other, as the plausibility of these different assumptions depends on the empirical setting.

Remark 3.3 (Comparison to Gandhi et al. (2020))**.** Gandhi et al. (2020) consider the identification of the nonparametric Hicks-neutral production function by inverting materials conditional on labor, which allows for labor as a dynamic input and other unobserved differences in labor choice. In contrast, my approach rules out these by assuming that labor is a flexible input, but it allows for two unobserved productivity shocks. With the flexible labor assumption, I can use the first-order conditions with respect to two inputs for identification.

4 Identification

This section presents the results for the identification of the output elasticities. First, I point out an identification problem by showing that the production function and output elasticities are not identified from variations in inputs and output. Then, I propose a solution to this problem by exploiting the first-order conditions of cost minimization to identify output elasticities.

4.1 A Non-identification Result

Taking the logarithm of output and denoting $f_t = \log(F_t)$, I write the logarithm of the production function in an additively separable form in ω_{it}^H and ϵ_{it} as:

$$y_{it} = f_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}.$$

Since $h_t(\cdot)$ is homogeneous of arbitrary degree in its second and third arguments by Assumption 2.1, I assume, without loss of generality, that it is homogeneous of degree one. Using this property, I rewrite the production function as follows:

$$y_{it} = f_t\big(K_{it}, L_{it} h_t(K_{it}, \omega_{it}^L, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{4.1}$$

This reformulation is convenient because ω_{it}^L is now an argument in $h_t(\cdot)$. In Section 2.3, I showed that $\omega_{it}^L = \bar{r}_t(K_{it}, \tilde{M}_{it})$. Substituting this into Equation (4.1),

$$y_{it} = f_t\Big(K_{it}, L_{it} h_t\big(K_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big)\Big) + \omega_{it}^H + \epsilon_{it}.$$

This representation of the production function reveals an identification problem.

Proposition 4.1. *Without further restrictions, h_t cannot be identified from variations in inputs and output.*

Proof. Note that for arbitrary values of $(K_{it}, \tilde{M}_{it})$, the second argument of the h_t function, $\bar{r}_t(K_{it}, \tilde{M}_{it})$, is uniquely determined. Therefore, it is not possible to independently vary $(K_{it}, \omega_{it}^L, \tilde{M}_{it})$ and trace out all dimensions of h_t. This implies that h_t is not identified from variations in inputs and output.[16,17] □

Most objects of interest, such as the output elasticities or elasticity of substitutions, are a function of h_t, underscoring the challenge for identification. To see this, we can write the output elasticities as (suppressing the function arguments):

$$\theta_{it}^K := (f_{t1} + f_{t2}h_{t1})K_{it}, \qquad \theta_{it}^L := f_{t2}h_{t2}L_{it}\bar{r}_t(K_{it}, \tilde{M}_{it}), \qquad \theta_{it}^M := f_{t2}h_{t3}M_{it},$$

where f_{tk} denotes the derivative of f_t with respect to its $k-$th component. I also use θ_{it}^j to denote the output elasticity of j. Note that all the output elasticities depend on the derivatives of h_t, which is not identified. This functional dependence

[16] Online Appendix Section C.2 shows the parametric analog of this non-identification problem for the CES production function where some parameters are not separately identified.

[17] Ekeland et al. (2004) show a similar nonidentification result for hedonic demand models.

problem breaks down when there is variation in input prices because $\bar{r}_t(K_{it}, \tilde{M}_{it})$ also depends on the price ratio. However, in this case, the identification requires exogenous variation in input prices. This case is analyzed in Online Appendix Section B.1.

Given this nonidentification result, I introduce another function, $\bar{h}_t(K_{it}, \tilde{M}_{it}) \equiv h_t(K_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$, and rewrite the production function as:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{4.2}$$

Here, $\bar{h}_t$ can be viewed as a *reduced form* function, which arises from the firm's optimal input choices. It combines the effects of ω_{it}^L and the ratio of the optimally chosen flexible inputs on output. The rest of the section investigates (i) what can be identified from the reduced form representation, that is, from $f_t(\cdot)$ and $\bar{h}_t(\cdot)$, and (ii) how first-order conditions of cost minimization help identification.

4.2 Identification of Output Elasticities

This section presents identification results that are relevant for markup estimation. Other identification results are presented in Appendix A.

4.2.1 Identifying the Ratio of Labor and Materials Elasticities

The multicollinearity problem presented in Section 4.1 implies that θ_{it}^L and θ_{it}^M cannot be identified from variation in the inputs and output. However, the data provides an additional source of information from cost minimization. Recall that cost minimization implies a link between the production function and optimally chosen flexible inputs through the first-order conditions. Thus, we can learn about the output elasticities from the level of flexible inputs. To show the information provided by the first-order conditions, I write the firm's cost minimization problem:

$$\min_{L_{it}, M_{it}} \quad p_t^l L_{it} + p_t^m M_{it} \qquad \text{s.t.} \quad F_t\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big)\exp(\omega_{it}^H)\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}] \geqslant \bar{Y}_{it},$$

where $\bar{Y}_{it}$ is planned output. The first-order condition associated with this optimization problem is $F_{tV}\lambda_{it} = p_t^V$, where $V \in \{M, L\}$, F_{tV} donates the marginal product of V, and λ_{it} is the Lagrange multiplier. Multiplying both sides by

$V_{it}/(Y_{it}p_{it})$ and rearranging gives,

$$\underbrace{\frac{F_{tV}\big(K_{it}, \omega_{it}^{L} L_{it}, M_{it}\big) V_{it}}{F_t\big(K_{it}, \omega_{it}^{L} L_{it}, M_{it}\big)}}_{\text{Elasticity}(\theta_{it}^{V})} \frac{\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}]\lambda_{it}}{\exp(\epsilon_{it})p_{it}} \quad = \quad \underbrace{\frac{V_{it}p_t^{v}}{Y_{it}p_{it}}}_{\text{Revenue Share of Input}(\alpha_{it}^{V})}, \tag{4.3}$$

where p_{it} is the output price. This expression holds for all flexible inputs. Therefore, taking the ratio of Equation (4.3) for $V = M$ and $V = L$ yields

$$\theta_{it}^{M}/\theta_{it}^{L} = \alpha_{it}^{M}/\alpha_{it}^{L}. \tag{4.4}$$

The ratio of labor and materials elasticities is identified as the ratio of revenue shares using the cost minimization assumption. Since the revenue shares are observed in the data, we can calculate the ratio of elasticities without estimation.[18]

Using the first-order conditions in production function estimation has long been recognized in the literature, but mostly under parametric assumptions. (See Doraszelski and Jaumandreu (2013) for Cobb-Douglas and Grieco et al. (2016) for CES). Gandhi et al. (2020) propose a method that uses Equation (4.3) nonparametrically under a perfectly competitive output market, which implies that the output elasticity of a flexible input equals that input's revenue share. My contribution is to exploit the first-order conditions nonparametrically in the presence of two flexible inputs, even if firms have market power in the output market.

4.2.2 Identification of Sum of Materials and Labor Elasticities

In this section, I show how to recover the sum of the labor and materials elasticities from the reduced form representation of the production function in Equation (4.2).

Proposition 4.2. *The sum of labor and materials elasticities is identified as*

$$\theta_{it}^{V} := \theta_{it}^{M} + \theta_{it}^{L} = f_{t2}\big(K_{it}, L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it})\big) L_{it}\bar{h}(K_{it}, \tilde{M}_{it}). \tag{4.5}$$

Proof. Using Equation (4.1), the materials and labor elasticities are written as:

$$\theta_{it}^{M} = f_{t2}h_{t3}M_{it}, \qquad \theta_{it}^{L} = f_{t2}(h_t - h_{t3}\tilde{M}_{it})L_{it}.$$

The sum of the elasticities depends only on h_t, but none of its derivatives:

$$\theta_{it}^{V} = \theta_{it}^{M} + \theta_{it}^{L} = f_{t2}h_t L_{it} = f_{t2}\bar{h}_t L_{it}.$$

□

[18]Doraszelski and Jaumandreu (2019) also use revenue shares to identify the ratio of elasticities.

From this proposition, we see that identifying f_t and $\bar{h}_t$ is sufficient for identifying the sum of flexible input elasticities. Importantly, we do not need to identify the structural part of the production function, h_t, and labor-augmenting productivity shock.[19] Using the ratio of elasticities identified in the previous subsection, we can obtain the labor and materials elasticities as

$$\theta_{it}^L = \theta_{it}^V \alpha_{it}^L / \alpha_{it}^V, \qquad \theta_{it}^M = \theta_{it}^V \alpha_{it}^M / \alpha_{it}^V, \tag{4.6}$$

where $\alpha_{it}^V = \alpha_{it}^L + \alpha_{it}^M$. This result shows that combining the first-order conditions with the sum of elasticities identifies the labor and materials elasticities separately. Overall, this result is important for markup estimation because variable input elasticity is the key input for markup estimation. Therefore, the researcher can identify markups even without identifying the entire production function.

4.2.3 Other Identification Results

While this section focused on the identification of objects related to markup estimation, Appendix A provides identification results for other objects, such as the elasticity of substitution, capital elasticity, and productivity shocks. It is also worth highlighting that my model can accommodate some commonly used models in the literature. Online Appendix C shows how to impose returns to scale restrictions and estimate the CES and Nested CES production functions using my method. Thus, researchers interested in estimating a more restricted production function with labor-augmenting productivity can use one of these models.

5 Empirical Model and Data

This section presents the empirical model, describes the estimation procedure, and introduces the datasets used in empirical estimation.

5.1 Empirical Model

The purpose of my empirical model is to estimate the output elasticities and to infer markups from those estimates. To ease the demand on data, I consider the weak homothetic separable production function:

$$y_{it} = f_t\big(K_{it}, h_t(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{5.1}$$

[19] Note that even if f_t and $\bar{h}_t$ are not uniquely identified, the sum of elasticities is identified.

This form is nested by Equation (2.2), but slightly less general in that h_t does not take K_{it} as an argument, which simplifies the estimation procedure. This form leads to the following estimating equation:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}. \tag{5.2}$$

In Section 4, I showed how to identify the output elasticities from f_t and $\bar{h}_t$, so the goal is to identify these functions.[20] To control for Hicks-neutral productivity, I use the control variables developed in Equation (3.8), $\omega^H_{it} = c_{2t}\left(W_{it-1}, u^1_{it}, u^2_{it}\right)$. Substituting this into Equation (5.2), the estimating equation can be written as:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}\left(W_{it-1}, u^1_{it}, u^2_{it}\right) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u^1_{it}, u^2_{it}] = 0.$$

Since all right-hand-side variables are orthogonal to the error term, this equation can be estimated by minimizing the sum of squared residuals. However, this equation is not the only moment restriction the model provides. Recall that capital is a predetermined input orthogonal to the innovation to productivity shocks at time t, which can be used to augment the moment restrictions. To see this, using the first-order Markov property of the productivity shocks, Hicks-neutral productivity can be expressed as

$$\omega^H_{it} \equiv \bar{c}_{3t}(\omega^L_{it-1}, \omega^H_{it-1}) + \xi_{it},$$

for an unknown function $\bar{c}_{3t}(\cdot)$, where ξ_{it} is the separable innovation to Hicks-neutral productivity with $\mathbb{E}[\xi_{it} \mid \mathcal{I}_{it-1}] = 0$. This innovation term is different from those defined in Section 3 because it is mean independent of $(\omega^H_{it-1}, \omega^L_{it-1})$ and separable, in contrast to (u^1_{it}, u^2_{it}), which are independent and non-separable. This representation is commonly used in the proxy variable approach for constructing moments.

Since $(\omega^L_{it-1}, \omega^H_{it-1})$ can be written as a function of W_{it-1}, I obtain a second representation of ω^H_{it} as $\omega^H_{it} \equiv c_{3t}(W_{it-1}) + \xi_{it}$, giving another estimating equation:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{3t}(W_{it-1}) + \xi_{it} + \epsilon_{it}.$$

The error term, $\xi_{it} + \epsilon_{it}$, is orthogonal to the firm's information set at time $t-1$, which includes K_{it}, so $\mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}] = 0$. I now summarize the estimation prob-

[20] Note that even though $\bar{h}_t$ is identified up to a scale, the elasticities are uniquely identified. I restrict the logarithm of h_t to have mean zero in the estimation to impose this normalization.

lem by combining the models and moment restrictions. We have two estimating equations:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}(W_{it-1}, u^1_{it}, u^2_{it}) + \epsilon_{it},\ \mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u^1_{it}, u^2_{it}] = 0 \quad (5.3)$$

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{3t}(W_{it-1}) + \xi_{it} + \epsilon_{it},\ \mathbb{E}[\xi_{it} + \epsilon_{it}, \mid K_{it}, W_{it-1}] = 0. \quad (5.4)$$

Identification of output elasticities requires identification of the functions, f_t and $\bar{h}_t$, using these moment restrictions. Therefore, one question is whether moment restrictions in Equation (5.3) and (5.4) identify these functions. I analyze this question in Online Appendix E and show that the moment restrictions in Equation (5.3) identify f_t and $\bar{h}_t$ except for special cases.[21] These cases include some support conditions on the conditional CDF in Equation (3.7) and conditions on the derivatives of the production function. Since Equation (5.3) identifies the output elasticities, the moment restrictions in Equation (5.4) provide efficiency gains.[22]

The estimation proceeds in two steps. In the first step, I estimate the control variable u^2_{it} by estimating the conditional CDF in Equation (3.7). In the model given in Equation 5.1, u^1_{it} corresponds to normalized $\tilde{M}_{it}$, so it does not require any estimation. After estimating the control variables, I approximate the nonparametric functions using polynomials and use the moment restrictions in Equations (5.3) and (5.4) for estimation.

5.1.1 Estimation Procedure

In this section, I provide an overview of the estimation procedure; the details are given in Online Appendix Section A.7. I estimate separate production functions for each industry. However, estimating the production function separately each year is not feasible for most industries due to the small sample size. To address this, I use an eight-year rolling-window estimation for Compustat and three-year rolling window estimation for other datasets following De Loecker et al. (2020).

The estimation involves two stages. The first stage estimates the CDF in Equation (3.7). To estimate this CDF, I partition the support into 500 equally

[21] This is called generic identification; see Lewbel (2019). One would ideally like to analyze the identification properties of moment restrictions in Equations (5.3) and (5.4) jointly. Since this is a difficult problem, I focus on the identification properties of Equation (5.3).

[22] One can also use the recursive representation of productivity shocks, $\omega^H_{it} = c_{2t}\left(W_{it-k}, \{u^1_{it-l}\}_{l=0}^{k-1}, \{u^2_{it-l}\}_{l=0}^{k-1}\right)$, and construct more moment functions using $\mathbb{E}[\epsilon_{it} \mid \{W^1_{it-l}\}_{l=0}^{k-1}\{u^1_{it-l}\}_{l=0}^{k-1}, \{u^2_{it-l}\}_{l=0}^{k-1}] = 0$ to increase efficiency.

sized grids and I use a logit model with third-degree polynomials to estimate the CDF at these points. I then approximate the CDF by interpolating between the points in the grid. In the second stage, I use polynomial approximations to estimate production function. I first approximate the logarithm of $\bar{h}_t$ by using third-degree polynomials:

$$\log(\widehat{\bar{h}}_t(\tilde{M}_{it})) = a_{1t} + a_{2t}\tilde{m}_{it} + a_{3t}\tilde{m}_{it}^2 + a_{4t}\tilde{m}_{it}^3, \tag{5.5}$$

where $\{a_{jt}\}_{j=1}^4$ are the parameters of the polynomial approximation. I set $a_{1t} = 0$ to impose the normalization for $\bar{h}_t(\tilde{M}_{it})$ described in Section 4. Letting $V_{it} := L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})$, the production function can be approximated as:

$$\begin{aligned}\widehat{f}_t(K_{it}, v_{it}) = b_{1t} + b_{2t}k_{it} + b_{3t}k_{it}^2 + b_{4t}k_{it}^3 + b_{5t}v_{it} + b_{6t}v_{it}^2 + b_{7t}v_{it}^3 \\ + b_{8t}k_{it}^2 v_{it} + b_{9t}k_{it}v_{it}^2 + b_{10t}k_{it}v_{it}\end{aligned} \tag{5.6}$$

where $\{b_{jt}\}_{j=1}^{10}$ are the parameters of the polynomial approximation.[23] I similarly approximate the control functions $c_{2t}(\cdot)$ and $c_{3t}(\cdot)$ using third-degree polynomials. I then construct an objective function using the moment restrictions in Equations (5.3) and (5.4). In particular, I use the sum of squared residuals from Equation (5.3) and timing moments from Equation (5.4) to obtain the following objective function to minimize:

$$J(\widehat{f}_t, \widehat{\bar{h}}_t, \widehat{c}_{2t}, \widehat{c}_{3t}) = \underbrace{\frac{1}{TN}\sum_{i,t}\widehat{\epsilon}_{1it}^2}_{\text{Sum of Squared Residuals}} + \underbrace{\left(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K_{it}\right)^2 + \left(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K_{it}^2\right)^2}_{\text{Timing Moments}}$$

Estimating $\widehat{c}_{2t}$ and $\widehat{c}_{3t}$ is computationally simple as they can be partialed out for a given $(\widehat{f}_t, \widehat{\bar{h}}_t)$. Thus, the estimation requires searching for $\widehat{f}_t$ and $\widehat{\bar{h}}_t$ to minimize the objective function. After obtaining the estimates for f_t and $\bar{h}_t$, I calculate the output elasticities as described in Equations (4.5), (4.6), and (A.3).

Deriving the large sample distribution of the output elasticities and other estimates used in the empirical applications is difficult. First, I need to account for estimation error in the first stage, and then I need to understand how estimation errors in the output elasticities translate into further stages. To avoid these complications, I use bootstrap to estimate standard errors. The bootstrap procedure

[23]For the US, the third-degree polynomial approximation gives very large standard errors and unstable estimates due to small sample size, especially in the beginning of the sample. For this reason, I use the second-degree polynomials for the US.

Table 1: Descriptive Statistics on Datasets

	US	Chile	Colombia	India	Turkey
Sample Period	1961-2014	1979-96	1978-91	1998-2014	1983-2000
Num. of Industries	3	5	9	5	8
Industry Level	2-dig NAICS	3-dig SIC	3-dig SIC	3-dig NIC	3-dig SIC
Num. of Obs/Year	1247	2115	3918	2837	4997

Notes: This table provides descriptive statistics for the dataset used in the empirical estimation.

treats firms as independent observations and resamples firms with replacement. I use 100 bootstrap repetitions to estimate the standard errors.

5.2 Data

I use panel data from manufacturing industries in five countries: Chile, Colombia, India, Turkey, and the US. Table 1 provides descriptive statistics of the dataset.

5.2.1 Chile, Columbia, India, and Turkey

The data for the four developing countries are plant-level production data collected through censuses. The Chilean dataset comes from the census of Chilean manufacturing plants with more than ten employees between 1979 and 1996. Similarly, the Colombian data come from the manufacturing census covering all manufacturing plants with more than ten employees from 1981 to 1991. The Turkish dataset is from the Annual Surveys of Manufacturing Industries and covers all establishments with ten or more employees between 1983 and 2000. Finally, the Indian data are from the Annual Survey of Industries conducted by the Indian Statistical Institute, covering plants with 100 or more employees from 1998 to 2014.

From these datasets, I obtain input and output measures for estimating the production functions. I obtain the materials input by deflating the materials cost using the appropriate deflators. The labor input equals the number of worker days or the number of workers. I obtain the capital input via the perpetual inventory method or deflated book values of capital. I remove outliers based on labor and materials' shares of the revenue. To improve the precision of the estimates, I limit my sample to industries with at least an average of 250 plants per year. I provide details about the datasets and descriptive statistics in Online Appendix A.

5.2.2 US

The Compustat sample contains all publicly traded manufacturing firms in the US between 1961 and 2014. It includes information from financial statements, including sales, total input expenditures, number of employees, capital stock formation, and industry classification. From this information, I obtain labor, materials, and capital inputs and output measures. My output measure is the net deflated sales, and my labor measure is the number of employees. Compustat does not report separate expenditures for materials. To address this, I follow Keller and Yeaple (2009) to estimate materials cost by netting out capital depreciation and labor costs from the cost of goods sold and administrative and selling expenses.

Even though Compustat is compiled from financial accounting data rather than manufacturing censuses, as in other countries, it has better coverage across industries and over time. Therefore, it is more suitable to study the change in markups and market power. For this reason, it played an important role in the growing evidence on the rise of market power in the US (De Loecker et al. (2020). I aim to revisit those findings and explore how using factor-augmenting production function technology affects the markup estimates.

An important typical data limitation is observing revenues and expenditures rather than physical quantities, a concern if sales reflect firm-level demand heterogeneity (Jaumandreu (2018), Doraszelski and Jaumandreu (2019) and Bond et al. (2021)). Although physical output or firm-level price indexes are available in some recent datasets, I aim to demonstrate my method on several commonly used datasets. As a robustness check, I use quantity input and output data from seven industries in India that produce relatively homogeneous products. I show that the main results of the paper on the importance of labor-augmenting productivity are robust to estimating quantity production functions (Section F.1).

6 Empirical Results: Production Function

This section presents results on production function estimates. I use the estimates to discuss several findings. First, I compare my estimates with other commonly used models and find that my estimates are significantly different from the estimates of these models. Second, my estimates uncover significant heterogeneity in output elasticities. This heterogeneity is meaningful: larger firms are more capital-intensive, and smaller firms are more labor- and variable input-intensive.

6.1 Output Elasticities

In this section, I estimate output elasticities with four different production functions: (i) Cobb-Douglas (CD), (ii) Translog with Hicks-neutral productivity (Translog), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) my model, a nonparametric production function with labor-augmenting productivity (FA).[24,25] The main goal is to compare the estimates from these models and study the potential biases in the output elasticities. Remember that my model introduces two forms of flexibility: functional form (nonparametric) and unobserved heterogeneity (labor-augmenting productivity). What are the contributions of these different forms of flexibilities to the estimates? Comparing the estimates from different models allows me to answer this question. For example, translog introduces functional form flexibility relative to CD, so comparing CD, Translog, and FA would identify the role of unobserved heterogeneity. On the other hand, CES-FA introduces unobserved heterogeneity only, so comparing CD, CES-FA, and FA would uncover the contribution of functional form flexibility.[26]

I report the sales-weighted economy-level output elasticities of capital, labor, and variable input in Figure 1. Looking at capital elasticity, I find that FA estimates a higher capital elasticity than other production functions in all countries except the US.[27,28] The results point out large biases in the CD estimates; for example, capital elasticity from FA is almost twice as large as the corresponding estimates from CD in Turkey and Chile. In most countries, translog estimates are close to the FA estimates, suggesting that functional form flexibility through Translog partially reduces the bias. However, the CES-FA estimates give lower

[24] Since the gross production functions are not identified with the Ackerberg et al. (2015) method, I estimate CD and Translog with the Blundell and Bond (2000) method. I estimate CES-FA using procedures described in Section 5.1.1, but I impose the CES functional form assumption. The details of these estimation procedures are given in Online Appendix Sections C.3 and A.8

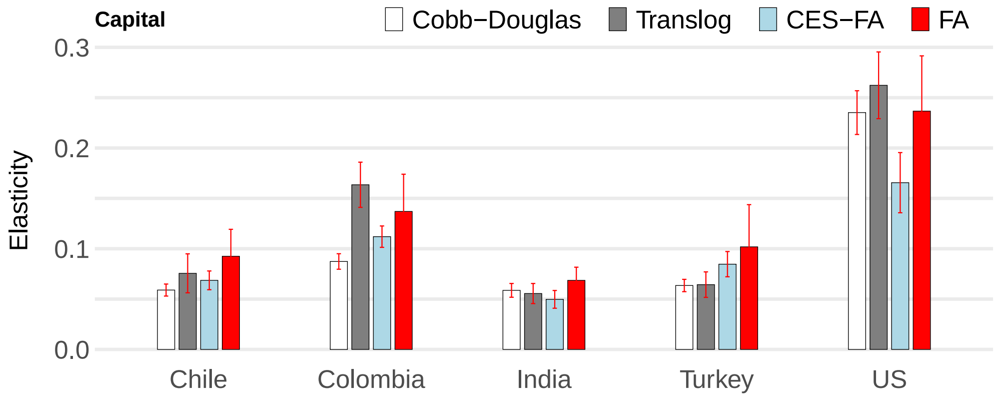
[25] Since Cobb-Douglas does not allow for a non-neutral productivity shock, I consider the CES as a parametric specification with labor-augmenting productivity.

[26] Other estimates, materials elasticity, and returns to scale are reported in Online Appendix Figure G.3. Supplemental Materials present the sales- and cost-weighted average elasticities for the three largest industries in each country.

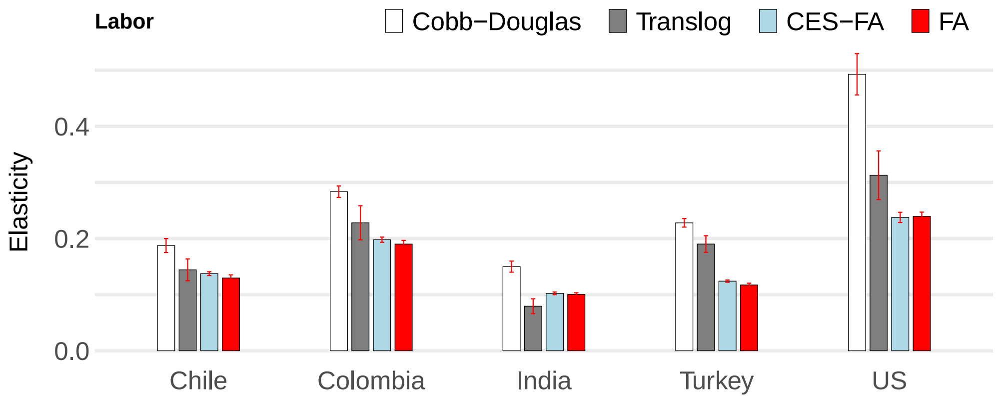
[27] The standard errors for differences between the FA estimate and other estimates are reported in Online Appendix Figure G.5

[28] Appendix A shows how to estimate capital elasticity.

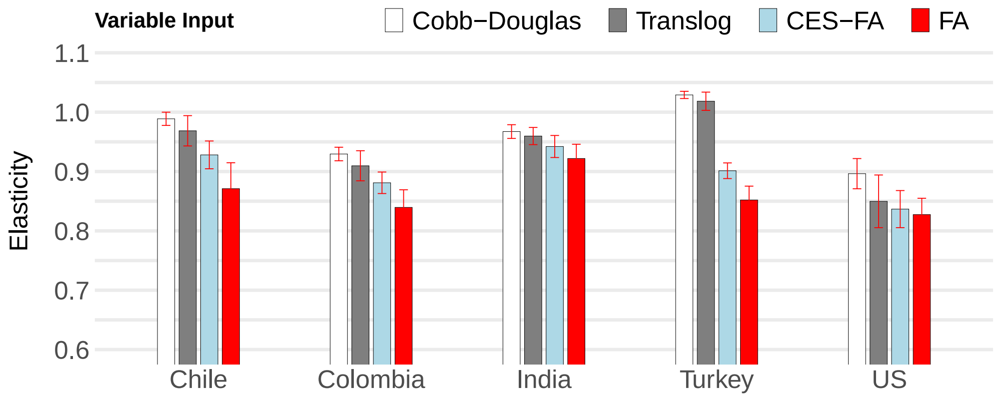
Figure 1: Average Capital and Labor Elasticities Comparison



(a) Capital Elasticity



(b) Labor Elasticity



(c) Variable Input Elasticity

Notes: Comparison of sales-weighted average elasticities produced by Cobb-Douglas (CD), (ii) Translog with Hicks-neutral productivity (Translog), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) nonparametric production function with factor-augmenting productivity (FA). For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95% confidence intervals calculated using bootstrap (100 iterations).

capital elasticities, suggesting that labor-augmenting productivity, alone, does not reduce the bias. Overall, this suggests that functional form flexibility is more important than labor-augmenting productivity for capital elasticity. Examining labor elasticities reveals a different pattern: the labor elasticity estimates decrease once we add labor-augmenting productivity to the production model. This time, CD and Translog estimates are similar, indicating that adding functional form flexibility does not change the estimates, whereas CES-FA results are close to FA, suggesting that labor-augmenting productivity alone, without the non-parametric form, eliminates the bias in labor elasticity. These results indicate that labor-augmenting productivity is more important than functional form flexibility when estimating labor elasticities.

I now turn to variable input elasticity, the key input for markup estimation. There is a clear pattern with variable input elasticity estimates. The variable elasticity estimates decrease from the least flexible (CD) to most flexible (FA) specification, suggesting that both forms of flexibility, functional form and unobserved heterogeneity, are important to correctly estimate the flexible input elasticity. This result has important implications for markups because variable input elasticity is the critical input for markup estimation. These results motivate my analysis in the next section, where I study how the production function specification affects markup estimates.[29]

To summarize, three conclusions can be drawn from these results: (i) flexible functional form is more important for capital elasticity estimation, (ii) labor-augmenting technology is more important for labor elasticity estimation, and (iii) both flexible functional form and factor-augmenting productivity are important to estimate variable input elasticity, the critical input for markup estimation.

6.2 Heterogeneity in Output Elasticities

This section examines the within-industry heterogeneity in the output elasticities. In particular, I first document heterogeneity in output elasticities across firms and then examine how production technology changes with firm size. The literature has found substantial firm-level heterogeneity in many firm outcomes and

[29]Some concerns in production function estimation are measurement error in capital and not observing capacity utilization. To address these concerns, Section 9.2 presents robustness checks and shows that my results are not explained by these factors.

Figure 2: Average Coefficient of Variation

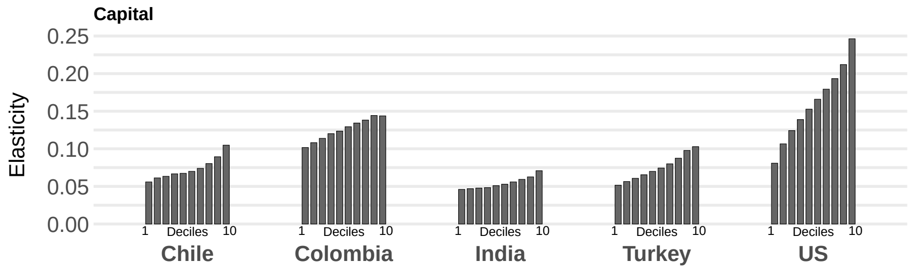
Notes: This figure shows the mean (gray) and 10th and 90th percentile (red) of the distribution of the average coefficient of variation cross industries and years for different output elasticities.

productivity (Syverson (2011), Van Reenen (2018)); however, there is limited evidence of heterogeneity in production technology. Moreover, by documenting this heterogeneity, this section demonstrates the added flexibility of my method.

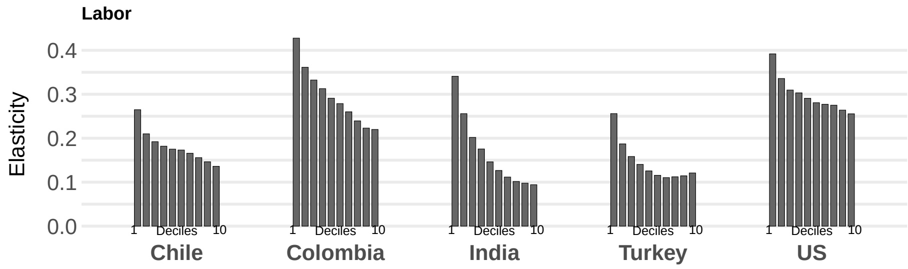
To measure heterogeneity, I estimate the coefficient of variation (CV) of the output elasticities within each industry-year. Figure 2 displays the average and 10–90th percentiles of the CV estimates for all countries. There is substantial heterogeneity in the output elasticities in all countries, as evidenced by the large average CV estimates. The heterogeneity is highest for labor and lowest for materials. This finding is consistent with the large heterogeneity in labor's revenue share and low heterogeneity in materials' revenue share observed in the data. Moreover, the 10–90th percentiles (red bars) show that the heterogeneity exists in most industries. Finally, I find little heterogeneity in returns to scale, a reasonable finding because too large or too small returns to scale would be inconsistent with optimal firm behavior.

Heterogeneity in production technology is an important finding, and it complements the existing evidence on large firm-level heterogeneity in other dimensions. Yet, a more interesting question is what explains this heterogeneity? To explore

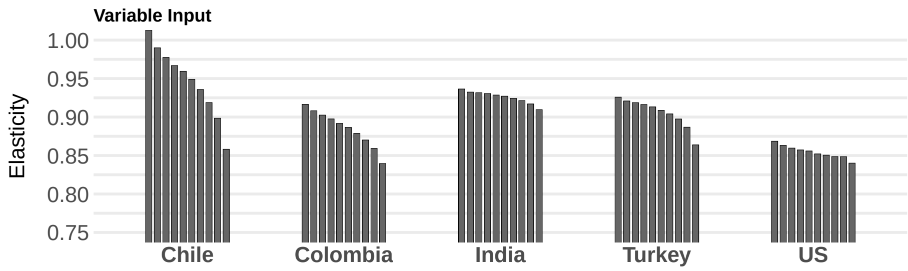
Figure 3: Output Elasticities by Firm Size Deciles



(a) Capital Elasticity



(b) Labor Elasticity



(c) Variable Input Elasticity

Notes: This figure shows output elasticities by firm decile. For each country, average elasticity for each decile within an industry-year is estimated first, then these estimates are averaged across industry-year bins.

this, I report the average capital, labor, and variable input elasticities across firm size deciles. In particular, for each industry-year-decile bin, I estimate the mean output elasticities and then report their averages across years and industries.

The result reveals a striking pattern. In all countries, as firm size increases, capital intensity of production increases and labor intensity of production decreases.[30] The differences are large: the largest firms are twice as capital-intensive as the smallest firms, and the smallest firms are twice as labor-intensive as the largest firms. These results have two important implications. First, it is crucial

[30]These findings agree with the literature, which finds that large firms use more capital and less labor than small firms (Holmes and Schmitz (2010), Kumar et al. (1999)).

to allow for flexible production technology to capture substantial heterogeneity in production technology across firms. Second and importantly for markup estimation, the relationship between variable input elasticity and firm size is key to correctly estimating aggregate markup, which we turn to next.

7 Inferring Markups from Production Function

There is a simple link between a firm's markup and its output elasticities, which has recently been widely used to estimate markups. This section describes this link and argues that the form of the production function has critical implications for markup estimates.

Building on Hall (1988), De Loecker and Warzynski (2012) propose an approach to estimate markups from production data under the assumptions that firms are cost-minimizers with respect to at least one flexible input and firms take input prices as given. In particular, markup is given by $\mu_{it} := \theta_{it}^V / \alpha_{it}^V$, where μ_{it} denotes the firm-level markup, which equals the output elasticity of a flexible input divided by its revenue share.[31] Since the revenue shares of flexible inputs are typically available in the data, an estimate of the flexible input elasticity is sufficient to estimate markups.

7.1 How Does the Form of the Production Function Affect Markups?

Output elasticity is the only estimated component in the markup formula when markups are estimated from production functions. As a result, the bias in output elasticities directly translates to markups, making the markup estimates sensitive to the elasticity estimates. We also know that elasticity estimates are sensitive to the production function specification. For example, Van Biesebroeck (2008) compares conventional production function estimation methods and finds that the elasticity estimates differ substantially. This suggests that the production function specification is critical when estimating markups from production data.[32] Motivated by this fact, this section discusses the implications of production function specifications on markups and then shows that labor-augmenting productivity

[31] As shown in De Loecker and Warzynski (2012), this revenue share is with respect to planned output, which requires correcting for $exp(\epsilon_{it})$ in the form of $p_{it} Y_{it} / exp(\epsilon)$. I implement this correction when estimating markups.

[32] This is in contrast to productivity estimates, which are shown to be robust to the specification; see Foster et al. (2017) and Van Beveren (2012).

provides a solution to some puzzling results in the literature.

Conflicting Markup Estimates from Different Flexible Inputs. Cost minimization implies that markup estimates from different flexible inputs should be the same. However, studies estimating markups from two flexible inputs have found that different flexible inputs give conflicting markup estimates (Doraszelski and Jaumandreu (2019), Raval (2020)). These papers suggest that at least one assumption required to estimate markups from production data is violated, and they point to a lack of factor-augmenting productivity.[33]

Consistent with these results in the literature, this paper shows that labor-augmenting productivity ensures identical markup estimates from labor and materials and provides a natural solution to this problem. To see this, Equation (4.4) immediately implies:

$$\frac{\theta_{it}^L}{\theta_{it}^M} = \frac{\alpha_{it}^L}{\alpha_{it}^M} \quad \Longrightarrow \quad \mu_{it}^L = \frac{\theta_{it}^L}{\alpha_{it}^L} = \frac{\theta_{it}^M}{\alpha_{it}^M} = \mu_{it}^M, \tag{7.1}$$

where μ_{it}^L and μ_{it}^M denote markup estimates from labor and materials. Two key components of my approach lead to this outcome: (i) using the ratio of revenue shares to identify the ratio of elasticities and (ii) the presence of labor-augmenting productivity. The presence of labor-augmenting productivity is key to the ability to use the ratio of revenue shares to identify the ratio of elasticities. As I will show in Section 8.1, without the labor-augmenting productivity, the identity in Equation (7.1) is typically rejected, suggesting that the model is not internally valid. The additional unobserved heterogeneity resolves this discrepancy by adding one more dimension of unobserved heterogeneity, making the model internally valid.

7.2 Markup Decomposition: The Role of Production Functions

This section presents a markup decomposition framework to understand how production function estimates affect markups. I show that production function estimation can bias the aggregate markup through two mechanisms: (i) bias in the average output elasticity and (ii) firm-level heterogeneity in the output elasticities.

[33] Raval (2020) formally tests the production function approach to markup estimation using the implication that markups from two flexible inputs should be identical. He finds that the two markup measures from labor and materials are negatively correlated and suggest different trends. He then examines possible mechanisms, such as heterogeneity in the production function, adjustment costs in labor, measurement error, and frictions in the labor market. He concludes that the most plausible explanation is the inability of the standard production functions to account for heterogeneity in production technology.

The aggregate markup is given by $\mu_t = \sum w_{it}\mu_{it}$, where w_{it} is the aggregation weight, usually a measure of firm size. To assess the influence of production functions on the estimated aggregate markup, I apply the Olley-Pakes (OP) decomposition, which decomposes a weighted average into an unweighted average and the covariance between the weight and variable of interest. To implement the OP decomposition, I use the aggregate log markup, $\tilde{\mu}_t = \sum w_{it} \log(\theta_{it}) - \sum w_{it} \log(\alpha_{it})$, which equals to the difference of two weighted averages. Applying the OP decomposition to both terms obtains:

$$\tilde{\mu}_t = \underbrace{\underbrace{\bar{\theta}_t}_{\text{Avg. Elas.}(1)} + \underbrace{\mathrm{Cov}(w_{it}, \log(\theta_{it}))}_{\text{Heterogeneity in Technology}(2)}}_{\text{Estimation}} \underbrace{- \underbrace{\bar{\alpha}_t}_{\text{Avg. Share}(3)} - \underbrace{\mathrm{Cov}(w_{it}, \log(\alpha_{it}))}_{\text{Heterogeneity in Shares}(4)}}_{\text{Data}} \tag{7.2}$$

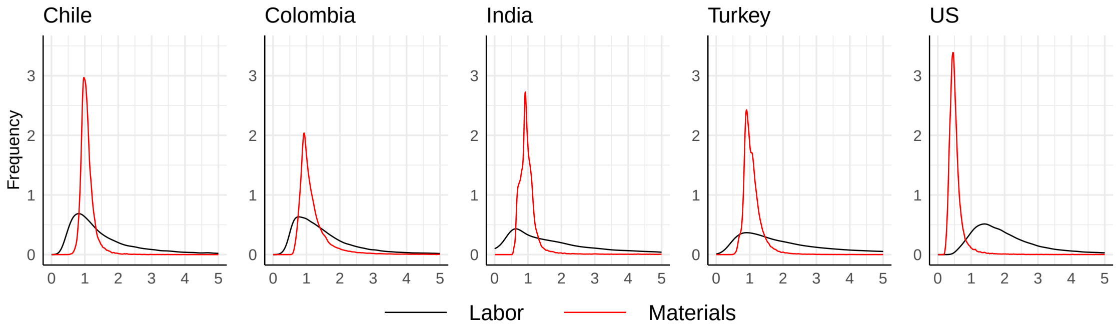
The aggregate log markup is composed of four parts. The first two parts involve output elasticities, and the last two involve revenue shares. This decomposition is useful for analyzing the aggregate markup because each component involves either the output elasticities, which are estimated, or the revenue shares, which come directly from data. Therefore, analyzing the first two components will reveal how biases in production function estimation translate to markup estimates.

Bias from the Average Output Elasticity. The first component in the decomposition is the average elasticity. Under misspecification, any bias in this component directly translates into bias in the aggregate markup. The elasticity estimates in the previous section suggested that Hicks-neutral production functions overestimate the flexible input elasticity. Therefore, we expect the bias from this source to be positive.

Bias from Heterogeneity in Production Technology. The second component is the covariance between firm size and flexible input's output elasticity. This component contributes to the aggregate markup when the elasticities are heterogeneous and correlated with firm size. The markup will be biased if the production function does not capture this heterogeneity. The bias is positive if large firms have lower flexible input elasticity than small firms. My estimates in Section 6.2 documented a negative correlation between firm size and the elasticities of flexible input, so we should expect this source of the bias to be positive. Moreover, if the first two components change over time, these biases affect the estimates of markup trends. This can happen, for example, if large firms become more capital-intensive over time, leading to a decrease in the second component.

Since the elasticity estimates in the previous section suggested that both sources of

Figure 4: Distribution of Markups Implied by Labor and Materials (Cobb-Douglas)



Notes: This figure compares the distribution of markups implied by labor (black) and materials (red) elasticities for each country from the Cobb-Douglas specification estimated using the Blundell and Bond (2000) method.

bias have positive signs, these biases do not cancel each other and markup estimates are expected to have upward bias with Hicks-neutral productivity. The next section presents markup estimates to quantify this bias.

8 Empirical Results: Markups

I estimate markups using the estimated output elasticities and investigate whether the form of the production function systematically affects the markup estimates. Then, I focus on the change in markups in the US and study how the production functions affect markup trends.

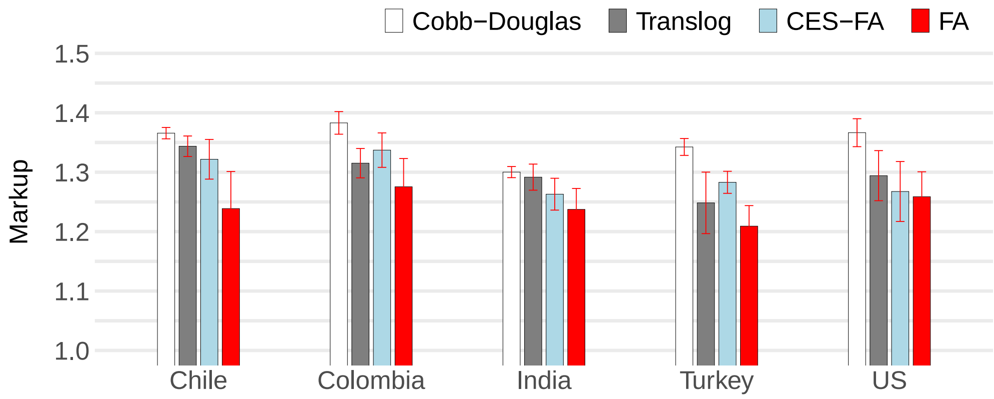
8.1 Testing the Cobb-Douglas Specification using Markups

As discussed in Section 7, testing the equality of markups from labor and materials serves as a specification test. This section applies this test to the Cobb-Douglas production function.

I use the output elasticity estimates produced by the Blundell and Bond (2000) method for markup estimation. Figure 4 plots the distributions of markup estimates inferred from the labor and materials elasticities. If the model is correct, the two distributions should overlap. However, the distributions are substantially different, with labor generating a more dispersed distribution than materials. Moreover, both distributions indicate that many firms have markups below one. These results provide strong evidence against the Hicks-neutral production functions.[34]

[34]Another reason markup estimates would be different is adjustment cost in labor. To test this hypothesis, Raval (2020) estimates markups from materials and energy inputs and finds different markups. From this result, he concludes that misspecification of the input type is not the main driver of conflicting markup estimates.

Figure 5: Average Markups Comparison



Notes: Comparison of sales-weighted markups produced by Cobb-Douglas (CD), (ii) Translog with Hicks-neutral productivity (Translog), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) nonparametric production function with factor-augmenting productivity (FA). For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95% confidence intervals calculated using bootstrap (100 iterations).

Since I reject the Cobb-Douglas specification with two flexible inputs, I estimate another production function with a single flexible input for comparison purposes, following De Loecker et al. (2020): $y_{it} = \beta_{kt}k_{it} + \beta_{vt}vr_{it} + \omega_{it}^H + \epsilon_{it}$. Here, vr_{it} is the combined flexible input of labor and materials, defined as the deflated sum of labor and materials cost. I estimate this model and calculate markups as $\mu_{it} = \beta_{vt}/\alpha_{it}^V$. For my model, I use the sum of flexible input elasticity divided by flexible input's revenue share as the markup measure.

8.2 Markups Comparison: Level

This section compares the aggregate markups produced by my method and other production functions considered in Section 6.1. For each country, I first calculate the annual sales-weighted markup and then take the average over the sample period.[35] Figure 5 displays the aggregate markups from four functional forms: CD, Translog, CES-FA, and FA, along with the 95% confidence interval. The FA generates aggregate markups significantly smaller than the CD estimates in all countries. The difference ranges from 0.1 to 0.2, an important magnitude when markups are interpreted as a measure of market power. The differences between markup estimates are statistically significant, as reported in Online Appendix Figure G.5. We can see the contribution of functional form flexibility and labor-augmenting productivity by looking at results from Translog and CES-FA.

[35]I also report cost-weighted estimates in Online Appendix G and find similar results.

Table 2: Sales-Weighted Average Markups for the Three Largest Industries

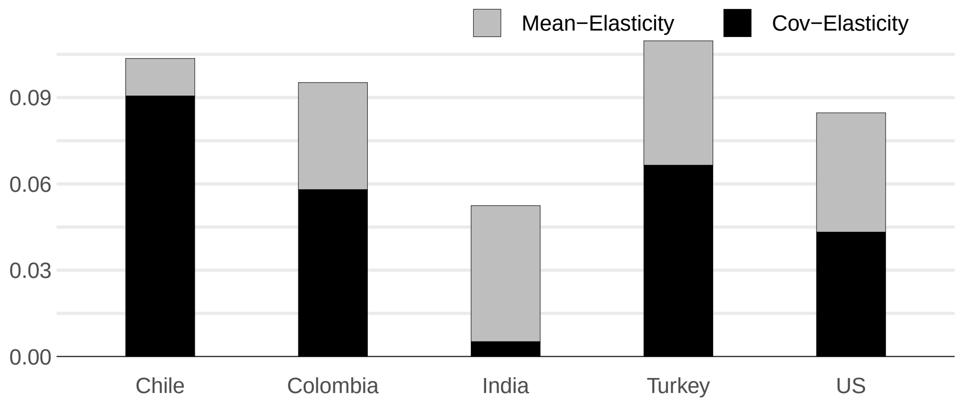
	Industry 1				**Industry 2**				**Industry 3**			
	CD	TR	CES	FA	CD	TR	CES	FA	CD	TR	CES	FA
					(311, 381, 321)							
Chile	1.35	1.33	1.29	1.21	1.43	1.42	1.39	1.28	1.32	1.29	1.31	1.27
	(0.01)	(0.01)	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)	(0.06)	(0.02)	(0.03)	(0.04)	(0.06)
					(311, 322, 381)							
Colombia	1.29	1.24	1.25	1.25	1.39	1.34	1.39	1.31	1.4	1.35	1.33	1.22
	(0.00)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)	(0.01)	(0.02)	(0.04)	(0.06)
					(230, 265, 213)							
India	1.2	1.2	1.14	1.15	1.17	1.16	1.17	1.16	1.42	1.44	1.33	1.26
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
					(321, 311, 322)							
Turkey	1.26	1.25	1.19	1.16	1.37	1.31	1.29	1.24	1.3	1.22	1.34	1.15
	(0.01)	(0.05)	(0.02)	(0.03)	(0.01)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)
					(33, 32, 31)							
US	1.59	1.42	1.4	1.32	1.4	1.37	1.3	1.26	1.27	1.19	1.25	1.24
	(0.03)	(0.07)	(0.11)	(0.07)	(0.02)	(0.04)	(0.05)	(0.04)	(0.01)	(0.02)	(0.03)	(0.02)

Notes: Sales-weighted average markup estimates for the three largest industries in each country. The estimates are obtained from Cobb-Douglas (CD), Translog (TR), CES with labor-augmenting productivity (CES), and nonparametric labor-augmenting productivity (FA). For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. Standard errors are obtained using bootstrap. Industry codes are given in parantheses. The corresponding industry names can be found in Supplemental Materials Section A.

The estimates from these specifications fall between the CD and FA estimates, reducing the bias from 20 to 80 percent depending on the country, but they do not eliminate it.

Industry-level markup estimates reported in Table 2 confirm these findings. This table reports the average sales-weighted markups in the three largest industries in each country. We see that except for a few industries, CD gives the highest markup estimates, FA gives the lowest markup estimates, and Translog and CES-FA markup estimates are between the two. These results clearly show the importance of both forms of flexibility in production functions when estimating markups: (i) functional form and (ii) unobserved heterogeneity. Allowing for only functional form flexibility or labor-augmenting productivity is not sufficient to eliminate biases in markups. Therefore, one needs both types of heterogeneity to estimate markups correctly. Furthermore, drawing similar conclusions from different datasets provides strong evidence that these results are robust to sample periods and country-specific characteristics.

Figure 6: Decomposition of the Difference between Aggregate Markups



Notes: This figure decomposes the difference between the aggregate log markups produced by my method and the Cobb-Douglas model.

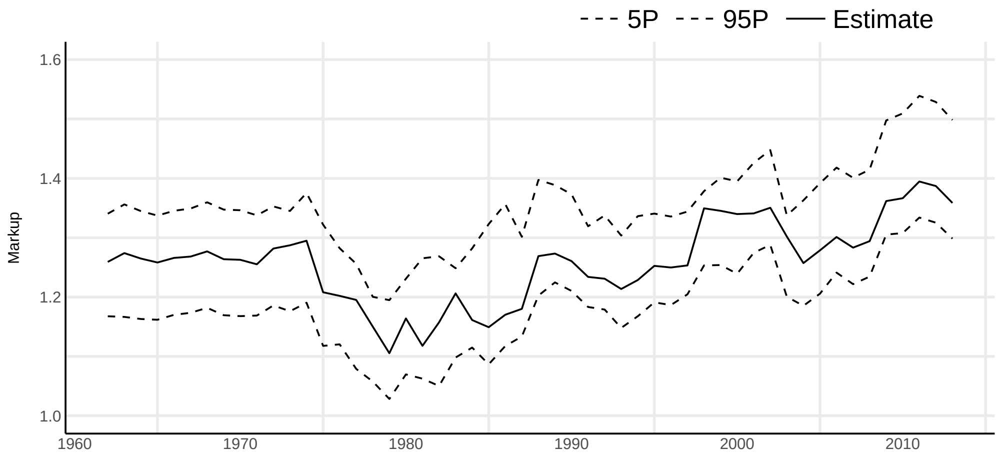
Differences in markup estimates when using more flexible production functions raise an important question: what drives these differences in markup estimates? My analysis in Section 7.2 suggested two mechanisms: (i) average output elasticities are estimated incorrectly, and (ii) output elasticity estimates do not capture heterogeneity in firms' production technology. To quantify the role of these mechanisms, I decompose the difference between markups estimated by CD and FA. According to Equation (7.2), the difference between the log aggregate markup estimates is decomposed as the difference between average elasticities and the covariances of firm size and elasticities. I plot this decomposition in Figure 6, which shows the difference between average output elasticities (gray) and the difference between covariance terms (black). The figure highlights two key reasons for the difference in markup estimates between the two methods. First, the Cobb-Douglas production function overestimates the variable input elasticity in all countries except Chile. Second, Cobb-Douglas does not capture the negative relationship between firm size and flexible input elasticity. This negative covariance is not surprising because both the literature and my analysis in Section 6 suggested that large firms are more capital-intensive and less flexible input-intensive, leading to a negative correlation between firm size and flexible input elasticity. Both of these factors generate upward bias in the Cobb-Douglas markup estimates.

After showing important differences in the level of markups across estimation methods, I now turn to the change in markups over time.

8.3 Markups Comparison: Trend

This section investigates the evolution of the aggregate markup in the US manufacturing sector. The estimates for other countries are reported in Online Appendix Figure G.6.

Figure 7: Change of Aggregate Markup in the US

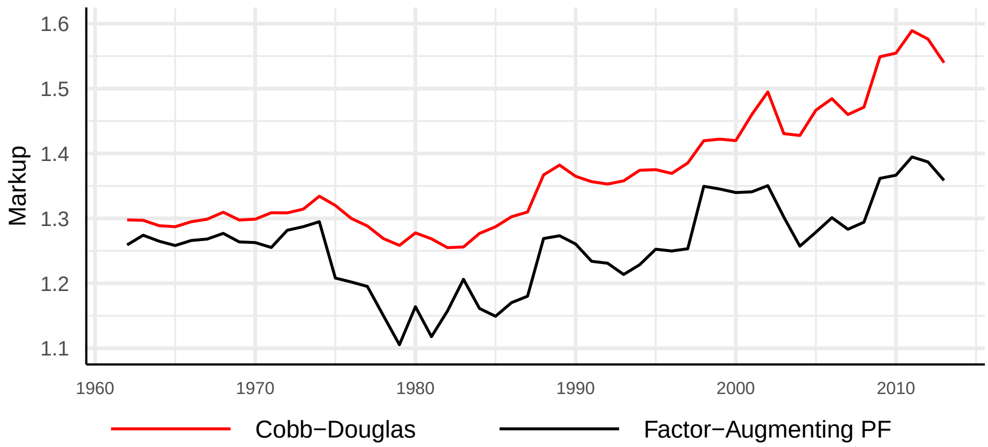


Notes: The evolution of markups in the US manufacturing industry. The dotted lines report the 5-95th percentile of the bootstrap distribution (100 iterations).

I focus on the US because there has been recent interest in understanding the change in market power in the US by measuring the change in markups over time. Figure 7 plots the sales-weighted aggregate markup from 1960 to 2012 along with the 5–95th percentile confidence band. In the 1960s, the aggregate markup is about 30 percent over marginal cost. It remains flat until 1970 and then declines gradually between 1970 and 1980, falling to about 15 percent in 1980. Then, markups start to rise with some cyclical patterns and reach 40 percent at the end of the sample period. We also see that the markup tends to decline during recessions. Overall, the manufacturing industry's aggregate markup has risen from 1.3 to 1.45 during the sample period.

Next, I compare my results with the Cobb-Douglas estimates. Cobb-Douglas and other Hicks-neutral production functions have been used in the literature to estimate markups, and the results suggest a dramatic rise in markups in the US economy since the 1960s (De Loecker et al. (2020)). I aim to study how estimating a labor-augmenting production function affects this conclusion. Figure 8 reports both markup measures, from Cobb-Douglas specification (red) and from nonparametric labor-augmenting production function (black). The Cobb-Douglas estimates suggest that markups rose more than 30 percentage points between 1960 and 2012, mirroring the findings in the literature. The markup estimates from the labor-augmenting production function also suggest a rise in markup, albeit a more modest one: around 15 percentage points between 1960 and 2012. The difference between the two series is statistically significant at the 95% confidence level as reported in Online Appendix Figure G.6. The two series are similar until the '70s and then start to diverge. The markup from the labor-augmenting production function

Figure 8: Sales-Weighted Markup (Compustat)



Notes: Comparisons of the evolution of markups in the US manufacturing industry produced by my method and the Cobb-Douglas production function.

declines in the '70s and then starts to rise, but reaches only 1.4. In contrast, Cobb-Douglas markup estimates indicate a steady rise in markups from the '80s to today, reaching 1.6.

This result is consistent with other evidence on markups from manufacturing industries. For example, Hsieh and Rossi-Hansberg (2019) find that in manufacturing, concentration has fallen rather than increased. Foster et al. (2021) find that the increase in the average sales-weighted markup declines across many industries when allowing for output elasticities that vary more flexibly. Some markup estimates from the demand approach also support this finding. For example, Grieco et al. (2021) estimate the change in markups in the US auto industry by estimating demand, and they find no increase in the average markups. Finally, it is important to note that these results are specific to the manufacturing industry and should not be extrapolated to the entire economy. The specific forces that affect market power likely vary by industry, and they should be analyzed separately.

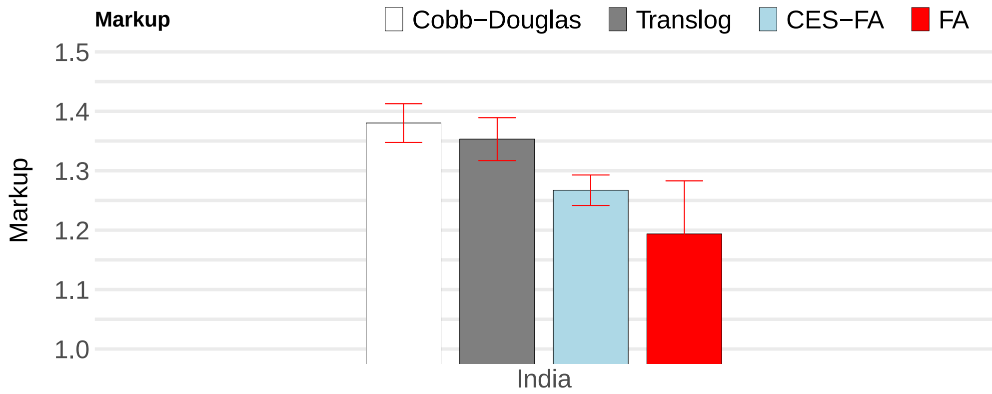
9 Robustness Checks and Extensions

This section discusses robustness checks and extensions provided in Online Appendix.

9.1 Robustness Check: Quantity Production functions

Most production datasets include revenue, not output quantity. If there is imperfect competition and unobserved heterogeneity firm's residual demand function, output elasticities estimated from revenues may fail to identify markups (Flynn et al. (2019), Bond

Figure 9: Markup Estimates from Quantity Production Functions



Notes: Comparison of sales-weighted markups from seven Indian industries that produce homogeneous products.

et al. (2021)). In this section, I investigate the robustness of my results to estimating a quantity production function.

I use data from seven industries in India that produce homogeneous products. These products include Brick Tiles, Cotton Shirts, Biri Cigarettes, Black Tea, Parboiled Non-Basmati Rice, Boxes, and Raw Non-Basmati Rice. I construct a sample of plants that derive at least 75 percent of their revenue from one of these products. For these firms, I observe the quantity of output produced, the quantity of materials consumed, and their prices. I estimate the production functions and calculate markups using these inputs and output measures. The estimates from this specification are reported in Figure 9. Factor-augmenting production functions suggest lower markup estimates than the Hicks-neutral production functions, and using a parametric labor-augmenting production function reduces bias only partially. The results mirror my main findings in Section 8.2, suggesting that the comparison results are robust to estimating quantity production functions. The details of this estimation procedure, the list of industries, and additional results are reported in Online Appendix Section F.1.

9.2 Robustness Check: Measurement Error in Capital

A concern in production function estimation is measurement error in capital and not observing capacity utilization, which could be more severe in a nonparametric model. To mitigate these concerns, I perform two exercises in Supplemental Materials Section C.1. First, using a simulation exercise, I show that measurement error cannot explain my results. Second, assuming capital and energy are perfect complements, I recover capacity utilization and show that it does not change the results.

9.3 Extension: Heterogeneous Input Prices

My model assumes that input prices are homogeneous, a standard assumption, mostly because traditional production datasets lack information on input prices. However, input prices are increasingly available in more recent datasets. To accommodate this case, I develop an extension where firms face different input prices in Online Appendix Section B.1. This extension requires incorporating heterogeneous input prices into input demand functions and accounting for them when constructing the control variables.[36]

10 Conclusions

Production function estimation plays a critical role in many policy discussions, including misallocation of inputs, rise in market power, and welfare effects of trade. Given this prevalence, it is increasingly important that our production functions capture the important aspects of production technology and firm behavior. This paper takes a step in this direction by comprehensively analyzing production function estimation with labor-augmenting productivity and documenting its impact on estimated output elasticities and markups.

Methodologically, I introduce an identification and estimation framework for production functions with labor-augmenting and Hicks-neutral productivity. Unlike previous methods, the identification strategy does not rely on parametric restrictions or variations in input prices. Empirically, I show that ignoring labor-augmenting productivity and imposing parametric restrictions generate biased output elasticity and markup estimates. These biases are economically significant. The commonly used specifications (i) underestimate capital elasticity, (ii) overestimate labor elasticity, and (iii) generate an upward bias in both the level and growth of markups. The estimates also document substantial firm-level heterogeneity in the output elasticities.

Although I focused on labor-augmenting productivity to introduce a richer heterogeneity in firm production, there are other dimensions of production and firm heterogeneity that might be equally important. Some examples include market power in the input market, labor market frictions, quality differences in inputs and output, and flexibly incorporating a demand model in the production framework. I believe that the techniques developed in this paper can help address these dimensions and develop even richer production function frameworks.

[36] Another important extension would be to account for firm exit. However, the selection problem is particularly difficult to deal with in the presence of multi-dimensional heterogeneity. I leave this as future work.

A Other Identification Results

This section examines the identification of the other important features of the production function. In particular, I ask what can be identified from $(f_t, \bar{h}_t)$ and the first-order conditions. All proofs are provided in Online Appendix D.

Proposition A.1. *Labor-augmenting productivity, the output elasticity of capital, and the elasticity of substitutions are not identified from* $(f_t, \bar{h}_t, \theta_{it}^L, \theta_{it}^M)$.

The intuition behind this result is that the first-order conditions are only informative about the flexible inputs' output elasticities and do not help identify other features of the production function. To solve this problem, I next ask what further restrictions are required to identify the labor-augmenting productivity, the output elasticity of capital, and the elasticity of substitutions.

A potential solution to non-identification results in the previous section is imposing additional structure on the production function. In this section, I consider a slightly more restrictive production function and establish that the capital elasticity and labor-augmenting productivity are identified, but the elasticity of substitution is not identified. Consider the following production function:

$$y_{it} = f_t\big(K_{it}, h_t(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{A.1}$$

This model differs from the main model in that h does not take K_{it} as an argument. Since this is a special case, Proposition 2.1 applies to this production function with $\omega_{it}^L = \bar{r}_t(\tilde{M}_{it})$. Substituting this into Equation (A.1), I obtain the *reduced form* for the production function in Equation (A.1) as follows:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{A.2}$$

Since K_{it} appears as an argument of f but not of h, this model is more convenient for identifying the capital elasticity and ω_{it}^L than the main model. The next proposition shows how to identify these objects.

Proposition A.2. *If we replace the production function in Assumption 2.1 with Equation (A.1), the capital elasticity is identified and labor-augmenting productivity is identified up to scale from* $(f_{t1}, \bar{h}_t, \theta_{it}^L, \theta_{it}^M)$ *as:*

$$\theta_{it}^K = f_{t1}(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})), \quad \log(\omega_{it}^L) = \log(\bar{r}_t(\tilde{M}_{it})) = \int_{\underline{\tilde{M}}}^{\tilde{M}_{it}} b(\bar{M}_{it}) d\bar{M}_{it} + a, \tag{A.3}$$

where $b(\cdot)$ *is a function provided in the proof, which depends on* f_t, $\bar{h}_t$, *and the output elasticities of flexible inputs.* f_{tj} *is the derivative of* f_t *with respect to its* j*-th argument* *and a is an unknown constant.*

θ_{it}^K is identified under the additional restriction because ω_{it}^L is not a function of capital, implying that we can learn the capital elasticity from f_{t1}. Identification of ω_{it}^L relies on the idea that we can obtain information about the first derivatives of h from the output elasticities of flexible inputs. In the proof, I show that information on the derivatives of h from the first-order conditions can be mapped back to ω_{it}^L.

My final result states the non-identification of the elasticity of substitutions.

Proposition A.3. *Under the conditions of Proposition A.2, the elasticity of substitutions is not identified from* $(f_t, \bar{h}_t, \theta_{it}^L, \theta_{it}^M)$.

The first-order conditions are only informative about the first derivatives of the production function, whereas the elasticity of substitution depends on the second derivatives of the production function. Thus, we can identify the output elasticities but not the elasticity of substitution.

This result extends the impossibility theorem of Diamond et al. (1978) to a model with firm-level data. They show that if the production function is at the industry-level, the elasticity of substitution is not identified from time series without exogenous variation in input prices. My result is similar in spirit because my model does not assume exogenous variation in input prices. In Online Appendix Section B.1, I extend my model to have variation in input prices. In this extension, if prices are exogenous, the elasticity of substitutions can potentially be identified.

An important implication of using first-order conditions for identification is that the output elasticities can only be identified for values of $(L_{it}, \omega_{it}^L, M_{it})$ on the surface $\{(\omega_{it}^L, M_{it}) \mid \omega_{it}^L = \bar{r}(\tilde{M}_{it})\}$. That is, I can identify the elasticities only at the observed input values realized in equilibrium. As a result, it is not possible to conduct counterfactual exercises, such as keeping ω_{it}^L constant and asking how a change in inputs affects the output. Nevertheless, this is not an important limitation in practice because most applications of production functions require elasticities and productivity only for the firms observed in the data.

B Proofs

Proof of Proposition 2.1

This proof builds on a classic result by Shephard (1953). Throughout the proof, I assume that the standard properties of production functions are satisfied (Chambers (1988, p.9)), so that the cost function exists and Shephard's Lemma holds. I also drop the time subscripts from functions for notational simplicity.

Part (i)

The firm minimizes the cost of flexible inputs for the level of planned output, $\bar{Y}_{it}$, which can be written as:

$$\min_{L_{it},M_{it}} p_t^l L_{it} + p_t^m M_{it} \quad \text{s.t.} \quad \mathbb{E}\big[F_t\left(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\right)\exp(\omega_{it}^H)\exp(\epsilon_{it}) \mid \mathcal{I}_{it}\big] \geqslant \bar{Y}_{it}.$$

Since the information set includes $(\omega_{it}^L, \tilde{\omega}_{it}^H)$, we can write the firm's problem as:

$$\min_{L_{it},M_{it}} p_t^l L_{it} + p_t^m M_{it} \quad \text{s.t.} \quad F_t\left(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\right)\exp(\omega_{it}^H)\mathcal{E}_{it}(\mathcal{I}_{it}) \geqslant \tilde{Y}_{it}, \tag{B.1}$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}]$. One can reformulate this problem as another cost minimization, where the firm chooses the effective labor facing the quality-adjusted input prices. To write this, let $\bar{L}_{it} := \omega_{it}^L L_{it}$ denote the effective (quality-adjusted) labor and $\bar{p}_{it}^l := p_t^l/\omega_{it}^L$ denote the quality-adjusted price of labor. Therefore, the cost minimization problem in Equation (B.1) can be rewritten as

$$\min_{M_{it},\bar{L}_{it}} \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_t \quad \text{s.t.} \quad F_t\big(K_{it}, h_t(K_{it}, \bar{L}_{it}, M_{it})\big)\exp(\omega_{it}^H) \geqslant \tilde{Y}_{it}(\mathcal{I}_{it}), \tag{B.2}$$

where $\tilde{Y}_{it}(\mathcal{I}_{it}) := \bar{Y}_{it}/\mathcal{E}_{it}(\mathcal{I}_{it})$. This problems is equivalent to Equation (B.1) since the firm takes ω_{it}^L as given. For what follows, I suppress $(\mathcal{I}_{it})$ and keep it implicit in $\tilde{Y}_{it}$. I next derive the cost function from this problem. Letting $\bar{p}_{it} = (\bar{p}_{it}^l, p_t^m)$ denote the (quality-adjusted) input price vector, the cost function is:

$$\begin{aligned}
C_t(\tilde{Y}_{it}, K_{it}, &\omega_{it}^H, \bar{p}_{it}) \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : \tilde{Y}_{it} \leqslant F_t\big(K_{it}, h_t(K_{it}, \bar{L}_{it}, M_{it})\big)\exp(\omega_{it}^H)\Big\}, \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : F_t^{-1}(\tilde{Y}_{it}/\exp(\omega_{it}^H), K_{it}) \leqslant h_t(K_{it}, \bar{L}_{it}, M_{it})\Big\}, \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{F_t^{-1}(\tilde{Y}_{it}/\exp(\omega_{it}^H), K_{it})\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h_t\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
&= F_t^{-1}(\tilde{Y}_{it}/\exp(\omega_{it}^H), K_{it}) \min_{\bar{L}_{it},M_{it}} \Big\{\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h_t\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
&\equiv C_{1t}(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_{2t}(K_{it}, \bar{p}_{it}^l, p_t^m).
\end{aligned} \tag{B.3}$$

The second line follows by the assumption that $F_t(\cdot,\cdot)$ is strictly monotone in its second argument. The third line uses $h_t(K_{it}, F_t(\cdot)\bar{L}_{it}, F_t(\cdot)M_{it}) = F_t(\cdot)h_t(\tilde{Y}_{it}/\exp(\omega_{it}^H), K_{it})$ due to homogeneity of h and the linearity of cost functions. The fourth lines follows from the exogeneity of $(K_{it}, \tilde{Y}_{it}, \omega_{it}^H)$. In the last line, I define two new functions that characterize the cost function. By Shephard's Lemma, the firm's optimal demands for flexible inputs are given by the derivatives of the cost function with respect to the input prices. Using this, the ratio of materials to effective labor can be obtained as:

$$\frac{M_{it}}{\bar{L}_{it}} = \frac{\partial C_{2t}(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial p^m_t}{\partial C_{2t}(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial \bar{p}^l_{it}} \equiv \frac{C_m(K_{it}, \bar{p}^l_{it}, p^m_t)}{C_l(K_{it}, \bar{p}^l_{it}, p^m_t)},$$

which does not depend on $(\tilde{Y}_{it}, \omega^H_{it})$. Using $\bar{L}_{it} = L_{it}\omega^L_{it}$, M_{it}/L_{it} is given by:

$$\frac{M_{it}}{L_{it}} = \frac{C_m(K_{it}, \bar{p}^l_{it}, p^m_t)\omega^L_{it}}{C_l(K_{it}, \bar{p}^l_{it}, p^m_t)}.$$

This function depends only on capital, ω^L_{it}, and input prices. Hence

$$\tilde{M}_{it} \equiv r_t(K_{it}, \omega^L_{it}, p^m_t, p^l_t) \equiv r_t(K_{it}, \omega^L_{it}), \tag{B.4}$$

for some function $r_t(K_{it}, \omega^L_{it})$. This completes the first part of the proof.

Part (ii)

In the second part of the proof, I will show that

$$\partial r_t(K_{it}, \omega^L_{it})/\partial \omega^L_{it} > 0 \quad \text{for all } (K_{it}, \omega^L_{it}) \quad \text{or} \quad \partial r_t(K_{it}, \omega^L_{it})\ \partial \omega^L_{it} < 0 \quad \text{for all } (K_{it}, \omega^L_{it}).$$

By the properties of the cost function, $C_m(\cdot)$ and $C_l(\cdot)$ are homogenous of degree of zero with respect to input prices (Chambers (1988, p.64)), implying that input ratio can be written as a function of quality-adjusted labor and materials prices:

$$\tilde{M}_{it} \equiv \frac{\tilde{C}_m(K_{it}, \tilde{p}_{it})\omega^L_{it}}{\tilde{C}_l(K_{it}, \tilde{p}_{it})}, \tag{B.5}$$

where $\tilde{p}_{it} := \bar{p}^l_{it}/p^m_t$, $\tilde{C}_m(K_{it}, \tilde{p}_{it}) := C_m(K_{it}, \tilde{p}_{it}, 1)$, and $\tilde{C}_l(K_{it}, \tilde{p}_{it}) := C_l(K_{it}, \tilde{p}_{it}, 1)$. Taking the logarithm of Equation (B.5), I obtain

$$\log(\tilde{M}_{it}) = \log(\tilde{C}_l(K_{it}, \tilde{p}_{it})/\tilde{C}_m(K_{it}, \tilde{p}_{it})) + \log(\omega^L_{it}).$$

Taking the derivative of $\log(\tilde{M}_{it})$ with respect to $\log(\omega^L_{it})$, I obtain

$$\begin{aligned}
\frac{\partial \log(\tilde{M}_{it})}{\partial \log(\omega^L_{it})} &= \frac{\partial \log\left(\tilde{C}_l(K_{it}, \tilde{p}_{it})/\tilde{C}_m(K_{it}, \tilde{p}_{it})\right)}{\partial \log(\tilde{p}_{it})}\left(\frac{\partial \log(\tilde{p}_{it})}{\partial \log(\omega^L_{it})}\right) + 1, \\
&= \frac{\partial \log\left(\tilde{C}_l(K_{it}, \tilde{p}_{it})/\tilde{C}_m(K_{it}, \tilde{p}_{it})\right)}{\partial \log(\tilde{p}_{it})} + 1 \equiv -\sigma(K, \omega^L L, M) + 1,
\end{aligned}$$

where the last line follows by the fact that the elasticity of substitution between two inputs equals the negative derivative of the logarithm of input ratio with respect to the logarithm of input price ratio (Chambers (1988, p.94)). To see this, note that the elasticity of substitution between two inputs is defined as elasticity of input ratio with respect to marginal rate of technical substitution: $\sigma = \partial \log(X_1/X_2)/\partial \log(F_1/F_2)$. Since by cost minimization, we have $F_1/F_2 = p_1/p_2$, we obtain $\sigma = -\partial \log(X_1/X_2)/\partial \log(p_1/p_2)$. By Assumption 2.1(iv), $\sigma(K, \omega^L L, M) < 1$ or $\sigma(K, \omega^L L, M) > 1$. From this I conclude that the flexible input ratio is strictly monotone in ω^L_{it}.

Proof of Lemma 3.1

By Assumption 2.2, we have that $\omega_{it}^L \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H$. Substituting ω_{it}^L from Equation (3.1), I obtain

$$g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{B.6}$$

Since $g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$ is strictly monotone in u_{it}^1, Equation (B.6) implies

$$u_{it}^1 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{B.7}$$

By normalization, u_{it}^1 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$ and by timing assumption $(K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H) \in \mathcal{I}_{it-1}$. Thus, Equation (B.7) implies

$$u_{it}^1 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0,1).$$

$(\omega_{it-1}^L, \omega_{it-1}^H)$ are functions of W_{it-1} by Equations (2.5) and (2.6). Using this

$$\begin{aligned} u_{it}^1 \mid K_{it}, W_{it-1}, \tilde{r}_t(K_{it-1}, \tilde{M}_{it-1}), \tilde{s}_t(K_{it-1}, \tilde{M}_{it-1}, M_{it-1}) \sim \text{Uniform}(0,1), \\ u_{it}^1 \mid K_{it}, W_{it-1} \sim \text{Uniform}(0,1). \end{aligned}$$

Therefore, u_{it}^1 is uniformly distributed conditional on (K_{it}, W_{it-1}).

Proof of Lemma 3.2

By Assumption 2.2, we have that $(\omega_{it}^L, \omega_{it}^M) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H$. Using the representations of productivity shocks in Equations (3.1) and (3.5) yields

$$g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1), g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H.$$

Monotonicity of g_1 and g_2 with respect to their last arguments and Lemma D.1 in the Online Appendix imply

$$u_{it}^2 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1. \tag{B.8}$$

It follows from Equation (B.8), the fact that u_{it}^2 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$, and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ that

$$u_{it}^2 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{t-1}^H, u_{it}^1 \sim \text{Uniform}(0,1).$$

$(\omega_{it-1}^L, \omega_{it-1}^H)$ are functions of W_{it-1} by Equations (2.5) and (2.6). Using this

$$u_{it}^2 \mid K_{it}, W_{it-1}, u_{it}^1 \sim \text{Uniform}(0,1).$$

References

Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451.

Ackerberg, Daniel, Hahn, Jinyong, and Pan, Qingsong. 2022. "Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity Shocks". *Working Paper.*

Autor, David, Dorn, David, Katz, Lawrence F, Patterson, Christina, and Van Reenen, John. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms". *The Quarterly Journal of Economics* 135 (2): 645–709.

Balat, Jorge, Brambilla, Irene, and Sasaki, Yuya. 2019. "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports". *Working Paper.*

Bernard, Andrew B, Jensen, J Bradford, Redding, Stephen J, and Schott, Peter K. 2007. "Firms in International Trade". *Journal of Economic Perspectives* 21 (3): 105–130.

Blundell, Richard and Bond, Stephen. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340.

Bond, Steve, Hashemi, Arshia, Kaplan, Greg, and Zoch, Piotr. 2021. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data". *Journal of Monetary Economics* 121:1–14.

Chambers, Robert G. 1988. *Applied Production Analysis: A Dual Approach.* Cambridge University Press.

Chirinko, Robert S. 2008. "σ: The long and Short of it". *Journal of Macroeconomics* 30 (2): 671–686.

De Loecker, Jan. 2011. "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity". *Econometrica* 79 (5): 1407–1451.

De Loecker, Jan, Eeckhout, Jan, and Unger, Gabriel. 2020. "The Rise of Market Power and the Macroeconomic Implications". *The Quarterly Journal of Economics* 135 (2): 561–644.

De Loecker, Jan and Warzynski, Frederic. 2012. "Markups and Firm-Level Export Status". *American Economic Review* 102 (6): 2437–2471.

Diamond, Peter, McFadden, Daniel, and Rodriguez, Miguel. 1978. "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change". In *Contributions to Economic Analysis*, 2:125–147. Elsevier.

Doraszelski, Ulrich and Jaumandreu, Jordi. 2018. "Measuring the Bias of Technological Change". *Journal of Political Economy* 126 (3): 1027–1084.

— . 2013. "R&D and Productivity: Estimating Endogenous Productivity". *Review of Economic Studies* 80 (4): 1338–1383.

— . 2019. "Using Cost Minimization to Estimate Markups". *Working Paper.*

Ekeland, Ivar, Heckman, James J., and Nesheim, Lars. 2004. "Identification and Estimation of Hedonic Models". *Journal of Political Economy* 112 (1): 60–109.

Flynn, Zach, Gandhi, Amit, and Traina, James. 2019. "Measuring Markups with Production Data". *Working Paper.*

Foster, Lucia, Grim, Cheryl, Haltiwanger, John, and Wolf, Zoltan. 2017. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights". *NBER Working Paper w23666.*

Foster, Lucia, Haltiwanger, John, and Tuttle, Cody. 2021. "Rising Markups or Changing Technology". *Working Paper.*

Gandhi, Amit, Navarro, Salvador, and Rivers, David A. 2020. "On the Identification of Gross Output Production Functions". *Journal of Political Economy* 128 (8): 2973–3016.

Grieco, Paul L. E., Li, Shengyu, and Zhang, Hongsong. 2016. "Production Function Estimation with Unobserved Input Price Dispersion". *International Economic Review* 57 (2): 665–690.

Grieco, Paul LE, Murry, Charles, and Yurukoglu, Ali. 2021. "The Evolution of Market Power in the US Auto Industry". *NBER Working Paper, w29013.*

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

Holmes, Thomas J. and Schmitz, James A. 2010. "Competition and Productivity: A Review of Evidence". *Annual Review of Economics* 2 (1): 619–642.

Hsieh, Chang-Tai and Rossi-Hansberg, Esteban. 2019. *The Industrial Revolution in Services* NBER Working Paper w25968.

Imbens, Guido and Newey, Whitney. 2009. "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity". *Econometrica* 77 (5): 1481–1512.

Jaumandreu, Jordi. 2018. "Dangerous Shortcuts: Ignoring Marginal Cost Determinants in Markup Estimation". *Working Paper.*

Kasahara, Hiroyuki, Schrimpf, Paul, and Suzuki, Michio. 2015. "Identification and Estimation of Production Function with Unobserved Heterogeneity". *Working Paper.*

Kasy, Maximilian. 2011. "Identification in Triangular Systems Using Control Functions". *Econometric Theory* 27 (3): 663–671.

Kehrig, Matthias and Vincent, Nicolas. 2021. “The Micro-level Anatomy of the Labor Share Decline”. Publisher: Oxford University Press, *The Quarterly Journal of Economics* 136 (2): 1031–1087.

Keller, Wolfgang and Yeaple, Stephen R. 2009. “Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States”. *The Review of Economics and Statistics* 91 (4): 821–831.

Kumar, Krishna B, Rajan, Raghuram G, and Zingales, Luigi. 1999. “What Determines Firm Size?” *Working Paper w7208, NBER.*

Levinsohn, James and Petrin, Amil. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables”. *The Review of Economic Studies* 70 (2): 317–341.

Lewbel, Arthur. 2019. “The Identification Zoo: Meanings of Identification in Econometrics”. *Journal of Economic Literature* 57 (4): 835–903.

Li, Tong and Sasaki, Yuya. 2017. “Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function”. *arXiv:1711.10031 [econ].*

Nadiri, M. Ishaq. 1982. “Producers Theory”. In *Handbook of Mathematical Economics*, 2:431–490. Elsevier.

Oberfield, Ezra and Raval, Devesh. 2021. “Micro Data and Macro Technology”. Publisher: Wiley Online Library, *Econometrica* 89 (2): 703–732.

Olley, G. Steven and Pakes, Ariel. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64 (6): 1263.

Raval, Devesh. 2020. “Testing the Production Approach to Markup Estimation”. *Working Paper.*

Raval, Devesh R. 2019. “The Micro Elasticity of Substitution and Non-neutral Technology”. *The RAND Journal of Economics* 50 (1): 147–167.

Shephard, Ronald W. 1953. *Cost and Production Functions.* Princeton University Press.

Syverson, Chad. 2011. “What Determines Productivity?” *Journal of Economic Literature* 49 (2): 326–365.

Van Beveren, Ilke. 2012. “Total Factor Productivity Estimation: A Practical Review”. *Journal of Economic Surveys* 26 (1): 98–128.

Van Biesebroeck, Johannes. 2008. “The Sensitivity of Productivity Estimates: Revisiting Three Important Debates”. *Journal of Business & Economic Statistics* 26 (3): 311–328.

Van Reenen, John. 2018. “Increasing Differences between Firms: Market Power and the Macro-Economy”. *Working Paper.*

Zhang, Hongsong. 2019. “Non-neutral Technology, Firm Heterogeneity, and Labor Demand”. *Journal of Development Economics* 140:145–168.

Online Appendix to "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups"

Mert Demirer

July 24, 2022

A Data and Estimation

This section describes the datasets. The summary statistics for each dataset is presented in the Supplemental Materials Section A.

A.1 Chile

Chile data are from the Chilean Annual Census of Manufacturing, Encuesta Nacional Industrial Anual (ENIA), covering the years 1979 through 1996. This dataset includes all manufacturing plants with at least ten employees. I restrict my sample to the industries with more than 250 firms per year. I drop observations at the bottom and top 2% of the distribution of revenue share of labor or revenue share of materials or combined flexible input for each industry to remove outliers. I report each industry's share in manufacturing in terms of sales and the number of plants for the first, last, and midpoint year of the sample in the Supplemental Materials Section A. The last row, labeled as "other industries", provides information about the industries excluded from the sample. After sample restrictions, five industries remain, covering around 30% of the manufacturing sector.

A.2 Colombia

The data for Colombia are from the annual Colombian Manufacturing census provided by the Departamento Administrativo Nacional de Estadistica, covering the years 1981 through 1991. This dataset contains all manufacturing plants with at least ten employees. I restrict my sample to the industries with more than 250 firms per year on average and follow the same steps as in the construction of the Chile Data to remove outliers. The number of industries after applying the sample restrictions is nine, larger than other datasets. The sample covers around 55% of the manufacturing sector.

A.3 India

The Indian data was collected by the Ministry of Statistics and Programme Implementation through the Annual Survey of Industries (ASI), which covers all factories with at least ten workers and use electricity, or those that do not use electricity but have at least 20 workers. The factories are divided into two categories: a census sector and a sample sector. The census sector consists of all large factories and all factories in states classified as industrially backward by the government. From 2001 to 2005, a large factory is defined as one with 200 or more workers, whereas from 2006 onward, it was changed to one with 100 or more workers. All factories in the census sector are surveyed every year. The remaining factories constitute the sample sector, from which a random sample is surveyed each year. India uses the National Industrial Classification (NIC) to classify manufacturing establishments, a similar industrial classifications to those used in other countries. The industry definition has changed multiple times over the sample period. I follow Allcott et al. (2016) to create a consistent industry definition at the NIC 87 level. For sample restrictions and data cleaning, I first follow Allcott et al. (2016). Then, I restrict my sample to the Census sample to be able to follow firms over time. My final sample includes industries with more than 250 firms per year. I follow the same steps as in the Chilian Data to remove outliers.

A.4 Turkey

Turkey data are provided by the Turkish Statistical Institute (TurkStat), which collects plant-level data for the manufacturing sector. Periodically, Turkstat conducts the Census of Industry and Business Establishments (CIBE), which collects information on all manufacturing plants in Turkey. In addition, TurkStat conducts the Annual Surveys of Manufacturing Industries (ASMI) that covers all establishments with at least ten employees. The set of establishments used for ASMI is obtained from the CIBE. I use a sample covering a period from 1983 to 2000. The data include gross revenue, investment, the book value of capital, materials expenditures, and the number of production and administrative workers. For variable construction, I follow Taymaz and Yilmaz (2015). I restrict my sample to industries with more than 250 firms per year on average and private establishments. I follow the same procedure as in Chilian Data to remove outliers.

A.5 Compustat

Compustat is obtained from Standard and Poor's Compustat North America database and covers the period from 1961 to 2012. Data after 2012 are available, but due to the unavailability of some deflators used in variable construction, I restrict my sample from 1961 to 2012. Since Compustat is compiled from firms' financial statements, it requires more extensive data cleaning than the other datasets. First, I drop the firms that are not incorporated in the US. Then, as is standard in the literature, I drop financial and utility firms with industry codes between 4900-4999 and 6000-6999. I also remove the firms with negative or nonzero sales, employment, cogs, xsga, and those with less than ten employees and firms that do not report an industry code. Finally, the sample is restricted to only manufacturing firms operating in industries with the NAICS codes 31, 32, and 33. To construct the inputs and output, I follow Keller and Yeaple (2009), who explain the procedure in detail in their Appendix B, page 831. Unlike other datasets, which are plant-level, Compustat is firm-level and contains only public firms. Also, the industry classification is based on NAICS, and industries are defined at the 2-digit level. For Compustat, I drop observations at the bottom and top 1%, instead of 2%, of the distribution to preserve the sample size.

A.6 Variable Construction

Labor: For Chile, Colombia, Turkey, and the US, I use the number of production workers as my measure of labor. For India, I use the total number of days worked. For the labor's revenue share, I use the sum of total salaries and benefits divided by total sales during the year.

Materials: For Chile, Colombia, India, and Turkey, I calculate materials cost as total spending on materials, with an adjustment for inventories by adding the difference between the end year and beginning year value of inventories. I deflate the nominal value of total material cost using the industry-level intermediate input price index. For Compustat, materials are calculated as the deflated cost of goods sold plus administrative and selling expenses less depreciation and wage expenditures. Materials' revenue share is materials cost divided by total sales during the year.

Capital: For Turkey, the capital stock series is constructed using the perpetual inventory method where investment in new capital is combined with deflated capital from period $t-1$ to form the capital level in period t. For Compustat, capital is calculated as the value of property, plant, and equipment and the net of depreciation deflated from the BEA satellite accounts. For India, the book value of capital is deflated by an implied national deflator calculated "Table 13: Sector-wise Gross Capital Formation" from the Reserve Bank of India's Handbook of Statistics on the Indian Economy. For Chile and Colombia, I follow Raval (2020).

Output: The output is calculated as deflated sales. For Compustat, it is net sales from the Industrial data file. For other countries, sales are total production value, plus the difference between the end year and beginning year value of finished goods inventories.

A.7 Estimation Algorithm

This section details the estimation algorithm. Apply the data cleaning and variable construction steps described in Subsection A.1 and denote the resulting sample by A. Remove the observations for which the previous period's inputs are missing and denote the resulting sample by B. Take the subset of observations in B that fall into the rolling window τ and denote this sample by B_{tau}. Estimate the control variable u_{it}^2 for each $it \in B_{tau}$ as follows. Construct a grid that partitions the support of M_{it} into 500 points so that each bin contains the same number of observations. Denote the set of these points by Q. For each $q \in Q$, estimate

$$\text{Prob}(M_{it} \leqslant q \mid K_{it} = k, W_{it-1} = w, u_{it}^1 = u) \equiv s_t(q, k, w, u)$$

using a logit model that includes third-degree polynomials. Then for each $it \in B_r$, obtain an estimate $\widehat{u}_{it}^2$ for $u_{it}^2 = s_t(M_{it}, K_{it}, W_{it}, u_{it}^1)$ by linearly interpolating the closest two points in Q to M_{it}. From this procedure obtain $\widehat{u}_{it}^2$ for all $it \in B_r$. For production function estimation, first approximate the logarithm of $\bar{h}_t$ by using third-degree polynomials

$$\log(\widehat{\bar{h}}_t(\tilde{M}_{it})) = a_{1t} + a_{2t}\tilde{m}_{it} + a_{3t}\tilde{m}_{it}^2 + a_{4t}\tilde{m}_{it}^3, \tag{A.1}$$

where $\{a_{jt}\}_{j=1}^4$ are the parameters of the polynomial approximation. Set $a_{t1} = 0$ to impose the normalization for $\widehat{\bar{h}}_t$ described in Section 4. Let $V_{it} := L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})$. Approximate the production function as

$$\begin{aligned}\widehat{f}_t(K_{it}, L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})) = b_{1t} + b_{2t}k_{it} + b_{3t}k_{it}^2 + b_{4t}k_{it}^3 + b_{5t}v_{it} + b_{6t}v_{it}^2 + b_{7t}v_{it}^3 \\ + b_{8t}k_{it}^2 v_{it} + b_{9t}k_{it}v_{it}^2 + b_{10t}k_{it}v_{it}\end{aligned} \tag{A.2}$$

where $\{b_{jt}\}_{j=1}^{10}$ are the parameters of the polynomial approximation. Approximate the control functions $c_{2t}(\cdot)$ and $c_{3t}(\cdot)$ using third-degree polynomials similarly. For given values $\{a_{jt}\}_{j=1}^{4}$, $\{b_{jt}\}_{j=1}^{10}$, $\widehat{c}_{2t}(\cdot)$ and $\widehat{c}_{3t}(\cdot)$ construct the objective function presented in Section 5.1.1. Minimize this objective function to estimate the production function using the following two step procedure. In the inner loop, for a candidate value of the parameter vector $\{a_{jt}\}_{j=1}^{4}$, estimate $\{b_{jt}\}_{j=1}^{10}$, $\widehat{c}_{2t}(\cdot)$ and $\widehat{c}_{3t}(\cdot)$ using the least squares regression. In the outer loop use an optimization routine to estimate $\{a_j\}_{j=1}^{4}$. Minimizing the objective function requires optimization over four parameters, so it is not computationally intensive. After estimating the production function parameters, the next step is elasticity and markups estimation.

Take the observations that are in the midpoint of the rolling window period in sample A and denote that sample by A_c. For each $it \in A_c$, calculate output elasticities and markups as follows. Obtain the estimates of f_t and $\bar{h}_t$ from the estimates of the parameters $\{a_{jt}\}_{j=1}^{4}$ and $\{b_{jt}\}_{j=1}^{10}$ in Equations (A.1) and (A.2). First, using the estimates of f_t and $\bar{h}_t$, calculate the output elasticity of capital and the sum of the materials and labor elasticities, given in Equations (4.5) and (A.3) by taking numerical derivatives. Then given an estimate of θ_{it}^{V} and revenue shares of materials and labor, use Equations (4.6) to estimate the output elasticity of labor and materials. Finally, estimate markups from $\widehat{\theta}_{it}^{V}$ and the revenue share of flexible input. For standard errors, resample firms with replacement from the sample A, then repeat the estimation procedure above. I use the same procedure to estimate the CES and Nested CES models after applying the appropriate parametric restrictions.

A.8 Blundell and Bond Method Estimation

This subsection describes the Blundell and Bond (2000) estimation method that is used to estimate the Hicks-neutral production functions. To apply the Blundell and Bond (2000) dynamic panel estimation model, I assume that productivity shock follows an AR(1) process: $\omega_{it}^{H} = \rho\omega_{it-1}^{H} + v_{it}$. Using this assumption and taking the first difference of the production function, one can obtain

$$\begin{aligned} y_{it} - \rho y_{it-1} = &\beta_k(k_{it} - \rho k_{it-1}) + \beta_l(l_{it} - \rho l_{it-1}) + \beta_m(m_{it} - \rho m_{it-1}) + \\ &\omega_{it}^{H} - \rho\omega_{it-1}^{H} + \epsilon_{it} - \epsilon_{it-1}. \end{aligned}$$

The composite error term $\nu_{it} := \omega_{it}^H - \rho\omega_{it-1}^H + \epsilon_{it} - \epsilon_{it-1}$ is orthogonal to firm's information set at $t-1$, that is, $\mathbb{E}[\nu_{it} \mid \mathcal{I}_{it-1}] = 0$. To make use of this orthogonality condition, construct a moment function as:

$$\gamma(\beta_k, \beta_l, \beta_m, \rho) = y_{it} - \rho y_{it-1} - \beta_k(k_{it} - \rho k_{it-1}) + \beta_l(l_{it} - \rho l_{it-1}) + \beta_m(m_{it} - \rho m_{it-1}).$$

The moment conditions to estimate the parameters are given by:

$$\mathbb{E}\Big[\gamma(\beta_k, \beta_l, \beta_m, \rho)(k_{it}, k_{it-1}, l_{it-1}, m_{it-1})'\Big] = 0. \tag{A.3}$$

The translog production function is estimated similarly after imposing the following translog functional form.

$$y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + \beta_4 k_{it}^2 + \beta_5 l_{it}^2 + \beta_6 m_{it}^2 + \beta_7 k_{it} m_{it} + \beta_8 l_{it} m_{it} + \beta_9 l_{it} k_{it} + \epsilon_{it}.$$

B Extensions

This section presents three extensions. All proofs are given in Section D.

B.1 Heterogeneous Input Prices

This extension assumes that input prices vary across firms. I denote labor and materials prices by p_{it}^l and p_{it}^m, respectively, and use $\bar{p}_{it}$ to denote the input price vector, so $\bar{p}_{it} := (p_{it}^l, p_{it}^m)$. I also use $p_{it}^{l/m} := (p_{it}^l / p_{it}^m)$ to denote the input price ratio. Differently from the main model, W_{it} now includes input prices, so $W_{it} = (K_{it}, L_{it}, M_{it}, \bar{p}_{it})$. I first modify the Markov and monotonicity assumptions to incorporate the input prices into the model. With variation in input prices, Assumptions 2.1 is replaced by the following assumption.

Assumption B.1. *The distribution of productivity shocks and input prices obey:*

$$P(\omega_{it}^L, \omega_{it}^H, \bar{p}_{it} \mid \mathcal{I}_{it-1}) = P(\omega_{it}^L, \omega_{it}^H, \bar{p}_{it} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}).$$

This assumption states that prices and productivity shocks jointly follow an exogenous first-order Markov process. Importantly, this assumption allows for a correlation between productivity shocks and input prices. Since we expect that more productive workers, as represented by higher ω_{it}^L, earn higher wages, the correlation between input prices and productivity is important to accommodate. It is possible to obtain stronger identification results with some additional structure, such as independence between the innovations to productivity shocks and input prices. However, I make minimal assumptions to develop a general framework.

Assumption B.2. *The firm's materials demand is given by* $M_{it} = s_t(K_{it}, \omega_{it}^L, \omega_{it}^H, \bar{p}_{it})$, *and* $s_t(K_{it}, \omega_{it}^L, \omega_{it}^H, \bar{p}_{it})$ *is strictly increasing in* ω_{it}^H.

This assumption is a natural extension of Assumption 2.4, as the materials demand now depends on both input prices. This section maintains the other assumptions in the model, namely Assumptions 2.2 and 2.3, and states the following proposition.

Proposition B.1.

(i) *Under Assumptions 2.2(i-iv) and with heterogeneity in input prices, the flexible input ratio, denoted by* $\tilde{M}_{it} = M_{it}/L_{it}$, *depends on* K_{it}, ω_{it}^L *and* $p_{it}^{l/m}$

$$\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^L, p_{it}^{l/m}). \tag{B.1}$$

(ii) *Under the Assumption 2.2(v),* $r_t(K_{it}, \omega_{it}^L, p_{it}^{l/m})$ *is strictly monotone in* ω_{it}^L.

The proof of this Proposition is a straightforward extension of the proof of Proposition 2.1, and therefore, is omitted. Compared to Proposition 2.1, the only difference is that the flexible input ratio depends also on the input price ratio. Note that the ratio of prices, not the price vector, affects the flexible input ratio due to the properties of cost functions. This property would reduce the dimension of the control variables. With this result, ω_{it}^L is invertible once we condition on the input price ratio and capital. By inverting the Equation (B.1), I can write productivity shocks as:

$$\omega_{it}^L = \bar{r}_t(K_{it}, \tilde{M}_{it}, p_{it}^{l/m}), \qquad \omega_{it}^H = \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}, \bar{p}_{it}). \tag{B.2}$$

The derivation of the control variables proceed similarly as in Section 3. I first use the Skorokhod's representation of ω_{it}^L to write:

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1), \quad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m} \sim \text{Uni.}(0,1). \tag{B.3}$$

Unlike Equation (3.1), I include the ratio of current and past input prices in $g_1(\cdot)$. This is because, as stated in Proposition B.1, the optimal input ratio depends on the ratio of input prices. Using Equations (B.1), (B.2) and (B.3), we have

$$\tilde{M}_{it} = r_t\big(K_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1), \bar{p}_{it}\big) \equiv \tilde{r}_t\left(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1\right).$$

where $\tilde{r}_t(\cdot)$ is strictly monotone in u_{it}^1.

Lemma B.1. *Under Assumptions 2.3, B.1, and B.2,* u_{it}^1 *is jointly independent of* $(K_{it}, W_{it-1}, \bar{p}_{it})$*:*

With this lemma and monotonicity, u_{it}^1 can be identified as:

$$u_{it}^1 = F_{\tilde{M}_{it}|K_{it},W_{it-1},\bar{p}_{it}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}, \bar{p}_{it}).$$

Next, we can use Equations (B.2) and (B.3) to write ω_{it}^L as: $\omega_{it}^L \equiv c_{1t}(W_{it-1}, p_{it}^{l/m}, u_{it}^1)$. Note that unlike the main model, the CDF for u_{it}^1 is conditional on the price vector $\bar{p}_{it}$ and control function includes the price ratio $p_{it}^{l/m}$ since prices are endogenous. The derivation for the control function for ω_{it}^H is similar to that of ω_{it}^L. We use

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, \sim \text{Uni.}(0,1).$$

Following the same steps in Equation (3.2) in Section 3, materials demand function can be obtained as $M_{it} \equiv \tilde{s}_t\big(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\big)$, where $\tilde{s}_t(\cdot)$ is strictly monotone in u_{it}^2.

Lemma B.2. *Under Assumptions 2.3, B.1 and B.2, u_{it}^2 is jointly independent of $(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1)$:*

By this lemma and monotonicity of M_{it} in u_{it}^2, we can recover u_{it}^2 as

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},\bar{p}_{it},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1),$$

and the control function is given by $\omega_{it}^H \equiv c_{2t}\left(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\right)$.

It follows that with variation in input prices control functions become $\omega_{it}^L = c_{1t}(W_{it-1}, p_{it}^{l/m}, u_{it}^1)$ and $\omega_{it}^H = c_{2t}\left(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\right)$. The main difference is that I must condition the current and previous period's input prices to derive control functions. The rest of the identification and estimation results remain the same with these modifications in the control variables.

C Application to Parametric Production Functions

The nonparametric approach I propose accommodates five models that are nested within each other. I list these models below, from least restrictive to most, to provide a complete picture.

$$y_{it} = f_t\big(K_{it}, h_t(K_{it}, \omega^L_{it} L_{it}, M_{it})\big) + \omega^H_{it} + \epsilon_{it} \tag{Homo. Sep.}$$

$$y_{it} = f_t\big(K_{it}, h_t(\omega^L_{it} L_{it}, M_{it})\big) + \omega^H_{it} + \epsilon_{it} \tag{Weak Homo. Sep.}$$

$$y_{it} = v k_{it} + f_t\big(\tilde{L}_{it} h_t(\omega^L_{it}, \tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it} \tag{Homogeneous}$$

$$y_{it} = \frac{v}{\sigma} \log\Big(\beta_k K^\sigma_{it} + (1-\beta_k)\big(\beta_l \big[\omega^L_{it} L_{it}\big]^{\sigma_1} + (1-\beta_l) M^{\sigma_1}_{it}\big)^{\frac{\sigma}{\sigma_1}}\Big) + \omega^H_{it} + \epsilon_{it} \tag{Nested CES}$$

$$y_{it} = \frac{v}{\sigma} \log\Big(\beta_k K^\sigma_{it} + \beta_l \big(\omega^L_{it} L_{it}\big)^\sigma + (1-\beta_l-\beta_m) M^\sigma_{it}\Big) + \omega^H_{it} + \epsilon_{it} \tag{CES}$$

This section describes how to apply the my method to parametric functional forms and how to impose a returns to scale restriction.

C.1 Imposing Returns to Scale Restrictions

My model can accommodate a return to scale restriction on the production function. Restricting the return to scale to an unknown constant v, the production function takes the form

$$y_{it} = v_t k_{it} + f_t\big(1, \tilde{L}_{it} h_t(\omega^L_{it}, \tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}, \tag{C.1}$$

where $\tilde{L}_{it} = L_{it}/K_{it}$. The reduced form of this production function is

$$y_{it} = v_t k_{it} + \tilde{f}_t\big(\tilde{L}_{it} \bar{h}_t(\tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}, \tag{C.2}$$

where $\tilde{f}_t = f_t\big(1, \tilde{L}_{it} \bar{h}_t(\tilde{M}_{it})\big)$. My identification results apply to this model after making the functional form restrictions.

C.2 Nested CES Production Function

This section studies the identification of the Nested CES production function in Example 2. We maintain the assumptions in Section 2.2. The logarithm of this production function is given by:

$$y_{it} = \frac{v}{\sigma} \log\Big(\beta_k K^\sigma_{it} + (1-\beta_k)\left(\beta_l \big[\omega^L_{it} L_{it}\big]^{\sigma_1} + (1-\beta_l) M^{\sigma_1}_{it}\right)^{\sigma/\sigma_1}\Big) + \omega^H_{it} + \epsilon_{it}.$$

Using the homotheticity of Nested CES, we can rewrite the production function:

$$y_{it} = v m_{it} + \frac{v}{\sigma} \log\left(\beta_k \tilde{K}^\sigma_{it} + (1-\beta_k)\big(\beta_l \big[\omega^L_{it} \tilde{L}_{it}\big]^{\sigma_1} + (1-\beta_l)\big)^{\sigma/\sigma_1}\right) + \omega^H_{it} + \epsilon_{it}$$

where $\tilde{K}_{it} := K_{it}/M_{it}$ and $\tilde{L}_{it} := L_{it}/M_{it}$ and $m_{it} := \log(M_{it})$. The FOCs of cost minimization imply that $\omega^L_{it} = \gamma \tilde{L}^{(1-\sigma_1)/\sigma_1}$, where is γ is a constant that depends on input prices and model parameters. Substituting this into the production function we obtain

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left(\beta_k \tilde{K}_{it}^{\sigma} + (1-\beta_k)\gamma_1\big(\tilde{L}_{it} + \gamma_2\big)^{\sigma/\sigma_1}\right) + \omega_{it}^{H} + \epsilon_{it},$$

where γ_1 and γ_2 are constants that depend on the model parameters. Note that ω_{it}^{L} disappeared from the model. This is the parametric analog of my nonparametric inversion result in Proposition 2.1. The parameters of the Nested CES functional form can be estimated using the control functions I develop with the following estimating equation:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left(\beta_k \tilde{K}_{it}^{\sigma} + (1-\beta_k)\gamma_1\big(\tilde{L}_{it} + (1-\beta_l)\big)^{\sigma/\sigma_1}\right) + c_t(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it}.$$

One can estimate the model parameters using the objective function in Equation (A.1). With some algebra, it is possible to show that the sum of the flexible input elasticites are identified from the model parameters as:

$$\theta_{it}^{V} = v\frac{(1-\beta_k)\gamma_1 x^{\sigma}}{(1-\beta_k)\gamma_1 x^{\sigma} + \beta_k K_{it}^{\sigma}}$$

where $x_{it} = M_{it}(\tilde{L}_{it} + \gamma_2)^{1/\sigma_1}$. Note that $(1-\beta_k)\gamma_1$ and β_k are not separately identified in the production function, but their ratio is identified. Therefore, θ_{it}^{V} is identified. This is the parametric analog of the non-identification result in Proposition 4.1, where γ_1 and β_k are not identified separately but the variable input elasticity is identified. Labor and materials elasticities are identified from θ_{it}^{V} using ratio of revenue shares as described in the paper. Finally, the output elasticity of capital is identified as

$$\theta_{it}^{K} = v\frac{\beta_k K_{it}^{\sigma}}{(1-\beta_k)\gamma x_{it}^{\sigma} + \beta_k K_{it}^{\sigma}}.$$

C.3 CES Production Function

In this section, I consider the CES production function: Using homotheticity we can divide all inputs by M_{it} and write the log production function as:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \beta_l\big[\omega_{it}^{L}\tilde{L}_{it}\big]^{\sigma} + \beta_m\right) + \omega_{it}^{H} + \epsilon_{it}.$$

The FOCs of cost minimization imply that $\omega_{it}^{L} = \gamma\tilde{L}_{it}^{(1-\sigma)/\sigma}$, where γ is a constant that depends on input prices and model parameters. Substituting this into the production function, we can obtain the following:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \gamma_1(\tilde{L}_{it} + \gamma_2)\right) + \omega_{it}^{H} + \epsilon_{it},$$

where $\gamma_1 := \gamma\beta_l$ and $\gamma_2 := \beta_m/\gamma\beta_l$. Note that ω_{it}^L disappeared from the model. The model parameters can be estimated using the control functions I developed. The estimation equation is:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \gamma_1(\tilde{L}_{it}+\gamma_2)\right) + c(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it},$$

with the same objective function as in Section 5. As in the Nested CES model, one can again show that the sum of the flexible input elasticites is identified from the model parameters as:

$$\theta_{it}^V = v\frac{\gamma_1 x_{it}^{\sigma}}{\gamma_1 x_{it}^{\sigma} + (1-\beta_l-\beta_m)K_{it}^{\sigma}},$$

where $x_{it} = M_{it}(\tilde{L}_{it}+\gamma_2)$. Note that $(1-\beta_l-\beta_m)$ and γ_1 are not separately identified from this production function but the ratio is identified. Since θ_{it}^V depends only on the ratio, the sum elasticity is identified. Labor and materials elasticities are identified from θ_{it}^V using the revenue shares as described in the paper. The capital elasticity is identified as

$$\theta_{it}^K = v\frac{(1-\beta_l-\beta_m)K_{it}^{\sigma}}{\gamma_1 x_{it}^{\sigma} + (1-\beta_l-\beta_m)K_{it}^{\sigma}}.$$

C.4 Cobb-Douglas Production Function

The control variable approach of this paper can be applied to Hicks-neutral production functions. This section presents this application and compares it with the proxy variable approach. Since the literature has shown that the gross Cobb-Douglas production function with two flexible inputs is not identified, I use the value-added production function studied in Ackerberg et al. (2015):

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it}^H + \epsilon_{it}$$

I consider the standard assumptions in the proxy variable literature: (i) the productivity shock follow an exogenous first-order Markov process $P(\omega_{it}^H \mid \mathcal{I}_{it-1}) = P(\omega_{it}^H \mid \omega_{it-1}^H)$, (ii) capital is a dynamic input, and labor is static input optimized every period, and (iii) the firm's intermediate input decision is given by $m_{it} = s(k_{it}, \omega_{it}^H)$, which is strictly increasing in ω_{it}^H. Using these assumptions, I construct a control variable using the steps in Section 3 as follows:

$$\omega_{it}^H = g(\omega_{it-1}^H, u_{it}) \qquad u_{it} \mid \omega_{it-1}^H \sim \text{Uniform}(0,1), \tag{C.3}$$

where $g(\omega^H_{it-1}, u_{it})$ is strictly increasing in u_{it}. By the Markov Assumption we have that $\omega^H_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^H_{it-1}$. Substituting ω^H_{it} using Equation (C.3) we have $g(\omega^H_{it-1}, u_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^H_{it-1}$, which implies that $u_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^H_{it-1}$. Using this,

$$m_{it} = s(k_{it}, \omega_{it}) = s(k_{it}, g(\omega_{it-1}, u_{it})) \equiv \tilde{s}(k_{it}, k_{it-1}, m_{it-1}, u_{it}).$$

Note that $s(k_{it}, \omega_{it})$ is strictly increasing in ω_{it} and $g(\omega_{it-1}, u_{it})$ is also strictly increasing in u_{it} by construction. Therefore, $\tilde{s}$ is strictly increasing in u_{it}. It follows from Lemma 3.1 that

$$u_{it} \mid k_{it}, m_{it-1}, k_{it-1} \sim \text{Uniform}(0,1). \tag{C.4}$$

Using this, we can recover u_{it} as the conditional CDF of m_{it}: $u_{it} = F_{m_{it}}(m_{it} \mid k_{it}, m_{it-1}, k_{it-1})$. This suggests that we can use a function of $(m_{it-1}, k_{it-1}, u_{it})$ to proxy ω^H_{it}:

$$\omega^H_{it} = g(\omega^H_{it-1}, u_{it}) = g(s^{-1}(k_{it-1}, m_{it-1}), u_{it}) \equiv c(m_{it-1}, k_{it-1}, u_{it}).$$

With this result, I obtain a partially linear model:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c(m_{it-1}, k_{it-1}, u_{it}) + \epsilon_{it}, \tag{C.5}$$

with $\mathbb{E}[\epsilon_{it} \mid I_{it}] = 0$. However, we can develop other moment restrictions using the first-order Markov property of ω^H_{it} as standard in the literature (Ackerberg et al. (2015)). In particular, using $\omega^H_{it} = c_2(\omega^H_{it-1}) + \xi_{it}$ with $\mathbb{E}[\xi_{it} \mid I_{it-1}] = 0$,

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_2(m_{it-1}, k_{it-1}) + \xi_{it} + \epsilon_{it}, \tag{C.6}$$

with $\mathbb{E}[\xi_{it} \mid I_{it-1}] = 0$. We can estimate the parameters (β_k, β_l) and unknown functions $c_1(\cdot), c_2(\cdot)$ in Equation (C.5) and (C.6) using the following moments.

$$\mathbb{E}[\epsilon_{it} \mid k_{it}, l_{it}, m_{it}, m_{it-1}, k_{it-1}, u_{it}] = 0, \qquad \mathbb{E}[\xi_{it} + \epsilon_{it} \mid k_{it}, m_{it-1}, k_{it-1}] = 0$$

In this estimation, the parameters might be identified even if labor is a flexible input and is written as $l_{it} = l(\omega_{it}, k_{it})$. The main distinction between this approach and proxy variable approach is the conditioning variables in the estimation. While the proxy variable approach conditions on an unknown function of (k_{it}, m_{it}), my method conditions on u_{it}, a known function of (k_{it}, m_{it}). Conditional on the control variable u_{it}, there might still be variation in l_{it} linearly independent of k_{it}, which can identify the production function. To see this, if labor is flexible, we can write it as $l_{it} = l(k_{it}, \omega^H_{it}) = l\big(k(k_{it-1}, \omega_{it-1}, \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})\big) =$

$l\big(k(k_{it-1}, s^{-1}(k_{it-1}, m_{it-1}), \nu_{it-1})), c(m_{it-1}, k_{it-1}, u_{it})\big)$, which gives $l_{it} =: \tilde{l}(k_{it-1}, m_{it-1}, u_{it}, \nu_{it-1})$, where ν_{it-1} corresponds to a vector of random variables that affects the firm's investment decision, such as investment prices and heterogeneous beliefs. So, in this example, conditional $(k_{it-1}, m_{it-1}, u_{it})$, ν_{it-1} could generate variation in labor independently of a linear function of capital depending on the data generating process for $l_{it} = l(\omega_{it}, k_{it})$. It is important to note that this example is valid only for Cobb-Douglas production function.

D Additional Proofs

Lemma D.1. *Suppose x, y and z are scalar and continuous random variables with a joint probability density function given by $f(x, y, z)$. Assume that (x, y) are jointly independent from z. Then x and z are independent conditional on y.*

Proof. Let $f(x \mid y)$ denote the conditional probability density function of x given y. Independence assumption implies that $f(x, y, z) = f(x, y)f(z)$. To achieve the desired result, I need to show that $f(x, z \mid y) = f(x \mid y)f(z \mid y)$. Using Bayes's rule for continuous random variables I obtain

$$\begin{aligned} f(x, z \mid y) &= \frac{f(x, y, z)}{f(y)} = \frac{f(x, y)f(z)}{f(y)} = \frac{f(x \mid y)f(y)f(z)}{f(y)} = f(x \mid y)f(z), \\ &= f(x \mid y)f(z \mid y), \end{aligned}$$

where in the last line $f(z \mid y) = f(z)$ follows by the independence assumption. □

Proof of Proposition A.1

For this proof, I drop the time subscripts from all functions for notational simplicity. One can assume that all functions are indexed by t. The proof consists of two parts. First, I will show that two different set of structural functions, lead to same $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. Then, I will show that labor-augmenting productivity, the output elasticity of capital, and the elasticity of substitutions depend on the structural functions h and $\bar{r}$, and, therefore, can not be identified. Looking at the elasticities first, θ^L_{it} and θ^M_{it} depend on the production function in the following way:

$$\theta^L_{it} = f_2 h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}(K_{it}, \tilde{M}_{it})L_{it}, \tag{D.1}$$

$$\theta^M_{it} = f_2 h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it}, \tag{D.2}$$

where arguments of the derivatives of f are omitted. Next, the derivatives of the reduced form function $\bar{h}$ can be written as:

$$\bar{h}_2(K_{it}, \tilde{M}_{it}) = h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}),$$
$$\bar{h}_1(K_{it}, \tilde{M}_{it}) = h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_1(K_{it}, \tilde{M}_{it}).$$

So the right-hand side of these equations are identified from $\bar{h}$ and the output elasticities θ^L_{it} and θ^M_{it} are identified from $\bar{h}$ the revenue shares. To give an intuition for the identification problem note that we have four equations, but structural functions $h(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$ and $\bar{r}(K_{it}, \tilde{M}_{it})$ include five dimensions in total. This suggests that it might not be possible to identify h and $\bar{r}$ from $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. More formally, consider two sets of functions $(h_1, h_2, h_3, \bar{r}_1, \bar{r}_2)$ and $(h'_1, h'_2, h'_3, \bar{r}'_1, \bar{r}'_2)$ such that

$$\begin{aligned}
\bar{r}'(K_{it}, \tilde{M}_{it}) &= \bar{r}(K_{it}, \tilde{M}_{it})T(K_{it}),\\
h'_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})/T(K_{it}),\\
h'_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) - \bar{r}(K_{it}, \tilde{M}_{it})T_1(K_{it})/T(K_{it}),\\
h'_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}),
\end{aligned}$$

where $T(K_{it})$ is an arbitrary function and $T_1(K_{it})$ denotes the derivative of $T(K_{it})$ with respect to K_{it}. These functions are equivalent for identification purposes since they lead to the same $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$ as I show below

$$\begin{aligned}
\theta^L_{it} &= f_2 h'_2(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r'(K_{it}, \tilde{M}_{it})L_{it} = f_2 h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r(K_{it}, \tilde{M}_{it})L_{it},\\
\theta^M_{it} &= f_2 h'_3(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it} = f_2 h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it},\\
\bar{h}_2(K_{it}, \tilde{M}_{it}) &= h'_2(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r'_2(K_{it}, \tilde{M}_{it}) + h'_3(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}),\\
&= h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}),\\
\bar{h}_1(K_{it}, \tilde{M}_{it}) &= h'_1(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h'_2(K_{it}, \bar{r}'(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r'_1(K_{it}, \tilde{M}_{it}),\\
&= h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_1(K_{it}, \tilde{M}_{it}).
\end{aligned}$$

This implies that we cannot distinguish between $(h_1, h_2, \bar{r}_1, \bar{r}_2)$ and $(h'_1, h'_2, \bar{r}'_1, \bar{r}'_2)$ from $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. Next, I will show that labor-augmenting productivity, capital elasticity and elasticity of substitutions depend on $(h_1, h_2, \bar{r}_1, \bar{r}_2)$, so they cannot be recovered from $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. Since $\omega^L_{it} = \bar{r}(K_{it}, \tilde{M}_{it})$, non-identification of $\bar{r}(K_{it}, \tilde{M}_{it})$ immediately implies that ω^L_{it} is not identified. The output elasticity of capital is

$$\theta_{it}^K = f_1 + f_2 h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}).$$

Since h_1 is not identified, θ_{it}^K is not identified. Finally, to see that the elasticity of substitution is not identified note that it is defined as $\sigma_{it}^{ML} = \partial \log(L_{it}/M_{it})/\partial \log(F_M/F_L)$. It depends on the ratio of marginal products, which can be written

$$\frac{F_L}{F_M} = \frac{h(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})}{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})} - \tilde{M}_{it}$$

Using this, the elasticity of substitution is given by

$$\sigma_{it}^{ML} = \frac{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})^2 - h(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) h_{33}(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})}{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})^2} - 1$$

which depends on $h_{33}(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$. This function is not identified because $\bar{r}(K_{it}, \tilde{M}_{it})$ and h are not identified. Therefore, σ_{it}^{ML} is not identified. The elasticity of substitutions for other input pairs can similarly be derived and it can be shown that they depend on the derivatives of h.

Proof of Proposition A.2

For this proof, I drop the time subscripts from all functions for notational simplicity. If production function takes the form given Equation (A.1) the output elasticities with respect to labor and materials, as a function of f and h, can be written as

$$\theta_{it}^L = f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) r(\tilde{M}_{it}) L_{it}, \qquad \theta_{it}^M = f_2 h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) M_{it}.$$

Since I already showed in Equation (4.6) that θ_{it}^L and θ_{it}^M are identified, the right-hand sides of these equations are identified. The identification of θ_{it}^M immediately implies that $h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$ is identified from (f_2, θ_{it}^M). Taking the derivative of the reduced form function $\bar{h}$ and using $\bar{h}(\tilde{M}_{it}) = h(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$, I obtain

$$\bar{h}_1(\tilde{M}_{it}) = h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it}) + h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}), \tag{D.3}$$

where $\bar{r}'(\tilde{M}_{it})$ denotes the derivative of $\bar{r}(\tilde{M}_{it})$. Therefore, the right-hand side of Equation (D.3) is identified from $\bar{h}(\tilde{M}_{it})$. Taking the ratio of θ_{it}^L/L_{it} and $f_2\bar{h}_1(\tilde{M}_{it}) - \theta_{it}^M/M_{it}$ gives

$$b(\tilde{M}_{it}) := \frac{\theta_{it}^L/L_{it}}{f_2\bar{h}_1(\tilde{M}_{it}) - \theta_{it}^M/M_{it}} = \frac{f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) r'(\tilde{M}_{it})}{f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) r(\tilde{M}_{it})} = \frac{\bar{r}'(\tilde{M}_{it})}{\bar{r}(\tilde{M}_{it})} = \frac{\partial \log(\bar{r}(\tilde{M}_{it}))}{\partial \tilde{M}_{it}}.$$

Hence, the derivative of $\log(r(\tilde{M}_{it}))$ with respect to $\tilde{M}_{it}$ are identified from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$

as $b(\tilde{M}_{it})$. So, we can recover $\log(r(\tilde{M}_{it}))$ up to a constant by integrating $b(\tilde{M}_{it})$:

$$\log(r(\tilde{M}_{it})) = \int_{\underline{\tilde{M}_{it}}}^{\tilde{M}_{it}} b(\tilde{M}_{it}) d\tilde{M}_{it} + a.$$

Since $\omega_{it}^L = r(\tilde{M}_{it})$, and $\log(r(\tilde{M}_{it}))$ is identified up to a constant, ω_{it}^L is identified up to a scale. Identification capital elasticity is easy to show since it depends on f and $\bar{h}$ only. We can recover the output elasticity of capital from f and $\bar{h}$ as: $\theta_{it}^K = f_1(K_{it}, L_{it}\tilde{h}(\tilde{M}_{it}))$

Proof of Proposition A.3

If production function takes the form in Equation (A.1), we can derive σ_{it}^{ML} as

$$\sigma_{it}^{ML} = \frac{h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})^2 - h(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})}{h_{22}(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})^2} - 1,$$

which depends on h_{22}. Since h_{22} is not identified, σ_{it}^{ML} is not identified. The elasticity of substitutions for other input pairs can similarly be derived and it can be shown that they depend on the second derivatives of h, which are not identified.

Proof of Lemma B.1

This proof closely follows the proof of Lemma 3.1. By Assumption B.1 we have

$$(\tilde{p}_{it}, \omega_{it}^L) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1},$$

$$\tilde{p}_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it}, \bar{p}_{it-1}, u_{it}^1) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}.$$

Monotonicity of g_1 with respect to its last argument and Lemma D.1 imply

$$u_{it}^1 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it}, \bar{p}_{it-1}.$$

Since u_{it}^1 has a uniform distribution conditional on $(\omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it}, \bar{p}_{it-1})$ by normalization and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ we have

$$u_{it}^1 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it}, \bar{p}_{it-1} \sim \text{Uniform}(0,1).$$

Using Equations (2.5) and (2.6) we substitute $(\omega_{it-1}^L, \omega_{it-1}^H)$ as functions of (W_{it-1}) to obtain

$$u_{it}^1 \mid K_{it}, W_{it-1}, \tilde{p}_{it} \sim \text{Uniform}(0,1).$$

Proof of Lemma B.2

This proof closely follows the proof of Lemma 3.2. By Assumption B.1 we have

$$(\bar{p}_{it}, \omega^L_{it}, \omega^H_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it-1},$$

$$\bar{p}_{it}, g_1(\omega^L_{it-1}, \omega^H_{it-1}, \tilde{p}_{it}, \bar{p}_{it-1}, u^1_{it}), g_2(\omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it-1}, \bar{p}_{it}, u^1_{it}, u^2_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it-1}.$$

Monotonicity of g_1 and g_2 with respect to their last arguments and Lemma D.1 imply that

$$u^2_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it}, \bar{p}_{it-1}, u^1_{it}.$$

Since u^2_{it} has a uniform distribution conditional on $(\omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it}, \bar{p}_{it-1}, u^1_{it})$ by normalization and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ we have

$$u^2_{it} \mid K_{it}, W_{it-1}, \omega^L_{it-1}, \omega^H_{it-1}, \bar{p}_{it}, u^1_{it} \sim \text{Uniform}(0, 1).$$

Using Equations (2.5) and (2.6) to substitute $(\omega^L_{it-1}, \omega^H_{it-1})$ as functions of W_{it-1}, I obtain

$$u^2_{it} \mid K_{it}, W_{it-1}, \tilde{p}_{it}, u^1_{it} \sim \text{Uniform}(0, 1).$$

Proof of Lemma C.2

By Assumption C.1 we have that

$$\omega^H_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^L_{it-1}, \omega^H_{it-1}.$$

Using the Skorokhod representation of ω^H_{it} in Equation (C.1) we write

$$g_2(\omega^L_{it-1}, \omega^H_{it-1}, u^2_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1}, g_1(\omega^L_{it-1}, \omega^H_{it-1}, u^1_{it}) \mid \omega^L_{it-1}, \omega^H_{it-1}. \tag{D.4}$$

By monotonicity of g_1 and g_2 in their last arguments, u^2_{it} is (conditionally) independent of $(\mathcal{I}_{it-1}, u^1_{it})$

$$u^2_{it} \perp\!\!\!\perp \mathcal{I}_{it-1}, u^1_{it} \mid \omega^L_{it-1}, \omega^H_{it-1}.$$

It follows from Equation (D.4) and the fact that u^2_{it} is uniformly distributed conditional on $(\omega^L_{it-1}, \omega^H_{it-1})$

$$u^2_{it} \mid \mathcal{I}_{it-1}, \omega^L_{it-1}, \omega^H_{it-1}, u^1_{it} \sim \text{Uniform}(0, 1).$$

Since $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ we have $u^2_{it} \mid K_{it}, W_{it-1}, \omega^L_{it-1}, \omega^H_{it-1}, u^1_{it} \sim \text{Uniform}(0, 1)$, which implies

$$u^2_{it} \mid K_{it}, W_{it-1}, u^1_{it} \sim \text{Uniform}(0, 1). \tag{D.5}$$

Next, I use the monotonicity condition for materials demand function to write

$$\begin{aligned}M_{it} &= s\left(K_{it}, \omega_{it}^{H}, \omega_{it}^{L}\right) = s\left(K_{it}, g_2\left(\omega_{it-1}^{L}, \omega_{it-1}^{H}, u_{it}^{2}\right), c_1\left(W_{it-1}, u_{it}^{1}\right)\right), \\ &= s\left(K_{it}, g_2\left(\tilde{r}\left(W_{it-1}\right), \tilde{s}\left(W_{it-1}\right), u_t^2\right), c_1\left(W_{it-1}, u_{it}^{1}\right)\right) \\ &\equiv \bar{s}\left(K_{it}, W_{it-1}, u_{it}^{1}, u_{it}^{2}\right). \end{aligned} \tag{D.6}$$

The intuition is similar to that of Lemma 3.1. Employing strict monotonicity of $\bar{s}$ in u_{it}^2 and Equation (D.5), we can use Equation (D.6) to identify u_{it}^2.

$$u_{it}^{2} = F_{M_{it}|K_{it},W_{it-1},u_{it}^{1}}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^{1}), \tag{D.7}$$

where $F_{M_{it}|K_{it},W_{it-1},u_{it}^{1}}$ denotes the CDF of M_{it} conditional on $(K_{it}, W_{it-1}, u_{it}^{1})$. Therefore, u_{it}^2 is identified from data and ω_{it}^H can be written as $\omega_{it}^H \equiv c_2\left(W_{it-1}, u_{it}^2\right)$.

E Identification

In this section, I show that the homogeneous and weak homothetic separable production functions are generically identified using the moment in Equation (5.3). First, I will present auxiliary lemmas and then analyze these cases separately.

Lemma E.1. *Let $f : \mathbb{R}_+^2 \to \mathbb{R}$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ are differentiable functions. If there exists a differentiable function $p : \mathbb{R}_+^3 \to \mathbb{R}$ with $f(w, zh(x)) = p(w, x, z)$, then $h(x)$ can be recovered from $p(w, x, z)$ up to a scale.*

Proof. Taking the derivatives of the both sides of $f(w, zh(x)) = p(w, x, z)$ with respect to z and x yields

$$f_2(w, zh(x))h(x) = p_2(w, z, x), \qquad f_2(w, zh(x))zh'(x) = p_3(w, z, x).$$

Taking the ratio between the two gives

$$\log'\big(h(x)\big) = \frac{p_2(w, z, x)z}{p_3(w, z, x)}. \tag{E.1}$$

Thus, $\log(h(x))$ is identified up to a constant and $h(x)$ is identified up to a scale. □

Lemma E.2. *Consider the following model*

$$y = f\big(zh(x)\big) + g(x) + \epsilon, \qquad \mathbb{E}[\epsilon \mid z, x] = 0.$$

where (y, x, z) are observed random variables and $f : \mathbb{R}_+ \to \mathbb{R}$, $h : \mathbb{R}_+ \to \mathbb{R}_+$ and $g : \mathbb{R}_+ \to \mathbb{R}$ are unknown functions. Let (f_0, h_0, g_0) denote the true functions. As-

sume that (i) $h_0'(x) > 0$ for all x in the support, where $h_0'(x)$ denotes the derivative of h_0, (ii) functions (f_0, h_0, g_0) are continuously differentiable and have non-zero derivatives almost everywhere, (iii) the joint distribution function of (y, z, x) is absolutely continuous with positive density everywhere on its support.

Let Ω be is the set of functions that obey the model restrictions and assumptions, so $(f_0, h_0, g_0) \in \Omega = \Omega_f \times \Omega_h \times \Omega_g$. Define the set of log-linear functions as $\Omega_{log} = \{f(x) : f(x) = a\log(x) + b, (a, b) \in \mathbb{R}^2\}$ and assume that they are excluded from Ω_f, i.e., $\Omega_{\log} \cap \Omega_f = \emptyset$.
I next provide some definitions based on Matzkin (2007). $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent if and only if

$$f\big(zh(x)\big) + g(x) = \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}(x),$$

for all $(z, x) \in \mathcal{X} \times \mathcal{Z}.(f_0, h_0, g_0) \in \Omega$ are identifiable if no other member of Ω is observationally equivalent to (f, h, g). If identification holds except in special or pathological cases the model is generically identified.

Based on these definitions and under my assumptions, g is identified up to a constant, h is identified up to a scale, and f is identified up to a constant and a normalization specified below in the proof. Since identification fails only in special cases, we say that the functions, (f, h, g), are generically identified. The special cases where identification fails are testable.

Proof. Note that from $\mathbb{E}[\epsilon \mid z, x] = 0$, we have

$$\mathbb{E}[y \mid z, x] = f\big(zh(x)\big) + g(x)$$

Since $\mathbb{E}[y \mid z, x]$ is identified from the distribution of observables, we can take it as known for identification purposes. This conditional expectation captures all the information from data based on the assumption of ϵ.

Assume there exists $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent. Using the definition of identification given above, we have:

$$f\big(zh(x)\big) + g(x) = \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}(x). \tag{E.2}$$

I will show that if Equation (E.2) holds, then (f, h, g) and $(\tilde{f}, \tilde{h}, \tilde{g})$ have to obey the normalization restrictions below

$$f(x) = \tilde{f}(\lambda x) + a, \qquad h(x) = \tilde{h}(x)/\lambda, \qquad g(x) = \tilde{g}(x) - a,$$

for $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}$. To show this, I will take the derivatives of Equation (E.2) with respect to x and z. Taking derivative with respect to z yields

$$f'\big(zh(x)\big)h(x) = \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x). \tag{E.3}$$

Next, taking derivative with respect to x gives

$$f'\big(zh(x)\big)zh'(x) + g'(x) = \tilde{f}'\big(z\tilde{h}(x)\big)z\tilde{h}'(x) + \tilde{g}'(x).$$

Rearranging this to collect similar terms, I obtain

$$f'\big(zh(x)\big)zh'(x) - \tilde{f}'\big(z\tilde{h}(x)\big)z\tilde{h}'(x) = \tilde{g}'(x) - g'(x).$$

Dividing and multiplying the two terms on the left hand side by $h(x)$ and $\tilde{h}(x)$, respectively,

$$f'\big(zh(x)\big)zh(x)\frac{h'(x)}{h(x)} - \tilde{f}'\big(z\tilde{h}(x)\big)z\tilde{h}(x)\frac{\tilde{h}'(x)}{\tilde{h}(x)} = \tilde{g}'(x) - g'(x).$$

Further rearranging and denoting $h'(x)/h(x)$ by $\log'(h(x))$, using assumption (i), we have

$$z\Big(f'\big(zh(x)\big)h(x)\log'(h(x)) - \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)\log'(\tilde{h}(x))\Big) = \tilde{g}'(x) - g'(x).$$

By Equation (E.3) we have that $f'\big(zh(x)\big)h(x) = \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)$. Using this

$$\begin{aligned} z\Big(f'\big(zh(x)\big)h(x)\log'(h(x)) - f'\big(zh(x)\big)h(x)\log'(\tilde{h}(x))\Big) &= \tilde{g}'(x) - g'(x) \\ zf'\big(zh(x)\big)h(x)\Big(\log'(h(x)) - \log'(\tilde{h}(x))\Big) &= \tilde{g}'(x) - g'(x). \end{aligned} \tag{E.4}$$

Now as a contradiction suppose $h(x) \neq \tilde{h}(x)/\lambda$ for $x \in \tilde{\mathcal{X}}$ such that $\Pr(x \in \tilde{\mathcal{X}}) > 0$. Then

$$f'\big(zh(x)\big) = \frac{\tilde{g}'(x) - g'(x)}{\big(\log'(h(x)) - \log'(\tilde{h}(x))\big)zh(x)},$$

which gives a differential equation. The only solution to this differential equation is

$$f'\big(zh(x)\big) = \frac{a}{zh(x)} \qquad \text{and} \qquad \frac{\tilde{g}'(x) - g'(x)}{h'(x)/h(x) - \tilde{h}'(x)/\tilde{h}(x)} = \frac{1}{a},$$

for some constant a. This solution gives

$$f(w) = a\log(w) + b,$$

which was excluded from Ω_f by assumptions. Thus, we cannot have $h(x) \neq$

$\tilde{h}(x)/\lambda$, implying

$$\log'(h(x)) = \log'(\tilde{h}(x)), \qquad \tilde{g}'(x) = g'(x).$$

Next, Equation (E.4) and $\log'(h(x)) = \log'(\tilde{h}(x))$ imply that

$$\tilde{g}'(x) = g'(x). \tag{E.5}$$

Integrating this, there exists λ and a such that

$$h(x) = \tilde{h}(x)/\lambda, \qquad g(x) = \tilde{g}(x) - a.$$

Now using these results and Equation (E.3) we solve for $f\big(zh(x)\big)$ and $\tilde{f}\big(zh(x)\big)$

$$f\big(zh(x)\big) = \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}(x) - g(x) = \tilde{f}\big(z\lambda h(x)\big) + a. \tag{E.6}$$

which obeys the stated normalization $f(x) = \tilde{f}(\lambda x) + a$. Therefore, I conclude that observationally equivalent functions $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ should satisfy

$$f(x) = \tilde{f}(\lambda x) + a, \qquad h(x) = \tilde{h}(x)/\lambda, \qquad g(x) = \tilde{g}(x) - a.$$

In this part of the proof, I will show that the assumption that $f \notin \Omega_{log}$ is testable. To see this, note that $f \in \Omega_{log}$ if and only if conditional expectation has the following form

$$y(x, z) := \mathbb{E}[y \mid z, x] = \lambda \log z + h(x) + g(x), \tag{E.7}$$

which is testable by estimating $\mathbb{E}[y \mid z, x]$. If part is trivial. To show the only if part, by the fundamental theorem of calculus, Equation (E.7) implies that $\partial t(x, z)/\partial \log z = \lambda$. Hence

$$\frac{\partial t(x, z)}{\partial \log z} = z\frac{\partial t(x, z)}{\partial z} = zf'(zh(x))h(x) = \lambda \implies f'(zh(x))h(x) = \lambda/z$$

The only solution to this equation is $f(w) = \lambda \log(w) + a$, which belongs to Ω_{log}. Thus, $f \in \Omega_{log}$ is testable by simply testing whether the derivative of $\mathbb{E}[y \mid z, x]$ with respect to $\log(z)$ is constant. □

Identification of Homogeneous Production Function

Under homotheticity assumption, the production function takes the following form

$$y_{it} = vk_{it} + \tilde{f}\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^{H} + \epsilon_{it}.$$

Substituting an unknown function of control variables for ω_{it}^H gives

$$y_{it} = vk_{it} + f\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + g\big(W_{it-1}, u_{it}^1, u_{it}^2)\big) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} \mid k_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}] = 0.$$

Under homothetic model the control variables are $u_{it}^1 = \tilde{M}_{it}$ and $u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. Substituting these, I obtain

$$y_{it} = vk_{it} + f\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + c_2\big(W_{it-1}, \tilde{M}_{it}, \tilde{s}(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it},$$

where $\tilde{s}(\cdot)$ equals the CDF given above, α and $(f, \bar{h}, g)$ are unknown parameter and functions to be estimated. Note that under the modelling assumptions, none of the random variables in $(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})$ are functionally dependent on others. To see this, note that, the inputs can be expressed as

$$\begin{aligned} M_{it} &= s(K_{it}, c_2(W_{it-1}, \tilde{M}_{it}, u_{it}^2), \bar{r}(\tilde{M}_{it})), \qquad K_{it} = k(K_{it-1}, c(W_{it-1}), \eta_{it-1}), \\ L_{it} &= s_2\big(K_{it}, s^{-1}(K_{it}, M_{it}, c_1(W_{it-1}, u_{it}^1)), c_1(W_{it-1}, u_{it}^1)\big), \end{aligned}$$

where η_{it-1} is a vector of random variables that affect the firm's investment decision besides the productivity shocks, $(s, \bar{r}, c_1, c_2)$ are functions defined in the main text, and (k, s_2) are capital and labor decision functions of the firm. This implies that there is variation in an input conditional on all other inputs. By transforming the arguments of $\tilde{s}$, we can rewrite this equation as:

$$y_{it} = vk_{it} + f\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + c_2\big(W_{it-1}, \tilde{M}_{it}, s(k_{it}, \tilde{L}_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it},$$

where $\tilde{s}(x_1, x_2, x_3, x_4) = s(\log(x_1), x_2/(x_3 x_1), x_3, x_4)$. To simplify the notation I relabel $(k_{it}, \tilde{L}_{it}, \tilde{M}_{it}, W_{it-1})$ as (w, z, x, t), relabel $\bar{h}$ by h, and drop the indices from the random variables. This gives

$$y = \alpha w + f\big(zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) + \epsilon, \qquad \mathbb{E}[\epsilon \mid w, z, x, t] = 0.$$

By the moment restriction in Equation (5.3), we have

$$\mathbb{E}[y \mid w, z, x, t] = \alpha w + f\big(zh(x)\big) + g\big(x, t, s(w, z, x, t)\big).$$

Therefore, the data identify $\mathbb{E}[y \mid w, z, x, t]$. Let Ω denote the set of functions that satisfy the restrictions imposed on the true parameter and functions, so $(\alpha_0, f_0, h_0, g_0) \in \Omega$. Using this, we say that $(\alpha, f, h, g) \in \Omega$ and $(\tilde{\alpha}, \tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent if and only if

$$\alpha w + f\big(zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) = \tilde{\alpha} w + \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}\big(x, t, s(w, z, x, t)\big). \tag{E.8}$$

We say that $(\alpha_0, f_0, h_0, g_0) \in \Omega$ are identifiable if no other member of Ω is observationally equivalent to $(\alpha_0, f_0, h_0, g_0)$. The following proposition establishes the generic identification of $(\alpha_0, f_0, h_0, g_0)$.

Proposition E.1. *Suppose that (i) Functions (f_0, h_0, g_0) are twice continuously differentiable and have non-zero derivatives almost everywhere, (ii) The joint distribution function of (w, z, x, t) is absolutely continuous with positive density everywhere on its support, (iii) $h_0'(x) > 0$ almost everywhere, (iv)$f_0 \notin \Omega_{\log}$, where $\Omega_{\log}$ is defined in Lemma E.2, and (v) the matrix defined below is full rank almost everywhere*

$$\begin{bmatrix} s_1^2(w, z, x, t) & s_{11}(w, z, x, t) \\ s_1(w, z, x, t)s_2(w, z, x, t) & s_{12}(w, z, x, t) \end{bmatrix}.$$

Then g_0 is identified up to constant, h_0 is identified up to scale and f_0 is identified up to constant and normalization given in Lemma E.2, and α_0 is identified.

Proof. I will show that if there exists observationally equivalent (α, f, h, g) and $(\tilde{\alpha}, \tilde{f}, \tilde{h}, \tilde{g})$, then they equal each other up to normalization described in the proposition. The proof adopts the notation that $r_i()$ denotes the derivative of function r with respect to its i-th argument and r' to denote the derivative if function r takes a single argument. To simplify the exposition, I will treat t as scalar, so the derivative with respect t should be considered as the derivative with respect to each element in t.

Taking the derivative of Equation (E.8) with respect to w we obtain

$$\alpha + g_3\big(x, t, s(w, z, x, t)\big)s_1(w, z, x, t) = \tilde{\alpha} + \tilde{g}_3\big(x, t, s(w, z, x, t)\big)s_1(w, z, x, t).$$

Rearranging this equation:

$$g_3\big(x, t, s(w, z, x, t)\big)s_1(w, z, x, t) - \tilde{g}_3\big(x, t, s(w, z, x, t)\big)s_1(w, z, x, t) = \tilde{\alpha} - \alpha. \tag{E.9}$$

As a contradiction suppose $\alpha \neq \tilde{\alpha}$ and define $\bar{g}_3 = g_3 - \tilde{g}_3$. We have that

$$\bar{g}_3\big(x, t, s(w, z, x, t)\big)s_1(w, z, x, t) = \tilde{\alpha} - \alpha. \tag{E.10}$$

Taking the derivatives of Equation (E.10) with respect to w and z

$$\bar{g}_{33}\big(x,t,s(w,z,x,t)\big)s_1^2(w,z,x,t)+\bar{g}_3\big(x,t,s(w,z,x,t)\big)s_{11}(w,z,x,t)=0.$$

$$\bar{g}_{33}\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t)s_2(w,z,x,t)+\bar{g}_3\big(x,t,s(w,z,x,t)\big)s_{12}(w,z,x,t)=0.$$

By the full rank assumption in (v) $\bar{g}_3=0$ is the only solution to this system of equations everywhere in the support. Therefore, we obtain

$$\alpha=\tilde{\alpha}, \qquad g_3\big(x,t,s(w,z,x,t)\big)-\tilde{g}_3\big(x,t,s(w,z,x,t)\big)=0. \tag{E.11}$$

This shows that α and g_3 are identified. Next, taking the derivative of Equation ($E.10$) with respect to t gives

$$\begin{aligned} g_2\big(x,t,s(w,z,x,t)\big)+g_3\big(x,t,s(w,z,x,t)s_4(w,z,x,t)=\\ \tilde{g}_2\big(x,t,s(w,z,x,t)\big)+\tilde{g}_3\big(x,t,s(w,z,x,t)s_4(w,z,x,t).\end{aligned}$$

Since I already showed that $g_3=\tilde{g}_3$, this gives:

$$g_2\big(x,t,s(w,z,x,t)\big)=\tilde{g}_2\big(x,t,s(w,z,x,t)\big). \tag{E.12}$$

Therefore $g_2\big(x,t,s(w,z,x,t)\big)$ is also identified. Taking the derivative of Equation ($E.10$) with respect to z to obtain

$$\begin{aligned} f'\big(zh(x)\big)h(x)+g_3\big(x,t,s(w,z,x,t)\big)s_2(w,z,x,t)=\\ \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)+\tilde{g}_3\big(x,t,s(w,z,x,t)\big)s_2(w,z,x,t)\end{aligned}$$

Using $g_3=\tilde{g}_3$ obtained in Equation in (E.11) gives

$$f'\big(zh(x)\big)h(x)=\tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x). \tag{E.13}$$

Finally, taking the derivative of Equation ($E.10$) with respect to x, we have

$$f'\big(zh(x)\big)h'(x)z+g_1'\big(x,t,s(w,z,x,t)\big)=\tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}'(x)z+\tilde{g}_1'\big(x,t,s(w,z,x,t)\big).$$

Rearranging,

$$z\Big(f'\big(zh(x)\big)h'(x)-\tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}'(x)\Big)=\tilde{g}_1\big(x,t,s(w,z,x,t)\big)-g_1\big(x,t,s(w,z,x,t)\big).$$

Using Equation (E.13) we can substitute $f'\big(zh(x)\big)h'(x)$ and, with some algebra

$$z\big(\tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)(\log'(h(x))-\log'(\tilde{h}(x)))\big)=\tilde{g}_1\big(x,t,s(w,z,x,t)\big)-g_1\big(x,t,s(w,z,x,t)\big). \tag{E.14}$$

Taking the derivative with respect to w, we have

$$g_{13}\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t) = \tilde{g}_{13}\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t).$$

This implies that $g_{13}\big(x,t,s(w,z,x,t)\big) = \tilde{g}_{13}\big(x,t,s(w,z,x,t)\big)$. Taking the derivative with respect to t

$$g_{12}\big(x,t,s(w,z,x,t)\big) + g_{13}\big(x,t,s(w,z,x,t)\big)s_4(w,z,x,t) = \\ \tilde{g}_{12}\big(x,t,s(w,z,x,t)\big) + \tilde{g}_{13}\big(x,t,s(w,z,x,t)\big)s_4(w,z,x,t)$$

Given that $g_{13} = \tilde{g}_{13}$, we have $g_{12}\big(x,t,s(w,z,x,t)\big) = \tilde{g}_{12}\big(x,t,s(w,z,x,t)\big)$. Now using these results and the fundamental theorem of calculus, we can define

$$\bar{g}_1(x) \equiv g_1\big(x,t,s(w,z,x,t)\big) - g_1'\big(x,t,s(w,z,x,t)\big) \tag{E.15}$$

Now as a contradiction suppose there exists with $\tilde{\mathcal{X}}$ such that $\Pr(x \in \tilde{\mathcal{X}}) > 0$, $h(x) \neq \tilde{h}(x)/\lambda$. Therefore, Equation (E.14) can be written as

$$f'\big(z\tilde{h}(x)\big) = \frac{\bar{g}_1'(x)}{\big(\log'(h(x)) - \log'(\tilde{h}(x)\big)\tilde{h}(x)z}.$$

The rest of the proof is an application of Lemma E.2 (see Equation E.4). Therefore, we obtain the desired result

$$f(x) = \tilde{f}(\lambda x) + a, \qquad h(x) = \tilde{h}(x)/\lambda, \qquad g(x) = \tilde{g}(x) - a, \qquad \alpha = \tilde{\alpha}.$$

□

Identification for Weak Homothetic Production Function

Under weak homothetic separability assumption, the function function takes the following form:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{E.16}$$

Substituting an unknown function of control variables for ω_{it}^H we obtain:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + c_2\big(W_{it-1}, u_{it}^1, u_{it}^2)\big) + \epsilon_{it}, \qquad \mathbb{E}[\epsilon_{it} \mid k_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0.$$

Under the weak homothetic separable model the control variables are $u_{it}^1 = \tilde{M}_{it}$ and $u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. Substituting these into Equation (E.16) gives:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + g\big(\tilde{M}_{it}, W_{it-1}, \tilde{s}(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it},$$

where $\tilde{s}(\cdot)$ equals the CDF given above, $(f, \bar{h}, g)$ are unknown functions to be estimated. By transforming the arguments of $\tilde{s}$, we can rewrite this equation as

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + g\big(\tilde{M}_{it}, W_{it-1}, s(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it},$$

where $\tilde{s}(x_1, x_2, x_3, x_4) = s(x_1, x_2/x_3, x_3, x_4)$. To simplify the notation, I relabel $(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})$ as (w, z, x, t), $\bar{h}$ as h, and drop indices from the random variables to obtain

$$y = f\big(w, zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) + \epsilon, \qquad \mathbb{E}[\epsilon \mid w, z, x, t] = 0.$$

By the moment restriction in Equation (5.3), we have

$$\mathbb{E}[y \mid w, z, x, t] = f\big(w, zh(x)\big) + g\big(x, t, s(w, z, x, t)\big).$$

From data, we can identify $\mathbb{E}[y \mid w, z, x, t]$. Let Ω denote the set of functions that satisfy the restrictions imposed on the functions, so $(f_0, h_0, g_0) \in \Omega$. Using this we say $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent if and only if

$$f\big(w, zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) = \tilde{f}\big(w, z\tilde{h}(x)\big) + \tilde{g}\big(x, t, s(w, z, x, t)\big). \quad \text{(E.17)}$$

$(f_0, h_0, g_0) \in \Omega$ are identifiable if no other member of Ω is observationally equivalent to (f_0, h_0, g_0).

Proposition E.2. *Suppose that (i) Functions (f_0, h_0, g_0, s) are twice continuously differentiable and have non-zero derivatives almost everywhere, (ii) The joint distribution function of (w, z, x, t) is absolutely continuous with positive density everywhere on its support, (iii) $h_0'(x) > 0$ almost everywhere. (iv) $\mathbb{E}[s_1(w, z, x, t)/s_2(w, z, x, t) \mid w, z, x] > 0$. (v) $\mathbb{E}[q^2 \mid x, s, t] > 0$ for all (x, s, t) where q is defined as $q := s_2(w, z, x, t) \log'\big(h_0(x)\big)z - s_3(w, z, x, t)$. (vi) $\mathbb{E}[f_1(w, zh(x))^2 \mid w, z] > 0$ or $\mathbb{E}[(f_2(w, zh(x))h(x))^2 \mid w, z] > 0$. Then g_0 is identified up to constant, h_0 is identified up to scale and f_0 is identified up to constant and normalization given in the proof.*

Proof. I will show that if there exists observationally equivalent (f, h, g) and $(\tilde{f}, \tilde{h}, \tilde{g})$, then they equal each other up to normalization given in the proposition. Denote $\mathbb{E}[y \mid w, x, z, t]$ by $y(w, z, x, t)$. Taking the derivative of $y(w, z, x, t)$

with respect to (w, z, x, t) we have

$$y_1(w, z, x, t) = f_1\big(w, zh(x)\big) + g_2\big(x, s(w, z, x, t), t\big)s_1(w, z, x, t), \tag{E.18}$$

$$y_2(w, z, x, t) = f_2\big(w, zh(x)\big)h(x) + g_2\big(x, s(w, z, x, t), t\big)s_2(w, z, x, t), \tag{E.19}$$

$$y_3(w, z, x, t) = f_2\big(w, zh(x)\big)h'(x)z + g_2\big(x, s(w, z, x, t), t\big)s_3(w, z, x, t) + g_1\big(x, s(w, z, x, t), t\big) \tag{E.20}$$

$$y_4(w, z, x, t) = g_2\big(x, s(w, z, x, t), t\big)s_4(w, z, x, t) + g_3\big(x, s(w, z, x, t), t\big) \tag{E.21}$$

Multiplying Equation (E.19) by $s_1(w, z, x, t)/s_2(w, z, x, t)$ and subtracting Equation (E.18) yields

$$y_2(w, z, x, t)\frac{s_1(w, z, x, t)}{s_2(w, z, x, t)} - y_1(w, z, x, t) = f_2\big(w, zh(x)\big)h(x)\frac{s_1(w, z, x, t)}{s_2(w, z, x, t)} - f_1\big(w, zh(x)\big)$$

The left-hand side of this equation is written in terms of identified functions. Now, denote $\tilde{f}_1(w, z, x) := f_1\big(w, zh(x)\big)$ and $\tilde{f}_2(w, z, x) := f_2\big(w, zh(x)\big)h(x)$ and denote the left-hand side by $\tilde{y}(w, z, x, t)$. This gives

$$\tilde{y}(w, z, x, t) = \tilde{f}_1(w, z, x) - \tilde{f}_2(w, z, x)\tilde{s}(w, z, x, t)$$

By Assumption (iv) there is variation in $\tilde{s}$ conditional on (w, z, x). This implies that $\tilde{f}_1(w, z, x)$ and $\tilde{f}_2(w, z, x)$ are identified from this equation. Using assumption (iv) and by applying Lemma E.1 $h(x)$ is identified up to a scale from $\tilde{f}_1(w, z, x)$. Next, multiplying Equation (E.19) by $\log'\big(h(x)\big)z$ and subtracting Equation (E.20) we obtain

$$\begin{aligned} &y_2(w, z, x, t)\log'\big(h(x)\big)z - y_3(w, z, x, t) = \\ &g_2\big(x, s(w, z, x, t), t\big)\big(s_2(w, z, x, t)\log'\big(h(x)\big)z - s_3(w, z, x, t)\big) - g_1\big(x, s(w, z, x, t), t\big). \end{aligned} \tag{E.22}$$

The left-hand side of this equation is identified because we already showed that $\log'\big(h(x)\big)$ is identified and y_2 and y_3 are identified functions. By assumption (v), conditional on (x, s, t) there is variation in $\big(s_2(w, z, x, t)\log'\big(\tilde{h}(x)\big)z - s_3(w, z, x, t)\big)$. Therefore, $g_2\big(x, s(w, z, x, t), t\big)$ and $g_1\big(x, s(w, z, x, t), t\big)$ can be identified from Equation (E.22). Using this $g_3\big(x, s(w, z, x, t), t\big)$ is identified from Equation (E.21), $f_1\big(w, zh(x)\big)$ is identified from Equation (E.18) and $f_2\big(w, zh(x)\big)$ is identified from Equation (E.20). Therefore, we obtain

$$f\big(w, zh(x)\big) = \tilde{f}\big(w, \lambda zh(x)\big) + a, \qquad h(x) = \tilde{h}(x)/\lambda, \qquad g\big(x, s, t\big) = \tilde{g}\big(x, s, t\big) - a.$$

Table F.0: List of Products

Product Category	*Unit*	*Obs*	*Products Included (Code)* 2000-09 (ASICC)	2010-14 (NPCMS)
Brick Tiles	Tonnes	7500	29101 29102	3732001 3732007
Black Tea	Kilograms	6902	12211 12212 12213 12214 12215	2391301 2391302 2391303 2391308
Rice, Parboiled Non-Basmati	Tonnes	6547	12311	2316107 2316202
Biri Cigarettes	Number of Cig.	5735	15323	2509001
Rice, Raw Non-Basmati	Tonnes	5057	12312	2316108 2316203
Shirts, Cotton	Number of Shirts	3515	63428 63428	2822203 2822299 2822406 2822408 2823499

Notes: This table presents the list of products that were used to estimate the quantity-based production function. The second column shows the unit of measurement and the third column shows the number of firm-year observations for each product. The final two columns list the product codes that are included in each product category. The name of the products for each code can be found here.

□

F Robustness Checks

F.1 Estimation with Quantity

In this section, I estimate production functions using data from six Indian homogeneous products for which I have the quantity produced and price of the good. The products include Brick Tiles, Cotton Shirts, Biri Cigarettes, Black Tea, Parboiled Non-Basmati Rice, and Raw Non-Basmati Rice. These products are relatively homogeneous products that are produced by a large number of firms in India.

Table F.0 lists the products, their units, and the product codes included in these product classifications during two different periods. These products are selected based on two different product classifications, ASICC and NPCMS, because the National Statistical Office of India revised their product characteristics in 2011. [1] The second column displays the unit of output recorded in each product, and the third column shows the number of firm-year observations for each product.

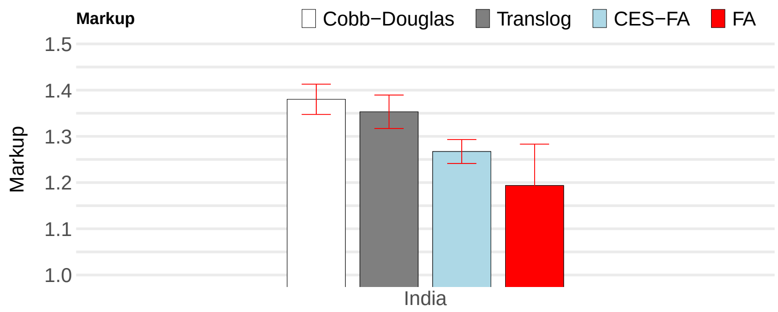
When forming the sample, I include the plants that earn at least 75% of their revenues from one of these products. This mitigates the issue of allocating inputs across products for multi-product firms because I focus mostly on the single-product firms. Following Raval (2020), I define the price of a product as the gross value minus any reported expenses (excise duty, sales tax, and other expenses) divided by the quantity sold. I then drop outlier plants whose price is greater than five times, or less than 20%, of the median price for a given product in a given year. Since the number of observations per industry is smaller than the main sample, I use a five-year rolling-window estimation instead of three for this estimation.

I estimate product-level quantity production functions for these products. The main difference between this estimation and the one presented in the paper is that, I control for input prices in the estimation because input prices are heterogeneous. See Section B.1 for the details of the estimation procedure. The rest of the estimation procedure follows Section A.7.

Figures F.1 and F.2 present the estimates of markup level and change from this production function estimation. Focusing on Figure F.1, we see the same pattern that we identified from other datasets: Hicks-neutral production functions estimate the highest markups, and as we go from Cobb-Douglas to non-parametric labor-augmenting productivity, the markup estimates decline. In Figure F.1, we observe that the factor-augmenting production function suggests lower markup estimates from 2000 to 2014. Overall, these findings suggest that observing revenues rather than quantities does not explain my main results.

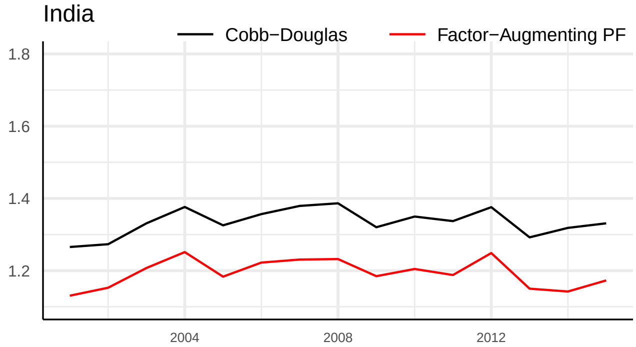
[1] The product codes after 2010 can be found here. The crosswalk between the two product categories can be found here.

Figure F.1: Estimates from Quantity Production Function



Notes: Estimates of aggregate markups from quantity production function in six Indian industries from 2000-2014.

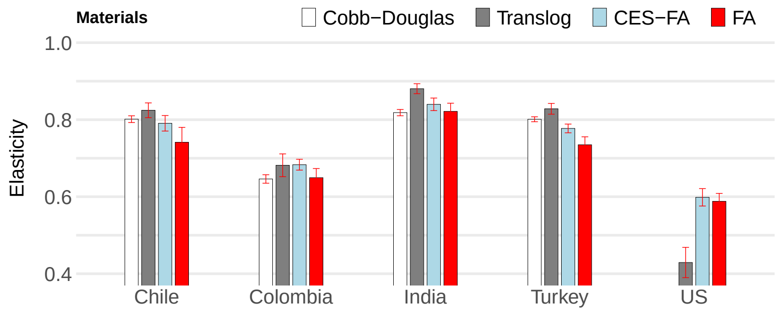
Figure F.2: Estimates from Quantity Production Function (Change in Markup)



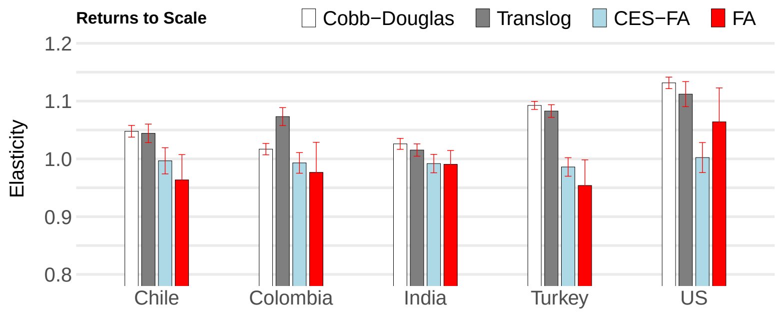
Notes: Estimates of the change in aggregate markups from quantity production function in six Indian industries from 2000-2014.

G Additional Figures

Figure G.3: Elasticity Comparison



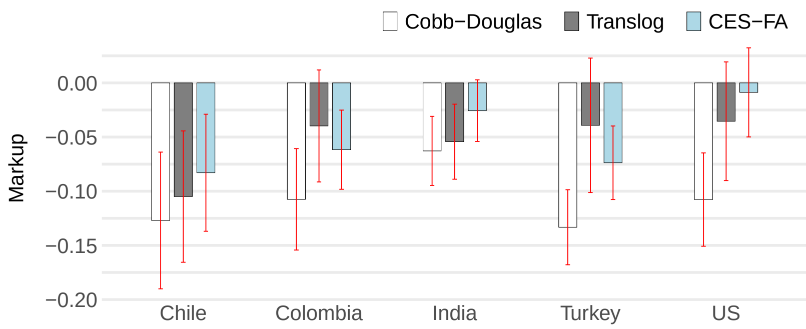
(a) Materials Elasticity



(b) Returns to Scale

Note: Comparison of sales-weighted average elasticities produced by Cobb-Douglas (CD), (ii) Translog (TR), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) nonparametric production function with factor-augmenting productivity (FA). The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95% confidence intervals calculated using bootstrap (100 iterations).

Figure G.4: Difference of Markup Estimates from FA and Other Models



Notes: Each bar shows the difference between the markups estimates from indicated method and FA. The red bars report corresponding 95 percent confidence intervals. Standard errors are calculated using bootstrap (100 iterations)

Figure G.5: Difference of Elasticity Estimates from FA and Other Models

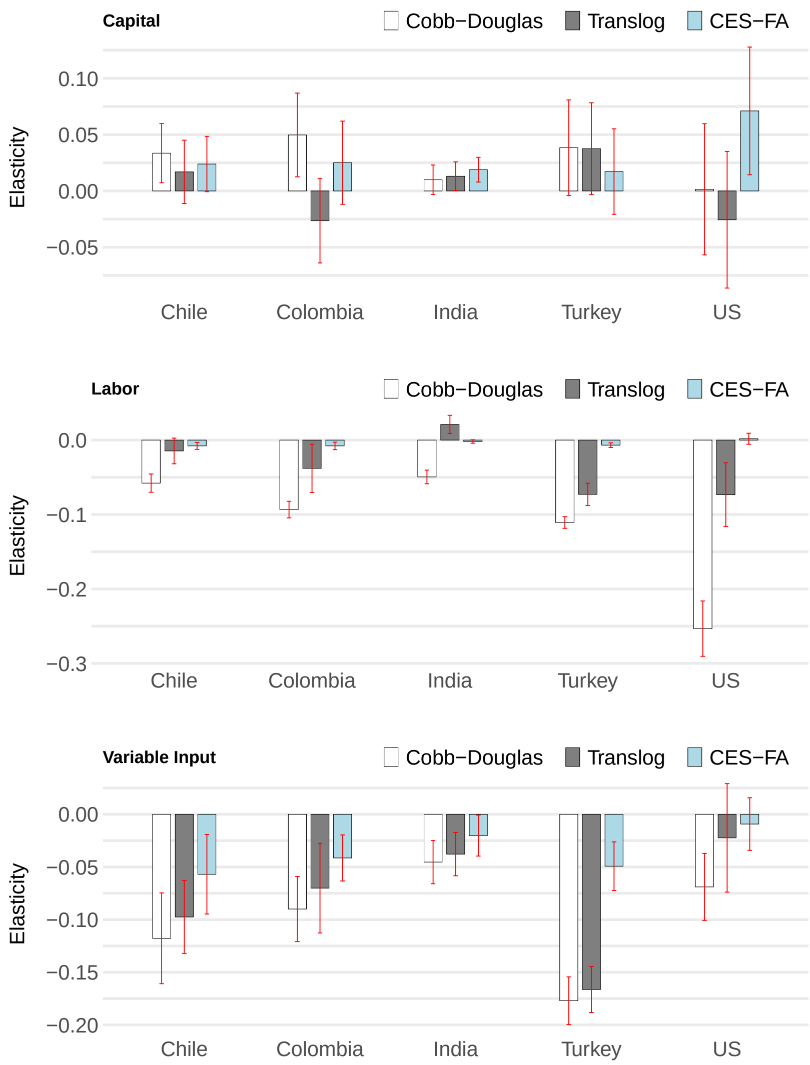
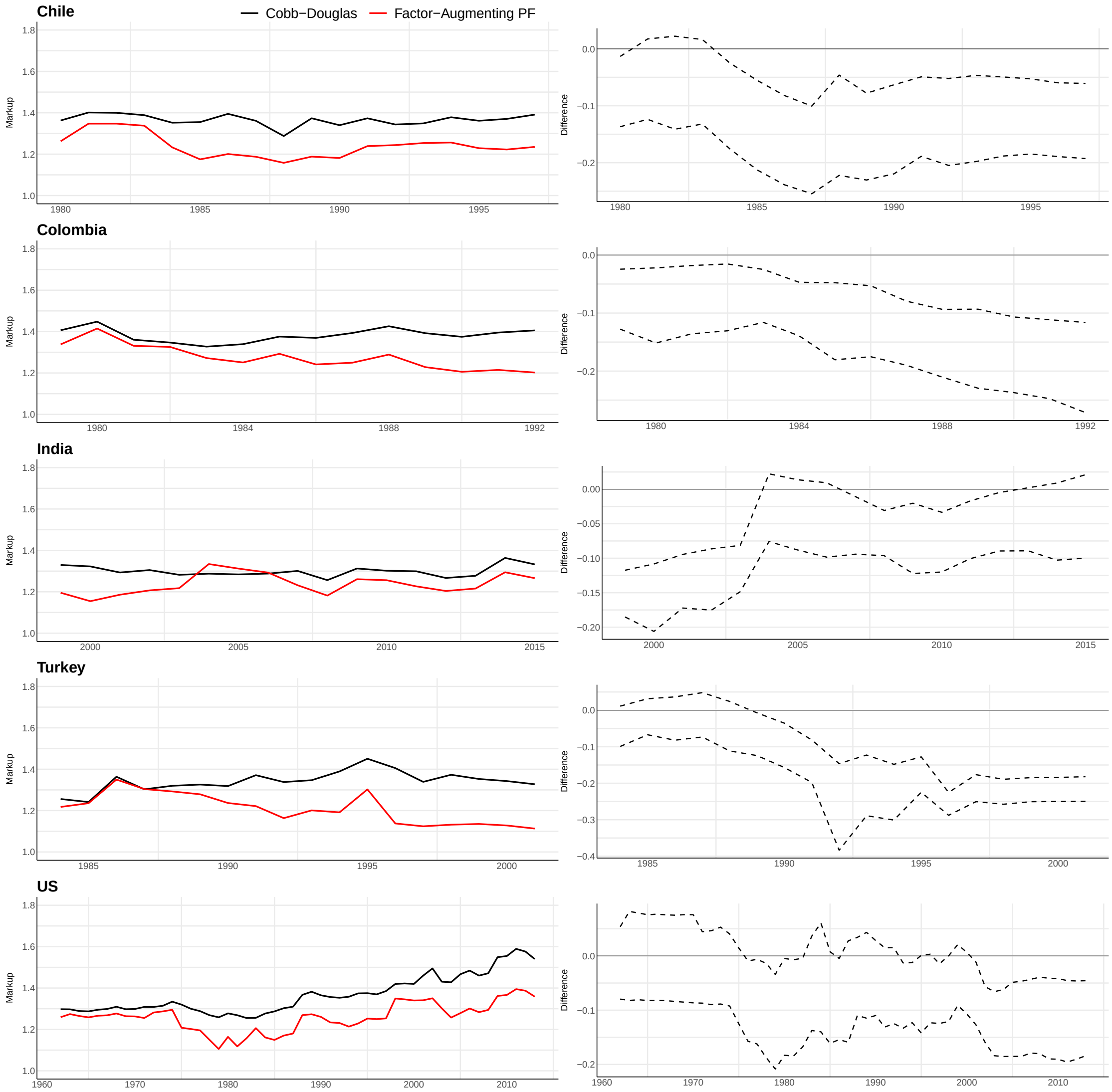


Figure G.6: Confidence Bands for Difference between Markup Estimates



Notes: This figure shows the evolution of the aggregate markups estimated from my method and Cobb-Douglas on left panel and 10-90th percentile of the bootstrap distribution (100 iterations) for the difference between the two estimates..

References

Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451.

Allcott, Hunt, Collard-Wexler, Allan, and O'Connell, Stephen D. 2016. "How do electricity shortages affect industry? Evidence from India". *American Economic Review* 106 (3): 587–624.

Blundell, Richard and Bond, Stephen. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340.

Keller, Wolfgang and Yeaple, Stephen R. 2009. "Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States". *The Review of Economics and Statistics* 91 (4): 821–831.

Matzkin, Rosa L. 2007. "Nonparametric identification". *Handbook of econometrics* 6:5307–5368.

Raval, Devesh. 2020. "Testing the Production Approach to Markup Estimation". *Working Paper.*

Taymaz, Erol and Yilmaz, Kamil. 2015. "Estimating Plant-level Marginal Costs and Mark-ups in the Turkish Manufacturing Industry". *Middle East Technical University, ERC Working Paper.*

Supplemental Materials (Not for Publication) "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups"

Mert Demirer

July 25, 2022

A Descriptive Statistics

Table A.0: Descriptive Statistics - Chile

		Share (Sales)			Number of Plants		
ISIC	Industry	1979	1988	1996	1979	1988	1996
311	Leather Tanning and Finishing	0.17	0.19	0.20	1245	1092	983
381	Metal Products	0.04	0.04	0.04	383	301	353
321	Textiles	0.05	0.04	0.02	418	312	257
331	Repair Of Fabricated Metal Products	0.03	0.02	0.03	353	252	280
322	Apparel	0.02	0.02	0.01	356	263	216
	Other Industries	0.69	0.69	0.69	2399	1957	1873

Note: Descriptive Statistics for Chile. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants. The last row provides information about the industries that are not included in the sample.

Table A.1: Descriptive Statistics - Colombia

		Share (Sales)			Number of Plants		
ISIC	Industry	1978	1985	1991	1978	1985	1991
311	Leather Tanning and Finishing	0.21	0.21	0.20	971	840	976
322	Apparel	0.03	0.03	0.03	666	862	842
381	Metal Products	0.04	0.04	0.03	593	478	534
321	Textiles	0.11	0.09	0.08	467	398	428
342	Cutlery, Hand Tools, and General Hardware	0.02	0.03	0.02	325	315	342
382	Laboratory Instruments	0.02	0.02	0.02	285	266	307
352	Farm and Garden Machinery and Equipment	0.06	0.07	0.09	287	257	305
369	Miscellaneous Electrical Machinery	0.03	0.04	0.03	299	257	267
356	General Industrial Machinery	0.02	0.03	0.04	197	252	341
	Other Industries	0.45	0.45	0.46	3893	3673	4001

Note: Descriptive Statistics for Colombia. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants. The last row provides information about the industries that are not included in the sample.

Table A.2: Descriptive Statistics - India

		Share (Sales)			Number of Plants		
NIC	Industry	1998	2007	2014	1998	2007	2014
230	Other non-metallic mineral products	0.09	0.05	0.08	596	1056	1386
265	Measuring and testing, equipment	0.01	0.02	0.02	272	877	750
213	Pharmaceuticals, medicinal chemical	0.01	0.01	0.01	186	479	670
304	Military fighting vehicles	0.04	0.03	0.07	118	383	704
206	Sugar	0.06	0.04	0.04	271	363	431
	Other Industries	0.79	0.86	0.78	1172	2795	3510

Note: Descriptive Statistics for India. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants. The last row provides information about the industries that are not included in the sample.

Table A.3: Descriptive Statistics - US

		Share (Sales)			Number of Firms		
NAICS	Industry	1961	1987	2014	1961	1987	2014
33	Manufacturing I	0.39	0.37	0.60	113	1092	752
32	Manufacturing II	0.51	0.53	0.25	84	392	222
31	Manufacturing III	0.10	0.10	0.15	36	138	104

Note: Descriptive Statistics for US. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants.

Table A.5: Descriptive Statistics - Turkey

		Share (Sales)			Number of Plants		
ISIC	Industry	1983	1991	2000	1983	1991	2000
321	Textiles	0.16	0.13	0.16	1017	945	1803
311	Food	0.12	0.12	0.11	1261	1120	1061
322	Apparel	0.02	0.05	0.04	300	831	800
381	Metal Products	0.04	0.04	0.04	650	542	834
382	Machinery	0.05	0.06	0.04	532	482	683
383	Electrical-Electronic Machinery	0.04	0.03	0.04	413	523	639
356	Plastic Products	0.08	0.07	0.07	309	312	402
352	Pharmaceuticals	0.08	0.09	0.12	331	286	428
371	Motor Vehicles and Motor Vehicle Equipment	0.02	0.02	0.03	287	261	383
312	Beverage and Tobacco Product Manufacturing	0.05	0.06	0.07	263	218	250
	Other Industries	0.33	0.34	0.29	5100	5302	7033

Note: Descriptive Statistics for Turkey. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants. The last row provides information about the industries that are not included in the sample.

B Output Elasticities

Table C.6 presents the sales-weighted average elasticities for the three largest industries in each country from three methods: (i) my approach (labeled "FA"), (ii) Cobb-Douglas estimated with Blundell and Bond (2000) and (iii) Cobb-Douglas estimated with OLS. My model generates output elasticities that are precisely estimated and reasonable: they are broadly in line with previous results, capital elasticities are positive, and returns to scales are around one. Materials have the highest elasticity, ranging from 0.50-0.67, across industry/county. The average labor and capital elasticities range from 0.22–0.52 and 0.04–0.16, respectively. The returns to scale estimates, measured by the sum of the elasticities, range from 0.93–1.1, indicating that firms, on average, operate close to constant returns to scale.

C Robustness Checks

This section considers four robustness checks. I look at how (i) measurement error in capital and (ii) correction for capacity utilization.

C.1 Measurement Error in Capital

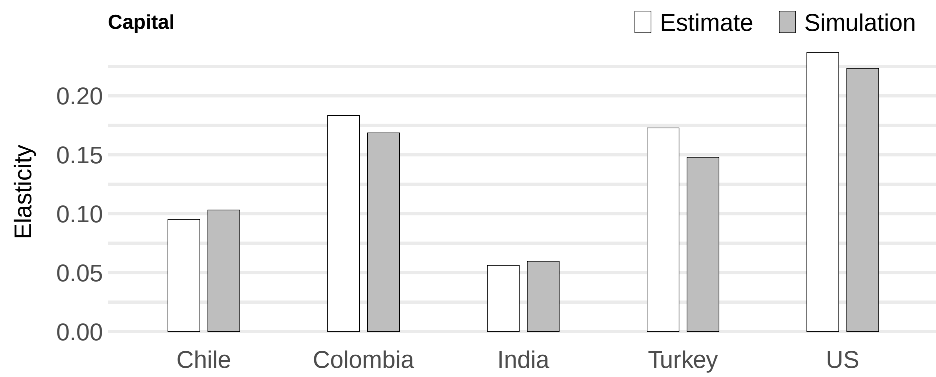
I analyze how measurement error in capital input affects my empirical estimates using a simulation study. In particular, I assume that the observed data are generated from the 'true' data generating process, and then to understand the impact of measurement error, I add independently distributed error to capital input. The error is drawn from a mean-zero normal distribution whose standard deviation equals one-tenth of the standard deviation of capital in the data. I simulate 100 datasets with measurement errors in capital, estimate output elasticities and markups using these dataset and report the average over 100 estimates.

Figure C.6 reports the original estimates together with the average of 100 estimates obtained from simulated data. As expected, measurement error in capital reduces the output elasticity of capital and increases the output elasticity of labor in most simulations. This observation suggests that the higher estimates of capital elasticity obtained using my model and reported in Subsection 6.1 cannot be explained by potential measurement error in capital.

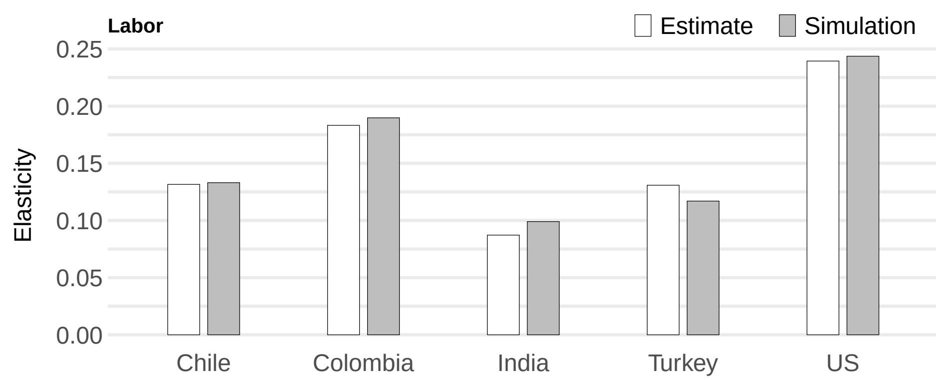
C.2 Capital Utilization

This section analyzes the effects of capacity utilization of capital on my estimates. For this I use firms' energy consumption under the assumption that capital energy takes a Leontief form in the production function. Under this assumption, one can recover the true amount

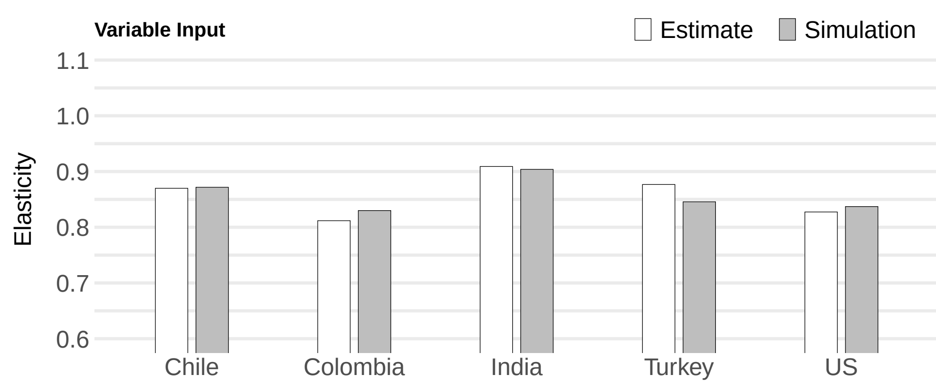
Figure C.6: Comparison of Estimates with and without Measurement Error



(a) Capital Elasticity



(b) Labor Elasticity



(c) Variable Input Elasticity

Notes: This figure presents results from a simulation exercise to understand the potential effects of measurement errors in the estimates. The white bars show elasticity estimates when the data is treated as the 'true' model. The grey bar is an average of 100 elasticity estimates that are obtained from datasets with added error to capital.

Table C.6: Sales-Weighted Average Output Elasticities for Three Largest Industries

	Industry 1			**Industry 2**			**Industry 3**		
	CD	TR	FA	CD	TR	FA	CD	TR	FA
				Chile (311, 381, 321)					
Capital	0.04	0.08	0.09	0.1	0.07	0.09	0.11	0.04	0.12
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.03)	(0.03)
Labor	0.14	0.09	0.1	0.23	0.25	0.18	0.32	0.25	0.19
	(0.01)	(0.01)	(0.00)	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.01)
Materials	0.86	0.88	0.79	0.72	0.72	0.65	0.7	0.75	0.69
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.04)	(0.01)	(0.03)	(0.04)
Rts	1.04	1.05	0.98	1.04	1.03	0.93	1.13	1.03	1
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.06)	(0.02)	(0.02)	(0.05)
				Colombia (311, 322, 381)					
Capital	0.06	0.1	0.13	0.13	0.1	0.12	0.08	0.3	0.19
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Labor	0.18	0.14	0.11	0.47	0.3	0.3	0.34	0.3	0.25
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Materials	0.79	0.81	0.78	0.56	0.72	0.63	0.54	0.52	0.56
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
Rts	1.03	1.05	1.01	1.16	1.12	1.05	0.96	1.12	1
	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.05)
				India (230, 265, 213)					
Capital	0.09	0.07	0.04	0.06	0.06	0.07	0.07	0.05	0.09
	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)
Labor	0.34	0.01	0.06	0.09	0.04	0.08	0.36	0.14	0.18
	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)	(0.00)
Materials	0.57	0.93	0.82	0.85	0.91	0.82	0.57	0.79	0.67
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Rts	1	1.01	0.93	1	1.01	0.96	1	0.99	0.94
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
				Turkey (321, 311, 322)					
Capital	0.03	0.05	0.08	0.05	0.06	0.14	0.04	0.03	0.07
	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Labor	0.16	0.15	0.08	0.22	0.2	0.14	0.27	0.18	0.11
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Materials	0.83	0.86	0.83	0.81	0.81	0.7	0.71	0.86	0.86
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Rts	1.02	1.06	0.99	1.09	1.07	0.98	1.02	1.06	1.05
	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)
				US (33, 32, 31)					
Capital	0.3	0.47	0.29	0.24	0.19	0.22	0.21	0.27	0.24
	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.04)	(0.01)	(0.02)	(0.02)
Labor	0.47	0.25	0.21	0.47	0.32	0.21	0.52	0.32	0.28
	(0.04)	(0.06)	(0.01)	(0.04)	(0.04)	(0.01)	(0.02)	(0.03)	(0.01)
Materials	0.25	0.32	0.56	0.31	0.5	0.61	0.27	0.39	0.58
	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)	(0.03)	(0.02)	(0.02)	(0.01)
Rts	1.02	1.05	1.05	1.01	1.02	1.04	1.01	0.98	1.09
	(0.02)	(0.03)	(0.05)	(0.01)	(0.02)	(0.04)	(0.00)	(0.01)	(0.02)

Note: Comparison of sales-weighted average output elasticities produced by different methods. FA refers to my estimates, BB refers to Blundell and Bond (2000) estimates and OLS is Cobb-Douglas estimated by OLS. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. Numbers in each panel correspond to the SIC codes of the three largest industries in each country. Bootstrapped standard errors in parentheses (100 iterations).

of capital used by the firm using energy consumption as capital input and energy should be proportional. I observe firms' energy consumption only in two datasets, Chile and Turkey, so I consider this robustness exercise only using dataset from those countries. For capacity utilization corrected estimates, I first recover the true capital used by the firm and then estimate output elasticities and markups.

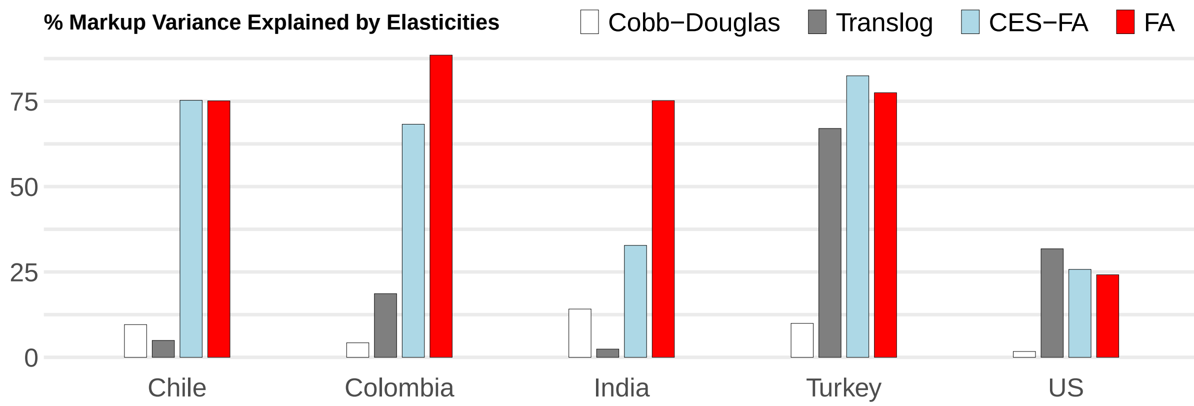
Figure C.7 reports the original estimates together with the estimates obtained with capacity utilization corrected capital. The results suggest that correcting for capacity utilization affect only capital elasticities, and for other elasticities and markups, the estimates remain the same with negligible differences. For the output elasticity of capital, correcting for capacity utilization changes the estimates in different directions in Chile and Turkey.

D Additional Analyses and Figures

D.1 Variance Decomposition of the Aggregate Markups

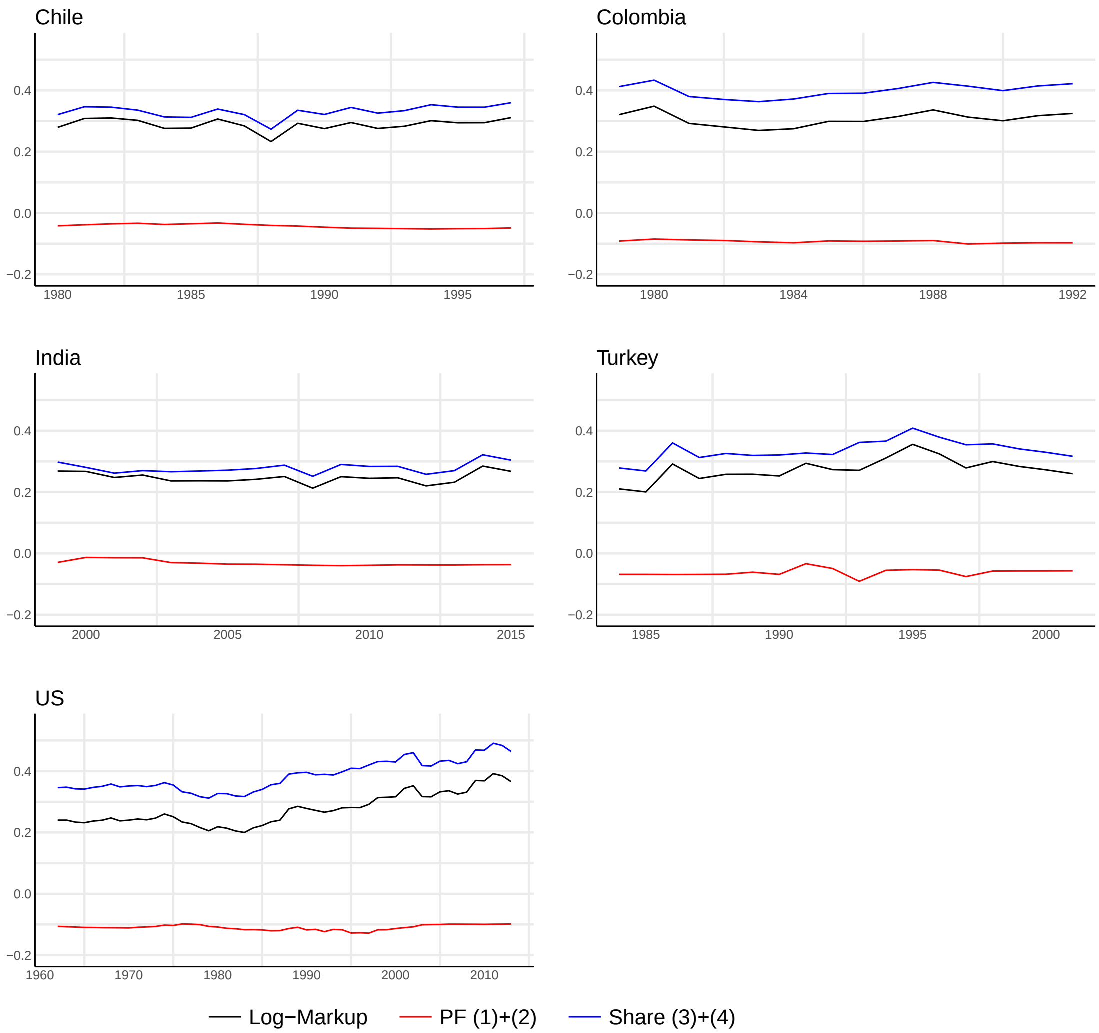
I decompose the time series variance of the aggregate log-markup into the variance of (1)+(2) and variance of (3)+(4) in the decomposition exercise presented in Subsection 7.2, ignoring the covariance between the two. Figure D.8 presents the results from this decomposition for both production functions. Focusing on the Cobb-Douglas model, we see that a large fraction of the variance is explained by the change in revenue shares. The result is particularly striking for the US, where the contribution of the change in output elasticity is only 1%. The decomposition from labor-augmenting productivity reveals a different picture. The change in the elasticity explains a significant fraction of the change in markups in all countries.

Figure D.8: Variance Decomposition of the Change in Markups



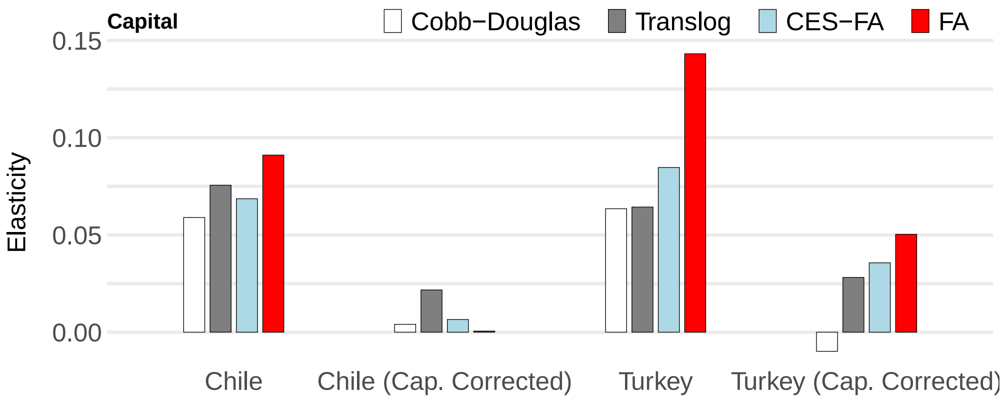
Notes: This figure shows the results by decomposing the annual aggregate log markups time series into the components obtained from elasticities (gray) and revenue shares (black). The covariance between the two components are subtracted from the total variance so that the two components sum to 100.

Figure D.9: Decomposition of Markup: Elasticity vs Share

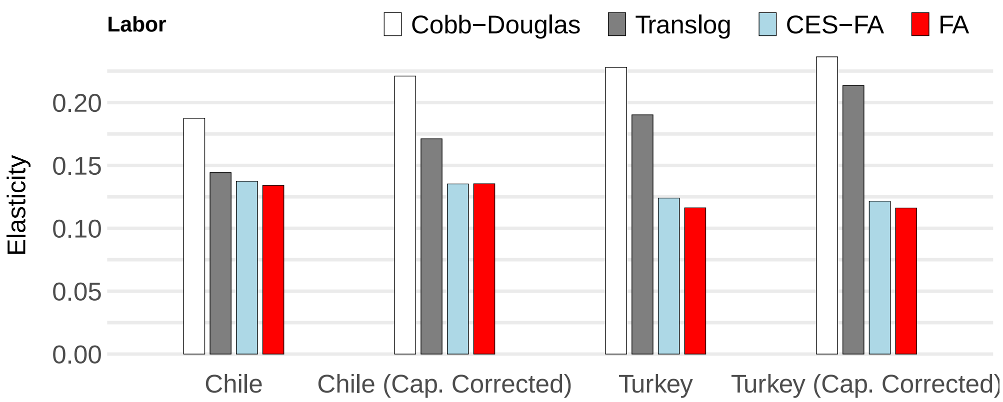


Notes: This figure shows the evolution of the two components of log aggregate markup from the decomposition exercise in Subsection 7.2. Black line displays the log aggregate markups, red line displays the component from production function estimation and, blue line displays the component from revenue share of flexible inputs.

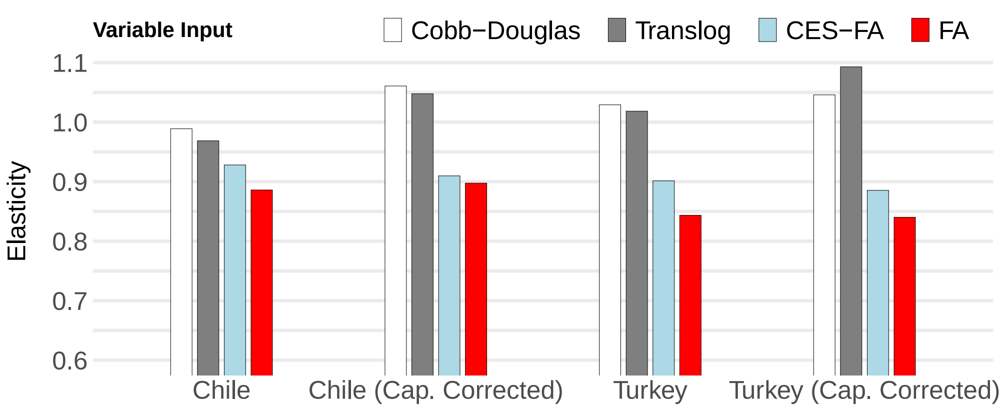
Figure C.7: Comparison of Estimates with and without Capacity Utilization Correction



(a) Capital Elasticity



(b) Labor Elasticity



(c) Variable Input Elasticity

Notes: This figure compares the original estimates with the ones obtained after capacity utilization correction in capital input. See section C.2 for the details.