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

Mert Demirer

December 20, 2020

**Abstract**

Traditional production function models rely on factor-neutral technology and functional form assumptions. These assumptions impose strong theoretical restrictions and are often rejected by the data. This paper develops a new method for estimating production functions with factor-augmenting technology. 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 to accommodate factor-augmenting technology. Then, I show how to identify output elasticities based on a novel control variable approach and the optimality of input expenditures. I use this method to estimate output elasticities and markups in manufacturing industries in the US and four developing countries. Neglecting labor-augmenting productivity and imposing parametric restrictions mismeasures output elasticities and heterogeneity in the production function. My estimates suggest that standard models (i) underestimate capital elasticity, and (ii) overestimate labor elasticity. These biases propagate into markup estimates inferred from output elasticities. My estimates point to a much more muted markup growth (about half) in the US manufacturing sector than recent estimates.

---

Post-Doctoral Researcher, Microsoft Research. Email: 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 Oguzhan Celebi, Allan Collard-Wexler, Max Cytrynbaum, Glenn Ellison, Kirill Evdekimov, Sara Fisher Ellison, Jerry Hausman, Stefan Hoderlein, Paul Joskow, Omer Karaduman, Ali Kakhbod, Christopher Knittel, Jing Li, Madeline Mckelway, Whitney Newey, Alex Olssen, Will Rafey, Nancy Rose, Tobias Salz, Karthik Sastry, Cansu Sen, Erol Taymaz, John Van Reenen, Michael Whinston, Samuel Yang, Kamil Yilmaz, Nathan Zorzi, as well as various seminar participants. I would like to thank State Institute of Statistics in Turkey for allowing us access to the data under the project "Growth and Technological Change in Turkey". I would also like to thank Davesh 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, gains from trade, and market power. The typical exercise requires researchers to specify a model of production function and estimate it using microdata. However, a misspecified production function may produce biased elasticity and productivity 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 markups 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.[1] Several papers have shown that these restrictions are rejected by data at the firm and industry levels. 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)).[2] 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 technology and examine its implications empirically. My model differs from standard models in two important ways. First, it includes non-separable labor-augmenting productivity in addition to standard Hicks-neutral productivity. These productivity shocks introduce unobserved firm-level heterogeneity in production technology. Second, the model does not rely on parametric assumptions to achieve identification; it only imposes a limited functional form structure, which nests the common parametric forms. Together, these features yield a more flexible production function than the standard models, with the ability to better explain the data.

This paper makes both methodological and empirical contributions. On the methodological side, I extend the canonical Olley and Pakes (1996) production function estimation framework to a model with multi-dimensional productivity, and then I study non-parametric identification of this model by building on the recent literature on factor-augmenting technical change (Doraszelski and Jaumandreu (2018), Raval (2019)). On the empirical side, my results indicate that neglecting factor-augmenting technology and imposing parametric restrictions mismeasure output elasticities and markups.

A major challenge in estimating production functions is the endogeneity of inputs. Firms choose inputs based on productivity shocks, but productivity shocks are unob-

[1]Hicks-neutral productivity implies no unobserved heterogeneity in the output elasticities. Cobb-Douglas restricts the elasticity of substitution to equal one and output elasticities to be common across firms.

[2]The decline in labor share, recently observed in developed countries, is also difficult to explain with Hicks-neutral production functions (Oberfield and Raval (2014)).

servable to researchers. This problem generates additional complications in my model due to the multi-dimensional unobserved productivity and the absence of parametric restrictions. To address this challenge, I make three novel methodological contributions.

My first result establishes the invertibility conditions for labor-augmenting productivity. In particular, I show how to express labor-augmenting productivity as a function of observed inputs by inverting input demand functions. This result is critical for controlling for labor-augmenting productivity and it generalizes the widely-employed parametric inversion (Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019)) to a non-parametric setup. I establish invertibility under the assumption that the production function satisfies homothetic separability in labor and materials. Homothetic separability is a weak condition that nests the most commonly used production functions such as CES, but it does not restrict the heterogeneity of output elasticities and elasticity of substitutions across firms. Most importantly, it is an economic rather than a statistical restriction, with clear implications on optimal firm behaviour.

My second contribution is to develop a novel control variable approach for production function estimation, building on Imbens and Newey (2009). In particular, I show how to construct control variables from inputs to control for productivity shocks. This result uses the timing assumption for capital and Markov property of productivity shocks, both of which are standard assumptions in the literature. The control variable approach differs from the standard "control function approach" in that it does not condition on the observed inputs directly. Instead, it identifies control variables from data in the form of conditional quantiles and provides efficiency gains. The control variable approach, in general, cannot be applied to multi-dimensional 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, leading to a two-step approach. First, I construct a control variable for labor-augmenting productivity using materials-to-labor ratio, and then, I construct a control variable for Hicks-neutral productivity using materials demand function.

The third methodological contribution is an identification strategy for output elasticities. After showing how to control for unobserved productivity using control variables, I study which features of the production function can be identified without parametric restrictions. I first establish a negative result: without exogenous variation in input prices, one cannot identify the output elasticity of flexible inputs due to a functional dependence problem; however the sum of the flexible input elasticities can still be identified. To separately identify the flexible input elasticities, I show how to use the first-order conditions of cost minimization. In particular, cost minimization implies that the ratio of two flexible inputs' elasticities is identified as the ratio of their expenditures, without any restrictions on the production function. Importantly for the purpose of markup estimation, the firm's market power is not restricted in the output market, in contrast to recent work that exploits first-order conditions of cost minimization (Gandhi et al. (2020)).[3]

Using cost minimization to identify the ratio of output elasticities has an appealing feature: markup estimates from two flexible inputs are identical, a condition required for the internal validity of the model.[4] This addresses the well-known problem that two different flexible inputs often give conflicting markups estimates (Raval (2020)). My results show that including labor-augmenting productivity into the production function provides a natural solution to this problem. Without labor-augmenting productivity, the model does not explain the large variation in relative input allocation across firms, leading to conflicting markups estimates. Labor-augmenting productivity provides an additional source of unobserved heterogeneity, explains this variation observed in the data, and generates identical markups estimates.

My final set of identification results look at the identification of elasticity of substitution and labor-augmenting productivity. I show that labor-augmenting productivity is not identified under the most general specification, but imposing strong separability between labor and materials allows for identification. In contrast, I show that the elasticity of substitutions cannot be identified without variation in input prices. This result extends the impossibility theorem of Diamond et al. (1978), which states that for industry-level production functions, the elasticity of substitution is not identified from time-series without exogenous variation in input prices. Overall, these results characterize what features of production function require additional restrictions for identification.

I use my method to estimate output elasticities in manufacturing industries in the US and four developing countries: Chile, Colombia, India, and Turkey. To document the biases in standard models, 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, in all countries, Hicks-neutral production functions estimate incorrect output elasticities. In particular, the Cobb-Douglas model underestimates the output elasticity of capital, on average, by 70 percent and overestimates the output elasticity of labor by 80 percent. Using a CES labor-augmenting production function only partially corrects these biases, pointing to the importance of relaxing parametric restrictions. My model also reveals substantial firm-level heterogeneity in the output elasticities. Large firms have a higher elasticity of capital and lower elasticity of flexible inputs than small firms, and exporting firms are more capital-intensive than domestic firms.

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

[3]However, differently from Gandhi et al. (2020), my approach requires two flexible inputs.

[4]Under cost minimization markup is given by output elasticity of a flexible input divided by its share in revenue. Since a firm has only one markup, markup estimates from two inputs should be equal.

study how these biases propagate into markups estimates.

Previous approaches yield severely biased estimates for markups. First, the Cobb-Douglas model overestimates markups in all countries, on average, by 10 to 20 percentage points, an important magnitude when markups are interpreted as a measure of market power. Second, the parametric CES production function with labor-augmenting technology overestimates markups by up to 10 percent. To explain what drives these biases in markup estimates, I present a decomposition exercise. Two sources of misspecification explain the bias in markups: (i) bias in the average output elasticity and (ii) unmodeled heterogeneity in output elasticities related to firm size. The existing empirical evidence and my elasticity estimates imply that both sources of bias are positive.

Next, I estimate the evolution of markups in US manufacturing with data from Compustat.[5] Recently, De Loecker et al. (2020) found that the aggregate markup in the US has risen by 40 percentage points using Hicks-neutral production functions. Their finding has drawn significant attention as it suggests an enormous increase in market power.[6] Using the same dataset, I instead find that the aggregate markup in US manufacturing has increased by only 15 percentage points, going up from 1.3 in 1960 to 1.45 in 2012. This difference arises because estimates from the Cobb-Douglas production function suggest a negligible change in production technology over the last fifty years. However, according to my production function estimates, flexible inputs' average output elasticity has declined, especially since the 1990s. My estimates also suggest important changes in the heterogeneity in output elasticities, which affects the evolution of markups.

This paper presents a general framework for labor-augmenting production functions, so it can easily incorporate economic restrictions, such as constant returns to scale. The framework also covers a family of specifications, ranging from the parametric CES to non-parametric weak homothetic separable production functions, which are nested within each other. Thus, the estimation method can be applied to parametric production functions if one has a priori knowledge about the functional form. Even though my main model analyses the most general production function, I show how to apply my identification results to parametric CES and nested CES production functions with labor-augmenting technology. The control variable approach developed in this paper is also applicable to parametric Hicks-neutral production functions and could be of independent interest. When applied to the Cobb-Douglas, it is robust to the functional dependence problem highlighted by Ackerberg et al. (2015) and it provides efficiency gains as it fully exploits the modeling assumptions.

I develop several extensions to my framework. I show how to extend the model and identification strategy to include observed, but not necessarily exogenous, firm-level input

[5] Although Compustat's data quality is lower than the other datasets in the sample, it has been an important source for the recent findings on the rise of markups.

[6] See, for example, Basu (2019), Berry et al. (2019), Traina (2018) for discussions.

prices. In another extension, I present a way of incorporating firm exit into the model under the assumption that firms exit when they receive a Hicks-neutral productivity shock below a threshold. Finally, I show how my results can be used to account for unobserved input prices when productivity is Hicks-neutral.

My paper contributes to the literature on production function estimation. The most common method for production function estimation is the proxy variables approach, which uses inputs to control for endogeneity. Olley and Pakes (1996) characterize the conditions under which investment can be used as a 'proxy' to control for unobserved productivity. Motivated by practical challenges to using investment as a proxy, Levinsohn and Petrin (2003) instead propose using materials. Ackerberg et al. (2015) point out a potential collinearity issue in these papers and introduce an alternative proxy variable approach that avoids the collinearity problem. More recently, Gandhi et al. (2020) study nonparametric identification of production functions using proxy variables. They show how to combine the proxy variable approach with first-order conditions of cost minimization. My approach builds on these papers but differs in three main respects. First, it allows for factor-augmenting productivity in addition to Hicks-neutral productivity. I characterize the conditions under which both productivity shocks can be expressed as a function of inputs by nonparametrically inverting input demand functions. Second, I use control variables identified from data to overcome the endogeneity instead of proxy variables directly observed in the data. Third, I use the first-order conditions of cost-minimization for identification. Unlike Gandhi et al. (2020), firms have market power in the output market, but my approach requires two flexible inputs.

Three recent papers highlighted the importance of incorporating factor-augmenting technology into production functions (Doraszelski and Jaumandreu (2018), Raval (2019), Zhang (2019)). These papers study the change in factor-augmenting productivity and its relation to other economic variables.[7] A common feature in these papers is 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. My contribution to this literature is to relax the CES assumption and analyze identification with or without variation in input prices. With this, I relax the arbitrary restrictions on the production technology, such as common returns to scale and elasticity of substitution across firms. Also, my empirical application focuses on markups rather than technological change.[8]

This paper contributes to the literature on markup estimation from production data (Hall (1988), De Loecker and Warzynski (2012), Raval (2020)). This literature demon-

[7] Raval (2019) estimates a value-added CES production function with labor-augmenting productivity. Doraszelski and Jaumandreu (2018) estimate a gross CES production function with labor-augmenting productivity. Zhang (2019) allows for the materials-augmenting productivity in a CES specification.

[8] Another strand of literature uses firm-specific coefficients in the Cobb-Douglas to introduce unobserved firm-level heterogeneity. See Kasahara et al. (2015), Balat et al. (2019), Li and Sasaki (2017).

strates how to estimate markups from output elasticities under a cost minimization assumption. Doraszelski and Jaumandreu (2019) extends this literature by studying markup estimation in the presence of unobserved demand shocks and adjustment costs in flexible inputs. I investigate the role of production function specification on markup estimates and argue that allowing for firm-level heterogeneity in production technology is crucial for markup estimation. Lastly, a growing empirical literature analyzes markup growth and market power in the US. This literature uses Hicks-neutral production functions and finds that markups have risen in the US (De Loecker et al. (2020), Autor et al. (2020)). I show that a production function with labor-augmenting productivity points to a more muted rise in markups in the US manufacturing sector.

## 2 Model

I start by introducing a production function model with labor-augmenting technology and specifying the assumptions on firm behavior. I then show how to express unobserved productivity shocks as functions of observed inputs under the modeling assumptions.

### 2.1 Production Function with Labor-Augmenting Technology

The defining feature of my production function is that it allows for both labor-augmenting and Hicks-neutral technology without parametric restrictions. In this way, the model can accommodate a rich heterogeneity in production technology across firms.

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 quantity of output produced by the firm. Two unobserved productivity terms affect production. 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 given input combination. Finally, $\epsilon_{it} \in \mathbb{R}$ is a random shock to planned output.

The factors of production are classified into two types: *flexible* and *predetermined*. I assume that labor and materials are *flexible inputs*, meaning that the firm optimize them each period, and their level do not affect future production. In contrast, I assume that capital is a *predetermined input*, that is, the firm chooses the level of capital one period in advance. Therefore, the firm's current capital decision affects future production.

In each period, the firm chooses the level of flexible inputs to minimize the production cost based on its information set. I use $\mathcal{I}_{it}$ to denote firm $i$'s information set at period $t$, which includes productivity ($\omega_{it}^L$, and $\omega_{it}^H$), capital, past information sets, and other signals observed by the firm that are related to production and revenue. The information set is orthogonal to the random shock, i.e., $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}] = 0$. Under this assumption, $\epsilon_{it}$ can be viewed as measurement error in output or an ex-post productivity shock not observed (or predicted) by the firm before production.

I assume that firms are price-takers in the input market, facing $p_t^l$ and $p_t^m$ as the prices of labor and materials. The input prices do not vary across firms, but they can vary over time. My model and identification strategy extends to the case with heterogeneous and observed input prices at the expense of more notational complexity.[9] I develop this extension in Supplemental Appendix B.1. The model does not assume that output markets are perfectly competitive.

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, as indicated by the index $t$ in Equation (2.1). Although the industry-specific production function is restrictive, firm-specific productivities and lack of parametric restrictions introduce firm-level heterogeneity in production technology. In particular, the nonparametric production function allows for *observed* heterogeneity based on the input mix, whereas labor-augmenting and Hicks-neutral productivity allow for *unobserved* heterogeneity. These features of the model are crucial for explaining the large cross-firm heterogeneity observed in the data.

Despite its flexibility, the production function comes with some restrictions. In the model, factor-augmenting productivity affects only the labor input, implying that the quality of capital and materials are homogeneous across firms. In general, my framework can accommodate only one factor-augmenting productivity, and that factor has to be a flexible input. The main reason for this limitation is that a non-flexible input has dynamic implications, making it difficult to model its unobserved productivity. Therefore, I do not consider capital-augmenting production technology. However, the framework and identification results can accommodate models with materials-augmenting technology instead of labor-augmenting technology.[10]

I choose to consider labor-augmenting technology for three reasons. First, labor-augmenting productivity is an essential component of growth models, and its changes are an important subject in the literature (Acemoglu (2003)). Second, heterogeneity in $\omega_{it}^L$ reflects firm-level differences in labor quality. Several sources of labor quality, such as firms managing labor differently and human capital, might lead to differences in labor productivity across firms. Because these measures of labor quality are typically not observed in the data, it is natural to model them as unobserved heterogeneity. Finally, in most production datasets, labor's cost share has the most across-firm variation among all inputs, suggesting potential unobserved heterogeneity in the labor input.

My model differs from standard models in two significant ways: (i) It contains factor-

[9]However, I cannot accommodate market power in the input markets for heterogeneous input prices.

[10]Modeling materials' productivity could be important in some industries as it might reflect heterogeneity in input quality; see Fox and Smeets (2011).

augmenting technology, and (ii) It does not impose a parametric structure. These features are not trivial, and they have important implications that are not captured by other production functions. For an illustration, consider the Cobb-Douglas production function:

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

where lowercase letters denote the logarithms of the corresponding uppercase variables. This specification is nested in Equation (2.1) and has two key restrictions: (i) The production function is log-linear and (ii) $\omega_{it}^H$ is the only source of unobserved heterogeneity in production technology. These are strong restrictions with strong implications. First, Cobb-Douglas implies that revenue shares of flexible inputs are the same across firms, a prediction rejected in the data (Raval (2019)). Second, the literature has documented large heterogeneity in capital and labor intensities of production, which contradicts constant elasticity.[11] Finally, overwhelming empirical evidence suggests that elasticity of substitution is less than one.[12]

A widely employed solution to these issues is to use a more flexible Hicks-neutral production function, such as translog. However, Hicks-neutral productivity as the only source of unobserved heterogeneity is still restrictive and contradicts several empirical facts. The literature has documented a large and increasing heterogeneity in labor shares at the firm-level and a significant decline in labor share at the economy-level in many advanced economies. Most important, these facts have been attributed to within-industry changes and reallocation across firms (Karabarbounis and Neiman (2014), Kehrig and Vincent (2018), Autor et al. (2020)) and heterogeneity in production technology have been proposed as a mechanism (Oberfield and Raval (2014)). Failing to account for this heterogeneity will lead to biased production function estimates.

In brief, the inability of commonly used production functions to explain the data suggests that we need a more flexible production function.

## 2.2 Assumptions

In this section, I present assumptions and discuss their implications. The first assumption imposes a homothetic separability restriction on the production function. This assumption allows me to invert the firm's inputs decisions to express $\omega_{it}^L$ as an unknown function of inputs. Other assumptions concern firm behavior and the distribution of productivity shocks. They generalize the standard firm 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

[11] For example, the literature finds that large firms are more capital-intensive and less labor-intensive than small firms (Holmes and Schmitz (2010), Bernard et al. (2009)) and exporting firms are more capital-intensive than domestic firms (Bernard et al. (2007)).

[12] Although estimates vary, the consensus is that the aggregate elasticity of substitution between capital and labor is less than one (Antras (2004), Alvarez-Cuadrado et al. (2018)).

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*

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

***(ii)*** $h_t(K_{it}, \cdot, \cdot)$ *is homogeneous of arbitrary degree for all* $K_{it}$*.*
***(iii)*** *The firm minimizes production cost with respect to* $(L_{it}, M_{it})$ *given* $K_{it}$*, productivity shocks* $(\omega_{it}^L, \omega_{it}^H)$ *and input prices* $(p_t^l, p_t^m)$*.*
***(iv)*** *The elasticity of substitution between effective labor* $(\omega_{it}^L L_{it})$ *and materials is either greater than 1 for all* $(K_{it}, \omega_{it}^L)$ *or less than 1 for all* $(K_{it}, \omega_{it}^L)$*.*

Assumption 2.1(i) requires that the production function is separable in $K_{it}$ and a composite input given by $h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})$. This assumption is without loss of generality unless further restrictions are imposed. Assumption 2.1(ii) states that $h_t(\cdot)$ is homogeneous of arbitrary degree in effective labor and materials for any capital level. Combined with Assumption 2.1(i), this property is called weak homothetic separability, first introduced by Shephard (1953). Weak homothetic separability is common in models of consumer preferences and production functions, and most parametric production functions satisfy this property. Its key economic implication is that the ratio of two input's marginal products is a function of the inputs only through their ratio, so the firm can optimize the ratio of inputs, rather than optimizing the inputs separately.

Assumption 2.1(iii) specifies that firms choose the level of flexible inputs to minimize (short-run) production cost. The production cost does not involve capital, as it is a predetermined input. Cost-minimization is weaker than profit maximization because the cost is minimized for an arbitrary output level, not necessarily for the profit-maximizing level. Moreover, it is a static problem, so this assumption is agnostic about the firm's dynamic decisions. In my model, cost-minimization does not give rise to parametric first-order conditions. As a result, this assumption is less restrictive in my model than in most models in the literature, which usually make parametric restrictions.

Assumption 2.1(iv) implies that 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 and $\omega_{it}^L$. Assumption 2.1(iv) precludes this possibility. Next, I provide two examples of parametric production functions that satisfy the restrictions in Assumption 2.1.

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

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

My framework nests the CES production function with $h\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big) = \beta_l\big[\omega_{it}^L L_{it}\big]^{\sigma} + (1-\beta_l-\beta_m)M_{it}^{\sigma}$. This function is homogeneous of degree one and the elasticity of substitution is $\sigma$. 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:

$$Y_{it} = \Big(\beta_k K_{it}^{\sigma} + (1-\beta_k)\left(\beta_l\big[\omega_{it}^L L_{it}\big]^{\sigma_1} + (1-\beta_l)M_{it}^{\sigma_1}\right)^{\sigma/\sigma_1}\Big)^{v/\sigma}\exp(\omega_{it}^H)\exp(\epsilon_{it}). \tag{2.2}$$

This is a special case of my model with $h\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big) = \left(\beta_l\big[\omega_{it}^L L_{it}\big]^{\sigma_1} + (1-\beta_l-\beta_m)M_{it}^{\sigma_1}\right)^{1/\sigma_1}$, which is is homogeneous of degree one and the elasticity of substitution between effective labor and materials is $\sigma_1$. Since the Nested CES is a special case, my approach can be used to estimate this model, if one is willing to make parametric assumptions. Supplemental Appendix E explains in detail how to employ my approach for the estimation of CES and Nested CES production functions.

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 $v$. Returns to scale vary across firms in my model.

Next, I show the invertibility of the materials-to-labor ratio under Assumption 2.1.

**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 $\omega_{it}^L$*

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

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

*Proof.* See Appendix A.

The first part of this proposition states that the flexible input ratio is a function of only one model unobservables: labor-augmenting productivity. To see the intuition for this result, observe that the firm's relative labor and materials allocation depend on these inputs' relative marginal products, which in turn depends on the ratio of inputs by the homotheticity of $h_t(\cdot)$. Formally, the proof relies on the multiplicative separability of the firm's cost function under Assumption 2.1. Under homothetic separability and cost minimization, the cost function can be derived as:

$$C(\bar{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, p_t^m, p_t^l) = C_1(K_{it}, \omega_{it}^L, p_t^m, p_t^l) C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H), \tag{2.4}$$

where $C(\cdot), C_1(\cdot)$ and $C_2(\cdot)$ are unknown functions that depend on the production function, and $\bar{Y}_{it}$ is planned output. By Shephard's Lemma, the optimal input demands equal the derivatives of the cost function with respect to input prices $(p_t^m, p_t^l)$, implying that the ratio of materials to labor input does not depend on $C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H)$.

The second part of Proposition 2.1 establishes that $r_t(K_{it}, \omega_{it}^L)$ is strictly monotone in $\omega_{it}^L$. For strict monotonicity, the flexible input ratio should always move in the same direction as $\omega_{it}^L$, which affects the ratio of marginal products of labor and materials. Because the relationship between the input ratio and the ratio of marginal products depends on whether the elasticity of substitution is below or above one, Assumption 2.1(iv) restricts the elasticity of substitution.[13] Together, these two results provide a function, $r_t(K_{it}, \omega_{it}^L)$, that is strictly monotone in a scalar unobserved variable.

To relate this result to parametric production functions, note that under the CES assumption $r_t(K_{it}, \omega_{it}^L)$ has a known functional form, which is log-linear in $\omega_{it}^L$: $\log(\tilde{M}_{it}) = \sigma \tilde{p}_{it} + \log(\omega_{it}^L)$, where $\tilde{p}_{it}$ is the ratio of input prices. A common strategy in the literature is to estimate this equation using instruments for input prices and recover $\omega_{it}^L$ (Doraszelski and Jaumandreu (2018)).[14] However, this strategy crucially relies on the linear separability of $\log(\omega_{it}^L)$ obtained from the CES parametric form, which does not necessarily hold in more flexible production functions. Therefore, one contribution of this paper is to generalize the CES production function to an arbitrary functional form (subject to Assumption 2.1), and show invertibility under more general conditions.[15]

### 2.2.2 Other Assumptions

The rest of the assumptions generalize the standard proxy variable framework assumptions to accommodate labor-augmenting technology.

**Assumption 2.2** (First-Order Markov). *Productivity shocks (jointly) follow an exogenous first-order Markov process,*

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

According to this assumption, the current productivity shocks are the only variables in the firm's information set that are informative about future productivity. It is a natural generalization of the standard first-order Markov assumption from Olley and

[13] In particular, if materials and effective labor are substitutes, firms increase materials-to-labor ratio as $\omega_{it}^L$ increases, otherwise firms decreases materials-to-labor ratio as $\omega_{it}^L$ increases.

[14] To estimate this equation, one needs to observe heterogeneous input prices at the firm level.

[15] Doraszelski and Jaumandreu (2018) discuss informally how to use $\tilde{M}_{it}$ to control for $\omega_{it}^L$ without parametric assumptions.

Pakes (1996) to accommodate two-dimensional productivity shocks.[16] This assumption does not restrict the joint distribution of productivity shocks, which can be arbitrarily correlated. For example, firms with high Hicks-neutral productivity can also have high labor-augmenting productivity. Furthermore, there is no restriction on the first-order dynamics of productivity shocks: higher $\omega_{it}^H$ this period might be associated with higher $\omega_{it+1}^L$ next period.

**Assumption 2.3** (Monotonicity). *Firms' materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^L, \omega_{it}^H), \tag{2.5}$$

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

Introduced by Levinsohn and Petrin (2003), the assumption that materials demand is monotone in Hicks-neutral productivity is pervasive in the literature. However, in my model, firms' materials demands also depend on $\omega_{it}^L$, as it affects the marginal product of materials. Therefore, the materials demand function takes capital and two unobserved productivity shocks as arguments.[17]

Verifying this assumption requires the primitives of the output market, such as the demand function and competition, which I do not model in this paper.[18] However, this assumption is intuitive and expected to hold under general conditions. It says that everything else constant, more productive firms have a lower marginal cost, leading to a decline in prices and an increase in output.

Implicit in this assumption is that there is no unobserved heterogeneity in firms' perceived residual demand curves in the output market; otherwise, the materials input demand function should include firm-specific demand shocks, violating two-dimensional unobserved heterogeneity. Even though it is restrictive, this assumption accommodates some commonly used demand models such as monopolistic and Cournot competitions. Moreover, it allows for ex-post demand shocks after the planned output is chosen.[19] And finally, one can include observed demand shifters as in De Loecker (2011) to introduce observed heterogeneity.[20]

**Assumption 2.4** (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 $t-1$.*

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

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

[18] This is different from the monotonicity result for $\omega_{it}^L$, which depends only on primitives of supply.

[19] 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.

[20] See also Jaumandreu (2018), Doraszelski and Jaumandreu (2019) and Bond et al. (2020).

According to this assumption, investment becomes productive in the next period, implying that firms choose capital one period in advance.[21] As a result, $K_{it}$ belongs to the information set at period $t-1$, that is, $K_{it} \in \mathcal{I}_{it-1}$.

### 2.3 Invertibility: Expressing Unobserved Productivity Using Inputs

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

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

Similarly, Assumption 2.3 provides a monotonicity result for $\omega_{it}^H$ using materials demand function in Equation (2.5). Inverting that function yields

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \omega_{it}^L). \tag{2.7}$$

This function contains the unobserved $\omega_{it}^L$. Substituting for it from Equation (2.6):

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \tilde{r}_t(K_{it}, \tilde{M}_{it})) \equiv \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}). \tag{2.8}$$

Equations (2.6) and (2.8) demonstrate that the modeling assumptions and optimal firm behavior allow me to write unobserved productivity shocks as unknown functions of inputs. The intuition is that, even though productivity shocks are unobservable to the researcher, firms observe them before making their input decisions. This makes it possible to use the firm's input decisions to obtain information about productivity.

Invertibility is a standard condition in the proxy variable approach, which uses observables, such as investment or materials, as a proxy to control for unobserved productivity. However, the proxy variable approach is infeasible in my production function model due to multi-dimensional productivity. I address this problem with two innovations. First, I will develop a control variable approach building on the invertibility results in this section. Then, I will show how to exploit first-order conditions of cost minimization as an additional source for identification.

## 3 A Control Variable Approach to Production Function Estimation

The control variable approach relies on constructing variables from data that can serve as proxy for unobservables (Imbens and Newey (2009), Matzkin (2004)). 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 the standard control variable framework presented in Imbens and Newey (2009). They show how to derive a control variable when a single-dimensional

[21] The approach is robust to a weaker timing assumption, which can potentially provide efficiency gains. For a discussion, see Ackerberg (2016).

unobserved variable is strictly monotone in an observed variable and satisfies an independence condition. To apply the control variable approach to production function estimation, I make two innovations. First, I show that Markov and timing assumptions provide the necessary independence condition to develop control variables. Second, my model involves two-dimensional unobserved heterogeneity, for which the standard control variable approach does not work (Kasy (2011). I overcome this challenge by using the triangular structure of input demand functions in Equations (2.3) and (2.5).[22],[23]

I derive control variables in two stages. In the first stage, I derive the control variable for $\omega_{it}^L$. In the second stage, building on the first control variable, I derive the control variable for $\omega_{it}^H$. For notational simplicity, I omit time subscripts from functions in the rest of the paper.

### 3.1 Derivation of the 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 $\omega_{it}^L$ 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 $\tau$-th conditional quantile of $\omega_{it}^L$ given $(\omega_{it-1}^L, \omega_{it-1}^H)$. As such, we can view $u_{it}^1$ as the productivity rank of firm $i$ relative to firms with the same past productivity level.

Another interpretation of $u_{it}^1$ is unanticipated *innovation* to $\omega_{it}^L$, which determines the current period productivity given previous period's productivity. Unlike the standard definition of "innovation" to productivity, which is separable from and mean independent of past productivity,[24] $u_{it}^1$ is non-separable and independent. These properties of $u_{it}^1$ are key for utilizing the modeling assumptions to construct the control variables. In the previous section, I showed that $\tilde{M}_{it} = r(K_{it}, \omega_{it}^L)$. Substituting for $\omega_{it}^L$ from Equation (3.1) and using Equations (2.6) and (2.8), I obtain

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

for some unknown function $\tilde{r}(\cdot)$ and $W_{it}$ denotes the input vector, $W_{it} = (K_{it}, M_{it}, L_{it})$. Note that $\tilde{M}_{it}$ is strictly monotone in $u_{it}^1$ because $r(\cdot)$ is strictly monotone in $\omega_{it}^L$ by

[22]The control variable approach has a long tradition in industrial organization. It has been used for estimating demand (Bajari and Benkard (2005), Ekeland et al. (2004)), dynamic discrete choice models (Hong and Shum (2010)) and auctions (Guerre et al. (2009)). To the best of my knowledge, this paper is the first application of the control variable approach to two-dimensional unobserved heterogeneity.

[23]See also Ackerberg and Hahn (2015) for an application of Imbens and Newey (2009) approach in a production function with a single, non-separable productivity.

[24]In particular, $\omega_{it}^L = g(\omega_{it-1}^L, \omega_{it-1}^H) + \xi_{it}$ with $\mathbb{E}[\xi_{it} \mid \omega_{it-1}^L, \omega_{it-1}^H] = 0$

Assumption 2.1, and $g_1(\cdot)$ is strictly monotone in $u_{it}^1$ by construction. Next, I establish an independence result that allows me to develop a control variable using Equation (3.2).

**Lemma 3.1.** *Under Assumptions 2.2 - 2.4, $u_{it}^1$ is jointly independent of* $(K_{it}, W_{it-1})$.

*Proof.* See Appendix A.

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 the 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 for deriving a control variable: (i) $\tilde{r}\left(K_{it}, W_{it-1}, u_{it}^1\right)$ 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 $(K_{it}, W_{it-1})$.[25] The main idea is that two firms, $i$ and $j$, with the same capital and previous period's inputs, but different materials-to-labor ratios, differ only in their innovations to labor-augmenting productivity. That is, conditional on $K_{it}$ and $W_{it-1}$, $\tilde{M}_{it} > \tilde{M}_{jt}$ if and only if $u_{it}^1 > u_{jt}^1$. Therefore, ranking of firms in terms of $\tilde{M}_{it}$ is the same as ranking in terms of $u_{it}^1$. As a result, I can recover $u_{it}^1$ by looking at a firm's rank in the flexible input ratio. Using this result, I can express $\omega_{it}^L$ as a function of the control variable and past inputs:

$$\begin{aligned}
\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) &= g_1\big((\bar{r}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1\big), \\
&\equiv c_1\left(W_{it-1}, u_{it}^1\right),
\end{aligned} \tag{3.4}$$

where $c_1(\cdot)$ is an unknown function.

### 3.2 Derivation of the Control Variable for Hicks-Neutral Technology

Control variable derivation for $\omega_{it}^H$ is similar. The Skorohod representation of $\omega_{it}^H$ is:[26]

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

I use the monotonicity of materials in $\omega_{it}^H$ given by Assumption 2.3 to write

$$M_{it} = s\left(K_{it}, c_1\left(W_{it-1}, u_{it}^1\right), g_2\left(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2\right)\right),$$

[25] 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. If $\tilde{M}_{it}$ is strictly decreasing in $u_{it}^1$, then $F_{\tilde{M}_{it}|K_{it},W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}) = 1 - u_{it}^1$, a monotone transformation.

[26] $u_{it}^1$ is included in this representation, in addition to $(\omega_{it-1}^L, \omega_{it-1}^H)$, to account for the correlation between $\omega_{it}^L$ and $\omega_{it}^H$. If one relaxes the joint Markov assumption and assumes that innovations to two productivity shocks are independent conditional on past productivity, I do not need to condition on $u_{it}^1$. See Supplemental Appendix B.3 for control variable derivation under this assumption.

$$= s\left(K_{it}, c_1\left(W_{it-1}, u^1_{it}\right), g_2\left(\bar{r}\left(W_{it-1}\right), \bar{s}\left(W_{it-1}\right), u^1_{it}, u^2_{it}\right)\right),$$
$$\equiv \tilde{s}\left(K_{it}, W_{it-1}, u^1_{it}, u^2_{it}\right), \tag{3.6}$$

where $\tilde{s}(\cdot)$ is an unknown function. Note that $\tilde{s}\left(K_{it}, W_{it-1}, u^1_{it}, u^2_{it}\right)$ is strictly increasing in $u^2_{it}$ because $s\left(K_{it}, \omega^L_{it}, \omega^H_{it}\right)$ is strictly increasing in $\omega^H_{it}$ by Assumption 2.3, and $g_2\left(\omega^L_{it-1}, \omega^H_{it-1}, u^1_{it}, u^2_{it}\right)$ is strictly increasing in $u^2_{it}$ by construction.

**Lemma 3.2.** *Under Assumptions 2.2 - 2.4, $u^2_{it}$ is jointly independent of $(K_{it}, W_{it-1}, u^1_{it})$.*

*Proof.* See Appendix A.

Having monotonicity and independence, we can use Equation (3.6) to identify $u^2_{it}$ as:

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

where $F_{M_{it}|K_{it},W_{it-1},u^1_{it}}$ denotes the CDF of $M_{it}$ conditional on $(K_{it}, W_{it-1}, u^1_{it})$. Therefore, by comparing firms' materials levels, conditional on $(K_{it}, W_{it-1}, u^1_{it})$, we can recover the innovation to Hicks-neutral productivity, $u^2_{it}$. With this result, $\omega^H_{it}$ can be written as:

$$\omega^H_{it} \equiv c_2\left(W_{it-1}, u^1_{it}, u^2_{it}\right) \tag{3.8}$$

for an unknown function $c_2(\cdot)$ whose derivation is the same as Equation (3.4). This result and Equation (3.4) obtained in the previous subsection imply that conditional on previous period's inputs and the two control variables, there is no variation in productivity shocks. Therefore, using these control variables, we can control for endogeneity in production function estimation.[27]

**Remark 3.1** (Application to the Cobb-Douglas Production Function)**.** Since my control variable approach relies only on the Markov and timing assumptions, it can be applied to other functional forms. Supplemental Appendix E.1 demonstrates its application to Cobb-Douglas production function. For an overview, consider a value added Cobb-Douglas production function $y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega^H_{it} + \epsilon_{it}$. Using a control variable, $\omega^H_{it}$ can be written as $\omega^H_{it} = c(m_{it-1}, k_{it-1}, u_{it})$, where $u_{it} = F_{m_{it}|k_{it},m_{it-1},k_{it-1}}(m_{it} \mid k_{it}, m_{it-1}, k_{it-1})$. Substituting this into the production function gives a partially linear model:

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

with $\mathbb{E}[\epsilon_{it} \mid k_{it}, l_{it}, m_{it-1}, k_{it-1}, u_{it}] = 0$. As I discuss in Supplemental Appendix E.1, estimating the production function using this partially linear model has two advantages over the standard proxy variable approach. First, estimation is robust to the functional dependence problem highlighted by Ackerberg et al. (2015). That is because even if labor

[27]Using the same procedure and substituting past productivities recursively, we can write productivity shocks as $\omega^L_{it} = c_1\left(W_{it-k}, \{u^1_{it-l}\}^{k-1}_{l=0}\right)$ and $\omega^H_{it} = c_2\left(W_{it-k}, \{u^1_{it-l}\}^{k-1}_{l=0}, \{u^2_{it-l}\}^{k-1}_{l=0}\right)$ for any integer $k$, where $u^1_{it-l}$ and $u^2_{it-l}$ are defined as in Equations (3.3) and (3.7). This would lead to more identifying variation at the expense of having to estimate more control variables.

is a flexible input, there is variation in labor conditional on $(m_{it-1}, k_{it-1}, u_{it})$.[28] Second, there are efficiency gains, as my approach fully uses the independence condition given by the Markov assumption.

**Remark 3.2** (Functional Dependence Problem)**.** It is well-known that in Hicks-neutral production functions with two flexible inputs, after conditioning on capital and one flexible input, there is no variation in the other flexible input (Bond and Söderbom (2005), Ackerberg et al. (2015)). My model is robust to this problem because the second productivity shock, $\omega_{it}^L$, generates additional variations in flexible inputs.

**Remark 3.3** (Comparison to the Proxy Variable Approach)**.** My approach differs from the standard proxy variable approach in that control variables condition on 'less' current period information than proxy variables. The proxy variable approach relies on the invertibility of productivity shocks shown in Section 2.3 to control for endogeneity: $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$, $\omega_{it}^H = \bar{s}(K_{it}, M_{it}, \tilde{M}_{it})$. The proxy variable approach requires conditioning on the observed inputs to control for productivity shocks, and then using the last period's inputs as instruments. However, as pointed out by Gandhi et al. (2020), after conditioning on the proxy variables, there might have no variation in the inputs. In contrast, the control variable approach relies on a different representation of productivity shocks: $\omega_{it}^L = c_1\left(W_{it-1}, u_{it}^1\right)$, $\omega_{it}^H = c_2\left(W_{it-1}, u_{it}^1, u_{it}^2\right)$, which requires past inputs and control variables, $u_{it}^1$ and $u_{it}^2$, to control for endogeneity. Consequently, I do not need to condition on any of the current period inputs directly, which reduces the dimension of the conditioning variables. I achieve this by exploiting the independence property given by the Markov assumption. Studies using the proxy variable framework, such as Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), have also assumed that productivity follows a first-order Markov process, but they have only used its mean independence implication. In contrast, I fully exploit the Markov assumption, which results in control variables and efficiency gains. However, if the mean independence holds but independence does not, then my method would give inconsistent estimates, whereas proxy variable estimator would remain consistent.

## 4 Identification

This section discusses the identification of the output elasticities, the elasticity of substitutions, and productivity shocks. First, I point out a fundamental identification problem by showing that the production function and output elasticities cannot be 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. Finally, I

[28] To see this, if labor is perfectly flexible, we can write it as $l_{it} = l(k_{it}, \omega_{it}^H) = l\big(k(k_{it-1}, \omega_{it-1}, \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})\big) = l\big(k(k_{it-1}, s^{-1}(k_{it-1}, m_{it-1}), \nu_{it-1})), c(m_{it-1}, k_{it-1}, u_{it})\big)$, which gives $l_{it} =: \tilde{l}(k_{it-1}, m_{it-1}, u_{it}, \nu_{it-1})$, where $\nu_{it-1}$ corresponds to a vector of random variables that affects the firm's investment decision, such as investment prices and heterogeneous beliefs.

examine the identification of the other features of the production function and explore what further restrictions can be imposed on the production function for identification.

### 4.1 A Non-identification Result

Taking the logarithm of output and denoting $f = \log(F)$, $y_{it} = \log(Y_{it})$, I write the logarithm of the production function in an additively separable form in $\omega_{it}^H$ and $\epsilon_{it}$ as:

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

Since $h(\cdot)$ 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:

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

This reformulation is convenient because $\omega_{it}^L$ and $\tilde{M}_{it}$ become arguments in $h(\cdot)$. In Subsection 2.3, I showed that $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$. Substituting this into Equation (4.1)

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

This representation of the production function reveals an identification problem.

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

*Proof.* To see this result note that for arbitrary values of $(K_{it}, \tilde{M}_{it})$, the second argument of the $h$ function, $\bar{r}(K_{it}, \tilde{M}_{it})$, 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 $h$.[29] This implies that $h$ is not identified from variations in inputs and output.[30] □

Most of the economically interesting objects, such as the output elasticities or elasticity of substitutions, are a function of $h$, which underscores the challenge for identification. Suppressing the arguments of the functions, we can write output elasticities as

$$\theta_{it}^K := (f_1 + f_2 h_1) K_{it}, \qquad \theta_{it}^L := f_2 h_2 L_{it} \bar{r}(K_{it}, \tilde{M}_{it}), \qquad \theta_{it}^M := f_2 h_3 M_{it},$$

where $f_k$ denotes the derivative of $f$ with respect to its $k-$th component. I also use $\theta_{it}^j$ to denote the output elasticity with respect to $j$. Note that all the output elasticities depend on the derivatives of $h$, which is not identified.

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

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

[29] As I show in Supplemental Appendix E.1, with variation in input prices $\bar{r}(K_{it}, \tilde{M}_{it})$ depends also on the price ratio and functional dependence brakes down. However, one needs exogenous input prices to be able to use input prices for identification.

[30] Ekeland et al. (2004) obtains a similar nonidentification result for hedonic demand model estimation.

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

## 4.2 Identification of Output Elasticities

This section investigates the identification of the output elasticities and labor-augmenting productivity and obtains both positive and negative results. I find that the output elasticity of labor and materials are identified by exploiting first-order conditions, but the output elasticity of capital and labor-augmenting productivity are not identified without further restrictions.

### 4.2.1 Identifying the Ratio of Labor and Materials Elasticities

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

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

The first-order condition associated with this optimization problem is $F_V \lambda_{it} = p_t^{V}$, where $V \in \{M, L\}$, $F_V$ donates the marginal product of $V$, and $\lambda_{it}$ corresponds to the Lagrange multiplier. Multiplying both sides by $V_{it}/(Y_{it} p_{it})$ and rearranging gives,

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

where $p_{it}$ is the price of output. This expression involves the output elasticity and revenue share of a flexible input, and it is satisfied for all flexible inputs. 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 the output elasticities of labor and materials is identified as the ratio of revenue shares using the cost-minimization assumption.[31] The revenue shares are often observed in the data so we can calculate the ratio of elasticities without estimation.

[31] For this result, I only need that firms are cost-minimizers, labor and materials are flexible inputs and firms are price takers in the input markets. Therefore, this result is robust to violations of other assumptions in the model.

An important implication of using the first-order conditions is that the identification of output elasticities is possible only at the observed input levels, precluding a counterfactual exercise. I provide further discussion on this in later sections.[32]

Using the first-order conditions to estimate production functions has long been recognized in the literature, but mostly under parametric assumptions. Doraszelski and Jaumandreu (2013) and Grieco et al. (2016) use first-order conditions to identify the Cobb-Douglas and CES production functions, respectively. Gandhi et al. (2020) propose a method that use Equation (4.3) nonparametrically. They assume a perfectly competitive output market, which implies that elasticity equals flexible inputs' 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 subsection, 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 from $f$ and $\bar{h}$:*

$$\theta_{it}^V := \theta_{it}^M + \theta_{it}^L = f_2\big(K_{it}, L_{it}\bar{h}(K_{it}, \tilde{M}_{it})\big) L_{it}\bar{h}(K_{it}, \tilde{M}_{it}), \tag{4.5}$$

*which equals the elasticity of $F\big(K_{it}, L_{it}\bar{h}(K_{it}, \tilde{M}_{it})\big)$ with respect to its second argument.*

*Proof.* Using Equation (4.1), the materials and labor elasticities can be obtained as:

$$\theta_{it}^M = f_2 h_3 M_{it}, \qquad \theta_{it}^L = f_2(h - h_3\tilde{M}_{it})L_{it}.$$

The sum of the elasticities depends only on $h$, but none of its derivatives. This yields

$$\theta_{it}^V = \theta_{it}^M + \theta_{it}^L = f_2 h L_{it} = f_2\bar{h}L_{it}.$$

□

From this proposition, we see that identification of $f$ and $\bar{h}$ is sufficient for identifying the sum of flexible input elasticities. Importantly, we do not need to identify the structural production function, $h$, and labor-augmenting productivity shock.[33] Given the sum of elasticities and the ratio identified in the previous subsection, the labor and materials elasticities can be written as

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

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 elasticity of labor and materials separately.

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

[33]Note that even if $f$ and $\bar{h}$ are not uniquely identified, the sum of elasticities is uniquely identified. Assume there exists $(f, \tilde{h})$ and $(f', \tilde{h}')$ such that $f(K_{it}, L_{it}\tilde{h}) = f'(K_{it}, L_{it}\tilde{h}')$. Taking the derivative of this expression with respect to $L_{it}$ I obtain $f_2\tilde{h} = f_2'\tilde{h}'$, giving the same sum of flexible input elasticities.

### 4.2.3 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 $(f, \bar{h})$ and first order conditions.

**Proposition 4.3.** *Labor-augmenting productivity, the output elasticity of capital and the elasticity of substitutions are not identified from* $(f, \bar{h}, \theta_{it}^L, \theta_{it}^M)$.

*Proof.* See Appendix A.

With this result, I conclude that we can learn only the elasticity of flexible inputs using the reduced form production function and first-order conditions. This makes sense because the first-order conditions are only informative about the flexible inputs' output elasticities and do not help identify other features of the production function. As a solution to 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.

## 4.3 Identification under Further Restrictions

A potential solution to non-identification of the capital elasticity and labor-augmenting productivity 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 still not identified. Consider the following production function:

$$y_{it} = f\big(K_{it}, h(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{4.7}$$

This model differs from the main model in that $h$ does not take $K_{it}$ as an argument.[34] Since this is a special case, Proposition 2.1 applies to this production function with $\omega_{it}^L = \bar{r}(\tilde{M}_{it})$. Substituting this into Equation (4.7), I obtain the *reduced form* for the production function in Equation (4.7) as follows:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{4.8}$$

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

**Proposition 4.4.** *If we replace the production function in Assumption 2.1 with Equation (4.7), the capital elasticity is identified and labor-augmenting productivity is identified up to scale from* $(f, \bar{h}, \theta_{it}^L, \theta_{it}^M)$ *as:*

$$\theta_{it}^K = f_1(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})), \qquad \log(\omega_{it}^L) = \log(\bar{r}(\tilde{M}_{it})) = \int_{\underline{\tilde{M}}}^{\tilde{M}_{it}} b(\bar{M}_{it})d\bar{M}_{it} + a. \tag{4.9}$$

[34] This function is called strongly separable with respect to partition of labor and materials. A production function is strongly separable if the marginal rate of substitution between two inputs do not depend on another input (Nadiri (1982)).

*where $b(\cdot)$ is a function provided in the proof, which depends on $f$, $\bar{h}$ and the output elasticities of flexible inputs, and $a$ is an unknown constant.*

*Proof.* See Appendix A.

$\theta_{it}^K$ is identified under the additional restriction because $\omega_{it}^L$ is not a direct function of capital, implying that we can learn the capital elasticity from $f_1$. Identification of $\omega_{it}^L$ 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 $\omega_{it}^L$.

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

**Proposition 4.5.** *Under the conditions of Proposition 4.4 the elasticity of substitutions are not identified from $(f, \bar{h}, \theta_{it}^L, \theta_{it}^M)$.*

*Proof.* See Appendix A.

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 setup 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 data without exogenous variation in input prices. My result is similar in spirit because my model does not assume exogenous variation in input prices. In Supplemental Appendix B.1, I extend my model to have variation in input prices. In this extension, if prices are exogenous, and the elasticity of substitutions can potentially be identified.

An important implication of using the 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 $\omega_{it}^L$ 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 function require elasticities and productivity only for the firms observed in the data.

### 4.4 Imposing A Returns to Scale Restriction

My model can accommodate a returns to scale restriction on the production function. If one wants to restrict the return to scale to an unknown constant $v$, the production function takes the form

$$y_{it} = vk_{it} + f\big(1, \tilde{L}_{it} h(\omega_{it}^L, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it},$$

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

$$y_{it} = vk_{it} + \tilde{f}\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}, \tag{4.10}$$

where $\tilde{f} = f\big(1, \tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big)$. My identification results apply to this model. In particular, after estimating the flexible input elasticities and $v$, the capital elasticity can be obtained using the returns to scale restriction, $\theta_{it}^K = v - \theta_{it}^L - \theta_{it}^M$.

### 4.5 Summary of Models

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

$$y_{it} = f\big(K_{it}, h(K_{it}, \omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it} \tag{Weak Homo. Sep.}$$

$$y_{it} = f\big(K_{it}, h(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it} \tag{Strong Homo. Sep.}$$

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

$$y_{it} = \frac{v}{\sigma}\log\left(\beta_k K_{it}^\sigma + (1-\beta_k)\big(\beta_l\left[\omega_{it}^L L_{it}\right]^{\sigma_1} + (1-\beta_l)M_{it}^{\sigma_1}\big)^{\frac{\sigma}{\sigma_1}}\right) + \omega_{it}^H + \epsilon_{it} \tag{Nested CES}$$

$$y_{it} = \frac{v}{\sigma}\log\left(\beta_k K_{it}^\sigma + \beta_l\big(\omega_{it}^L L_{it}\big)^\sigma + (1-\beta_l-\beta_m)M_{it}^\sigma\right) + \omega_{it}^H + \epsilon_{it} \tag{CES}$$

Even though the paper focuses on the nonparametric models, Supplemental Appendix E shows how to estimate the CES and Nested CES production functions using the techniques developed in this paper. Therefore, a researcher interested in estimating a more restricted production function with labor-augmenting technology can use one of the nested models. When applied to these special cases, the identification strategy based on cost minimization and control variable approach is new.

There are two advantages to providing a family of models. First, comparing the results from a nested model and a general model tests the economic restrictions imposed by the nested model. Second, we can impose regularization based on economic theory. One can start with the least restrictive model, and if it is too demanding on the data, a nested model can be considered to improve precision. This is especially relevant for industries with a small number of firms.

## 5 Empirical Model and Data

### 5.1 Empirical Model

The purpose of my empirical model is to estimate the output elasticities and to infer markups from those estimates. To avoid the identification problems described above and to ease the demand on data, I use the strong homothetic production function in Equation

(4.7), which leads to the following estimating equation:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}. \tag{5.1}$$

In Section 4, I showed how to identify the output elasticities from $f$ and $\bar{h}$, so the goal is to identify these functions.[35] To control for Hicks-neutral productivity, I use the control variables developed in Equation (3.8), $\omega^H_{it} = c_2\left(W_{it-1}, u^1_{it}, u^2_{it}\right)$. Substituting this into Equation (5.1), the estimating equation can be written as:

$$y_{it} = f\big(K_{it}, L_{it}\tilde{h}(\tilde{M}_{it})\big) + c_2\left(W_{it-1}, u^1_{it}, u^2_{it}\right) + \epsilon_{it}. \tag{5.2}$$

Since $\epsilon_{it}$ is orthogonal to the firm's information set, we have:

$$\mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u^1_{it}, u^2_{it}] = 0. \tag{5.3}$$

Since all right-hand-side variables are orthogonal to the error term, Equation (5.2) can be estimated by minimizing the sum of squared residuals. However, Equation (5.3) is not the only moment restriction provided by the model. Recall that capital is a predetermined input that is orthogonal to the innovation to productivity shocks at time $t$, which can be used to augment the moment restriction in Equation (5.3). To see this, using the first-order Markov property of the productivity shocks, Hicks-neutral productivity can be expressed as

$$\omega^H_{it} \equiv \tilde{c}_3(\omega^L_{it-1}, \omega^H_{it-1}) + \xi_{it},$$

for an unknown function $\tilde{c}_3(\cdot)$, where $\xi_{it}$ 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. $\xi_{it}$ is commonly used in the proxy variable approach for constructing moments.

Since $(\omega^L_{it-1}, \omega^H_{it-1})$ can be written as functions of $W_{it-1}$, I obtain a second representation of $\omega^H_{it}$ as $\omega^H_{it} \equiv c_3(W_{it-1}) + \xi_{it}$. This representation gives another estimating equation:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + c_3(W_{it-1}) + \xi_{it} + \epsilon_{it}. \tag{5.4}$$

The error term, $\xi_{it} + \epsilon_{it}$, is orthogonal to the firm's information set at time $t-1$, which includes $K_{it}$ so we have $\mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}] = 0$, an additional moment restriction. Now I summarize the estimation problem by combining the models and moment restrictions. We have two estimating equations

$$\begin{aligned} y_{it} &= f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + c_2\left(W_{it-1}, u^1_{it}, u^2_{it}\right) + \epsilon_{it}, \\ y_{it} &= f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + c_3(W_{it-1}) + \xi_{it} + \epsilon_{it}, \end{aligned}$$

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

with two conditional moment restrictions:

$$\mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0, \tag{5.5}$$

$$\mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}, W_{it-1}] = 0. \tag{5.6}$$

Identification of output elasticities requires identification of the unknown functions $f$, $\bar{h}$ using these moment restrictions. Therefore, one question is whether moment restrictions in Equation (5.5) and (5.6) identify these functions. I analyze this question in Supplemental Appendix D, and show that the moment restriction in Equation (5.5) identifies $f$ and $\bar{h}$ except for special cases.[36] These cases include some testable support conditions on the derivatives of conditional CDF in Equation (3.7) and conditions on the derivatives of the production function.[37] Since Equation (5.5), by itself, generically identifies the output elasticities, the moment restriction in Equation (5.6) provides more identifying variations and efficiency gains.[38]

The estimation proceeds in two steps. In the first step, I estimate the control variable $u_{it}^2$ by estimating the conditional CDF in Equation (3.7). In the strongly separable model, $u_{it}^1$ 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 (5.5) and (5.6) for estimation.

### 5.1.1 Estimation Procedure

In this section, I provide an overview of the estimation procedure. A detailed estimation algorithm is given in Supplemental Appendix 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 eight-year rolling-window estimation for Compustat and three-year rolling window estimation for other datasets following De Loecker et al. (2020).[39]

The estimation involves two stages. In the first stage, I estimate the conditional distribution function in Equation (3.7). For this estimation, I first choose a grid of values in the support of $M$ and estimate the CDF at each point using a flexible logit model. For the second stage, I use a polynomial series approximation for the unknown functions. In particular, I use second-degree polynomials to approximate the production function and third-degree polynomials to approximate the control functions. Replacing the true

[36] This is called generic identification; see Lewbel (2019). One would ideally like to analyze the identification properties of moment restrictions in Equations (5.5) and (5.6) jointly. Since this is a difficult problem in a non-parametric model, I focus on the identification properties of Equation (5.5).

[37] I also provide these conditions for the homogeneous production function in Supplemental Appendix Proposition D.1.

[38] One can also use the recursive representation of productivity shocks, $\omega_{it}^H = c_2\left(W_{it-k}, \{u_{it-l}^1\}_{l=0}^{k-1}, \{u_{it-l}^2\}_{l=0}^{k-1}\right)$ and construct more moments using $\mathbb{E}[\epsilon_{it} \mid \{W_{it-l}^1\}_{l=0}^{k-1} \{u_{it-l}^1\}_{l=0}^{k-1}, \{u_{it-l}^2\}_{l=0}^{k-1}] = 0$ to increase efficiency.

[39] The window size is higher for the US than other countries because of the smaller sample size.

functions with the approximations yields

$$y_{it} = \widehat{f}\big(K_{it}, L_{it}\widehat{\bar{h}}(\tilde{M}_{it})\big) + \widehat{c}_2(W_{it-1}, \widehat{u}^1_{it}, \widehat{u}^2_{it}) + \widehat{\epsilon}_{1it},$$
$$y_{it} = \widehat{f}\big(K_{it}, L_{it}\widehat{\bar{h}}(\tilde{M}_{it})\big) + \widehat{c}_3(W_{it-1}) + \widehat{\xi}_{it} + \widehat{\epsilon}_{2it}.$$

I construct an objective function using the moment restrictions in Equations (5.5) and (5.6). In particular, I use the sum of squared residuals from Equation (5.5) and timing moments from Equation (5.6) to obtain the following objective function:

$$J(\widehat{f}, \widehat{\bar{h}}, \widehat{c}_2, \widehat{c}_3) = \underbrace{\frac{1}{TN}\sum_{i,t}\widehat{\epsilon}^2_{1it}}_{\text{Sum of Squared Residuals}} + \underbrace{\Big(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K_{it}\Big)^2 + \Big(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K^2_{it}\Big)^2}_{\text{Timing Moments}}$$

I minimize the objective function for estimation[40]. Estimating $\widehat{c}_2(W_{it-1})$ and $\widehat{c}_3(W_{it-1})$ is computationally simple as they can be partialed out for a given $(\widehat{f}, \widehat{\bar{h}})$. Thus, the estimation requires searching for $\widehat{f}$ and $\widehat{\bar{h}}$ to minimize the objective function. After obtaining the estimates for $f$ and $\bar{h}$, I calculate the output elasticities as described in Equations (4.5), (4.6) and (4.9).

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 procedures treat firms as independent observations and resample firms with replacement.

## 5.2 Data

I use panel data from manufacturing industries in five countries: Chile, Colombia, India, Turkey, and the US. The data source is Compustat for the US and manufacturing censuses for other countries. Table 1 provides some descriptive statistics.

### 5.2.1 Chile, Columbia, India, Turkey

The data for the four developing countries are traditional plant-level production data collected through censuses. The first dataset comes from the census of Chilean manufacturing plants conducted by Chile's Instituto Nacional de Estadística (INE). It covers all firms from 1979-1996 with more than ten employees. Similarly, the Colombian data come from the manufacturing census covering all manufacturing plants with more than ten employees from 1981-1991. These datasets have been used extensively in previous studies. The Turkish dataset is from the Annual Surveys of Manufacturing Industries, conducted by the Turkish Statistical Institute, and covers all establishments with ten or more employees between 1983 and 2000. Finally, the Indian data are from the Annual

[40] This estimation strategy is similar to Wooldridge (2009)'s joint estimation framework.

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 |

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

Survey of Industries conducted by the Indian Statistical Institute, covering the plants with 100 or more employees from 1998 to 2014.

From these datasets, I obtain the measures of inputs and output for estimating the production functions. I obtain materials inputs by deflating the materials cost using the appropriate deflators. To calculate the materials cost, I add separate measures of materials for non-energy raw materials and energy. The labor input measure is the number of manufacturing days for India and the number of workers for Chile, Colombia, and Turkey. I obtain capital either via the perpetual inventory method or from deflated book values. I remove outliers based on labor's share of revenue, materials' share of revenue, and the combined variable input share of the revenue for each industry. To obtain precise estimates, I limit my sample to industries with at least an average of 250 plants per year. I provide details about the data collection, industries, and descriptive statistics in Supplemental Appendix A.

#### 5.2.2 US

The Compustat sample contains all publicly traded manufacturing firms in the US between 1961–2014. It includes information compiled from firm-level financial statements, including sales, total input expenditures, number of employees, capital stock formation, and industry classification. From this information, I obtain measures of labor, materials, and capital inputs and produced output. My output measure is the net sales deflated by a common 3-digit deflator, and my labor measure is the number of employees. Compustat does not report separate expenditures for materials. To address this issue, 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. For the details of the variables' construction, see Supplemental Appendix A.5

Some concerns about Compustat data are worth mentioning in the context of production function estimation. Compustat is not representative of the general economy as it only includes publicly traded firms. These firms are bigger, older, and more capital intensive. Also, Compustat is compiled from accounting data, which is a low-quality data source compared to manufacturing censuses. Despite these concerns, I use Compustat dataset because some of the recent findings on the rise of market power in the US have been obtained using Compustat (De Loecker et al. (2020)). I aim to revisit those findings and explore how using flexible production function technology affects the results. To alleviate the concerns on Compustat, I use high-quality datasets from four developing countries and check whether my results hold in these datasets.

Another limitation of the data is the lack of physical output quantities, a concern if sales reflect firm-level demand heterogeneity independent of product quality. Even though physical output or firm-level price indexes are available in some recent datasets, I aim to demonstrate my method on several commonly used datasets. It is worth noting that some features of my model mitigate concerns about using sales-based output measures. Methodologically, identifying the ratio of elasticities and controlling for labor-augmenting productivity do not require physical output measures, so they are robust to using sales-based output measures. Empirically, my focus is to understand how ignoring labor-augmenting productivity biases the results. If data quality issues affect different models similarly, they should not affect the comparisons. Finally, Raval (2020) shows that using physical output does not address the problems with the Hicks-neutral production functions I document in the empirical section.

# 6 Empirical Results: Production Function

This section presents results from the empirical model. I use production function estimates to discuss several findings. First, my model generates different elasticity estimates compared to the Cobb-Douglas model in all countries. Second, I find significant substantial heterogeneity in output elasticities, which are related to firm size and export. This section focuses on the comparisons with Cobb-Douglas as it is the most commonly used specification. Supplemental Appendix F compares my results with the translog production function and Nested CES production function with labor-augmenting productivity.

## 6.1 Output Elasticities

Table 2 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 Ackerberg et al. (2015) (henceforth, ACF) 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.

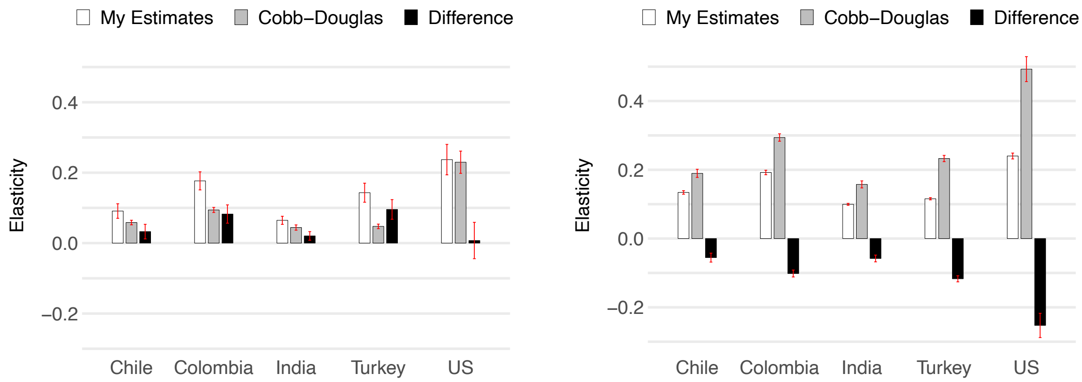
The estimates point out large differences in the average elasticity estimates between specifications. Cobb-Douglas generates higher labor elasticities and lower capital elastic-

Table 2: Sales-Weighted Average Output Elasticities for Three Largest Industries

| | **Industry 1** | | | **Industry 2** | | | **Industry 3** | | |
|---|---|---|---|---|---|---|---|---|---|
| | FA | ACF | OLS | FA | ACF | OLS | FA | ACF | OLS |
| | | | | *Chile (311, 381, 321)* | | | | | |
| Capital | 0.09 | 0.04 | 0.05 | 0.12 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| | (0.01) | (0.00) | (0.00) | (0.03) | (0.01) | (0.01) | (0.03) | (0.01) | (0.01) |
| Labor | 0.1 | 0.14 | 0.14 | 0.19 | 0.31 | 0.31 | 0.19 | 0.23 | 0.23 |
| | (0.00) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Materials | 0.79 | 0.87 | 0.88 | 0.69 | 0.69 | 0.69 | 0.66 | 0.72 | 0.72 |
| | (0.02) | (0.01) | (0.01) | (0.04) | (0.01) | (0.01) | (0.05) | (0.01) | (0.01) |
| Rts | 0.98 | 1.06 | 1.06 | 1 | 1.09 | 1.09 | 0.94 | 1.04 | 1.04 |
| | (0.02) | (0.01) | (0.01) | (0.04) | (0.01) | (0.01) | (0.06) | (0.01) | (0.01) |
| | | | | *Colombia (311, 322, 381)* | | | | | |
| Capital | 0.13 | 0.07 | 0.07 | 0.12 | 0.07 | 0.08 | 0.19 | 0.13 | 0.13 |
| | (0.02) | (0.00) | (0.00) | (0.02) | (0.02) | (0.02) | (0.03) | (0.01) | (0.01) |
| Labor | 0.11 | 0.18 | 0.18 | 0.3 | 0.46 | 0.45 | 0.25 | 0.36 | 0.36 |
| | (0.00) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.01) | (0.01) | (0.01) |
| Materials | 0.78 | 0.8 | 0.8 | 0.63 | 0.56 | 0.54 | 0.56 | 0.61 | 0.61 |
| | (0.02) | (0.00) | (0.00) | (0.02) | (0.01) | (0.01) | (0.04) | (0.01) | (0.01) |
| Rts | 1.01 | 1.05 | 1.05 | 1.05 | 1.09 | 1.06 | 1 | 1.1 | 1.09 |
| | (0.03) | (0.00) | (0.00) | (0.03) | (0.01) | (0.01) | (0.05) | (0.01) | (0.01) |
| | | | | *India (230, 265, 213)* | | | | | |
| Capital | 0.07 | 0.05 | 0.05 | 0.09 | 0.02 | 0.04 | 0.04 | 0.03 | 0.07 |
| | (0.01) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) |
| Labor | 0.08 | 0.09 | 0.09 | 0.18 | 0.43 | 0.34 | 0.06 | 0.37 | 0.33 |
| | (0.00) | (0.01) | (0.01) | (0.00) | (0.02) | (0.02) | (0.00) | (0.04) | (0.04) |
| Materials | 0.82 | 0.84 | 0.84 | 0.67 | 0.54 | 0.56 | 0.82 | 0.65 | 0.56 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.04) | (0.04) |
| Rts | 0.96 | 0.98 | 0.98 | 0.94 | 1 | 0.94 | 0.93 | 1.05 | 0.97 |
| | (0.01) | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) | (0.03) |
| | | | | *Turkey (321, 311, 322)* | | | | | |
| Capital | 0.14 | 0.03 | 0.03 | 0.08 | 0.03 | 0.03 | 0.07 | 0.03 | 0.03 |
| | (0.02) | (0.00) | (0.00) | (0.03) | (0.00) | (0.00) | (0.03) | (0.01) | (0.01) |
| Labor | 0.14 | 0.22 | 0.22 | 0.08 | 0.17 | 0.17 | 0.12 | 0.29 | 0.29 |
| | (0.00) | (0.01) | (0.01) | (0.00) | (0.01) | (0.01) | (0.00) | (0.01) | (0.01) |
| Materials | 0.7 | 0.79 | 0.78 | 0.83 | 0.84 | 0.84 | 0.9 | 0.72 | 0.71 |
| | (0.02) | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) | (0.02) | (0.01) | (0.01) |
| Rts | 0.98 | 1.04 | 1.04 | 0.99 | 1.04 | 1.04 | 1.09 | 1.04 | 1.03 |
| | (0.03) | (0.00) | (0.00) | (0.03) | (0.00) | (0.00) | (0.04) | (0.01) | (0.01) |
| | | | | *US (33, 32, 31)* | | | | | |
| Capital | 0.24 | 0.21 | 0.2 | 0.22 | 0.24 | 0.23 | 0.31 | 0.28 | 0.29 |
| | (0.03) | (0.01) | (0.01) | (0.05) | (0.03) | (0.03) | (0.07) | (0.05) | (0.05) |
| Labor | 0.28 | 0.52 | 0.52 | 0.21 | 0.47 | 0.46 | 0.21 | 0.44 | 0.45 |
| | (0.01) | (0.02) | (0.02) | (0.01) | (0.03) | (0.03) | (0.01) | (0.05) | (0.05) |
| Materials | 0.58 | 0.26 | 0.26 | 0.6 | 0.31 | 0.3 | 0.55 | 0.23 | 0.24 |
| | (0.01) | (0.02) | (0.02) | (0.04) | (0.06) | (0.06) | (0.03) | (0.06) | (0.06) |
| Rts | 1.1 | 0.99 | 0.98 | 1.03 | 1.02 | 0.99 | 1.07 | 0.95 | 0.98 |
| | (0.03) | (0.01) | (0.01) | (0.05) | (0.01) | (0.01) | (0.07) | (0.02) | (0.02) |

Note: Comparison of sales-weighted average output elasticities produced by different methods. FA refers to my estimates, ACF refers to Ackerberg et al. (2015) 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 code of the largest, second largest and third largest industries, respectively, in each country. Industry codes are provided in parentheses in each panel. Corresponding industry names are Food Manufacturing (311), Equipment Manufacturing (381), Paper Manufacturing (322), Glass Manufacturing (311), Cotton ginning (230), Textile (265). Bootstrapped standard errors in parentheses (100 iterations).

Figure 1: Average Capital and Labor Elasticities Comparison




(a) Capital Elasticity

(b) Labor Elasticity

Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. 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).

ities than my model for most industries. Low labor' share is consistent with labor's low revenue share in the data, as we will see in the next section. Lastly, looking at the OLS estimates, I find small and insignificant differences between the ACF and OLS methods, whereas my estimates are significantly different from the OLS estimates. This suggests that my method corrects the biases in the OLS estimates.

To see the differences in estimates across methods more clearly, I report the economy-level output elasticities of capital and labor in Figure 1.[41] The results suggest that I estimate a higher capital elasticity and lower labor elasticity in all countries. The difference is statically significant in all countries for labor and in all countries except the US for capital[42]. Drawing the same conclusions in all datasets provides strong evidence that these results are robust to sample periods and country-specific characteristics.[43]

### 6.2 Heterogeneity in Output Elasticities

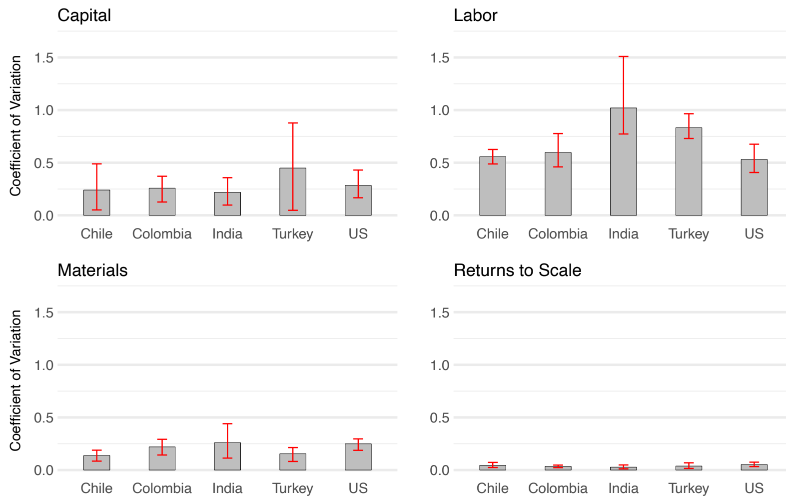
This section examines the within-industry heterogeneity in the output elasticities and relates it to other economic variables. In particular, I test: (i) Are large firms more capital-intensive and less flexible input-intensive? (ii) Are exporters more capital-intensive? The literature has found firm-level heterogeneity along many dimensions, including productivity, labor share, and size (Van Reenen (2018)); however, there is limited evidence on heterogeneity in production technology. Moreover, this section provides evidence for the

[41]Other elasticity estimates are reported in Supplemental Appendix Figure G.6.

[42]One reason US results differ from other countries is the US data includes only big and public firms.

[43]A concern in production function estimation is measurement error in capital, which could be more severe in a nonparametric model. With measurement error, the capital elasticity estimates will be biased towards zero, and other elasticities will be biased upwards since other inputs are positively correlated with capital. I verify this prediction using simulations in Supplemental Appendix F.1. Since I find larger capital elasticity and lower labor elasticity, measurement error cannot explain my results.

Figure 2: Average Coefficient of Variation




Note: This figure shows the average coefficient of variation for the output elasticities averaged across industries over years. In each panel, each bars reports the average CoV of the output elasticity of the corresponding input for all countries. The error bars indicate the 10th and 90th percentile of the distribution.

model's external validity since firm size and export are outside the model.

I estimate the coefficient of variation (CV) of the output elasticities within each industry-year group to measure heterogeneity. 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 show that this result is not driven by only a small number of industries. Finally, I find little heterogeneity in returns to scale, a reasonable finding because too large or too small returns to scale would not be consistent with the economic theory.

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? Although the evidence on heterogeneity in production technology is scarce, the literature has produced two findings on the relationship between production functions and other economic variables: (i) large firms are more capital-intensive than small firms (Holmes and Mitchell (2008), Kumar et al. (1999)), (ii) exporting firms are more capital-intensive than domestic firms (Bernard et al. (2009)). I use my elasticity estimates to revisit these findings.

To understand the relationship between output elasticities and firm size, I estimate:

$$d_{ijt} = \alpha_0 + \gamma \times \text{Firm Size}_{ijt} + \delta_{jt} + \epsilon_{it}, \tag{6.1}$$

where $j$ indexes the 4-digit industry, so $\delta_{jt}$ denotes the industry-year interaction fixed

Table 3: Regressions of the Output Elasticities on Firm Size

| | Chile | Colombia | India | Turkey | US |
|---|---|---|---|---|---|
| Capital Elasticity | 0.008 | 0.02 | 0.006 | 0.016 | 0.025 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Flexible Input | -0.023 | -0.02 | -0.011 | -0.012 | -0.004 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Capital Intensity | 0.253 | 0.303 | 0.387 | 0.396 | 0.228 |
| | (0.003) | (0.002) | (0.002) | (0.002) | (0.001) |

Notes: Regressions control for 4 digit industry-year fixed effects. Firm size is proxied by log sales. Each row corresponds to a separate regression with left-hand side variable given in the first column. Standard errors are clustered at the firm level.

effects. $\gamma$ is the coefficient of interest. I estimate separate regressions for three outcomes: the flexible input elasticity, capital elasticity, and capital intensity. Following the literature, I define capital intensity as log capital elasticity divided by log labor elasticity. I use log-sales to proxy for firm size.

Table 3 reports the coefficient estimates. Focusing on capital intensity, I find that large firms are more capital intensive than small firms in all countries. This finding remains similar when I use the capital elasticity as the outcome variable. Finally, negative and statistically significant coefficient estimates in the second row suggest that flexible input elasticity is negatively associated with firm size. Overall, these findings agree with the literature, which finds that large firms use more capital and less labor than small firms.

The second estimation concerns the relationship between capital intensity and exports. I consider the same model as above, replacing firm size with an indicator variable that equals one if the firm exports and zero otherwise. I estimate this model on the Chilean and Indian datasets since firm-level export data are available only for these countries. The outcome variables are capital intensity and capital elasticity. The coefficient of interest reflects the average difference of the outcome variable between exporters and non-exporters. Results in Table 4 suggest that exporting firms are more capital intensive in both countries. This finding is also consistent with the existing empirical evidence.

In brief, this section documents substantial heterogeneity in production technology related to firm size and export status. This analysis can also be seen as external validation for my model because the explanatory variables, firm size, and export, are outside the production function model. I show that these variables explain the output elasticities in a way that is consistent with the literature.

## 7 Inferring Markups from Production Function

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

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-

Table 4: Regression of Capital Intensity on Export Status

| | Chile | India |
|---|---|---|
| Capital Elasticity | 0.014 | 0.005 |
| | (0.000) | (0.000) |
| Capital Intensity | 0.136 | 0.4 |
| | (0.018) | (0.016) |

Notes: The regressions control for 4 digit industry-year fixed effects. Each row corresponds to a separate regression with the left-hand side variable given in the first column. Standard errors are clustered at the firm level.

minimizers with respect to at least one flexible input and firms take input prices as given. In particular, markup is given by $\mu_{it} := \theta^V_{it}/\alpha^V_{it}$, where $\mu_{it}$ denotes the firm-level markup, and it equals the output elasticity of a flexible input, divided by its revenue share. 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. Moreover, this identity holds for all flexible inputs, so we need an estimate of only one input's elasticity.

In recent years, markup estimation using this approach has become popular. Since this method does not require a model of competition, researchers estimated markups at the macro level using production data. (De Loecker et al. (2016), Autor et al. (2020), Traina (2018)). The evidence from this literature ignited a debate over the rise in market power in the US and other developed countries (Basu (2019), Berry et al. (2019)).

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

Output elasticity is the only estimated component of markup when markup is estimated from output elasticities. As a result, the bias in output elasticity directly translates into 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 find that the elasticity estimates differ substantially. This suggests that the production function specification is critical when estimating markups from production data.[44] Motivated by this, this section discusses implications of functional form assumptions on markups and then shows that labor augmenting productivity provides a solution to some puzzling results in the literature.

**Heterogeneity in Markups.** The most common production function specification in the literature is Cobb-Douglas. Under this assumption, output elasticities are equal across firms in the same industry, so the cross-section variation in markups comes only from revenue shares. If the true output elasticities vary across firms, then Cobb-Douglas would give an incorrect markup distribution. This point is particularly important for

[44] This is in contrast to productivity estimates, which are shown to be robust to the specification; see Foster et al. (2017), Van Biesebroeck (2008) and Van Beveren (2012).

studies that relate markups to other firm-level observables. In fact, if the true production function is Cobb-Douglas, then industry fixed-effects in a regression of (log) markups on another variable are sufficient to account for variation output elasticities.

**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 markups estimates (Doraszelski and Jaumandreu (2019), Raval (2020)). This evidence suggests that at least one assumption required to estimate markups from production data is violated.

Raval (2020) formally tests the production function approach using its implication that markups from two flexible inputs should be identical. He estimates markups from labor and materials under the Cobb-Douglas specification in five datasets. He finds that the two markup measures are negatively correlated and suggest different trends. He then examines the 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.[45]

Raval (2020)'s results suggest unobserved heterogeneity in elasticities as a potential solution to conflicting markups estimates. This paper builds on Raval (2020) and show that labor-augmenting productivity ensures identical markup estimates from labor and materials and provides a natural solution to this problem. Two key components of my approach lead to this outcome: (1) the presence of labor-augmenting productivity and (2) using the ratio of revenue shares to identify the ratio of elasticities. The latter immediately implies that the two markups estimates are the same:

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

where $\mu_{it}^L$ and $\mu_{it}^M$ denote markup estimates from labor and materials. However, the presence of labor-augmenting productivity is key to being able to use the ratio of revenue shares to identify the ratio of elasticities. As shown in Section 8.1, without the labor augmenting productivity the identity in Equation (7.1) is rejected, suggesting that the model is not internally valid. The additional unobserved heterogeneity explains the discrepancy and makes the model internally valid. Summarizing, this section, together with the evidence in Section 6.2, establishes both internal and external validity of the model.

### 7.2 Decomposing Markups: The Role of Production Function Estimation

This section presents a markup decomposition framework to quantify the role of the production function. I show that production function estimation can bias the aggregate

[45] To account for factor-augmenting productivity, he uses the quintile cost-share method where quantiles are based on labor cost to materials cost ratio. This method yields positively correlated markups.

markup through two channels: (i) bias in the average output elasticity and (ii) firm-level heterogeneity in the output elasticities.

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. Recently, researchers used the aggregate markup to measure the change in market power in the US and other economies (De Loecker et al. (2020), Diez et al. (2018)). To assess the influence of production function estimation on the estimated aggregate markup, I apply the Olley-Pakes decomposition, which separates a weighted average into two parts: (1) an unweighted average and (2) covariance between the weight and variable of interest. To implement the OP decomposition, I look at the aggregate log markup, $\tilde{\mu}_t = \sum w_{it} \log(\theta_{it}) - \sum w_{it} \log(\alpha_{it})$, a difference of two weighted averages. Applying the OP decomposition to both terms:

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

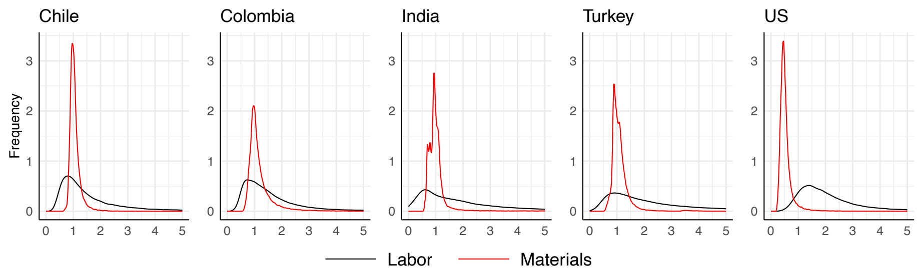
The aggregate log markup is composed of four parts. The first two parts involve the output elasticity: (1) is the average of log elasticity, and (2) is the covariance between firm size and log elasticity. The last two parts involve the revenue share: (3) is the average revenue share, and (4) is the covariance between firm size and log revenue share. 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 the data. Therefore, analyzing the first two components will reveal how biases in production function estimation translate into markup estimates.

**Bias from the Average Output Elasticity.** The first component in the decomposition is the average elasticity, reflecting the economy's underlying production technology. Under misspecification, this component will be estimated with bias, which directly translates into bias in the aggregate markup. The elasticity estimates in the previous section suggested that Cobb-Douglas overestimates the flexible input elasticity. Therefore, the bias from this source should be positive.

**Bias from Heterogeneity in Production Technology.** The second component in the decomposition 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. If the production function does not account for this heterogeneity, then the aggregate markup will be biased. The bias is positive when large firms have lower flexible input elasticity than small firms and negative otherwise. My estimates and existing empirical evidence suggest that this source of bias is also positive.

If the first two components change over time, we should also expect bias in the change in markups. This can happen, for example, if large firms become more capital-intensive over time, leading to an increase in the second component. Failing to capture this trend

Figure 3: 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 from the Cobb-Douglas specification estimated using the Ackerberg et al. (2015) procedure for each country.

in production technology would overestimate the change in the aggregate markup.

Together, this section makes two arguments that motivate a flexible production function for markup estimation. It is critical to (i) estimate the average output elasticity correctly and (ii) account for firm-level heterogeneity in the output elasticities.

## 8 Empirical Results: Markups

I estimate markups using the output elasticities reported in Section 6 and investigate whether my markup estimates are systematically different from those generated by Cobb-Douglas and other production functions. I find that my aggregate markup estimates are lower than the Cobb-Douglas estimates in all countries. Two factors drive this difference: (1) Cobb-Douglas overestimates the average output elasticity, and (2) Cobb-Douglas does not capture the negative correlation between firm size and the output elasticity of flexible input. Then I look at whether the differences in production function estimates affect the estimates for the trend in markups in the US.

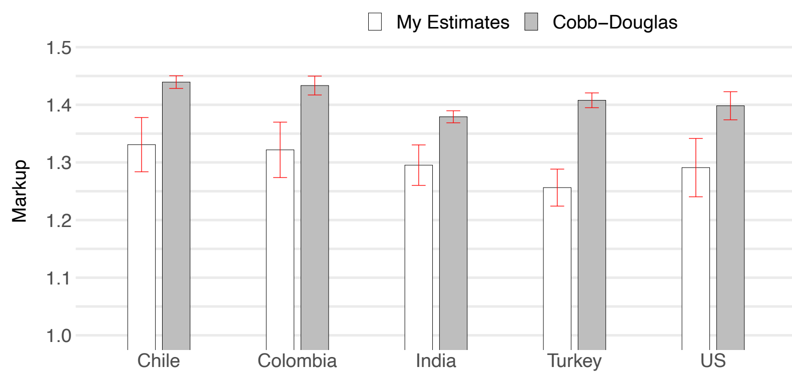
### 8.1 Testing the Cobb-Douglas Specification using Markups

As discussed in Section 7, testing the equality of markups from labor and materials elasticities 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 ACF method for markup estimation. Figure 3 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 in all countries.[46] Moreover, both distributions indicate that a significant fraction of firms has markups below one. These results provide strong evidence against the Hicks-neutral production functions.[47]

[46]Since this is a test for the model, violations of other assumptions such as firm-specific output prices, mismeasured inputs can generate this result. However, Raval (2020) tests these alternative possibilities

Figure 4: Average Markups Comparison




Notes: Comparison of sales-weighted average markups produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95 percent 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): $y_{it} = \beta_k k_{it} + \beta_v v_{it} + \omega_{it} + \epsilon_{it}$. Here, $v_{it}$ is the combined flexible input of labor and materials, defined as the deflated sum of labor and materials cost. I estimate this model using the ACF method and calculate markups as $\mu_{it}^{CD} = \beta_v / \alpha_{it}^V$. 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 by the Cobb-Douglas production function. After finding significant differences between the two estimates, I use the markup decomposition framework presented in Section 7.2 to understand what drives this difference.

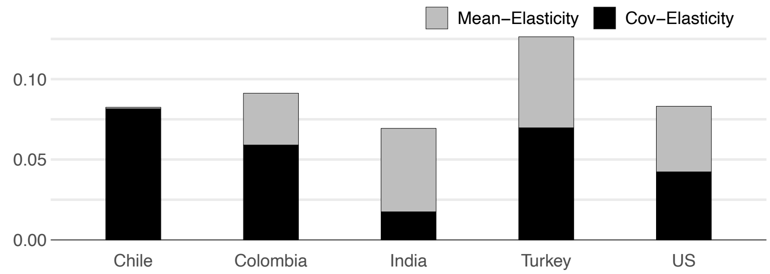
For each country, I first calculate the sales-weighted markup for every year and then take the average over the years.[48] Figure 4 displays the aggregate markups from the two methods, along with the 95% confidence interval. My model generates aggregate markups that are significantly smaller than the Cobb-Douglas estimates in all countries. The difference ranges from 0.1 to 0.2, an important magnitude when markups are interpreted as market power. Furthermore, reaching the same conclusion in all countries provides

and concludes that the likely reason is the Hicks-neutral productivity assumption.

[47] An alternative explanation the model ruled out is labor market frictions, which might exist in some industries and countries. However, observing similar patterns in all datasets suggests that labor market frictions cannot explain these results. Moreover, Raval (2020) obtains similar results when using materials and energy, which are less likely to be subject to frictions.

[48] Edmond et al. (2018) argue that cost-weighting is more appropriate for understanding the welfare implications of the change in markups. I report cost-weighted estimates in Supplemental Appendix G and find qualitatively similar results.

Figure 5: 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 estimated using the ACF procedure.

compelling evidence that the results are not driven by country-specific characteristics.[49]

What drives these differences in markup estimates? I answer this question by decomposing markups into its components, as presented in Section 7.2. To focus on the first two components, I take the difference between markup estimates. The third and fourth components cancel out, so the difference in markups is explained by the differences in the mean elasticity and covariance between firm size and elasticity. I plot these differences in Figure 5. This result highlights two key reasons behind the difference in markup estimates between the two methods. First, the Cobb-Douglas production function overestimates the flexible 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 suggest that large firms are more capital-intensive and less flexible input-intensive, leading to a negative correlation between firm size and the flexible input elasticity. Both of these factors generate upward bias in the Cobb-Douglas markup estimates.

### 8.3 Markups Comparison: Trend

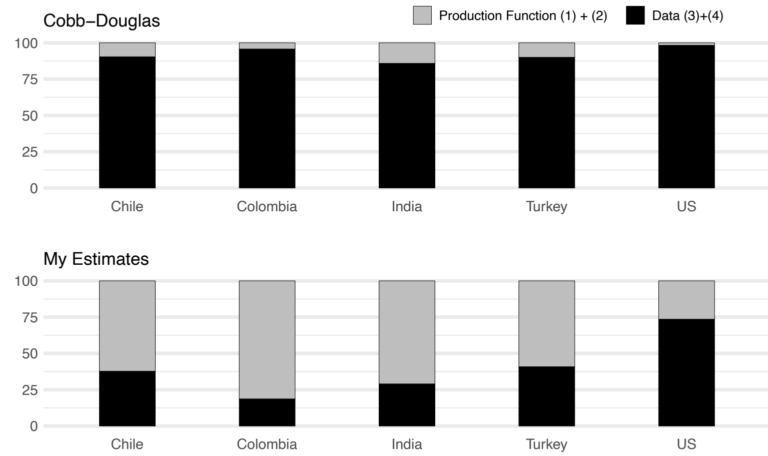
After showing important differences in the level of markups across estimation methods, I now turn to the change in markups over time. I start by looking at what explains the time-series variation in markups. Then I focus on the markup growth in US manufacturing.

#### 8.3.1 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 Equation (7.2), ignoring the covariance between the two. Figure 6 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

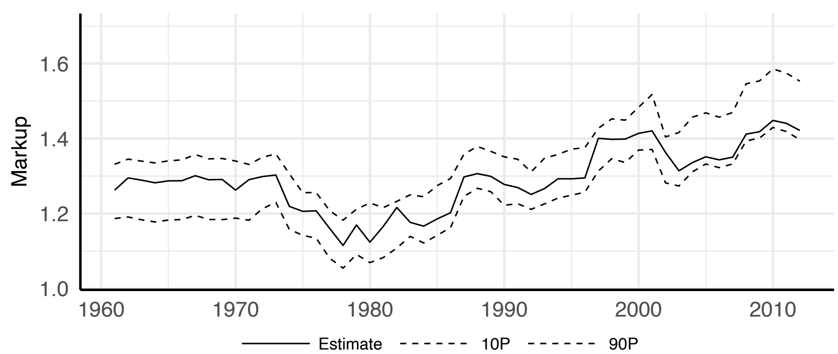
[49] Supplemental Appendix Figure G.8 presents the evolution of markups based on two production function models and the 10-90the percentile of the bootstrap distribution for the difference in estimates.

Figure 6: 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 7: Evolution of the Aggregate Markup




Notes: The evolution of markups in the US manufacturing industry. The dotted lines report the 10-90the percentile of the bootstrap distribution (100 iterations).

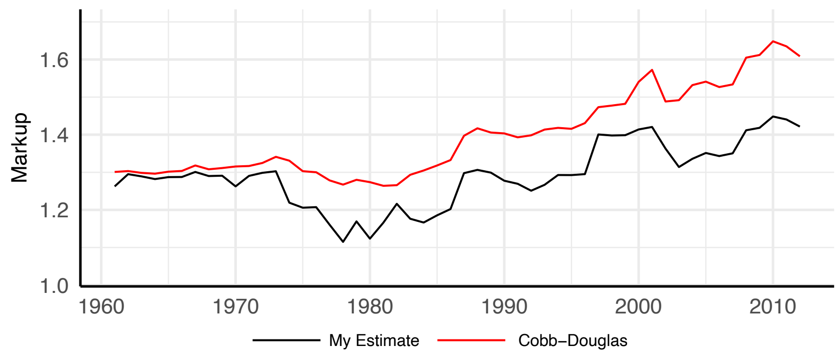
elasticity explains a significant fraction of the change in markups in all countries.

If the true production function is Cobb-Douglas, then aggregate markups are almost entirely driven by the change in revenue shares. As a result, if we want to understand the evolution of markups, looking at the change in revenue shares is sufficient. Is the role of change in technology really minimal? For the rest of this section, I answer this question.

### 8.3.2 Change in Markups in the US Manufacturing Sector

This section investigates the evolution of the aggregate markup in the US manufacturing sector. Figure 7 plots the sales-weighted aggregate markup from 1960 to 2012 along with the 10-90th 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

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 model estimated using the ACF procedure.

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 estimation essentially replicates one of the specifications in De Loecker et al. (2020), who estimated a Cobb-Douglas production function with a single flexible input. They find a dramatic rise in markups in the US economy since 1960 and interpret this finding as a large increase in market power. My goal is to understand how a labor-augmenting production function affects this conclusion. Figure 8 reports both markups measures. The Cobb-Douglas estimates suggest that markups rose more than 30 percentage points between 1960 and 2012. This finding mirrors De Loecker et al. (2020)'s finding. The markups 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. This rise is even smaller when markups are weighted by cost shares, reported in Supplemental Appendix Figure G.7.

This result is also consistent with other evidence in the literature. For example, Hsieh and Rossi-Hansberg (2019) find that in manufacturing concentration has fallen rather than increased. Cheng et al. (2019) incorporate unobserved heterogeneity through random coefficients, and they also find that a more flexible production function points to a lower increase in markups relative to Cobb-Douglas. Markups estimates from the demand approach also support this finding. For example, Grieco et al. (2020) estimates the change in markups in the US auto industry by estimating demand, and they find no increase in the average markups.

# 9 Extensions

## 9.1 Heterogeneous Input Prices

My model assumes that input prices are common across firms. This assumption is standard in the literature, 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. This extension requires incorporating heterogeneous input prices into input demand functions and accounting for them when constructing the control variables. For details, see Supplemental Appendix B.1 presents this extension.

### 9.2 Unobserved Materials Prices

My framework can also be used for estimating production functions when materials prices are unobserved, and productivity is Hicks-neutral. This situation may arise if firms use different quality inputs at different prices. The key in this extension is to show that unobserved materials-augmenting productivity is observationally equivalent to a model with unobserved materials prices under my assumptions. Under this equivalence, the toolkit developed in this paper can be used to account for unobserved materials prices. I show this extension in Supplemental Appendix B.2.

### 9.3 Accounting for Firm Selection

In Supplemental Appendix B.3, I incorporate non-random firm exit into my estimation framework. This extension requires two simplifying assumptions: (i) that productivity shocks are independent conditional on the previous period's productivity, and (ii) firms decide whether to exit based on only Hicks-neutral productivity. With these assumptions, I rely on Olley and Pakes (1996)'s insight that there is a cutoff in Hicks-neutral productivity conditional on observables and firms that draw Hicks-neutral productivity below the cutoff exit. I estimate the propensity of exit conditional on observables, which allows me to control for selection. The empirical results from implementing this selection correction are provided in Supplemental Appendix F.3.

## 10 Conclusions

Production function estimation plays a critical role in many policy discussions, including misallocation of inputs, rise in market power, and the 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 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. To overcome the challenges due to two sources of unobserved heterogeneity and absence of parametric restrictions, I first show how to express labor-augmenting productivity using firms' input demand function. Then I develop a novel control variable approach to control for productivity shocks. Finally, I show how to exploit first-order conditions of cost minimization without restricting market power in the output market.

Empirically, I show that ignoring labor-augmenting productivity and imposing parametric restrictions generate biased output elasticity and markups estimates. These biases are economically significant. The commonly used specifications underestimate capital elasticity by up to 70 percent and overestimate labor elasticity by up to 80 percent. The estimates also document substantial firm-level heterogeneity in the output elasticities. These biases and heterogeneity in output elasticities translate into biases in the inferred markups. The estimates suggest that the standard methods generate an upward bias in both the level and growth of markups. Finally, I also revisit the recent findings on the rise of US markups. I find that markup growth in the US manufacturing sector is 15 percentage points, in contrast to 30 percentage points suggested by recent papers over the last five decades.

Even though I focus 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 (Rubens (2020), Morlacco (2020)), 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 other dimensions and develop even richer production function frameworks.

# A Auxiliary Lemmas and Proofs

**Lemma A.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 continous 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 indepedence assumption. □

**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)**

With some abuse of notation, I use $\omega_{it}^H$ and $\epsilon_{it}$ in place of $\exp(\omega_{it}^H)$ and $\exp(\epsilon_{it})$ in the production function. With this, the production function becomes:

$$Y_{it} = F\left(K_{it}, h(K_{it}, \omega_{it}^L L_{it}, M_{it})\right) \omega_{it}^H \epsilon_{it}.$$

The firm minimizes the cost of flexible inputs for a given level of planned output, $\bar{Y}_{it}$:

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

Since the firm's information set includes productivity shocks we can write the firm's problem as:

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

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}]$. 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 (A.1) can be rewritten as

$$\min_{M_{it}, \bar{L}_{it}} \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_t \qquad \text{s.t.} \quad F\big(K_{it}, h(K_{it}, \bar{L}_{it}, M_{it})\big)\omega_{it}^H \geqslant \bar{Y}_{it}(\mathcal{I}_{it}), \tag{A.2}$$

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

$$\begin{aligned}
&C(\tilde{Y}_{it}, K_{it}, \omega_{it}^H, \bar{p}_{it}) = \\
&\min_{\bar{L}_{it}, M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : \tilde{Y}_{it} \leqslant F\big(K_{it}, h(K_{it}, \bar{L}_{it}, M_{it})\big)\omega_{it}^H\Big\}, \\
=&\min_{\bar{L}_{it}, M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}) \leqslant h(K_{it}, \bar{L}_{it}, M_{it})\Big\}, \\
=&\min_{\bar{L}_{it}, M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : 1 \leqslant h\big(K_{it}, \bar{L}_{it}/F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}), M_{it}/F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it})\big)\Big\}, \\
=&\min_{\bar{L}_{it}, M_{it}} \Big\{F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it})\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
=&F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}) \min_{\bar{L}_{it}, M_{it}} \Big\{\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
\equiv&C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_2(K_{it}, \bar{p}_{it}^l, p_t^m).
\end{aligned} \tag{A.3}$$

The second line follows by the assumption that $F(\cdot, \cdot)$ is strictly monotone in its second argument. The third and fourth lines are due to homotheticity property of $h(K_{it}, \cdot, \cdot)$. In the last line, I define two new functions that characterize the cost function.

Equation (A.3) implies that the cost function can be expressed as a product of two functions, one of which depends only on capital and input prices. 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:

$$\bar{L}_{it} = C_1(K_{it}, \tilde{Y}_{it}, \omega^H_{it})\big(\partial C_2(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial \bar{p}^l_{it}\big), \quad M_{it} = C_1(K_{it}, \tilde{Y}_{it}, \omega^H_{it})\big(\partial C_2(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial p^m_t\big).$$

The ratio of materials to the effective labor equals:

$$\frac{M_{it}}{\bar{L}_{it}} = \frac{\partial C_2(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial p^m_t}{\partial C_2(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial \bar{p}^l_{it}} \equiv \frac{C_m(K_{it}, \bar{p}^l_{it}, p^m_t)}{C_l(K_{it}, \bar{p}^l_{it}, p^m_t)},$$

which does not depend on $(\tilde{Y}_{it}, \omega^H_{it})$. Using $\bar{L}_{it} = L_{it}\omega^L_{it}$ the ratio of materials to labor is:

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

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

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

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)). This implies that the input ratio can be written as a function of quality-adjusted labor and materials prices:

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

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 (A.5), the logarithm of input is given by

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

Taking the derivative of this expression with respect to $\log(\omega^L_{it})$ and with some algebra, I obtain

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

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)). So, $\sigma(K_{it}, \tilde{p}_{it})$ equals the elasticity of subtitution between effective labor and materials. By Assumption 2.1(iv) $\sigma(K_{it}, \tilde{p}_{it}) > 1$ for all $(K_{it}, \omega_{it}^L)$ or $\sigma(K_{it}, \tilde{p}_{it}) < 1$ for all $(K_{it}, \omega_{it}^L)$. From this I conclude that the flexible input ratio is strictly monotone in $\omega_{it}^L$.

**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 $\omega_{it}^L$ from Equation (3.1) I obtain

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

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

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

Note that by normalization $u_{it}^1$ is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$ and by timing assumption $(K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H) \in \mathcal{I}_{it-1}$. Therefore, Equation (A.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 (2.6) and (2.8). 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, the $u_{it}^1$ 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 Equation (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 A.1 imply

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

It follows from Equation (A.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 (2.6) and (2.8). Using this

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

**Proof of Proposition 4.3**

The proof consists of two parts. First, I will show that two different set of structural functions, lead to same $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. Then, I will show that labor-augmenting productivity, the output elasticity of capital and elasticity of substitutions depend on the structural functions $h$ and $\bar{r}$, and therefore can not identified. Looking at the elasticities first, $\theta_{it}^L$ and $\theta_{it}^M$ depend on the production function in the following way:

$$\theta_{it}^L = f_2 h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) \bar{r}(K_{it}, \tilde{M}_{it}) L_{it}, \tag{A.9}$$
$$\theta_{it}^M = f_2 h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) M_{it}, \tag{A.10}$$

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

$$\bar{h}_2(K_{it}, \tilde{M}_{it}) = h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) r_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}), \tag{A.11}$$
$$\bar{h}_1(K_{it}, \tilde{M}_{it}) = h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) r_1(K_{it}, \tilde{M}_{it}). \tag{A.12}$$

So the right-hand side of these equations are identified from $\bar{h}$ and the output elasticities $\theta_{it}^L$ and $\theta_{it}^M$ 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 $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. 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 $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$ as I show below

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

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 $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. 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 $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. Since $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$, non-identification of $\bar{r}(K_{it}, \tilde{M}_{it})$ immediately implies that $\omega_{it}^L$ 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, $\theta_{it}^K$ 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, $\sigma_{it}^{ML}$ 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 4.4**

If production function takes the form given Equation (4.7) 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 $\theta_{it}^L$ and $\theta_{it}^M$ are identified, the right-hand sides of these equations are identified. The identification of $\theta_{it}^M$ immediately implies that $h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$ is identified from $(f_2, \theta_{it}^M)$. 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

$$\bar{h}_1(\tilde{M}_{it}) = h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it}) + h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}), \tag{A.13}$$

where $\bar{r}'(\tilde{M}_{it})$ denotes the derivative of $\bar{r}(\tilde{M}_{it})$. Therefore, the right-hand side of Equation (A.13) is identified from $\bar{h}(\tilde{M}_{it})$. Taking the ratio of $\theta_{it}^L/L_{it}$ 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 $b(\tilde{M}_{it})$ as

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

Since $\omega_{it}^L = r(\tilde{M}_{it})$, and $\log(r(\tilde{M}_{it}))$ is identified up to a constant, $\omega_{it}^L$ 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 4.5**

If production function takes the form in Equation (4.7), we can derive $\sigma_{it}^{ML}$ 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, $\sigma_{it}^{ML}$ is not identified. he elasticity of substitutions for other input pairs can similarly be derived and it can be shown that they depend on the second derivatives of $h$.

## References


**Acemoglu, Daron**. 2003. "Labor-and capital-augmenting technical change". *Journal of the European Economic Association* 1 (1): 1–37.

**Ackerberg, Daniel A**. 2016. "Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context". *Working Paper*: 39.

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

**Ackerberg, Daniel and Hahn, Jinyong**. 2015. "Some Non-Parametric Identi. . . cation Results using Timing and Information Set Assumptions". *Working Paper*: 22.

**Alvarez-Cuadrado, Francisco, Long, Ngo Van, and Poschke, Markus**. 2018. "Capital-labor Substitution, Structural Change and the Labor Income Share". *Journal of Economic Dynamics and Control* 87:206–231.

**Antras, Pol**. 2004. "Is the US Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution". *Contributions in Macroeconomics* 4 (1).

**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.

**Bajari, Patrick and Benkard, C. Lanier**. 2005. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach". *Journal of Political Economy* 113 (6): 1239–1276.

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

**Basu, Susanto**. 2019. "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence". *Journal of Economic Perspectives* 33 (3): 3–22.

**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.

**Bernard, Andrew B., Redding, Stephen J., and Schott, Peter K.** 2009. "Products and Productivity". *Scandinavian Journal of Economics* 111 (4): 681–709.

**Berry, Steven, Gaynor, Martin, and Scott Morton, Fiona**. 2019. “Do Increasing Markups Matter? Lessons from Empirical Industrial Organization”. *Journal of Economic Perspectives* 33 (3): 44–68.

**Bond, Stephen and Söderbom, Mans**. 2005. “Adjustment Costs and the Identification of Cobb Douglas Production Functions”. *Working Paper, Institute for Fiscal Studies (IFS).*

**Bond, Steve, Hashemi, Arshia, Kaplan, Greg, and Zoch, Piotr**. 2020. *Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data* w27002. National Bureau of Economic Research.

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

**Cheng, Xu, Schorfheide, Frank, and Shao, Peng**. 2019. “Clustering for Multi-Dimensional Heterogeneity”. *Working Paper.*

**Chirinko, Robert S**. 2008. “: 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, Goldberg, Pinelopi K., Khandelwal, Amit K., and Pavcnik, Nina**. 2016. “Prices, Markups, and Trade Reform”. *Econometrica* 84 (2): 445–510.

**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.

**Diez, Federico, Leigh, Daniel, and Tambunlertchai, Suchanan**. 2018. “Global Market Power and its Macroeconomic Implications”. *IMF Working Papers* 18 (137).

**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.*

**Edmond, Chris, Midrigan, Virgiliu, and Xu, Daniel Yi**. 2018. “How Costly Are Markups?” *Working Paper 24800 National Bureau of Economic Research.*

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

**Foster, Lucia, Grim, Cheryl, Haltiwanger, John, and Wolf, Zoltan**. 2017. *Macro and Micro Dynamics of Productivity: From Devilish Details to Insights* w23666. National Bureau of Economic Research.

**Fox, Jeremy T. and Smeets, Vallerie**. 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review* 52 (4): 961–989.

**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 L. E., Murry, Charles, and Yurukoglu, Ali**. 2020. "The Evolution of Market Power in the US Auto Industry". *Slide Deck*.

**Guerre, Emmanuel, Perrigne, Isabelle, and Vuong, Quang**. 2009. "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions". *Econometrica* 77 (4): 1193–1227.

**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 Mitchell, Matthew F**. 2008. "A Theory of Factor Allocation and Plant Size". *The RAND Journal of Economics* 39 (2): 329–351.

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

**Hong, Han and Shum, Matthew**. 2010. "Pairwise-Difference Estimation of a Dynamic Optimization Model". *The Review of Economic Studies* 77 (1): 273–304.

**Hsieh, Chang-Tai and Rossi-Hansberg, Esteban**. 2019. *The Industrial Revolution in Services* w25968. National Bureau of Economic Research.

**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*.

**Karabarbounis, L. and Neiman, B.** 2014. "The Global Decline of the Labor Share". *The Quarterly Journal of Economics* 129 (1): 61–103.

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

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

**Kehrig, Matthias and Vincent, Nicolas**. 2018. "The Micro-Level Anatomy of the Labor Share Decline". *Working Paper 25275 National Bureau of Economic Research*.

**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, National Bureau of Economic Research*.

**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]*.

**Matzkin, Rosa L**. 2004. "Unobservable Instruments". *Working Paper*.

**Morlacco, Monica**. 2020. "Market Power in Input Markets:Theory and Evidence from French Manufacturing". *Working Paper*.

**Nadiri, M. Ishaq**. 1982. "Producers theory". In *Handbook of Mathematical Economics*, 2:431–490. Elsevier.

**Oberfield, Ezra and Raval, Devesh**. 2014. "Micro Data and Macro Technology". *Working Paper 20452, National Bureau of Economic Research*.

**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.

**Rubens, Michael**. 2020. "Market Structure, Oligopsony Power and Productivity".

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

**Traina, James**. 2018. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements". *Working Paper*.

**Van Beveren, Ilke**. 2012. "Total Factor Productivity Estimamtion: 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*.

**Wooldridge, Jeffrey M.** 2009. "On Estimating firm-level Production Functions Using Proxy Variables to Control for Unobservables". *Economics Letters* 104 (3): 112–114.

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

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

Mert Demirer

December 20, 2020

# A Data and Estimation

## 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 10 employees. I restrict my sample to industries that have more than 250 firms per year on average. I drop observations that are at the bottom and top 2% of the distribution of revenue share of labor or revenue share of materials or revenue share of combined flexible input for each industry to remove outliers. Table G.0 lists the names and SIC codes of the industries in the final sample. I report each industry's share in manufacturing in terms of sales, and the number of plants operating in each industry for the first, last and midpoint year of the sample. The last row labeled as "other industries" provides information about the industries that are excluded from the sample. After sample restrictions, five industries remains, 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 10 employees. I restrict my sample to industries that have more than 250 firms per year on average and follow the same step as in Chile Data to remove outliers. Table G.1 provides summary statistics. The number of industries after sample restrictions is 9, relatively higher than the number of industries in 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 10 workers and use electricity, or 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, the definition of a large factory is one with 200 or more workers, whereas from 2006 onward, the definition 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, which is similar to industrial classifications in other countries. The industry definition repeatedly changes over the sample period. I follow Allcott et al. (2016) to create a consistent industry definition at the NIC 87 level. For sample restriction 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 that have more than 250 firms per year on average. I follow the same steps as in Chilian Data to remove outliers. Table G.2 provides summary statistics. The Indian sample is the least representative as five industries in the sample make up only 20% of the manufacturing sector.

### 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 10 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 includes 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. Table G.5 provides summary statistics.

### A.5 Compustat

Compustat is obtained from Standard and Poor's Compustat North America database and covers the period from 1961 to 2012. Data from more recent years 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 firm's 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 code between 4900-4999 and 6000-6999. I also remove the firms with negative or nonzero sales, employment, cogs, xsga and less than 10 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 variables used in production function estimation, 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. Table G.3 provides summary statistics. My sample covers the entire population of public manufacturing firms. There is a large increase in sample size from 1961 to 2012, reflecting the enormous increase in the number of public firms in the US over the sample periods. I drop observations that are 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 is calculated as 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, 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 capital in period $t$. For Compustat, capital is calculated as the value of property, plant, and equipment, net of depreciation deflated using 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 Industrial data file. For other countries, sales are total production value, plus the difference between the end year and beginning year value of inventories of finished goods.

### A.7 Estimation Algorithm

This section details the estimation procedure. Apply data cleaning and variable construction 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 corresponding rolling window and denote this sample by $B_r$. Estimate control variables $u^2_{it}$ for each $it \in B_r$ 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

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

using a flexible logit model. Then for each $it \in B_r$, estimate $u^2_{it} = s(M_{it}, K_{it}, W_{it}, u^1_{it})$ as $\widehat{u}^2_{it} = s(\bar{q}, K_{it}, W_{it}, u^1_{it})$ where $\bar{q}$ denotes the closest point to $M_{it}$ in $Q$.[1] From this procedure obtain $\widehat{u}^2_{it}$ for all $it \in B_r$. For production function estimation, first approximate the logarithm of $\bar{h}$ by using second-degree polynomials

$$\log(\widehat{h}(\tilde{M}_{it})) = a_1 + a_2\tilde{m}_{it} + a_3\tilde{m}^2_{it}, \tag{A.1}$$

where $\{a_1, a_2, a_3\}$ are the parameters of the polynomial approximation. Set $a_1 = 0$ to impose the normalization for $\widehat{h}(\tilde{M}_{it})$ described in Section 4. Let $V_{it} := L_{it}\widehat{h}(\tilde{M}_{it})$. Approximate the production function as

$$\widehat{f}(K_{it}, L_{it}\widehat{h}(\tilde{M}_{it})) = b_1 + b_2k_{it} + b_3k^2_{it} + b_3k_{it}v_{it} + b_4v_{it} + b_5v^2_{it}, \tag{A.2}$$

where $\{b_1, b_2, b_3, b_4, b_5\}$ are the parameters of the polynomial approximation. Approximate the control functions $c_2(\cdot)$ and $c_3(\cdot)$ using third-degree polynomials similarly. For given values $\{a_j\}^3_{j=1}$, $\{b_j\}^5_{j=1}$, $\widehat{c}_2(\cdot)$ and $\widehat{c}_3(\cdot)$ construct the objective function 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 $\{a_j\}^3_{j=1}$, estimate $\{b_j\}^5_{j=1}$, $\widehat{c}_2(\cdot)$ and $\widehat{c}_3(\cdot)$ using the least squares regression. In the outer loop use an optimization routine to estimate $\{a_j\}^3_{j=1}$. Minimizing the objective function requires optimization only over three parameters, so it is not computationally intensive. After estimating the production function parameters, the next step is elasticity and markups estimation.

Take 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 $f$ and $\bar{h}$ from the estimates of the parameters $\{a_j\}^3_1$ and $\{b_j\}^5_1$ in Equations (A.1) and (A.2). First, using the estimates of $f$ and $\bar{h}$, calculate the output elasticity of capital and the sum of the materials and labor elasticities, given in Equations (4.5) and (4.9) by taking numerical derivatives. Then given an estimate of $\theta^V_{it}$ and revenue shares of materials and labor use Equations (4.6) to estimate output elasticity of labor and materials. Finally estimate markups from $\widehat{\theta}^V_{it}$ and the revenue share of flexible

[1] One can estimate $s(m, k, w, u)$ for every $M_{it}$ observed in the data with additional computational cost.

input. For standard errors, resample firms with replacement from sample A,then repeat the estimation procedure above. For estimation of the CES and Nested CES models in Section F.4, I use the same procedure except that I impose the parametric restrictions.

# B Extensions

This section presents three extensions to my model. All proofs are provided in Section C.

## B.1 Heterogeneous Input Prices

This extension assumes that input prices vary across firms. I denote labor and materials prices by $p^l_{it}$ and $p^m_{it}$, respectively, and use $\bar{p}_{it}$ to denote the input price vector, so $\bar{p}_{it} := (p^l_{it}, p^m_{it})$. I also use $\tilde{p}_{it} := (p^l_{it}/p^m_{it})$ to denote the input price ratio. Differently from the main model, $W_{it}$ includes also 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:*

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

This assumption states that prices and productivity shocks jointly follow an exogenous first-order Markov process. Importantly, this assumption allows for correlation between productivity shocks and input prices. Since we expect that more productive workers, as represented by higher $\omega^L_{it}$, earn higher wages, 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.** *Firm's materials demand is given by $M_{it} = s_t(K_{it}, \omega^L_{it}, \omega^H_{it}, \bar{p}_{it})$, and $s_t(K_{it}, \omega^L_{it}, \omega^H_{it}, \bar{p}_{it})$ is strictly increasing in $\omega^H_{it}$.*

This assumption is a natural extension of Assumption 2.3, as the materials demand should depend on both input prices. In this section, I maintain the other assumptions in the model, namely Assumptions 2.2 and 2.4, and state 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 $K_{it}$, $\omega^L_{it}$ and $\tilde{p}_{it}$*

$$\tilde{M}_{it} = r_t(K_{it}, \omega^L_{it}, \tilde{p}_{it}). \tag{B.1}$$

***(ii)*** *Under Assumptions 2.2(v), $r_t(K_{it}, \omega_{it}^L, \tilde{p}_{it})$ is strictly monotone in $\omega_{it}^L$.*

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 proposition, $\omega_{it}^L$ is invertible once we condition on the input price ratio and capital. By inverting Equations (B.1), and omitting the time subscripts in functions, I can write productivity shocks as:

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

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

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it}, u_{it}^1), \qquad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it} \sim \text{Uniform}(0,1). \tag{B.3}$$

Unlike Equation (3.1), I include the ratio of current and past input prices in $g_1(\cdot)$ function. This is needed 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

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

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

**Lemma B.1.** *Under Assumptions 2.4, B.1, and B.2, $u_{it}^1$ is jointly independent of $(K_{it}, W_{it-1}, \tilde{p}_{it})$:*

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

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

Next, we can use Equations (B.2) and (B.3) to write $\omega_{it}^L$ as: $\omega_{it}^L \equiv c_1(W_{it-1}, \tilde{p}_{it}, u_{it}^1)$. Note that unlike the main model, the CDF when estimating $u_{it}^1$ is conditional on the price vector $\bar{p}_{it}$ and control function includes the price ratio $\tilde{p}_{it}$ since prices are endogenous. The procedure for deriving the control function for $\omega_{it}^H$ is similar to that of $\omega_{it}^L$. We use

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

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

**Lemma B.2.** *Under Assumptions 2.4, 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 $\omega_{it}^H \equiv c_2\left(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\right)$.

It follows that with variation in input prices control functions become $\omega_{it}^L = c_1\left(W_{it-1}, \tilde{p}_{it}, u_{it}^1\right)$ and $\omega_{it}^H = c_2\left(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\right)$. The main difference is that I need to condition on 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.

### B.2 Unobserved Materials Prices under Hicks-Neutral Productivity

The approach in this paper can be adopted to a model with Hicks-neutral productivity, where only cost of materials is observed, but not the quantity. This is important because in standard production datasets, average wages are typically observed, however materials prices are not. Since, only the cost of materials is typically available in the data, quantity of materials cannot be recovered from its cost with heterogeneous in input prices. This is especially the case if there are differences in quality of materials used by firms.

The researcher observes $(K_{it}, L_{it})$ but not $M_{it}$. Instead, materials expenditure, denoted by $R_{it}^m = M_{it}p_{it}^m$, is observed. Due to heterogeneity in materials prices, as indicated by $p_{it}^m$, we cannot recover $M_{it}$ from $R_{it}^m$. But we can replace materials with its expenditure as follows:

$$Y_{it} = F_t\left(K_{it}, L_{it}, R_{it}^m \omega_{it}^M\right) \exp(\omega_{it}^H) \exp(\epsilon_{it}), \tag{B.4}$$

where I define $\omega_{it}^M := 1/(p_{it}^m)$. In Equation (B.4), one can interpret materials cost as an input in the production function and the inverse of materials prices as unobserved materials-augmenting productivity shock. Given this equivalence, I will show that the tools developed in this paper can be used to estimate this model using $R_{it}^m$ in place of materials. First, I modify the assumptions to accommodate unobserved materials prices. I maintain the assumption that firms face the same wages in the labor market.

**Assumption B.3.** *Productivity shock and materials price jointly follow an exogenous joint first-order Markov process:* $P(\omega_{it}^H, p_{it}^m \mid \mathcal{I}_{it-1}) = P(\omega_{it}^H, p_{it}^m \mid \omega_{it-1}^H, p_{it-1}^m)$

This assumption does not restrict the correlation between materials prices and firm productivity, so firms that use higher quality materials can be more productive. The next assumption includes the unobserved materials prices into the materials demand function.

**Assumption B.4.** *Firm's materials decision is given by* $M_{it} = s_t(K_{it}, p_{it}^m, \omega_{it}^H)$*, where* $s_t(K_{it}, p_{it}^m, \omega_{it}^H)$ *is strictly increasing in* $\omega_{it}^H$*.*

We can write the materials expenditure, $R_{it}^m$, using this assumption and materials prices as:

$$R_{it}^m = s_t(K_{it}, p_{it}^m, \omega_{it}^H)/p_{it}^m \equiv s_t^M(K_{it}, p_{it}^m, \omega_{it}^H).$$

Since $s_t(K_{it}, p_{it}^m, \omega_{it}^H)$ is strictly monotone in $\omega_{it}^H$ conditional on $(K_{it}, p_{it}^m)$, $R_{it}^m$ is also strictly monotone in $\omega_{it}^H$ conditional on $(K_{it}, p_{it}^m)$. This shows that the monotonicity with respect to materials implies monotonicity with respect to materials expenditure. Next, I define a version of Assumptions 2.1 to accommodate heterogeneous materials prices.

**Assumption B.5.** *Suppose that*

***(i)*** *Production function is* $Y_{it} = F_t\big(K_{it}, h(K_{it}, L_{it}, M_{it})\big)\exp(\omega_{it}^H)\exp(\epsilon_{it})$.

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

***(iii)*** *The firm minimizes the production cost with respect to* $(L_{it}, M_{it})$ *given* $K_{it}$*, productivity shock* $\omega_{it}^H$ *and input prices* $(p_t^l, p_{it}^m)$.

***(iv)*** *The elasticity of substitution between labor and materials is either greater than 1 for all* $(K_{it}, p_{it}^m)$ *or less than 1 for all* $(K_{it}, p_{it}^m)$.

Next, using this assumption, I show that the ratio of labor and materials cost, $L_{it}/R_{it}^M$, depends only on $K_{it}$ and unobserved materials prices $p_{it}^m$.

**Proposition B.2.**

***(i)*** *Under Assumptions B.5(i-iii), the ratio of labor and materials cost, denoted by* $\tilde{L}_{it} = L_{it}/R_{it}^M$*, is given by* $\tilde{L}_{it} \equiv r_t(K_{it}, \omega_{it}^M)$*, where* $r_t(\cdot)$ *is an unknown function.*

***(ii)*** *Under Assumptions B.5(iv)* $r_t(K_{it}, \omega_{it}^M)$ *is strictly monotone in* $\omega_{it}^M$.

With this result, I have two monotonicity conditions analogous to those in the main model. The difference is that I replace $M_{it}$ with $R_{it}^m$ and $\omega_{it}^L$ with $1/p_{it}^m$. Also, this model involves materials-augmenting producticity instead of labor-augmenting productivity. Therefore, following the same steps in the main model I can write $\omega_{it}^H \equiv s_t(K_{it}, R_{it}^m, \tilde{L}_{it})$ and $\omega_{it}^M \equiv r_t(K_{it}, \tilde{L}_{it})$. Given the equivalence of this model and the main model, the procedure for developing the control functions and identification analysis are the same as the main model. Thus, the rest of the derivations and proofs are omitted.

### B.3 Selection

This section presents a method of incorporating non-random firm exit, which generates a selection problem, into my framework under two simplifying assumptions: (i) firms decide whether to exit based only on Hicks-neutral productivity, and (ii) productivity shocks are independent conditional on previous period's productivity. Under these assumptions, I show how to adjust my control variables to account for selection by relying on Olley and Pakes (1996)'s insight that there is a cutoff in productivity level below which firms exit. In this section, I maintain the assumptions in Section 2.2, and impose some additional restrictions.

**Assumption B.6.** *Productivity shocks are independent conditional on* $(\omega_{it-1}^H, \omega_{it-1}^L)$*:*

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

**Assumption B.7.** *The firm's exit decision depends only on* $\omega_{it}^H$ *and* $K_{it}$*. In particular, the firm exits if and only if* $\omega_{it}^H \leqslant \bar{\omega}(K_{it})$*, where* $\bar{\omega}$ *specifies the exit threshold conditional on* $K_{it}$*.*

The control variable derivation remains the same as in Subsection 3.1. However, for $\omega_{it}^H$, differently from Equation (3.5), I use the following representation:

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0,1). \tag{B.5}$$

In contrast to Equation (3.5), $g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2)$ does not include $u_{it}^1$. This follows by Assumption B.6, which implies independence of innovations to productivity shocks. To introduce exit to the model, let $I_{it}$ denote an indicator variable which equals one if firm $i$ exits and zero otherwise. By Assumption B.7, $I_{it} = 1$ if and only if $\omega_{it}^H \leqslant \bar{\omega}(K_{it})$. So, firm $i$'s exit decision at time $t$ depends on its capital level and current Hicks-neutral productivity. Using the representation of $\omega_{it}^H$ in Equation (B.5) I can write the exit rule as:

$$g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2) \leqslant \bar{\omega}(K_{it}) \implies u_{it}^2 \leqslant g_2^{-1}\big(\omega_{it-1}^L, \omega_{it-1}^H, \bar{\omega}(K_{it})\big) \implies u_{it}^2 \leqslant \tilde{\omega}(W_{it-1}, K_{it}) \tag{B.6}$$

where first I use the invertibility of $g_2$ in $u_{it}^2$, and then invertibility of demand functions. $\tilde{\omega}$ is a function that defines exit rule based on a cutoff value in $u_{it}^2$. This formulation of exit suggests conditional on $(W_{it-1}, K_{it})$ the firm's exit decision depends only on the realization of $u_{it}^2$. By Lemma 3.2, I have

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

This is useful because the variable that determines whether a firm exits, $u_{it}^2$, is uniform and independent from the variables I need to condition on in Equation (B.6), $(W_{it-1}, K_{it})$. Thus, I can estimate the cutoff in $u_{it}^2$ conditional on $(W_{it-1}, K_{it})$ from the fraction of firms that exit conditional on $(W_{it-1}, K_{it})$. In particular, this cutoff value equals the conditional exit probability observed in the data and can be written as:

$$\tilde{\omega}(W_{it-1}, K_{it}) = \text{Prob}(I_{it} = 1 \mid W_{it-1}, K_{it}) \equiv p(W_{it-1}, K_{it}). \tag{B.7}$$

This suggests that conditional on $(W_{it-1}, K_{it})$ firms that receive $u_{it}^2$ that is greater than $p(W_{it-1}, K_{it})$ stay and other firms exit. As a result, the distribution of $u_{it}^2$ conditional on $(W_{it-1}, K_{it})$ and $(I_{it} = 1)$ can be written as another uniform distribution:

$$u_{it}^2 \mid W_{it-1}, K_{it}, (I_{it} = 1) \sim \text{Uniform}\big(p(W_{it-1}, K_{it}), 1\big). \tag{B.8}$$

As shown in Equations (3.6) and (3.7), to control for $\omega_{it}^H$, I need the distribution of $M_{it}$ conditional on $(W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1))$ since $M_{it}$ depends on $u_{it}^1$. This creates a problem because even though $(W_{it-1}, K_{it})$ is observed for all firms (stayed and exited), $u_{it}^1$ cannot be estimated for the firms that exit since we do not observe $\tilde{M}_{it}$ for those firms. The following lemma will overcome this problem.

**Lemma B.3.**

$$\text{Prob}(I_{it} = 1 \mid W_{it-1}, K_{it}, u_{it}^1) = \text{Prob}(I_{it} = 1 \mid W_{it-1}, K_{it}).$$

*Proof.* The probability of exit conditional on $(W_{it-1}, K_{it}, u_{it}^1)$ equals

$$\begin{aligned}
\text{Prob}(I_{it} = 1 \mid W_{it-1}, K_{it}, u_{it}^1) &= \text{Prob}\big(g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2) \geqslant \bar{\omega}(K_{it}) \mid W_{it-1}, K_{it}, u_{it}^1\big), \\
&= \text{Prob}\big(g_2(\bar{r}(W_{it-1}), \bar{s}(W_{it-1}), u_{it}^2) \geqslant \bar{\omega}(K_{it}) \mid W_{it-1}, K_{it}, u_{it}^1\big), \\
&= \text{Prob}\big(g_2(\bar{r}(W_{it-1}), \bar{s}(W_{it-1}), u_{it}^2) \geqslant \bar{\omega}(K_{it}) \mid W_{it-1}, K_{it}\big), \\
&= p(W_{it-1}, K_{it}),
\end{aligned}$$

where the third line follows because $u_{it}^1$ and $u_{it}^2$ are independently distributed conditional on $W_{it-1}$. To see this, note that we have $\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$ and $\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2)$. By assumption B.6, we have $\omega_{it}^L \perp\!\!\!\perp \omega_{it}^H \mid (\omega_{it-1}^L, \omega_{it-1}^H)$. The monotonicity of $g_1$ and $g_2$ in their last arguments imply that $u_{it}^1$ and $u_{it}^2$ are independently distributed conditional on $W_{it-1}$. $\square$

From this result, I obtain

$$u_{it}^2 \mid W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 0) \sim \text{Uni}(p(W_{it-1}, K_{it}, u_{it}^1), 1) \sim \text{Uni}(p(W_{it-1}, K_{it}), 1). \tag{B.9}$$

This allows me to recover $u_{it}^2$ conditional on $(I_{it} = 1)$ using observables because I can estimate $p(W_{it-1}, K_{it})$ from data. After showing the effects of non-random firm exit on control variables, the next lemma provides the control variable and control function under these new assumptions.

**Lemma B.4.** *Under the assumptions, we have that*

$$\omega_{it}^H \equiv c_2\left(W_{it-1}, u_{it}^2\right), \qquad u_{it}^2 = F_{M_{it} \mid K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1). \tag{B.10}$$

Since I do not observe $F_{M_{it} \mid K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$ but only observe the distribution conditional on selection $F_{M_{it} \mid K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0)$, $u_{it}^2$ cannot be recovered using this lemma. However, I can use use Lemma B.3, which gives the distribution of $u_{it}^2$ for the firms that stay, to write $u_{it}^2$ as:.

$$u_{it}^2 = p(W_{it-1}, K_{it})\big(1 - F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0)\big) + F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0)$$

This result uses the distribution of $u_{it}^2$ conditional on firms that stay, given in Equation (B.9). It says that $u_{it}^2$ can be recovered from the observed distribution function of $M_{it}$ conditional on $(I_{it} = 0)$ and the propensity score, both of which are identified from data. After recovering $u_{it}^2$ from data, the rest of the estimation procedure follows the main model.

# C Proofs

### 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 A.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 (2.6) and (2.8) 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

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

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

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

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

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

Using Equations (2.6) and (2.8) to substitute $(\omega_{it-1}^L, \omega_{it-1}^H)$ as functions of $W_{it-1}$, I obtain

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

**Proof of Lemma B.4**

By Assumption B.6 we have that

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

Using the Skorokhod representation of $\omega_{it}^H$ in Equation (B.5) we write

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

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

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

It follows from Equation (C.1) and the fact that $u_{it}^2$ is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$

$$u_{it}^2 \mid \mathcal{I}_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \text{Uniform}(0,1).$$

Since $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ we have $u_{it}^2 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \text{Uniform}(0,1)$, which implies

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

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

$$\begin{aligned} M_{it} &= s\left(K_{it}, \omega_{it}^H, \omega_{it}^L\right) = s\left(K_{it}, g_2\left(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2\right), c_1\left(W_{it-1}, u_{it}^1\right)\right), \\ &= s\left(K_{it}, g_2\left(\tilde{r}\left(W_{it-1}\right), \tilde{s}\left(W_{it-1}\right), u_t^2\right), c_1\left(W_{it-1}, u_{it}^1\right)\right) \equiv \bar{s}\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right). \end{aligned} \tag{C.3}$$

The intuition is similar to that of Lemma 3.1. Employing strict monotonicity of $\bar{s}$ in $u_{it}^2$ and Equation (C.2), we can use Equation (C.3) to identify $u_{it}^2$. In particular, $u_{it}^2$ equals

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1), \tag{C.4}$$

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 $\omega_{it}^H$ can be written as $\omega_{it}^H \equiv c_2\left(W_{it-1}, u_{it}^2\right)$.

**Proof of Proposition B.2**

This proof closely follows the proof of Proposition 2.1. I maintain the same conditions and notation. The main difference is that production function involves only Hicks-neutral productivity, but materials prices vary at the firm-level. The firm minimizes the cost of flexible inputs for a given level of planned output, $\bar{Y}_{it}$. This problem, under Assumption B.5, can be written as:

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

Since the firm's information set includes $\omega_{it}^H$, we have

$$\min_{L_{it},M_{it}} \quad p_t^l L_{it} + p_{it}^m M_{it} \qquad \text{s.t.} \quad F\left(K_{it}, h(K_{it}, L_{it}, M_{it})\right)\omega_{it}^H \mathcal{E}_{it} \geqslant \bar{Y}_{it}, \tag{C.5}$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}]$. The cost minimization problem in Equation (C.5) is given by

$$\min_{M_{it},L_{it}} p_t^l L_{it} + p_{it}^m M_t \qquad \text{s.t.} \quad F(K_{it}, h(K_{it}, L_{it}, M_{it})\omega_{it}^H \geqslant \tilde{Y}_{it},$$

where $\tilde{Y}_{it} := Y_{it}/\mathcal{E}_{it}(\mathcal{I}_{it})$. Suppress the argument $\mathcal{I}_{it}$ of $\tilde{Y}_{it}$ and let $\bar{p}_{it} = (p_t^l, p_{it}^m)$ denote the price vector. Following the steps I used to obtain Equation (A.3), the cost function can be expressed as:

$$C(\bar{p}_{it}, \tilde{Y}_{it}, K_{it}, \omega_{it}^H) = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_2(K_{it}, p_t^l, p_{it}^m). \tag{C.6}$$

By Shephard's Lemma the input demands are given by the derivatives of cost function:

$$M_{it} = C_1(K_{it}, \bar{Y}_{it}, \omega_{it}) / \big(\partial C_2(K_{it}, p_t^l, p_{it}^m)/\partial p_{it}^m\big), \qquad L_{it} = C_1(K_{it}, \bar{Y}_{it}, \omega_{it}) / \big(\partial C_2(K_{it}, p_t^l, p_{it}^m)/\partial p_t^l\big).$$

Using optimal labor and materials demand the ratio of labor to materials is

$$\frac{L_{it}}{M_{it}} \equiv \frac{C_l(K_{it}, p_t^l, p_{it}^m)}{C_m(K_{it}, p_t^l, p_t^m)}.$$

Since $M_{it}$ is not observed we cannot use $L_{it}/M_{it}$ to control for $\omega_{it}^L$. Therefore, I next define the ratio in terms of the observed variables. Using $R_{it}^m \omega_{it}^M = M_{it}$, the ratio of materials cost to labor is

$$\frac{L_{it}}{R_{it}^m} = \frac{C_l(K_{it}, p_t^l, p_{it}^m)\omega_{it}^M}{C_m(K_{it}, \bar{p}_t^l, p_{it}^m)}.$$

This equation has the same structure as Equation (A.4) in the proof of Proposition 2.1, with $\omega_{it}^L = \omega_{it}^M$, $\bar{p}_{it}^l = 1/\omega_{it}^M$, $p_t^m = p_t^l$ and $\tilde{M}_{it} = L_{it}/R_{it}^m$. Therefore, we can treat $R_{it}^m$ as materials input and treat $\omega_{it}^M = 1/p_{it}^M$ as the materials-augmenting productivity. This solves the problem that materials quantity, $M_{it}$, is unobserved since we can replace it with $R_{it}^m$ and introduce a materials-augmenting productivity to the model. Given this equivalence, the rest of proof proceeds similarly to the proof of Proposition 2.1 and, therefore, is omitted.

# D Identification

In this section, I show that the homogeneous and strong homothetic separable production functions in Section 4.5 are generically identified using the moment restriction in Equation (5.5). First, I will present auxiliary lemmas, and then analyze these cases separately.

**Lemma D.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 derivative 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

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

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

**Lemma D.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. Assume 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 $\Omega$ 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 $\Omega_f$, 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 $\Omega$ 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 on $\epsilon$.

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:

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

I will show that if Equation (D.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 (D.2) with respect to $x$ and $z$. Taking derivative with respect to $z$ yields

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

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

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

$$zf'\big(zh(x)\big)h(x)\Big(\log'(h(x)) - \log'(\tilde{h}(x))\Big) = \tilde{g}'(x) - g'(x). \tag{D.4}$$

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 $\Omega_f$ 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 (D.4) and $\log'(h(x)) = \log'(\tilde{h}(x))$ imply that

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

Integrating this, there exists $\lambda$ and $a$ such that

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

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

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

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

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

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 (D.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 $\Omega_{log}$. 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. $\square$

### 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 $\omega_{it}^{H}$ 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, $\alpha$ 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 $\eta_{it-1}$ 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 (5.5), 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 $\Omega$ 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

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

We say that $(\alpha_0, f_0, h_0, g_0) \in \Omega$ are identifiable if no other member of $\Omega$ 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 D.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 D.2. (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 D.2, and $\alpha_0$ is identified.*

*Proof.* I will show that if there exists observationally equivalent $(\alpha, f, h, g)$ 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 (D.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:

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

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

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

Taking the derivatives of Equation (D.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

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

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

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

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

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

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

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

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

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

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

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

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

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

Taking the derivative with respect to $w$

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

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

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

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 (D.14) can be written as

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

The rest of the proof is an application of Lemma D.2 (see Equation D.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 Strong Homothetic Production Function**

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

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

Substituting an unknown function of control variables for $\omega_{it}^H$ 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 strong 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 (D.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 (5.5), 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 $\Omega$ 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

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

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

**Proposition D.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

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

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

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

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

Multiplying Equation (D.19) by $s_1(w, z, x, t)/s_2(w, z, x, t)$ and subtracting Equation (D.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 D.1 $h(x)$ is identified up to a scale from $\tilde{f}_1(w, z, x)$. Next, multiplying Equation (D.19) by $\log'(h(x))z$ and subtracting Equation (D.20) we obtain

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

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

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

□

# E Application to Parametric Production Functions

## E.1 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 discusses its advantages over 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) Productivity shocks 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, (iii) The firm's intermediate input decision is given by $m_{it} = s(k_{it}, \omega_{it}^H)$, which is strictly increasing in $\omega_{it}^H$. Using these assumptions, I construct a control variable using the steps in Section 3. Productivity can be represented as:

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

where $g(\omega^H_{it-1}, u_{it})$ is strictly increasing in $u_{it}$ by construction. By the Markov Assumption we have that $\omega^H_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega^H_{it-1}$. Substituting $\omega^H_{it}$ using Equation (E.1) 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 a strictly in $\omega_{it}$ 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

$$u_{it} \mid k_{it}, m_{it-1}, k_{it-1} \sim \text{Uniform}(0, 1). \tag{E.2}$$

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 an unknown function of $(m_{it-1}, k_{it-1}, u_{it})$ to proxy $\omega^H_{it}$.

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

With this result, I obtain a partially linear model

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c(m_{it-1}, k_{it-1}, u_{it}) + \epsilon_{it} \tag{E.3}$$

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

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_2(m_{it-1}, k_{it-1}) + \xi_{it} + \epsilon_{it} \tag{E.4}$$

with $\mathbb{E}[\xi_{it} \mid I_{it-1}] = 0$. Now we can estimate the parameters $(\beta_k, \beta_l)$ and unknown functions $c_1(\cdot), c_2(\cdot)$ in Equation (E.3) and (E.4) using the following moment restictions.

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

I emphasize that the control variable approach does not suffer from the functional dependence problem studied in Ackerberg et al. (2015) even if labor is a flexible input and can be written as $l_{it} = l(\omega_{it}, k_{it})$. The main distinction between my 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 a known function of $(k_{it}, m_{it})$, a single dimensional variable. Conditional on the control variable $u_{it}$, there is still variation in $k_{it}$ and $l_{it}$, 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 $\nu_{it-1}$ corresponds to a vector of random variables that affects the firm's investment decision, such as investment prices and heterogeneous beliefs. As a result, conditional on $(k_{it-1}, m_{it-1}, u_{it})$, $\nu_{it-1}$ generates variation in labor.

## E.2 Nested CES Production Function

This section studies the identification of Nested CES production function in Equation 2.2. We maintain the assumptions in Section 2.2. The logarithm of this production function is

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

Using the homotheticity property of Nested CES we can rewrite this production function as:

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

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

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

where $\gamma_1$ and $\gamma_2$ are constants that depend on the model parameters. Note that $\omega_{it}^L$ disappeared from the model. This is the parametric analog of my nonparametric inversion result in Proposition 2.1. The model parameters can be estimated using the control functions I develop with the following estimating equation

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

where $u_{it}^1 = \tilde{L}_{it}$ in the Nested CES model because it falls into the model in Equation (4.7). One can estimate the model using the objective function in Equation (4.7). One can 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 $x = M(\tilde{L}_{it} + \gamma_2)^{1/\sigma_1}$. Note that $(1-\beta_k)\gamma_1$ and $\beta_k$ are not separately identified in the production function, but their ratio is identified. Since $\theta_{it}^V$ depends only on the ratio, it is identified. Labor and materials elasticities are identified from $\theta_{it}^V$ and the ratio of revenue shares as described in the paper. Finally, the output elasticity of capital is identified as

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

## E.3 CES Production Function

In this section, I consider the CES production function: Using homotheticity we can write

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

By using the FOCs of cost minimization, one can show that $\omega_{it}^L = \gamma \tilde{L}^{(1-\sigma)/\sigma}$, where $\gamma$ is a constant that depends on input prices and model parameters. Substituting this into the production function:

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

where $\omega_{it}^L$ disappeared from the model. The model parameters can be estimated using the control functions I develop

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

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

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

where $x = M_{it}(\tilde{L}_{it}+\gamma_2)$. Note that $(1-\beta_l-\beta_m)$ and $\gamma_1$ are not separately identified from this production function but the ratio is identified. Since $\theta_{it}^V$ depends only on the ratio, the sum elasticity is identified. Labor and materials elasticities are identified from $\theta_{it}^V$ and the ratio of revenue shares as described in the paper. The capital elasticity is identified as

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

# F Robustness Checks

## F.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 G.12 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. 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.

### F.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 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 G.13 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.

### F.3 Selection

In this section, I estimate output elasticities and markups after accounting for non-random firm exit as described in Subsection B.3. Figure G.14 reports the estimates with and without selection correction using my method. The results suggest that selection correction does not have a significant impacts on the results.

### F.4 Comparison to Nested CES with Labor-Augmenting Technology

I estimate the nested CES production function given in Equation (2.2). Figure G.10 presents the results from comparing the output elasticities estimated from the two models. The capital and materials elasticity estimates of the nested CES are significantly lower than my estimates; however, the labor elasticity estimates are similar. This suggests that, although estimates from a parametric model with labor-augmenting technology are closer to my results than Cobb-Douglas, allowing for a nonparametric model still gives quantitatively different results. I next turn to the markups in Figure G.11. Estimated markups are significantly different for the four developing countries between two models, but the two methods produce similar results for the US. I conclude that differences in the output elasticity estimates affect markups estimates, showing the implications of the parametric restrictions.

# G Additional Tables and Figures

Table G.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 G.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.

Table G.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.

Table G.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 |

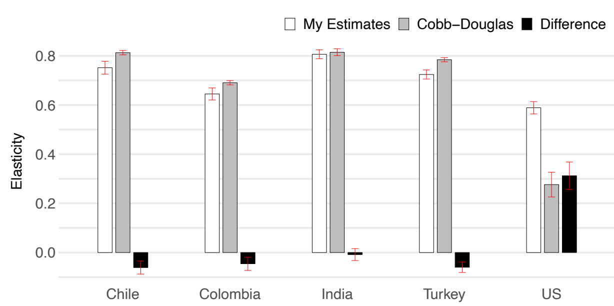
Note: Descriptive Statistics for US. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants.

Table G.5: Descriptive Statistics - Turkey

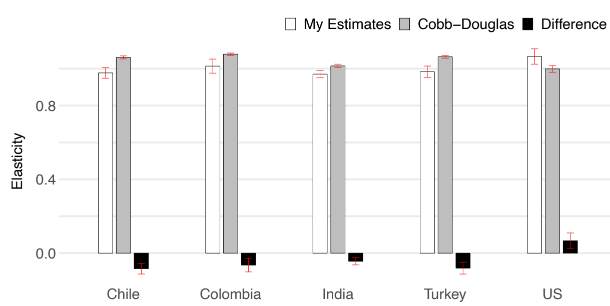
| | | Share (Sales) | | | Number of Plants | | |
|---|---|---|---|---|---|---|---|
| ISIC | Industry | 1983 | 1991 | 2000 | 1983 | 1991 | 2000 |
| 321 | Textiles | 0.16 | 0.13 | 0.16 | 1017 | 945 | 1803 |
| 311 | Food | 0.12 | 0.12 | 0.11 | 1261 | 1120 | 1061 |
| 322 | Apparel | 0.02 | 0.05 | 0.04 | 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.

Figure G.6: Comparison of Output Elasticities



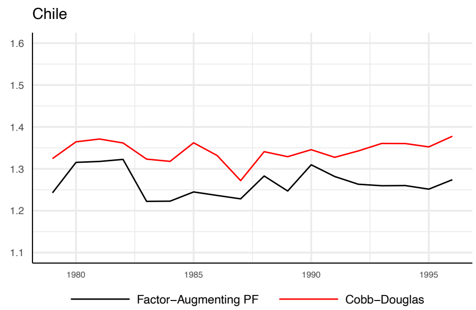

(a) Materials Elasticity



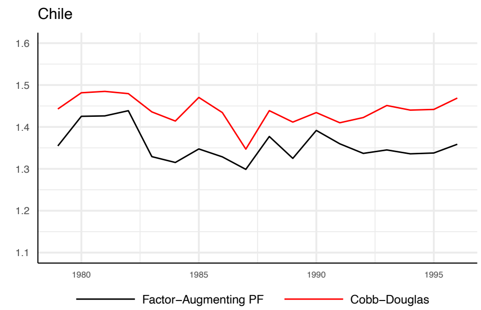

(b) Returns to Scale

Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. 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).

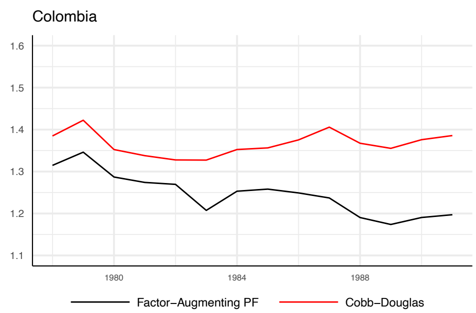
Figure G.7: Evolution of Markups-Other Countries



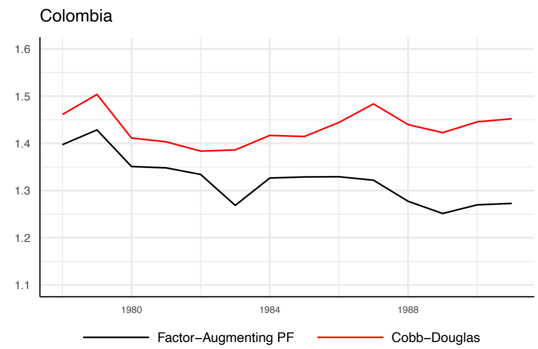

(a) Cost-Weighted Markup (Chile)



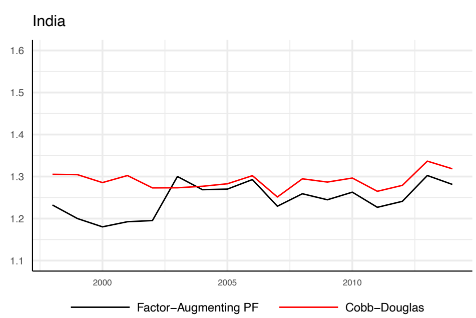

(b) Sales-Weighted Markup (Chile)



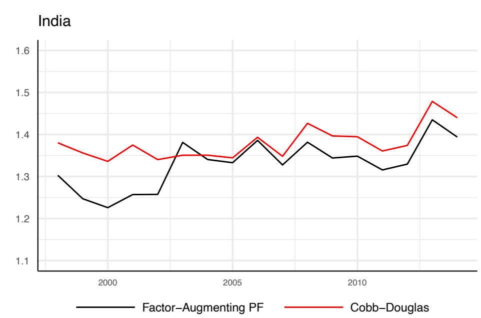

(c) Cost-Weighted Markup (Colombia)



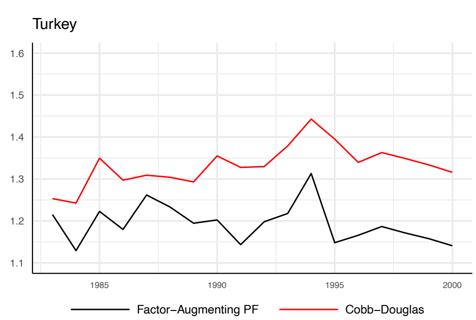

(d) Sales-Weighted Markup (Colombia)



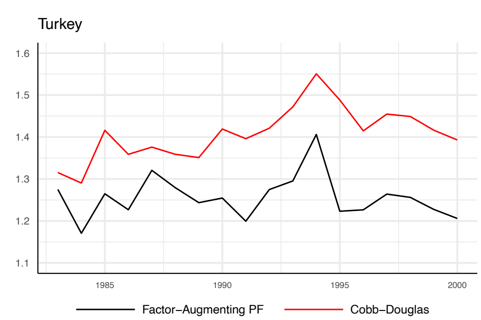

(e) Cost-Weighted Markup (India)




(f) Sales-Weighted Markup (India)



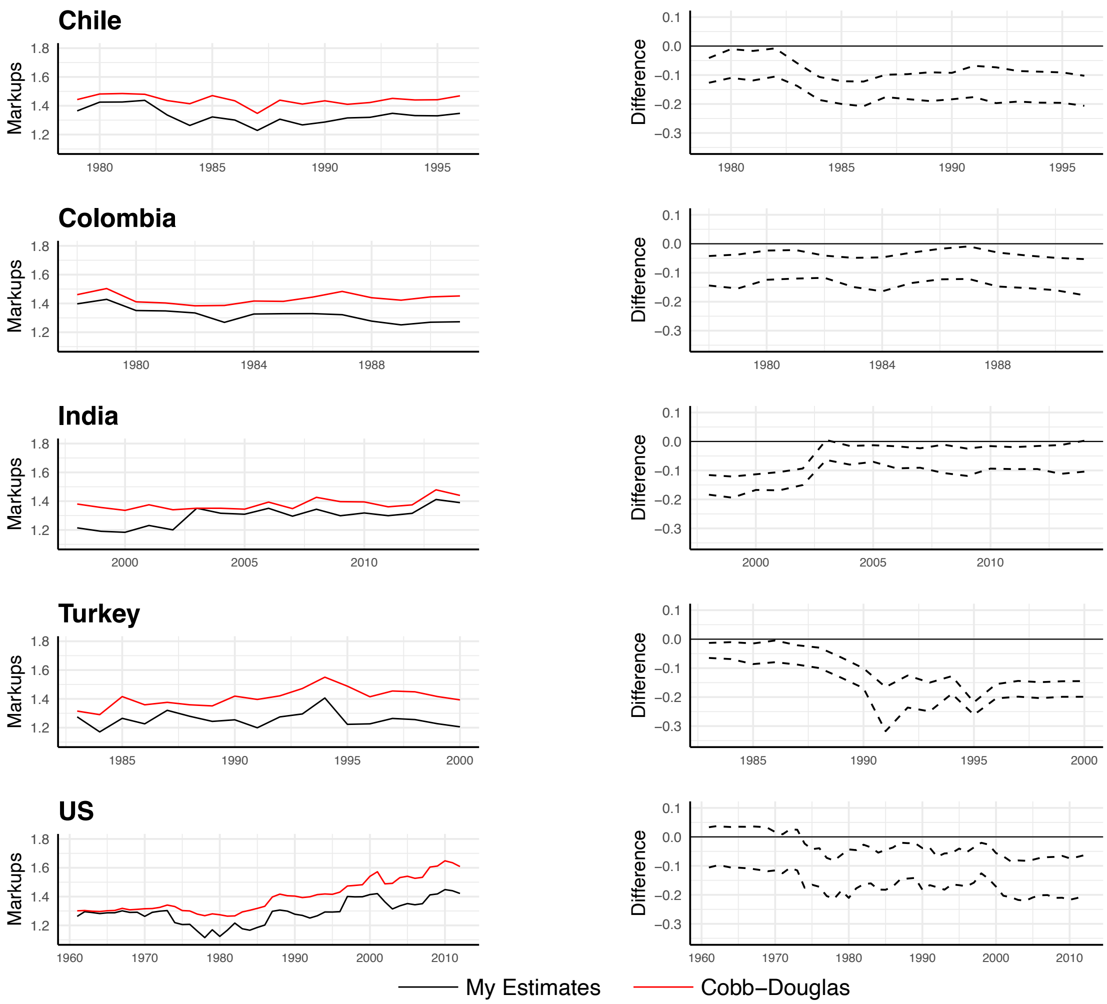

(g) Cost-Weighted Markup (Turkey)




(h) Sales-Weighted Markup (Turkey)

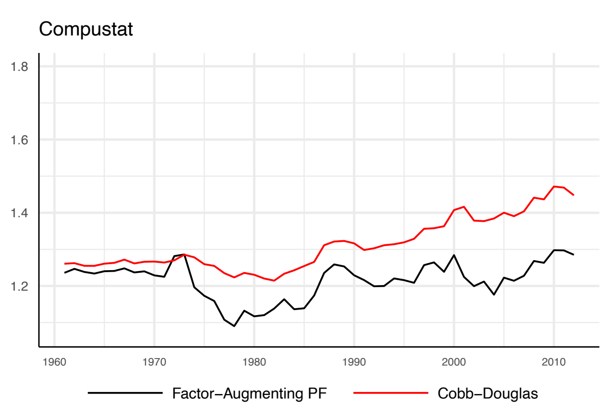
Notes: Comparisons of the evolution of markups in manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure G.8: Confidence Bands for Difference



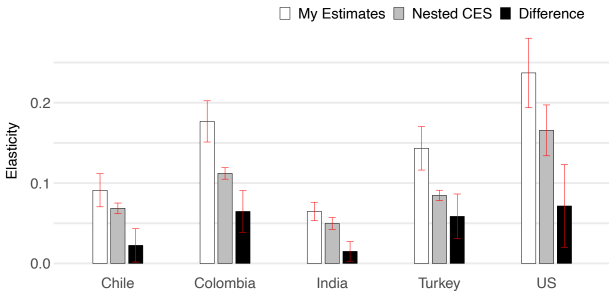

Notes: This figure shows the evolution of the aggregate markups estimated from my method and Cobb-Douglas on left panel and 10-90th percentile of the bootstrap distribution (100 iterations) for the difference between the two estimates.

Figure G.9: Aggregate Markup in the US - Cost Weighted

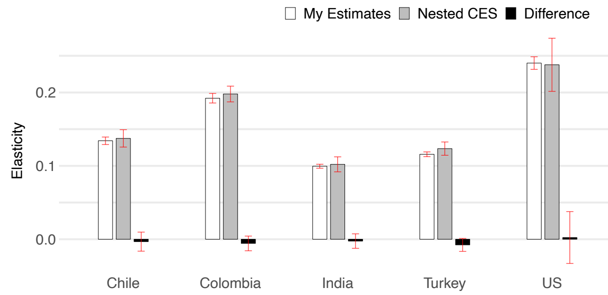



Notes: This figure shows the evolution of aggregate markup in the US calculated by firm's cost shares.

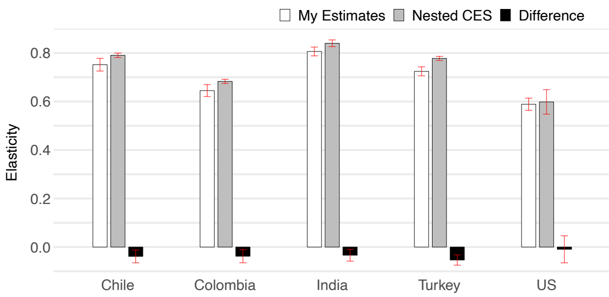
Figure G.10: Comparison with Nested CES



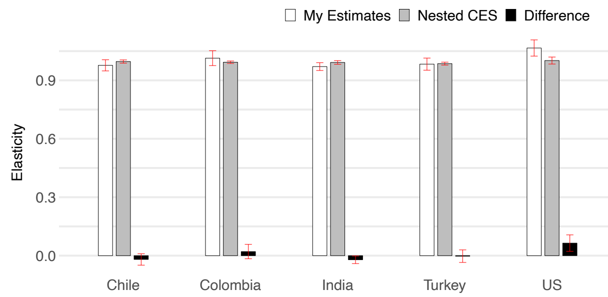

(a) Capital




(b) Labor




(c) Materials




(d) Returns to Scale

Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Nested CES estimated by procedure given in Subsection E.2 for each country. 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 percent confidence intervals calculated using bootstrap (100 iterations).

Figure G.11: Markups Comparison with Nested CES

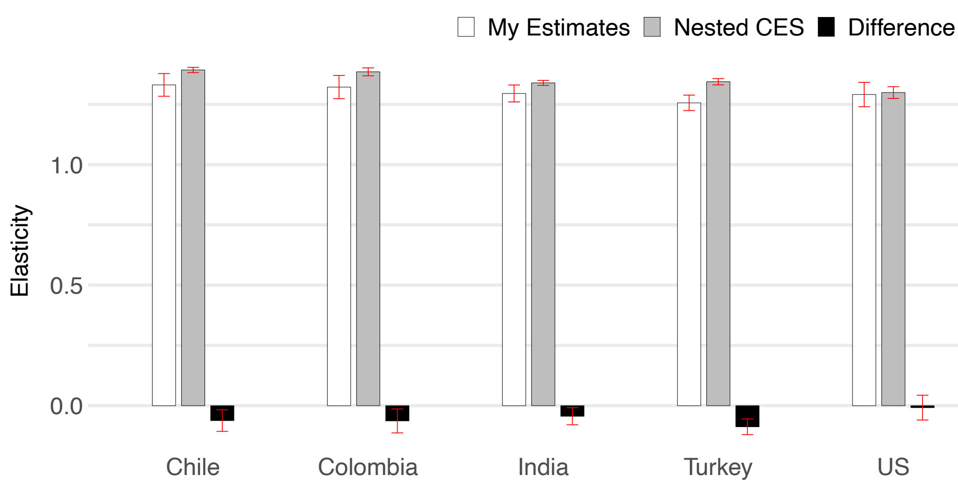


Figure G.12: Comparison of Estimates with and without Measurement Error

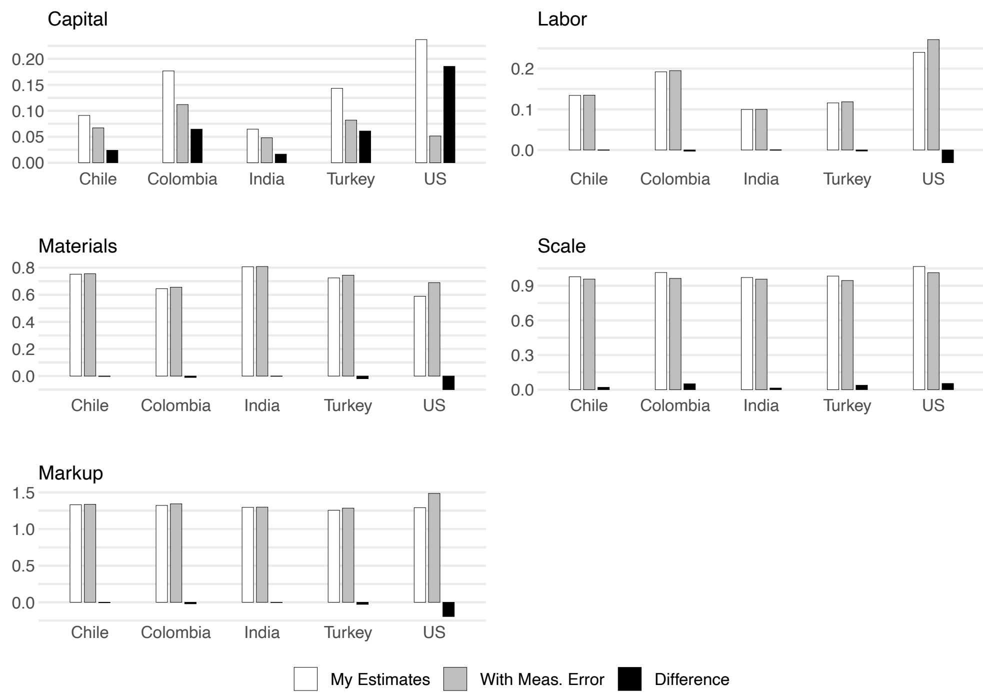



Figure G.14: Comparison of Estimates with and without Selection Correction

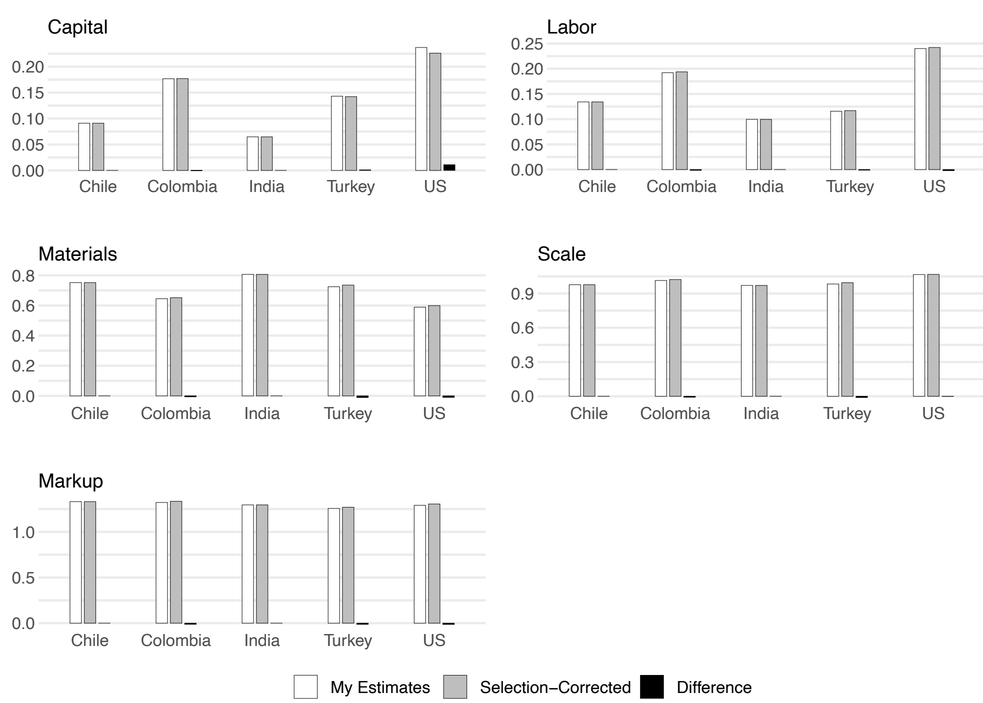


Figure G.13: Comparison of Estimates with and without Capacity Utilization Correction

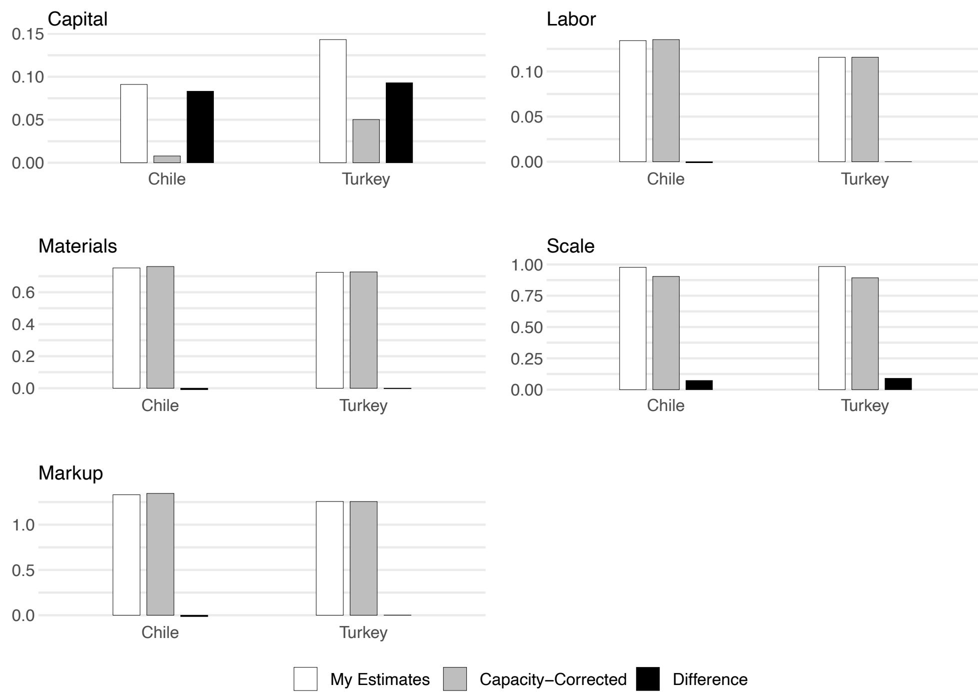
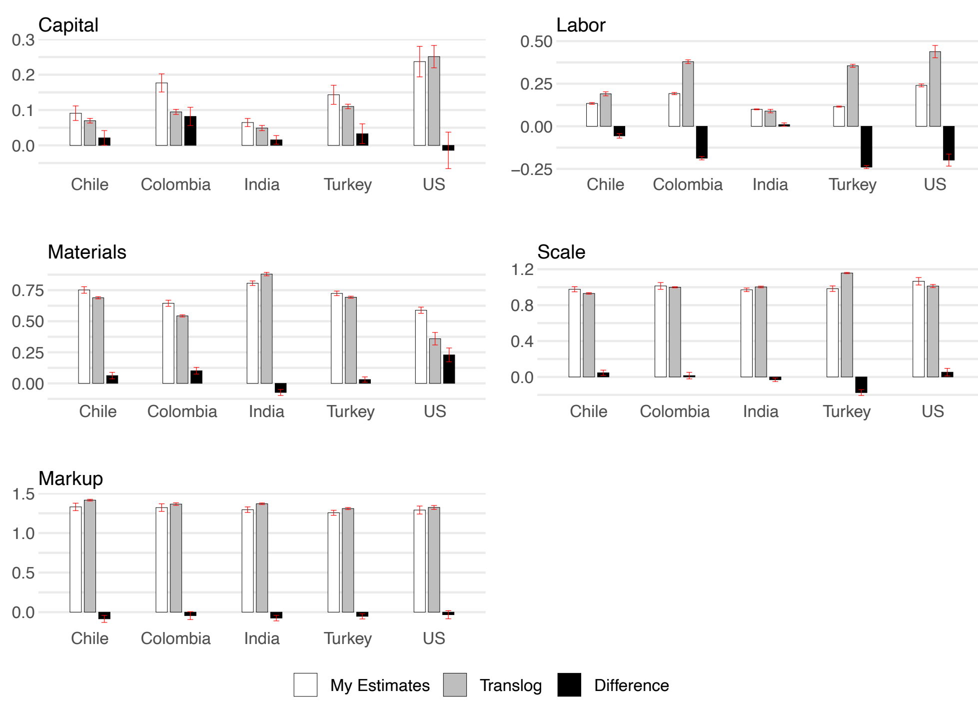



Figure G.15: Comparison of my Estimates with Translog Production Function



## References


**Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth**. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451. (Cit. on pp. 22, 23).

**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. (Cit. on p. 2).

**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. (Cit. on p. 2).

**Matzkin, Rosa L**. 2007. "Nonparametric identification". *Handbook of econometrics* 6:5307–5368. (Cit. on p. 14).

**Olley, G. Steven and Pakes, Ariel**. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64 (6): 1263. (Cit. on p. 9).

**Raval, Devesh**. 2020. "Testing the Production Approach to Markup Estimation". *Working Paper.* (Cit. on p. 3).

**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.* (Cit. on p. 2).