

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

Mert Demirer

September 22, 2020

Abstract

Traditional production function models rely on factor-neutral technology and functional form assumptions, such as Cobb-Douglas. 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 and assesses its economic implications. 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 by up to 70 percent (ii) overestimate labor elasticity by up to 80 percent. These biases propagate into markup estimates inferred from output elasticities: markups are overestimated by 20 percentage points. Finally, heterogeneity in output elasticities also affects estimated trends in markups: 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, Max Cytrynbaum, Glenn Ellison, Kirill Evdekimov, Sara Fisher Ellison, Jerry Hausman, Stefan Hoderlein, Paul Joskow, Omer Karaduman, Ali Kakhdoud, Christopher Knittel, Jing Li, Madeline Mckelway, Whitney Newey, Alex Olszen, 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 seminar participants at the MIT Industrial Organization and Econometrics lunches. 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 its parameters 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, such as Cobb-Douglas, for production function estimation. These two elements of standard practice impose strong theoretical restrictions.¹ Indeed, several papers have shown that these restrictions are strongly 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 (2019a)). Also, 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 ways. First, it includes two unobserved technology shocks: labor-augmenting productivity, which changes the productivity of labor, and Hicks-neutral productivity, which changes the productivity of all inputs. These productivity shocks introduce unobserved firm-level heterogeneity in the 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 match the data.

This paper makes both methodological and empirical contributions. On the methodological side, I first extend the standard Olley and Pakes (1996) framework to accommodate labor-augmenting technology. Then, I show how to identify output elasticities by developing a novel control variable approach and exploiting the first-order conditions of the firm's cost minimization problem.³ On the empirical side, my results indicate that neglecting factor-augmenting technology and imposing parametric restrictions mismeasure output elasticities and markups. I first present the empirical results, and then explain how I deal with methodological challenges.

I use my method to estimate output elasticities in manufacturing industries in the US and

¹For example, in the absence of input price variation, Hicks-neutral productivity implies no unobserved heterogeneity in the output elasticities. The Cobb-Douglas specification restricts the elasticity of substitution to equal one and output elasticities to be common across firms.

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

³My approach does not rely on variation in input prices. Instead, I use optimal expenditure on flexible inputs. However, the model can accommodate variation in input prices.

four developing countries: Chile, Colombia, India and Turkey. To document the biases in standard models, I compare my results with estimates from two production functions with Hicks-neutral technology, Cobb-Douglas and translog. The results suggest that, in all countries, the Cobb-Douglas model estimates incorrect output elasticities. In particular, it underestimates the output elasticity of capital by 70 percent and overestimates the output elasticity of labor by 80 percent. Allowing for labor-augmenting productivity 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. Comparing my estimates with a more flexible Hicks-neutral production function, such as translog, gives quantitatively similar results.

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. (2018)). After documenting biases in output elasticities, I 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 by 10 to 20 percentage points, an important difference 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. This finding highlights the importance of relaxing parametric assumptions. To explain what drives these biases in markup estimates, I present a decomposition exercise. Standard models generate biased markup estimates due to two sources of misspecification: (i) bias in the average output elasticity and (ii) unmodeled heterogeneity in output elasticities. The existing empirical evidence and my elasticity estimates imply that both sources of bias are positive.

The output elasticity estimates matter not only for the level but also for the trend of markups. This is especially true when there is a change in a flexible input's revenue share, as markup estimates are inversely related to revenue shares.⁴ If production technology does not change over time, a decline in the flexible input's share immediately implies an increase in the markup. Therefore, for correct markup estimation, it is crucial to account for the change in production technology. The recent literature, using Hicks-neutral technology, has found little change in output elasticities for the last fifty years, so the decline in labor share in advanced countries has been interpreted as an increase in markups.

Among the countries analyzed, the change in the revenue share of labor is notable only in the US, so I focus on the change in markups in the US. I estimate the evolution of markups in US manufacturing with data from Compustat. 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. In particular, De Loecker et al. (2018) finds that the aggregate markup in the US has risen by 40 percentage points in the US using a Cobb-Douglas production function.⁵ Their finding

⁴More precisely, markup equals the elasticity of a flexible input divided by that input's share in revenue.

⁵They also use a translog production function and obtain similar results.

has drawn significant attention recently 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, about half of the estimates in De Loecker et al. (2018), 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, so the increase in markups found in the literature comes entirely from the decline in revenue shares of flexible inputs. However, according to my production function estimates, the average output elasticity of flexible inputs has declined since the 1990s. Also, my estimates suggest important changes in the heterogeneity in output elasticities, which affects the evolution of markups.

On the methodological side, a major challenge in estimating production functions is the endogeneity of inputs. Firms' input choices are related to productivity shocks, but productivity shocks are unobservable. This problem generates an additional complication in my model due to the multi-dimensional unobserved productivity and absence of parametric assumptions.

To address this challenge I make three methodological contributions. First, I impose a homothetic separability restriction on the production function, which enables me to express labor-augmenting productivity as a function of inputs by inverting input demand functions. Homothetic separability is a necessary and sufficient condition to achieve this; therefore, it is the minimal assumption to control for labor-augmenting productivity. This result generalizes the widely-employed parametric inversion (Doraszelski and Jaumandreu (2018), Raval (2019b), Zhang (2019)) to a non-parametric setup. The rest of the assumptions extend the standard proxy variable framework of Oley and Pakes (1996) to a model with multi-dimensional productivity.

The second contribution is to develop a novel control variable approach for production function estimation, building on Imbens and Newey (2009). In particular, I show that under standard assumptions on firm behavior, one can construct variables from inputs to control for productivity shocks.⁷ This result overcomes two challenges that are not present in other applications of control variables: (i) the model contains two structural unobserved variables and (ii) the independence restriction required for a control variable derivation is not available. I address the first challenge by showing that productivity shocks form a triangular structure under the modeling assumptions. For the second challenge, I show that the modeling assumption provides a conditional independence restriction, which I use to derive the control variables.

The third methodological contribution is an identification strategy for output elasticities and markups. After developing control variables to address endogeneity, I study which features of the production function can be identified from data. I first establish a negative result: without exogenous variation in input prices, one cannot identify the output elasticity of flexible inputs from variation in inputs and output; only the sum of the flexible input elasticities is identified. To separately identify the flexible input elasticities, I use the first-order conditions of optimal input choices. Cost minimization implies that the ratio of two flexible inputs' elasticities is identified as the ratio of their

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

⁷In particular, I use the timing assumption and joint first-order Markov property of productivity shocks, both of which are standard assumptions in the production function estimation literature.

expenditures, without further 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 (Gandhi et al. (2018)).

The model has an especially attractive feature for markup estimation: estimates from two flexible inputs are numerically identical. This feature addresses the well-known problem that two different flexible inputs often give conflicting markups estimates (Raval (2019a), Doraszelski and Jaumandreu (2019)). Obtaining identical markup estimates is the direct implication of using the ratio of expenditures to identify the ratio of elasticities. However, allowing for labor-augmenting technology is still essential for this result. With labor-augmenting technology, the output elasticities cannot be separately identified from variation in inputs and output; only the sum of the flexible elasticities is identified. This feature of the model makes it possible to use the ratio of expenditures to identify the ratio of flexible input elasticities, ensuring markups from two flexible inputs are equal.

My framework can incorporate many economic restrictions on the production function, such as constant returns to scale. This is possible because my model covers a family of specifications, ranging from parametric CES to nonparametric weak homothetic separable production functions, that are nested within each other. The nested structure provides three advantages. First, the estimation method can be applied to the CES production function, if one is willing to make functional form assumptions. Second, it is possible to test the restrictions of a model by comparing its results with a more general model. For example, getting significantly different estimates from a CES production function and a nonparametric model would suggest rejecting constant elasticity of substitution in the production technology. Third, the nested structure makes it possible to impose regularization based on economic theory. One can start with the most general model with as few restrictions as possible. If the estimates are too noisy, then a nested model can be employed to improve precision. Regularization is especially relevant for industries with a small number of firms, for which nonparametric estimation is often infeasible.

The control variable approach developed in this paper is applicable to parametric production functions, including CES and Cobb-Douglas. When applied to the Cobb-Douglas production function, this approach provides some advantages over the standard methods. For example, it is robust to the functional dependence problem highlighted by Ackerberg et al. (2015). Also, it conditions on less information, and therefore provides efficiency gains in estimation.

In terms of data requirements, I focus on the common data scheme in the production function literature, which in general lacks firm-level input prices. Therefore, variation in input prices is not required for identification. However, I show how to extend the model and identification strategy to include observed firm-level input prices. In another extension, I present a way of incorporating non-random firm exit into the estimation method under a simplifying assumption that firms exit when they receive a Hicks-neutral productivity draw below a threshold.

1.1 Related Literature

The most common method for production function estimation is the proxy variables approach, which uses inputs to control for endogeneity (Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), Ackerberg et al. (2015), Gandhi et al. (2018)).⁸ Olley and Pakes (1996) find 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. (2018) study nonparametric identification of production functions using proxy variables. They show how to combine the proxy variable approach with first-order conditions.

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 determine 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 of productivity shocks, as opposed to variables directly observed in the data. Third, I use the first-order conditions of cost-minimization within the proxy variable framework for identification. Unlike Gandhi et al. (2018), firms have market power in the output market, but my approach requires two flexible inputs.

Three recent papers have also highlighted the importance of incorporating factor-augmenting technology into production functions (Raval (2019b), Zhang (2019), Doraszelski and Jaumandreu (2018)). These papers study the change in factor-augmenting productivity and its relation to other economic variables.⁹ The 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 labor-augmenting productivity. I relax the CES assumption and generalize the parametric inversion to a nonparametric inversion. Also, my paper does not require variation in input prices, but it can accommodate it. Finally, the focus of my empirical application is different. I analyze how labor-augmenting technology affects output elasticities and markups.¹⁰

This paper benefits from and contributes to the literature on markup estimation from production data (Hall (1988), De Loecker and Warzynski (2012), Raval (2019a)). This literature demonstrates how to estimate markups from output elasticities under a cost minimization assumption. In a recent

⁸The production function estimation literature goes back to Marschak and Andrews (1944), who first recognized the endogeneity problem. First attempts to address the endogeneity problem have used panel data methods (Mundlak (1961), Mundlak and Hoch (1965)). However, these methods do not give satisfactory answers in practice, as summarized by Griliches and Mairesse (1995). See also Blundell and Bond (2000).

⁹Raval (2019b) 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.

¹⁰Another strand of literature uses the random coefficients framework to model firm-level heterogeneity in production technology. See Kasahara et al. (2015), Balat et al. (2016), Li and Sasaki (2017) and Fox et al. (2019).

paper, Doraszelski and Jaumandreu (2019) extends this literature by studying markup estimation in the presence of unobserved demand shocks and adjustment cost in flexible inputs. I investigate the role of production function specification on markup estimates and argue that correct estimation of output elasticities and firm-level heterogeneity is crucial for markup estimation. Specifically, I show that estimating a flexible production technology leads to lower markup estimates.

Lastly, a growing empirical literature analyzes markup growth and market power in the US. Much of this literature assumes a Cobb-Douglas production function and finds that markups have risen in the US and other developed countries (Diez et al. (2018), De Loecker et al. (2018), Autor et al. (2019)). I emphasize the importance of a flexible production function by showing that labor-augmenting technology points to a more muted rise in markups in the US.

2 Model

I start by introducing a production function model with labor-augmenting technology. I then show how this model explains the data better by comparing it to commonly-used production function models.

2.1 Nonparametric 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:

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

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 chooses them each period, and they do not affect future production. Materials consist of intermediate inputs used for production, such as raw materials and energy. In contrast, I assume that capital is a *predetermined input*, that is, the firm chooses the level of capital to use during period t in period $t - 1$. Therefore, the firm's current capital decision affects future production.

In each period, the firm chooses the level of flexible inputs to minimize the total cost of production based on its information set. I use \mathcal{I}_{it} to denote firm i 's information set at period t , which includes productivity, ω_{it}^L , and ω_{it}^H , past information sets, and other signals related to production and profit. The information set is orthogonal to the random shock, i.e., $\mathbb{E}[\epsilon_{it} | \mathcal{I}_{it}] = 0$, the only

orthogonality restriction imposed on the information set. Under this assumption, ϵ_{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 the input markets for labor and materials are *perfectly competitive*. The input prices do not vary across firms, but they can vary over time. Therefore, firms are price-takers facing p_t^L and p_t^M as the prices of labor and materials, respectively. My model and identification strategy extends to the case where input prices are heterogeneous and observed, but firms do not have market power in the input markets.¹¹ 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 heterogeneity based on the input mix, whereas labor-augmenting and Hicks-neutral productivity allow for unobserved heterogeneity in labor productivity and total factor productivity. 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 inputs are homogeneous across firms. In general, my framework can accommodate only one factor-augmenting productivity, and that factor should be a flexible input. The main reason for this limitation is that a non-flexible input has dynamic implications, which makes 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.¹²

I choose to consider labor-augmenting technology for three reasons. First, labor-augmenting productivity is an essential component of endogenous growth models and its changes are an important subject in the literature (Acemoglu (2003)). Second, heterogeneity in ω_{it}^L reflects firm-level differences in labor quality. Several sources of labor quality, such as firms managing labor differently, human capital, and experience might lead to differences in labor productivity across firms. Labor-augmenting productivity can account for these unobserved sources of productivity differences. Finally, in most production datasets, labor has the most across-firm variation among all inputs, so intuitively we should expect most unobserved heterogeneity in labor input.

My production function differs from standard models in two significant ways: (i) It contains factor-augmenting technology and (ii) It does not impose a parametric structure. These features are not trivial and that my flexible production function has important implications not captured

¹¹I present this extension in Supplemental Appendix 3.1.

¹²Modeling materials' productivity could be important in some industries as it might reflect heterogeneity in input quality; see Fox and Smeets (2011).

by other production functions. For an illustration, a common specification is 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) ω_{it}^H is the only source of unobserved heterogeneity in production technology. These are strong restrictions with strong implications. The log-linear functional form constrains the output elasticities to be common across firms. This constraint is not consistent with some of the empirical findings in the literature. First, it implies that all firm-level heterogeneity in flexible input allocation comes from variation in input prices since a cost-minimizing firm sets marginal products equal to prices for the flexible inputs.¹³ Second, the literature has documented large heterogeneity in capital and labor intensities of production, which contradicts constant elasticity.¹⁴ Another implication of the log-linear functional form is unitary elasticity of substitution between all input pairs. This prediction is also inconsistent with empirical findings in the literature.¹⁵

A solution to these issues, commonly employed in the literature, is to assume a more flexible Hicks-neutral production function, such as translog. However, assuming that Hicks-neutral productivity is the only source of unobserved heterogeneity is still restrictive and is not consistent with several observed patterns. 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 rather than across-industry changes (Karabarbounis and Neiman (2014), Kehrig and Vincent (2018), Autor et al. (2019)). Changes and heterogeneity in production technology have been proposed as a mechanism (Oberfield and Raval (2014)). Labor-augmenting productivity in my production function model captures this heterogeneity. Failing to account for this 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.¹⁶

¹³Raval (2019b) tests and rejects this prediction using data from the Census of US manufacturing.

¹⁴For 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)).

¹⁵Although estimates vary, the consensus is that the aggregate elasticity of substitution between capital and labor is less than one (Antras (2004), Klump et al. (2007), Herrendorf et al. (2015), Alvarez-Cuadrado et al. (2018)). Furthermore, Chirinko et al. (2011) show that the elasticity of substitution at the firm-level must be lower than the aggregate elasticity of substitution, providing further evidence that Cobb-Douglas is not an accurate representation of the firms' production technology.

¹⁶There are well-known identification problems with Hicks-neutral production functions. Gandhi et al. (2018) study such production functions and show that the standard proxy variable approach from Olley and Pakes (1996) does not identify the production function nonparametrically. They instead propose a method that exploits the first-order conditions under the assumption that the output market is perfectly competitive. However, this assumption rules out markups, one of the objects of interest in this paper. Empirical studies often estimate the translog functional form to allow for a flexible production technology. However, the translog production function is subject to the same identification problems studied in Gandhi et al. (2018).

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 ω_{it}^L as an unknown function of inputs. Other assumptions concern firm behavior and the distribution of productivity shocks. They generalize the standard proxy variable framework to a model with two productivity shocks. Throughout the paper, I assume that all functions are continuously differentiable as needed and all random variables have a continuous and strictly increasing distribution function.

2.2.1 A Homothetic Separability Restriction

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

Assumption 2.1 (Homothetic Separability). *Suppose that*

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

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

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

(iii) *The firm minimizes 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}, ω_{it}^L) or less than 1 for all (K_{it}, ω_{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 a homothetic function 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 implication is that the ratio of the marginal products of two inputs does not depend on ω_{it}^H . Note that the homotheticity of h_t does not imply that the production function is homothetic.

Assumption 2.1(iii) specifies that firms choose the level of flexible inputs to minimize their (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 output level does not have to maximize profit; the cost is minimized for an arbitrary level of output. Moreover, it is a static problem, so this assumption is agnostic about the firm's dynamic problem.¹⁷

¹⁷In my model, cost-minimization does not give rise to parametric first-order conditions. As a result, this assumption is not equivalent to profit maximization.

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 ω_{it}^L . Assumption 2.1(iv) precludes this possibility.

Next, I state a result establishing the properties of the ratio of flexible inputs using 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 ω_{it}^L

$$\tilde{M}_{it} \equiv r_t(K_{it}, \omega_{it}^L), \quad (2.2)$$

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 ω_{it}^L .

Proof. See Appendix B.

The first part of this proposition states that the flexible input ratio is a function of only one of the model unobservables: labor-augmenting productivity. To see the intuition for this result, observe that the firm's relative labor and materials allocation depends on the relative marginal products of these inputs. By the homotheticity of $h_t(\cdot)$, the ratio of the marginal products does not change with ω_{it}^H , so as the ratio of flexible inputs. 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:

$$C(\bar{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, p_t^m, p_t^l) = C_1(K_{it}, \omega_{it}^L, p_t^m, p_t^l)C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H), \quad (2.3)$$

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). This implies that the ratio of materials to labor input does not depend on $C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H)$ and its arguments.

The second part of Proposition 2.1 establishes that $r_t(K_{it}, \omega_{it}^L)$ is strictly monotone and invertible in ω_{it}^L . For strict monotonicity, the flexible input ratio should always move in the same direction as ω_{it}^L , which affects the ratio of marginal products of labor and materials. Since 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.¹⁸ Together, these two results provide a function, $r_t(K_{it}, \omega_{it}^L)$, that is strictly monotone in a scalar unobserved variable.

Next, I provide two examples of parametric production functions that satisfy the restrictions in Assumption 2.1.

tion is less restrictive in my model than in most models in the literature, which usually make parametric restrictions.

¹⁸In particular, if materials and effective labor are substitutes, firms increase materials-to-labor ratio as ω_{it}^L increases, otherwise firms decreases materials-to-labor ratio as ω_{it}^L increases.

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

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

My framework nests the CES production function with $h(K_{it}, \omega_{it}^L L_{it}, M_{it}) = \beta_l [\omega_{it}^L L_{it}]^\sigma + (1 - \beta_l - \beta_m) M_{it}^\sigma$. This function is homogeneous of degree one and the elasticity of substitution between effective labor and materials is σ . The CES specification has been widely used in the literature to study factor-augmenting technology (Doraszelski and Jaumandreu (2018), Zhang (2019), Raval (2019b)). Under the CES assumption $r_t(K_{it}, \omega_{it}^L)$ has a known form, which is log-linear in ω_{it}^L :

$$\log(\tilde{M}_{it}) = \sigma \tilde{p}_t + \log(\omega_{it}^L), \quad (2.4)$$

where \tilde{p}_t is the ratio of input prices. A common strategy in the literature is to estimate this linear equation using instruments for input prices and recover ω_{it}^L .¹⁹ However, this strategy relies on the CES functional form, because first-order conditions are, in general, not separable in ω_{it}^L and prices. Therefore, one contribution of this paper is to generalize the CES production function with labor-augmenting technology to an arbitrary functional form, subject to Assumption 2.1. I show that ω_{it}^L is invertible under more general conditions.²⁰

Example 2 (Nested CES). A more flexible parametric form is the nested CES:

$$Y_{it} = \left(\beta_k K_{it}^\sigma + (1 - \beta_k) (\beta_l [\omega_{it}^L L_{it}]^{\sigma_1} + (1 - \beta_l) M_{it}^{\sigma_1})^{\sigma/\sigma_1} \right)^{v/\sigma} \exp(\omega_{it}^H) \exp(\epsilon_{it}). \quad (2.5)$$

This is a special case of my model with $h(K_{it}, \omega_{it}^L L_{it}, M_{it}) = (\beta_l [\omega_{it}^L L_{it}]^{\sigma_1} + (1 - \beta_l - \beta_m) M_{it}^{\sigma_1})^{1/\sigma_1}$. $h(\cdot)$ is homogeneous of degree one and the elasticity of substitution between effective labor and materials is σ_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 Section 4 explains in detail how to employ my approach for the estimation of CES and Nested CES production functions.

My production function differs from these parametric models 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 varies across firms without restriction in my model.²¹

¹⁹In order to estimate Equation (2.4), one needs to observe heterogeneous input prices at the firm level.

²⁰Doraszelski and Jaumandreu (2018) discuss informally how to use \tilde{M}_{it} to control for ω_{it}^L without parametric assumptions. However, their model is more restricted than my model and they do not show invertibility of \tilde{M}_{it} in ω_{it}^L .

²¹See Section 4.5 for how to impose common returns to scale across firms.

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 Pakes (1996) to accommodate two-dimensional productivity shocks.²² 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 ω_{it}^H this period might be associated with higher ω_{it+1}^L next period.²³

Assumption 2.3 (Monotonicity). *Materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^L, \omega_{it}^H), \quad (2.6)$$

where $s_t(K_{it}, \omega_{it}^L, \omega_{it}^H)$ is an unknown function that is strictly increasing in ω_{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 ω_{it}^L , as it affects the marginal product of materials. Therefore, the materials demand function takes capital and two unobserved productivity shocks as arguments. ²⁴

Verifying this assumption requires the primitives of the output market, such as the demand function and competition structure, which I do not model in this paper.²⁵ However, this assumption is intuitive and expected to hold under general conditions. It says that keeping everything else the same, more productive firms have a lower marginal cost, leading to a decline in prices and an increase in output.²⁶

Implicit in this assumption is that there is no unobserved heterogeneity in firms' residual demand curves in the output market; otherwise, the materials input demand function should include firm-

²²The model can be extended to have a controlled Markov process, where some observed variables such as R&D and export can affect the joint distribution of productivity (Doraszelski and Jaumandreu (2013)).

²³Note that this assumption does not restrict the firm's beliefs about future productivity shocks, which can be incorrect and heterogeneous across firms. This is because I am agnostic about the firm's dynamic problem.

²⁴As discussed in Gandhi et al. (2018), this assumption imposes an implicit restriction on ϵ_{it} , i.e. $\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}] = \mathbb{E}[\exp(\epsilon_{it}) \mid K_{it}, \omega_{it}^L, \omega_{it}^H]$.

²⁵Note that this is different from the monotonicity condition for ω_{it}^L , which depends only on the form of production function and not on any other parts of the firm's problem.

²⁶Using standard demand models, Melitz (2000) and DeSouza (2006) verify this assumption for Hicks-neutral productivity.

specific demand shocks, violating two-dimensional unobserved heterogeneity. Even though it is restrictive, it covers commonly some commonly used demand models such as monopolistic and Cournot competitions. Moreover, it allows for ex-post demand shock after the firm decides on its planned output. For more discussion; see Jaumandreu (2018) and Doraszelski and Jaumandreu (2019).

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 $t - 1$.

According to this assumption, investment becomes productive in the next period, implying that firms choose capital one period in advance. As a result, K_{it} belongs to the information set at period $t - 1$, that is, $K_{it} \in \mathcal{I}_{it-1}$ ²⁷.

2.3 Invertibility: Expressing Unobserved Productivity Using Inputs

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

$$\omega_{it}^L = r_t^{-1}(K_{it}, \tilde{M}_{it}) \equiv \bar{r}_t(K_{it}, \tilde{M}_{it}). \quad (2.7)$$

Similarly, Assumption 2.3 provides a monotonicity result for ω_{it}^H using materials demand function in Equation (2.6). Inverting that function yields

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \omega_{it}^L). \quad (2.8)$$

This function contains another unobservable, ω_{it}^L , as an argument. Substituting for it from Equation (2.7) gives

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it})) \equiv \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}). \quad (2.9)$$

Equations (2.7) and (2.9) 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. To see why, if we use Equation (2.9) to control for ω_{it}^H , we have to condition on all the inputs, leaving no variation for identification. To address this problem, I will first develop a

²⁷The approach is robust to a weaker timing assumption, which can potentially provide efficiency gains. For a discussion, see Ackerberg (2016).

control variable approach building on the invertibility results in this section. Then, I will show how to exploit first-order conditions of cost minimization for production function estimation.

3 A Control Variable Approach to Production Function Estimation

The control variable approach relies on constructing variables from data to control for endogeneity. In particular, by conditioning on control variables, one can isolate the exogenous variation and achieve identification (Imbens and Newey (2009), Matzkin (2004)). In this section, I construct a control variable for each productivity shock using the first-order Markov process 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 unobserved variable is strictly monotone in an observed variable and satisfied an independence condition. To apply their approach in production function estimation, I make two innovations. First, I show that the modelling assumption provides an independence condition for innovation to productivity. Second, my model involves two-dimensional unobserved heterogeneity, for which standard control variable approach does not work (Kasy (2011)). I overcome this challenge by using the triangular structure of productivity shocks using Equations (2.7) and (2.8)²⁸.

I derive control variables in two stages. In the first stage, I derive the control variable for ω_{it}^L . In the second stage, building on the first control variable, I derive the control variable for ω_{it}^H . For notational convenience, 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:

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

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

Another interpretation of u_{it}^1 is unanticipated *innovation* to ω_{it}^L , which determines the current period productivity given previous period's productivity. Unlike the standard definition of "innovation", which is separable from and mean independent of past productivity, 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 ω_{it}^L

²⁸The 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), Kim and Petrin (2010)), dynamic discrete choice models (Hong and Shum (2010)) and auctions (Guerre et al. (2009)). To the best of my knowledge, my paper is the first application of the control variable approach to a model with two-dimensional unobserved heterogeneity.

from Equation (3.1) and using Equations (2.7) and (2.9), I obtain

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

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 ω_{it}^L by Assumption 2.1, and $g_1(\cdot)$ is strictly monotone in the last argument by construction. Next, I establish an independence condition so that I can use Equation (3.2) to derive the control variable for ω_{it}^L .

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

Proof. See Appendix B.

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}) belongs to \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) $\tilde{r}(K_{it}, W_{it-1}, u_{it}^1)$ 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 it from data as:

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

where $F_{\tilde{M}_{it}|K_{it}, W_{it-1}}$ denotes the CDF of \tilde{M}_{it} conditional on (K_{it}, W_{it-1}) .²⁹ 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 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 ω_{it}^L as a function of the control variable and past inputs:

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

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

²⁹For notational simplicity, 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 $F_{\tilde{M}_{it}|K_{it}, W_{it-1}}(\tilde{M}_{it} | K_{it}, W_{it-1}) = 1 - u_{it}^1$, a monotone transformation.

3.2 Derivation of the Control Variable for Hicks-Neutral Technology

Control variable derivation for ω_{it}^H proceeds similarly. The Skorohod representation of ω_{it}^H is³⁰:

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

Next, I use the monotonicity of materials in ω_{it}^H given by Assumption 2.3 to write

$$\begin{aligned} M_{it} &= s(K_{it}, c_1(W_{it-1}, u_{it}^1), g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2)) \\ &= s(K_{it}, c_1(W_{it-1}, u_{it}^1), g_2(\bar{r}(W_{it-1}), \bar{s}(W_{it-1}), u_{it}^1, u_{it}^2)) \\ &\equiv \tilde{s}(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2), \end{aligned} \quad (3.6)$$

where $\tilde{s}(\cdot)$ is an unknown function. Note that $\tilde{s}(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2)$ is strictly monotone in u_{it}^2 because $s(K_{it}, \omega_{it}^L, \omega_{it}^H)$ is strictly monotone in ω_{it}^H by Assumption 2.3, and $g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2)$ is strictly monotone in u_{it}^2 by construction.

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

Proof. See Appendix B.

Having strict monotonicity and independence, we can use Equation (3.6) to identify u_{it}^2 . In particular,

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

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

$$\omega_{it}^H \equiv c_2(W_{it-1}, u_{it}^1, u_{it}^2), \quad (3.8)$$

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.^{31,32}

Remark 3.1 (Application to the Cobb-Douglas Production Function). Since my control variable approach relies only on timing and Markov assumptions, it can be applied to other functional forms.

³⁰Unlike the previous subsection, 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 ω_{it}^L and ω_{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 Section Supplemental Appendix 3.3 for control variable derivation under this assumption.

³¹Using the same procedure and substituting past productivities recursively, we can write productivity shocks as $\omega_{it}^L = c_1(W_{it-k}, \{u_{it-l}^1\}_{l=0}^{k-1})$ and $\omega_{it}^H = c_2(W_{it-k}, \{u_{it-l}^1\}_{l=0}^{k-1}, \{u_{it-l}^2\}_{l=0}^{k-1})$ for any integer k , where u_{it-l}^1 and u_{it-l}^2 are defined as in Equation (3.3) and (3.7). This would lead to more identifying variation at the expense of having to estimate more control variables.

³²I show in Supplemental Appendix 3.1 how to extend the control variable approach when there is heterogeneity in input prices.

Supplemental Appendix 4.1 demonstrates its application to Cobb-Douglas production function and discusses its properties. For an overview, consider a value added Cobb-Douglas production function $y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it}^H + \epsilon_{it}$. Using a control variable, ω_{it}^H can be written as $\omega_{it}^H = c(m_{it-1}, k_{it-1}, u_{it})$, where $u_{it} = F_{m_{it}|m_{it-1}, k_{it-1}}(m_{it} | k_{it}, m_{it-1}, k_{it-1})$. 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 the condition $\mathbb{E}[\epsilon_{it} | k_{it}, l_{it}, m_{it-1}, k_{it-1}, u_{it}] = 0$. As I discuss in Supplemental Appendix 4.1, 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 is a flexible input, there is variation in labor conditional on $(m_{it-1}, k_{it-1}, u_{it})$ ³³. 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 (Ackerberg et al. (2015), Bond and Söderbom (2005)). My model is robust to this problem because the second productivity shock, ω_{it}^L , generates additional variation in inputs.

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,

$$\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it}), \quad \omega_{it}^H = \bar{s}(K_{it}, M_{it}, \tilde{M}_{it}). \quad (3.9)$$

Applying the proxy variable approach would require conditioning on (K_{it}, \tilde{M}_{it}) and $(K_{it}, M_{it}, \tilde{M}_{it})$ to control for ω_{it}^L and ω_{it}^H , and then using the last period’s inputs as instruments. However, as pointed out by Gandhi et al. (2018), after conditioning on the proxy variables, there is no variation in any of the inputs. In contrast, the control variable approach relies on a different representation of productivity shocks:

$$\omega_{it}^L = c_1(W_{it-1}, u_{it}^1), \quad \omega_{it}^H = c_2(W_{it-1}, u_{it}^1, u_{it}^2),$$

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 result by exploiting the Markov assumption. Papers using the proxy variable framework, such as Olley and Pakes (1996), Levinsohn and Petrin

³³To see this, if labor is perfectly flexible, we can write it as $l_{it} = l(k_{it}, \omega_{it}^H) = l(k(k_{it-1}, \omega_{it-1}, \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})) = l(k(k_{it-1}, s^{-1}(k_{it-1}, m_{it-1}), \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})) =: \tilde{l}(k_{it-1}, m_{it-1}, u_{it}, \nu_{it-1})$, where ν_{it-1} corresponds to a vector of random variables that affects the firm’s investment decision, such as investment prices and heterogeneous belief about future.

(2003), and Ackerberg et al. (2015), have also assumed that productivity follows a first-order Markov process, but they have not exploited all the information provided by that assumption; they have only used its mean independence implication. In contrast, I fully exploit the Markov assumption, which results in stronger identification results and efficiency gains. However, if 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 identification of the output elasticities, the elasticity of substitution, 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 examine identification of the other features of the production function and explore how further economic restrictions can be imposed on the production function.

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 ω_{it}^H as

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

Since $h(\cdot)$ is homothetic 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

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

This reformulation is convenient because ω_{it}^L becomes an argument 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) gives

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

This equation reveals an identification problem.

Proposition 4.1. *Without further restrictions, h cannot be identified from variation in (W_{it}, Y_{it}) .*

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 .³⁴ Therefore, h is not identified from the relationship between the

³⁴As I show in Supplemental Appendix 4.1, with variation in input prices $\bar{r}(K_{it}, \tilde{M}_{it})$ depends also on the price ratio and functional dependence brakes down.

inputs and output.³⁵ 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. To see this, suppressing the arguments of the functions, we can write output elasticities as

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

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

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

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

Here, \bar{h} can be viewed as an (ex-post) *reduced form* function, which arises as a result of the firm's optimal input choices in equilibrium. It combines the effects of ω_{it}^L and the ratio of the optimally chosen flexible inputs on output. In the rest of this section, I propose a solution to nonidentification of output elasticities by investigating (i) what can be identified from first-order conditions of cost minimization, and (ii) what can be identified from the functions $f(\cdot)$ and $\bar{h}(\cdot)$.

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 θ_{it}^L and θ_{it}^M cannot be identified from variation in the inputs and output. However, the model 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 as:

$$\begin{aligned} \min_{L_{it}, M_{it}} \quad & p_t^L L_{it} + p_t^M M_{it} \\ \text{s.t.} \quad & F(K_{it}, \omega_{it}^L L_{it}, M_{it}) \exp(\omega_{it}^H) \mathbb{E}[\exp(\epsilon_{it}) | \mathcal{I}_{it}] \geq \bar{Y}_{it}. \end{aligned}$$

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 λ_{it} corresponds to the Lagrange multiplier.

³⁵A similar nonidentification result is obtained by Ekeland et al. (2004) in the context of hedonic demand estimation.

Multiplying both sides by $V_{it}/(Y_{it}p_{it})$ and rearranging gives,

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

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

$$\frac{\theta_{it}^M}{\theta_{it}^L} = \frac{\alpha_{it}^M}{\alpha_{it}^L}. \quad (4.4)$$

The ratio of the output elasticities of labor and materials is identified as the ratio of revenue shares using the cost-minimization assumption.³⁶ The revenue shares are often observed in the data so that we can calculate the ratio of elasticities without estimation. An important implication of using the first-order conditions is that identification of output elasticities is possible only at the observed input levels. This situation precludes a counterfactual exercise. I provide further discussion on this in later sections.³⁷

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 and McDevitt (2016) use first-order conditions to identify the Cobb-Douglas and CES production functions, respectively. Gandhi et al. (2018) propose a method that employs Equation (4.3) in a nonparametric fashion. They assume perfect competition in the output market, which implies elasticity equals the revenue share. My contribution is to show how to exploit the first-order conditions nonparametrically in the presence of two flexible inputs, even if firms have market power.

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} as*

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

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

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

$$\theta_{it}^M = f_2 h_3 M_{it}, \quad \theta_{it}^L = f_2 \left(h - h_3 \frac{M_{it}}{L_{it}} \right) L_{it}.$$

³⁶For 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.

³⁷In a recent paper, Doraszelski and Jaumandreu (2019) also used the ratio of revenue shares to identify the ratio of elasticities.

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

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

□

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 functions and labor-augmenting productivity shock.³⁸ The intuition is the following. If labor and materials simultaneously increase by the same factor, $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$ remains the same because it is a function of labor and materials only through their ratio. Thus, any change in the output would not be confounded by the change in ω_{it}^L , and therefore, this change corresponds to the sum of the flexible input elasticities.

Given the sum of elasticities, θ_{it}^V , and the ratio identified in the previous subsection, the labor and materials elasticities can be written as

$$\theta_{it}^L = \theta_{it}^V \frac{\alpha_{it}^L}{\alpha_{it}^V}, \quad \theta_{it}^M = \theta_{it}^V \frac{\alpha_{it}^M}{\alpha_{it}^V}, \quad (4.6)$$

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.

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 from the output elasticity of flexible inputs.

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 B.

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 output elasticities with respect to flexible inputs. Identification of other features suffers from the non-identification problem due to multicollinearity described in Subsection 4.1. As a solution to this problem, I next ask what further restrictions are required to identify the objects in Proposition 4.3.

³⁸Note that even if f and \bar{h} are not uniquely identified, the sum of elasticities is uniquely identified. Assume that 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 $f_2 h = f'_2 \tilde{h}'$. Therefore, the observationally equivalent (f, \tilde{h}) and (f', \tilde{h}') give the same sum of flexible input elasticities.

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 not identified. Consider the following production function:

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

This model differs from the main model in that h does not take K_{it} as an argument.³⁹ Since this is a special case, Proposition 2.1 applies to this production function with $\omega_{it}^L = \bar{r}(\tilde{M}_{it})$. Substitution this into Equation (4.7), I obtain the *reduced form* for the production function in Equation (4.7) as

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

Since K_{it} appears as an argument of f but not of h , this model is more convenient for identification than the main model. The next proposition shows how to identify the output elasticity of capital and the labor-augmenting productivity shock.

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:*

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

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

Proof. See Appendix B.

θ_{it}^K is identified under the additional restriction because ω_{it}^L is not a direct function of capital, implying that we can learn capital elasticity from f_1 . Identification of ω_{it}^L relies on the idea that we can obtain information about the first derivatives of h from the output elasticities of flexible inputs. In the proof, I show that information on the first derivatives of h from the first-order conditions can be mapped back to ω_{it}^L . The identification of ω_{it}^L up to scale is standard in the literature. My final result states the non-identification of elasticity of substitution.

Proposition 4.5. *Under the conditions of Proposition 4.4 the elasticity of substitution between effective labor and materials is not identified from $(f, \bar{h}, \theta_{it}^L, \theta_{it}^M)$.*

Proof. See Appendix B.

³⁹This function is called strongly separable with respect to partition of labor and materials. A production function is called strongly separable if the marginal rate of substitution between two inputs is independent of another input (Nadiri (1982)).

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 I also assume no variation in input prices. In Supplemental Appendix 3.1, I extend my model to have variation in input prices at the firm level. With this extension, the multicollinearity problem disappears, and the elasticity of substitution 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 $\{(\omega_{it}^L, M_{it}) \mid \omega_{it}^L = \bar{r}(\tilde{M}_{it})\}$. This means that I can identify the output elasticities only at the observed input values realized in equilibrium. Therefore, it is not possible to conduct counterfactual exercises, such as keeping ω_{it}^L constant and asking how change in inputs affects output.⁴⁰ However, this is not an important limitation in practice because the majority of the applications of production function require output elasticities and productivity only at the observed values.

4.4 Imposing A Returns to Scale Restriction

My model can easily accommodate a returns to scale restriction on the production function. In particular, if one is willing to restrict the return to scale to an unknown constant v , the production function takes the form

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

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

$$y_{it} = v k_{it} + \tilde{f}(\tilde{L}_{it} \bar{h}(\tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it}, \quad (4.10)$$

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

⁴⁰Note that this problem does not arise in a production function with only Hicks-neutral productivity when first-order conditions are used; see Gandhi et al. (2018). This is because ω_{it}^L is non-separable from the production function, so output elasticities depend on an unobserved variable.

4.5 Summary of Models

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

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

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

$$y_{it} = v k_{it} + f(\tilde{L}_{it} h(\omega_{it}^L, \tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it} \quad (\text{Homogeneous})$$

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

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

Even though I analyze the most general model, a researcher interested in estimating a more restricted production function with labor-augmenting technology can use one of the nested models. The identification strategy and control variable approach, when applied to these special cases, are new.

There are two advantages of providing a family of models, where models are nested within each other. First, comparing the results from a nested model and a general model tests the restrictions imposed by the nested model. For example, we can test the restrictions of the CES model by comparing its estimates with the estimates of the strong homothetic separable model. Second, we can impose regularization based on economic theory. One can start with the most general model to impose as few restrictions as possible. If the estimates are too noisy, then a nested model can be considered to improve precision. This is especially relevant for industries with a small number of firms, for which nonparametric estimation is often not feasible.

5 Empirical Model and Data

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

5.1 Empirical Model

The purpose of my empirical model is to estimate the output elasticities and to infer markups from those estimates. To 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:

$$y_{it} = f(K_{it}, L_{it} \bar{h}(\tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it}. \quad (5.1)$$

In Section 4, I showed how to identify the output elasticities from f and \bar{h} , so the goal is to identify these functions.⁴¹ To control for Hicks-neutral productivity, I use the control variables developed

⁴¹Note that \bar{h} is identified up to a scale since its scale is not identified separately from f . However, the elasticities are uniquely identified. I restrict the logarithm of h to have mean zero in the estimation to impose this normalization.

in Equation (3.8), $\omega_{it}^H = c_2(W_{it-1}, u_{it}^1, u_{it}^2)$. Substituting this into Equation (5.1), the estimating equation can be written as

$$y_{it} = f(K_{it}, L_{it}\tilde{h}(\tilde{M}_{it})) + c_2(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it}. \quad (5.2)$$

Since ϵ_{it} is orthogonal to the firm's information set, we have the conditional moment restriction

$$\mathbb{E}[\epsilon_{it} | W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0. \quad (5.3)$$

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

$$\omega_{it}^H \equiv \tilde{c}_3(\omega_{it-1}^H, \omega_{it-1}^L) + \xi_{it},$$

for an unknown function $c_3(\cdot)$, where ξ_{it} is the separable innovation to Hicks-neutral productivity with $\mathbb{E}[\xi_{it} | \mathcal{I}_{it-1}] = 0$. This innovation term is different from those defined in Section 3 because it is mean independent of $(\omega_{it-1}^H, \omega_{it-1}^L)$ and separable, in contrast to (u_{it}^1, u_{it}^2) , which are independent and non-separable. ξ_{it} is commonly used in the proxy variable approach for constructing moments.

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

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

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 $\mathbb{E}[\xi_{it} + \epsilon_{it} | K_{it}] = 0$, additional moment restrictions. Now I summarize the estimation problem by combining the models and moment restrictions. We have two estimating equations

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

with two conditional moment restrictions:

$$\mathbb{E}[\epsilon_{it} | W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0, \quad (5.5)$$

$$\mathbb{E}[\xi_{it} + \epsilon_{it} | K_{it}, W_{it-1}] = 0. \quad (5.6)$$

Estimating output elasticities requires estimates of the unknown functions f , \bar{h} , c_2 and c_3 using these moment restrictions. In Supplemental Appendix C, I analyze the identification of f and \bar{h} based on the moment restriction in Equation (5.5) and show that it identifies f and \bar{h} except for special cases.⁴² These cases include some support conditions on the derivatives of conditional CDF in Equation (3.7),

⁴²This is sometimes called generic identification; see Lewbel (2016).

so they are testable.⁴³ Since Equation (5.5), by itself, generically identifies the output elasticities, the moment restriction in Equation (5.6) provides efficiency gains and overidentifying restrictions.⁴⁴

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. Then, I approximate the nonparametric functions using polynomials and use the moment restrictions in Equations (5.5) and (5.6).

5.1.1 Estimation Procedure

In this section, I provide an overview of the estimation procedure. A more detailed estimation algorithm is given in Supplemental Appendix 1.7.

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. (2018).⁴⁵

The estimation involves two stages. In the first stage, I learn 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 follow Chen and Pouzo (2012) and 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 functions with the approximations yields

$$\begin{aligned} y_{it} &= \hat{f}(K_{it}, L_{it}\hat{h}(\tilde{M}_{it})) + \hat{c}_2(W_{it-1}, \hat{u}_{it}^1, \hat{u}_{it}^2) + \hat{\epsilon}_{1it}, \\ y_{it} &= \hat{f}(K_{it}, L_{it}\hat{h}(\tilde{M}_{it})) + \hat{c}_3(W_{it-1}) + \hat{\xi}_{it} + \hat{\epsilon}_{2it}. \end{aligned}$$

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:

$$J(\hat{f}, \hat{h}, \hat{c}_2, \hat{c}_3) = \underbrace{\frac{1}{TN} \sum_{i,t} \hat{\epsilon}_{1it}^2}_{\text{Sum of Squared Residuals}} + \underbrace{\left(\frac{1}{TN} \sum_{i,t} (\hat{\xi}_{it} + \hat{\epsilon}_{2it}) K_{it} \right)^2 + \left(\frac{1}{TN} \sum_{i,t} (\hat{\xi}_{it} + \hat{\epsilon}_{2it}) K_{it}^2 \right)^2}_{\text{Timing Moments}} \quad (5.7)$$

I minimize this objective function for estimation. The estimation of $\hat{c}_2(W_{it-1})$ and $\hat{c}_3(W_{it-1})$ are computationally simple as they can be partialled out for a given (\hat{f}, \hat{h}) . So the estimation requires searching for \hat{f} and \hat{h} to minimize the objective function. After obtaining the estimates for f and \hat{h} , I calculate the output elasticities as described in Equations (4.5), (4.6) and (4.9).

⁴³I also provide these conditions for homothetic production function in Appendix Proposition C.1 and for strong homothetic separable production function in Appendix Proposition C.2.

⁴⁴As described in Section 3.1, it is possible to construct other moment restrictions to increase efficiency at the expense of a more complicated estimation procedure.

⁴⁵The number of rolling windows is higher for the US than other countries because the US sample size is significantly smaller than those of other countries. The results are robust to different rolling window size but they are less precise.

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 the bootstrap to estimate standard errors. The bootstrap procedures treat firms as independent observations and resample firms with replacement.

5.2 Data

For the empirical model, I use panel data from manufacturing industries in five countries: Chile, Colombia, India, Turkey, and the United States. The data source for the US is Compustat, compiled from firms' financial statements. For other countries, I use plant-level production datasets. The sample periods are given in Table 1, which vary across countries based on data availability. The US data covers the longest period, from 1961 to 2014. The Indian sample covers a recent period, while the Chilean, Colombian, and Turkish samples end before 2000.

5.2.1 Chile, Columbia, India, Turkey

The datasets 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 dataset comes 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.⁴⁶ The Turkish dataset is from the Annual Surveys of Manufacturing Industries (ASMI), conducted by the Turkish Statistical Institute, and covers all establishments with ten or more employees.⁴⁷ Finally, the Indian data come from the Annual Survey of Industries conducted by the Indian statistical institute for plants with 100 or more employees.⁴⁸

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. For materials cost, I construct separate measures of materials for non-energy raw materials and energy (which includes electricity and fuels) for the manufacturing datasets. Materials cost is the sum of the cost of 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. The number of industries ranges from five to eight across datasets. I provide details about the data collection, industries, and summary statistics in Supplemental Appendix 1.

⁴⁶Some examples are Gandhi et al. (2018), Eslava et al. (2010) and Pavcnik (2002), and Liu (1993).

⁴⁷This dataset has previously been used by Levinsohn (1993) and Taymaz and Yilmaz (2015).

⁴⁸The survey also includes a sample of firms with less than 100 employees. I exclude these firms from my sample.

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
Level Of Estimation	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.

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 1.4

Some concerns about Compustat data are worth mentioning in the context of production function estimation. First, Compustat is not representative of the general economy as it only includes publicly traded firms. These firms are bigger, older and more capital intensive. Second, firms drop out of the sample due to mergers and acquisitions and enter the sample as they become public. Finally, it is from accounting data, which is low-quality compared to traditional manufacturing censuses.

The concerns on Compustat cast doubt on the suitability of Compustat data for production function and estimation. 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 and Scott (2016)). 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 given above and check whether I obtain similar results using these datasets.

6 Empirical Results: Production Function

This section presents results from the empirical model. I use production function estimates to discuss several findings. First, I find that my model generates different output 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 in a way that is consistent with previous

findings.

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.

There are large differences in the average elasticity estimates between my model and Cobb-Douglas estimated with ACF. Cobb-Douglas generates higher labor elasticities and lower capital elasticities than my model for most industries. Lower labor elasticity estimates from my method are consistent with labor’s low revenue share in the data. 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 transmission bias 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 from my model and ACF, along with the difference in Figure 1.⁴⁹ The results suggest that I estimate a higher output elasticity of capital and lower elasticity of labor in all countries. The difference is statistically significant in all countries for the labor elasticity and in all countries except the US for the capital elasticity. Drawing the same conclusions in all datasets provides strong evidence that these results are robust to the sample period and country-specific characteristics.⁵⁰

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 labor- and flexible-input intensive? (ii) Are exporters more capital-intensive? The literature has found heterogeneity at the firm-level along many dimensions, including productivity, labor share, and size (Van Reenen (2018)). However, there is limited evidence on firm-level heterogeneity in production technology. Moreover, this section provides some evidence for the external validation of

⁴⁹Other elasticity estimates are reported in Supplemental Appendix Figure 6.8.

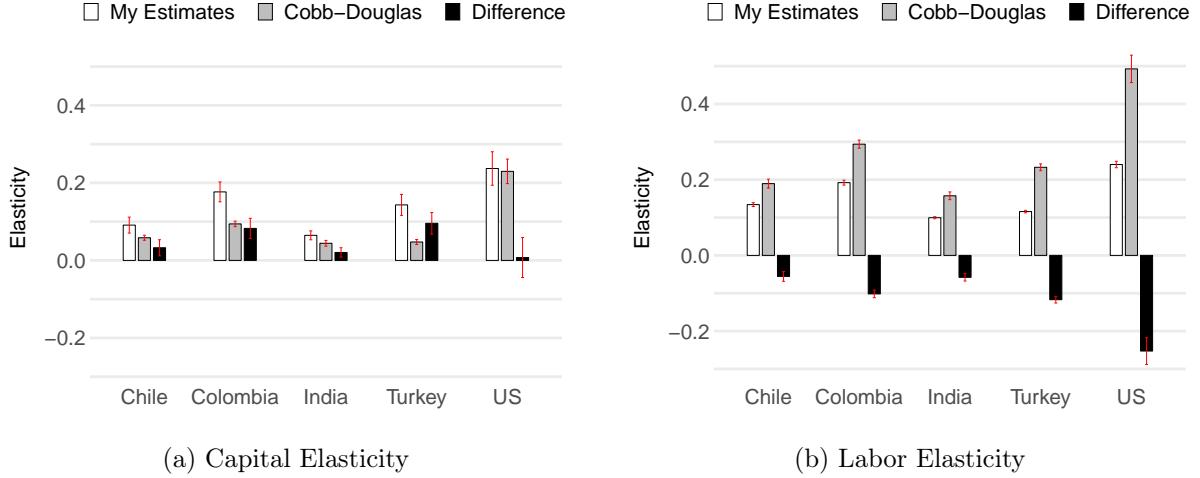
⁵⁰A common problem in production function estimation is measurement error in capital, which could be more severe in a nonparametric model. With measurement error in capital, the capital elasticity estimates will be biased towards zero, and other elasticities will be biased upwards since other inputs are usually positively correlated with capital. I verify this prediction using a simulation exercise in Section 5.5. Since my results suggest larger capital elasticity and lower labor elasticity, they cannot be driven by measurement error. See Hu et al. (2011), Collard-Wexler and De Loecker (2016) and Kim et al. (2016) for attempts to address measurement error in capital.

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

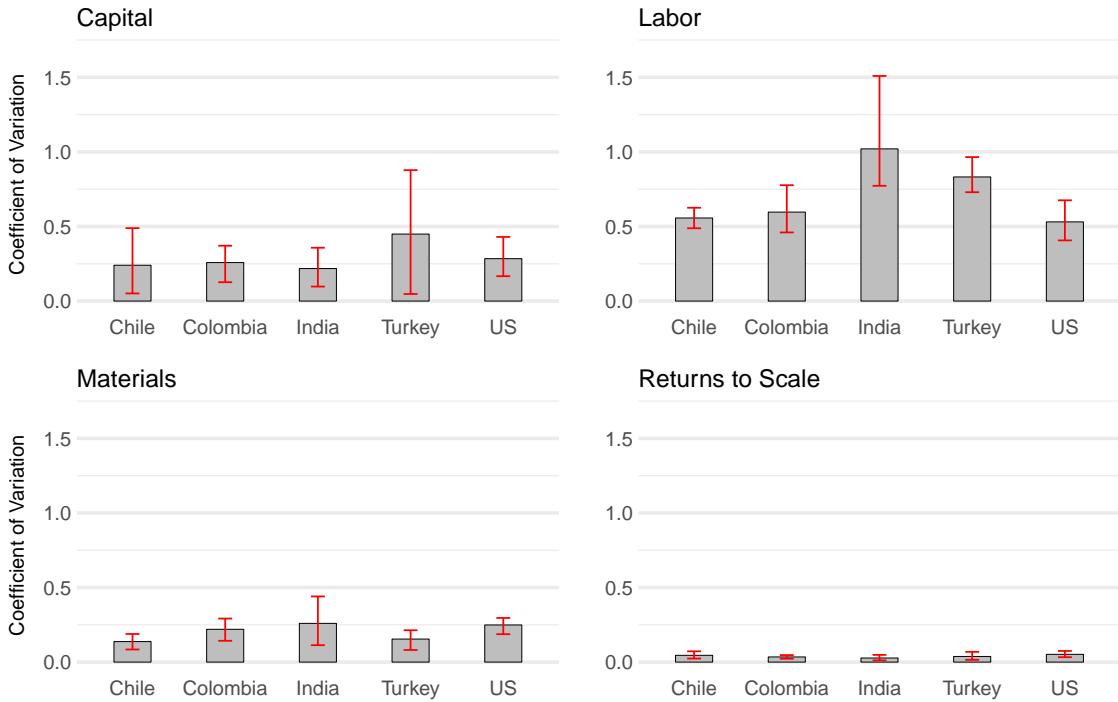
Note: Comparison of sales-weighted average output elasticities produced by different methods. FA refers to my estimates, ACF refers to Ackerberg et al. (2015) estimates and OLS is Cobb-Douglas estimated by OLS. For each year and industry, sales-weighted averages are calculated, and then simple averages 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



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 percent confidence intervals calculated using bootstrap (100 iterations).

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.

Table 3: Regressions of the Output Elasticities on Firm Size

	Chile	Colombia	India	Turkey	US
Capital Elasticity	0.008 (0.000)	0.02 (0.000)	0.006 (0.000)	0.016 (0.000)	0.025 (0.000)
Labor Elasticity	-0.021 (0.000)	-0.037 (0.000)	-0.053 (0.000)	-0.024 (0.000)	-0.016 (0.000)
Flexible Input	-0.023 (0.000)	-0.02 (0.000)	-0.011 (0.000)	-0.012 (0.000)	-0.004 (0.000)
Capital Intensity	0.253 (0.003)	0.303 (0.002)	0.387 (0.002)	0.396 (0.002)	0.228 (0.001)

Notes: Regressions of firm size on the output elasticities and capital intensity controlling for 4 digit industry-year fixed effects based on Equation (6.1). Firm size is proxied by the logarithm of sales. Each row corresponds to a separate regression where left-hand side variable is given in the first column. Standard errors are clustered at the firm level and reported in parentheses.

the model since firm size and export are outside the model.

To measure heterogeneity, I estimate the coefficient of variation (CV) of the output elasticities within each industry-year group. 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 the labor elasticity and lowest for the materials elasticity. This finding is consistent with the large heterogeneity in labor's revenue share and low heterogeneity in materials' revenue share observed in the data. Also, the 10-90th percentiles show that this result is not driven by only a small number of industries. Also, 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.

The presence of 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 literature on heterogeneity in production technology is scarce, there are two findings on the relationship between production functions and other economic variables. First, the literature has found that large firms are more capital-intensive than small firms (Holmes and Mitchell (2008), Kumar et al. (1999)). Second, the literature has documented that 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:

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

where j indexes the 4-digit industry, so δ_{jt} denotes the industry-year interaction fixed effects. γ 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

Table 4: Regression of Capital Intensity on Export Status

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

Notes: Regressions of capital elasticity on a dummy of whether the firm exports, controlling for 4 digit industry-year fixed effects. Each row corresponds to a separate regression where the left-hand side variable is given in the first column. Standard errors are clustered at the firm level and reported in parentheses.

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 coefficients estimates in the second and third row suggest that flexible input elasticity and labor elasticity are negatively associated with firm size. Overall, these findings agree with the literature, which finds that large firms use more capital, and less labor relative to 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 only for these countries firm-level export data are available. The outcome variables are capital intensity and capital elasticity.⁵¹ 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 than domestic firms in both countries. This finding is also consistent with the existing empirical evidence.

In brief, this section documents substantial heterogeneity in production technology that is related to firm size and export status. This analysis can also be seen as an external validation exercise 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 predicted by theoretical literature and the results agrees with the existing empirical evidence.

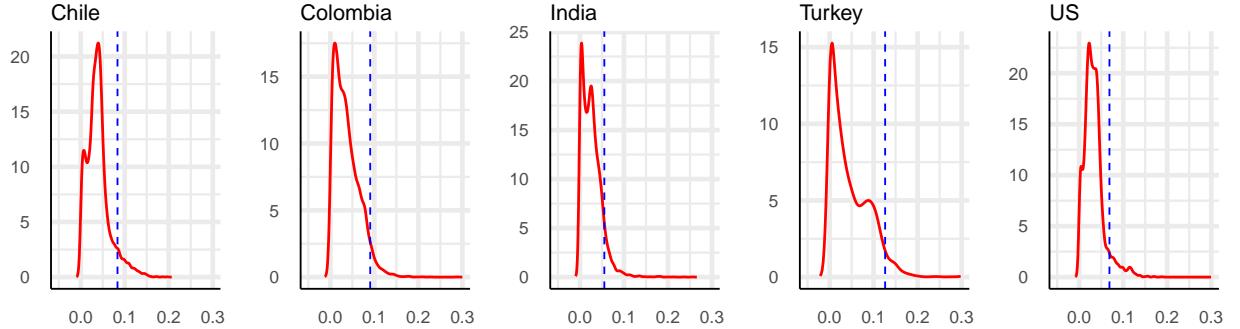
7 Inferring Markups from Production

There is a simple link between a firm's markup and its output elasticities, which has been widely used to estimate markups recently. In this section, I first describe this link and then argue 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

⁵¹I focus on these variables because trade literature finds an association between capital intensity and export.

Figure 3: Distribution of Coefficient of Variation of Sum Elasticity



Notes: This figure compares the distribution of coefficient of variation of the sum elasticity within firm (red) with the estimates of unconditional coefficient of variation in the entire sample (blue) for each country.

markups from production data under the assumption that firms are cost-minimizers with respect to at least one flexible input and they take input prices as given. In particular, under these assumptions, markup is given by

$$\mu_{it} := \frac{\theta_{it}^V}{\alpha_{it}^V}, \quad (7.1)$$

where μ_{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 enough to estimate markups. Moreover, since Equation (7.1) holds for all flexible inputs, we need an estimate of only one flexible input's elasticity.

In recent years, estimating markups from production data 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. (2019) and 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 Markup Estimates?

Output elasticity is the only estimated component of markup in Equation (7.1). Therefore, when the production approach is used for markup estimation, the bias in output elasticity directly translates into markups, making the markup estimates sensitive to the form of the production function.⁵² In this section, I first discuss implications of functional form assumptions on markups and then argue that labor augmenting productivity provides solution to some common problems in the literature.

Heterogeneity in Markups. Much of the empirical research estimating markups assumes a Cobb-

⁵²Van Biesebroeck (2003) compares conventional production function estimation methods and finds that they give broadly similar productivity measures, but significantly different output elasticities. Therefore, how we estimate production functions is particularly important for markup estimation.

Douglas production function. 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 is particularly important for 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 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 often give conflicting markups estimates (De Loecker et al. (2018), Doraszelski and Jaumandreu (2019), Raval (2019a)). This evidence suggests that at least one assumption required to estimate markups from production data is violated.

Raval (2019a) formally tests the production function approach using its implication that two flexible inputs should give the same markups. 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 that explain this result, such as heterogeneity in the production function, adjustment costs in labor, measurement error, and violation of cost minimization assumption. He concludes that the most plausible explanation is the inability of the standard production functions to account for heterogeneity in production technology.⁵³

Raval (2019a)'s results suggest unobserved heterogeneity in the output elasticities as a potential solution to conflicting markups estimates. One contribution of this paper is to show that labor-augmenting productivity ensures identical markup estimates from labor and materials. 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 in Subsection 4.2.1. The latter immediately implies that the two markups estimates are the same:

$$\frac{\theta_{it}^L}{\theta_{it}^M} = \frac{\alpha_{it}^L}{\alpha_{it}^M} \implies \mu_{it}^L = \frac{\theta_{it}^L}{\alpha_{it}^L} = \frac{\theta_{it}^M}{\alpha_{it}^M} = \mu_{it}^M, \quad (7.2)$$

where μ_{it}^L and μ_{it}^M denote markup estimates from labor and materials, respectively.⁵⁴ However, the presence of labor-augmenting productivity is crucial to be able to use the ratio of revenue shares to identify the ratio of output elasticities. As shown in Section 6.1 and argued by Raval (2019b), without the labor augmenting productivity the identity in Equation (7.2) is rejected by the model. This identification result also provides some intuition for identification: the over-identifying restrictions already available in the Hicks-neutral production function allow me to add another unobserved productivity and identify the model.

⁵³To account for labor-augmenting productivity he uses the quintile cost share method, where quantiles correspond to labor cost to materials cost ratio. He finds that this method gives positively correlated markups from labor and materials.

⁵⁴Doraszelski and Jaumandreu (2019) also make this observations.

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 markup through two channels: (i) bias in the average output elasticity and (ii) firm-level heterogeneity in the output elasticities. After estimating firm-level markups, researchers often compute the aggregate markup, μ_t , for an industry or economy using:

$$\mu_t = \sum_{i=1}^{N_t} w_{it} \mu_{it},$$

where w_{it} is the aggregation weight, usually a measure of firm size. Recently, researchers have used the aggregate markup to measure the change in market power in the US and other developed economies (De Loecker et al. (2018), Diez et al. (2018)).

To assess the influence of production function estimation on the estimated aggregate markup, I apply a decomposition method proposed by Olley and Pakes (1996). This method decomposes a weighted average into two parts: (1) an unweighted average and (2) covariance between the weight and variable of interest. To implement the Olley-Pakes decomposition, I look at the logarithm of markup. Using the firm-level markups and weights, the aggregate log markup can be expressed as

$$\tilde{\mu}_t = \sum_{i=1}^{N_t} w_{it} \log(\theta_{it}) - \sum_{i=1}^{N_t} w_{it} \log(\alpha_{it}),$$

which equals the difference of two weighted averages. Therefore, we can apply the Olley-Pakes decomposition to both terms to obtain:

$$\tilde{\mu}_t = \underbrace{\bar{\theta}_t}_{\substack{\text{Average Elasticity (1)} \\ \text{Estimation}}} + \underbrace{\text{Cov}(w_{it}, \log(\theta_{it}))}_{\substack{\text{Heterogeneity in Technology (2)}}} - \underbrace{\bar{\alpha}_t}_{\substack{\text{Average Share (3)} \\ \text{Data}}} - \underbrace{\text{Cov}(w_{it}, \log(\alpha_{it}))}_{\substack{\text{Heterogeneity in Shares (4)}}} \quad (7.3)$$

The aggregate log markup is composed of four parts. The first two parts involve the output elasticity: (1) is the unweighted average of log elasticity, denoted by $\bar{\theta}_t$ and (2) is the covariance between firm size and log elasticity. The last two parts involve the revenue share: (3) is the unweighted average revenue share, denoted by $\bar{\alpha}_t$, 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 elasticity, which is estimated, or the revenue share, which comes directly from the data. Therefore, we can disentangle the role of the elasticity estimates from the revenue shares in markup estimation. More precisely, since production function estimates appear only in the first two components, analyzing those components reveals how biases in production function estimates translate into markup estimates.

7.2.1 Bias from the Average Output Elasticity

The first component in the decomposition is the average elasticity, which reflects the underlying production technology in the economy. Under misspecification, this component will be estimated with bias, which directly translates into bias in the aggregate markup.⁵⁵ My output elasticity estimates in the previous section suggested that Cobb-Douglas overestimates the flexible input elasticity. Therefore, the bias from this source should be positive.

7.2.2 Bias from Heterogeneity in Production Technology

The second component in the decomposition is the covariance between firm size and the output elasticity of flexible input. 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 magnitude of the second component in the markup decomposition. A production function that fails to capture this trend in production technology overestimates the change in the aggregate markup.

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

8 Empirical Results: Markups

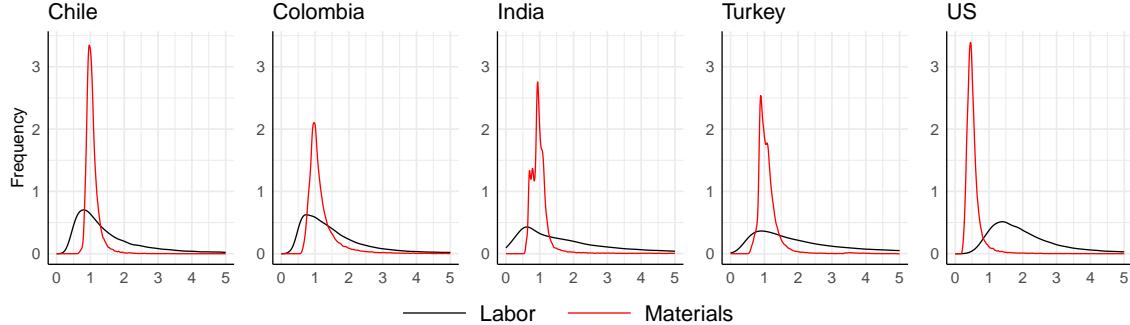
I estimate markups using the output elasticities reported in Section 6. With these estimates in hand, I look at whether my markup estimates are systematically different from those generated by Cobb-Douglas and other production functions.⁵⁶ My aggregate markup estimates are lower than the Cobb-Douglas estimates in all countries. I find that 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 trend in markups. For this analysis, I focus only on the US, given the recent empirical findings on the rise

⁵⁵It is difficult to evaluate the direction or magnitude of this bias, besides some special cases, as it comes from misspecification rather than from an omitted variable. Therefore, I rely on the empirical model to understand the bias by comparing the average output elasticities across different methods.

⁵⁶This section mainly focuses on the comparison with Cobb-Douglas since it is the most commonly used specification. In Supplemental Appendix 5, I compare my results with the translog production function with Hicks-neutral productivity and in Subsection 5.4) with Nested CES production function with labor-augmenting productivity.

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



Notes: This figure compares the distribution of markups implied by labor (black) and materials (red) elasticities from the Cobb-Douglas specification estimated using the Ackerberg et al. (2015) procedure for each country. Each plot shows the result for the country given in the top-left corner.

in markups in the US. I find that the markup growth is lower according to my estimates.

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 4 plots the distributions of markup estimates inferred from the labor and materials elasticities. If the model is correct, the two distributions should overlap. However, the distributions are quite different, with labor generating a more dispersed distribution than materials in all countries. This result is driven by high dispersion in labor's revenue share in the data, as Cobb-Douglas model assumes constant output elasticities. Moreover, both markup measures indicate that a significant fraction of firms have markups below one. These results provide strong evidence against the Cobb-Douglas specification.

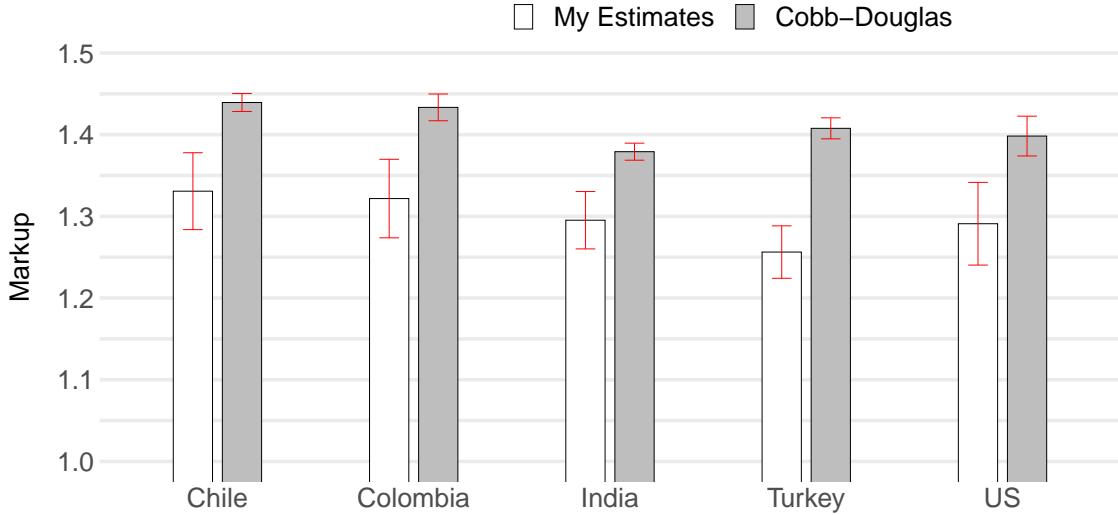
Since I reject the Cobb-Douglas specification with two flexible inputs, I estimate another production function with a single flexible input for comparison purposes, following De Loecker et al. (2018):

$$y_{it} = \beta_k k_{it} + \beta_v v_{it} + \omega_{it} + \epsilon_{it}.$$

Here, v_{it} is the combined flexible input of labor and materials, defined as the deflated sum of labor and materials cost.⁵⁷ I estimate this model using the ACF method and calculate markups as $\mu_{it}^{CD} = \frac{\beta_v}{\alpha_{it}^V}$. For my model, I use the sum of flexible input elasticity divided by flexible input's revenue share as my markup measure. This markup measure equals the markups obtained from labor and materials elasticities.

⁵⁷Having a single flexible input avoids conflicting markups estimates. However, this model implicitly assumes that labor and materials are perfect substitutes because only under that assumption β_v equals the output elasticity of the flexible input.

Figure 5: 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. 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).

8.2 Markups Comparison: Level

This section compares the aggregate markups produced by my method and by 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.

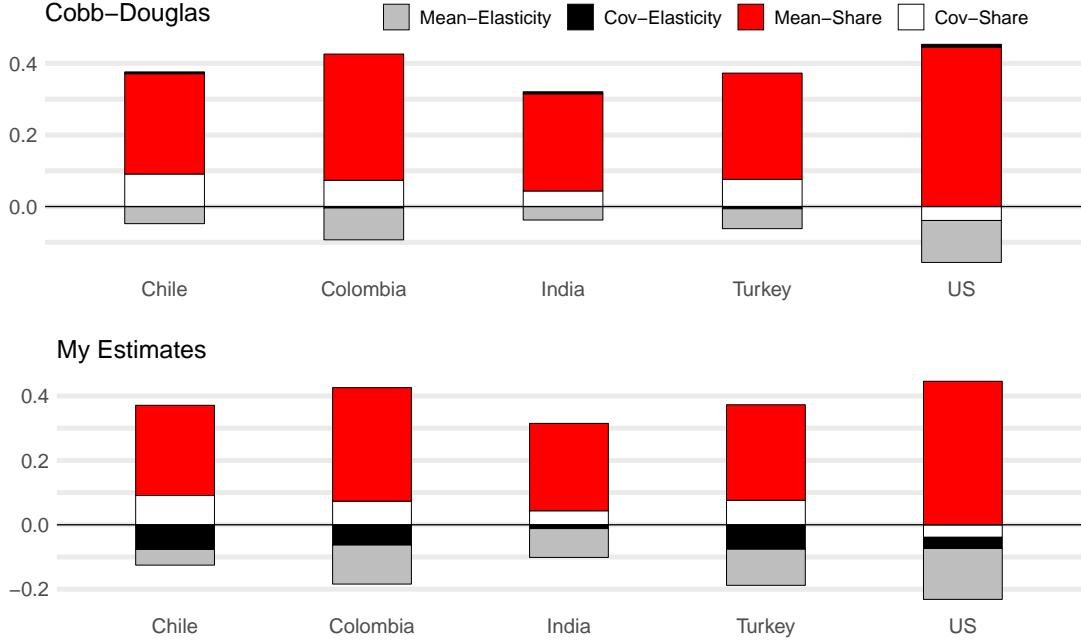
For each country, I first calculate the sales-weighted markup for every year and then take the average over years.⁵⁸ Figure 5 displays the aggregate markups from the two methods, along with the 95 percent confidence interval. My model generates aggregate markups that are significantly smaller than the 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 compelling evidence that the results are not driven by country-specific characteristics.⁵⁹

What drives these differences in markups estimates? I answer this question by decomposing markups into its four components, as presented in Section 7.2. These components, averaged over time, are presented in Figure 6. The red and white bars come directly from the data, and their

⁵⁸Edmond et al. (2018) argue that weighting by cost, instead of sales is more appropriate for understanding the welfare implications of markups. I report cost-weighted estimates in Supplemental Appendix 6 and find qualitatively similar results.

⁵⁹Supplemental Appendix Figure 6.15 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. I find that the Cobb-Douglas markup estimates are always higher than my markups estimates, and the difference is statistically significant. So this difference is not driven by a small number of years.

Figure 6: Markup Decomposition



Notes: This figure compares the four components of the aggregate log-markups given in Equation (7.3) produced by my method and Cobb-Douglas estimated with ACF procedure. For each country, each component is averaged over years and indicated by a different color.

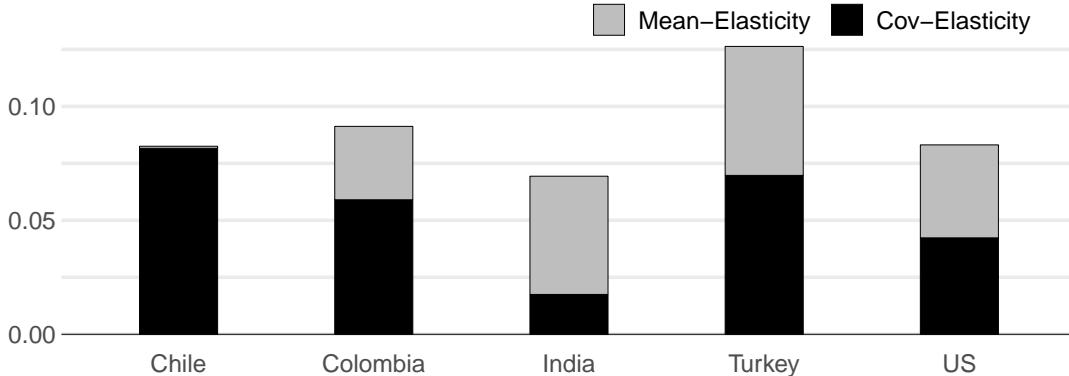
magnitudes are the same for both estimation methods. Therefore, markup estimates between the two methods differ only through the grey and black bars in Figure 6. The largest difference is from the black bar, the covariance between firm size and elasticity. While this component is negligible under the Cobb-Douglas assumption, my estimates suggest that it is negative. This 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.

To focus on the first two components I take the difference between my markup measure and the Cobb-Douglas markup measure:

$$\tilde{\mu}_t^{CD} - \tilde{\mu}_t = \underbrace{\bar{\theta}_t^{CD} - \bar{\theta}_t}_{\text{Mean-Elasticity}} + \underbrace{\text{Cov}(w_{it}, \log(\theta_{it}^{CD})) - \text{Cov}(w_{it}, \log(\theta_{it}))}_{\text{Cov-Elasticity}}, \quad (8.1)$$

where 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 output elasticity across two methods. I plot these differences in Figure 7. Except for Chile, both components are positive for all countries. This result reveals two key reasons behind the difference in markup estimates between 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. Both of these factors generate upward bias in the

Figure 7: 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 (Equation 8.1).

Cobb-Douglas markup estimates.

8.3 Markups Comparison: Trend

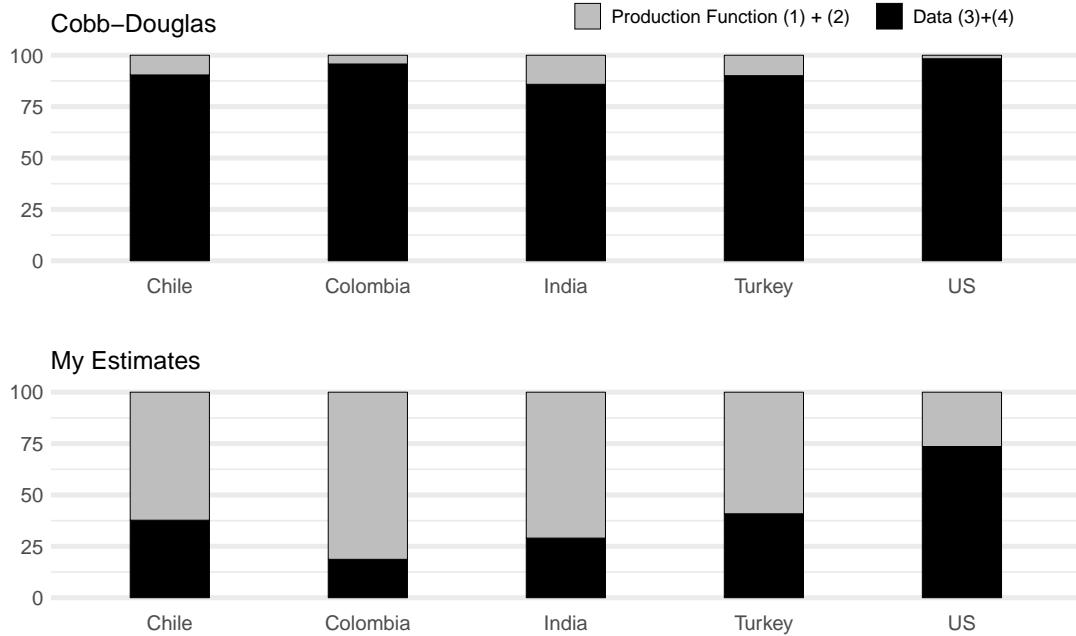
After showing important differences in the level of markups across estimation methods, I now turn to the change in markups over time. I start by looking at what explains the time series variation in markups. Then I focus on the markup growth in US manufacturing.

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 (7.3), ignoring the covariance between the two. Figure 8 presents the results from this decomposition for both production functions. Focusing on the Cobb-Douglas model first, we see that a large fraction of the variance is explained by the change in revenue shares, consistently in all datasets. The result is particularly striking for the US, where the contribution of the change in output elasticity is only 1%. The decomposition results from my method reveal a different picture. The change in the output 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.⁶⁰ Is the role of change in technology really minimal? For the rest of this section, I seek to answer this question.

Figure 8: Variance Decomposition of the Change in Markups



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

8.3.2 Change in Markups in the US Manufacturing Sector

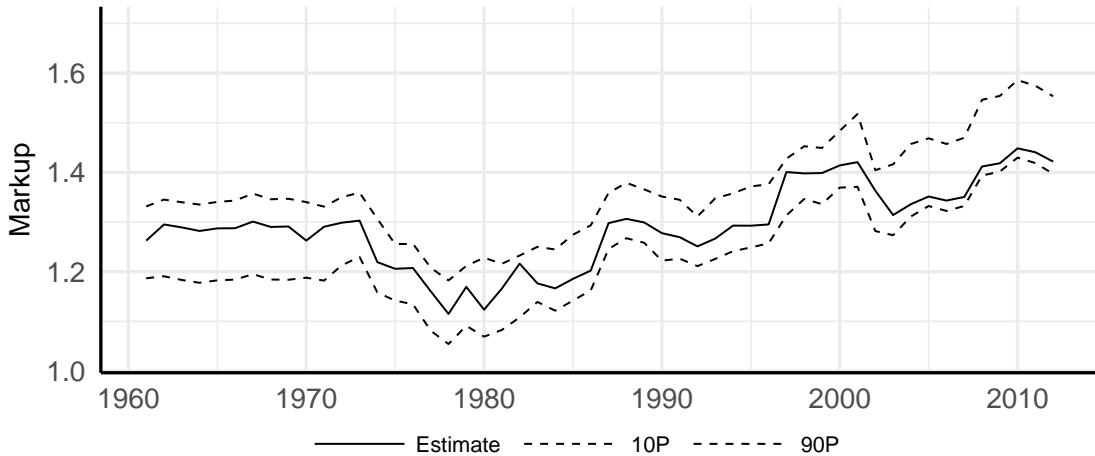
This section investigates the evolution of the aggregate markup in the US manufacturing sector. Figure 9 plots the sales-weighted aggregate markup from 1960 to 2012 along with the 10-90th percentile confidence band. In the 1960s, the aggregate markup is about 30 percent over marginal cost. It remains flat until 1970 and then declines gradually between 1970 and 1980, falling to about 15 percent in 1980. Starting from this point, markups rise with some cyclical pattern and reach 40 percent at the end of the sample period. We also see that the aggregate markup tends to decline during recessions. Overall, the aggregate markup in the manufacturing industry has risen from 30 percent to 40 percent during the sample period.⁶¹

One concern about a nonparametric model is precision because a nonparametric model trades off flexibility for precision, generating noisier estimates than the parametric models. The narrow confidence band reported in Figure 9 suggests that this is not a concern. Note that the estimate is not centered around the confidence band because the aggregate markup is a non-linear function of the output elasticities. The change in the sample size affects the width of the confidence band—the

⁶⁰This is also evident in Supplemental Appendix Figure 6.14, which displays the evolution of markups along with its two components. The aggregate markups closely track the revenue share in all countries.

⁶¹To explore the importance of weighing in aggregation, Supplemental Appendix Figure 6.9 compares the sales-weighted and cost-weighted markup series. Although they exhibit similar trends, the sales-weighted markup is always above the cost-weighted markup. Moreover, the change in the sales-weighted markup is larger than the change in the cost-weighted markup.

Figure 9: Evolution of the Aggregate Markup



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

sample size of Compustat changes with mergers and acquisitions over the estimation period. The sample size is small at first, with few publicly-traded companies in the 1960s. The sample size increases until the 1990s and then declines again. We see the impact of this on the width of the confidence band: The most precise markup estimates are obtained in the 1990s.

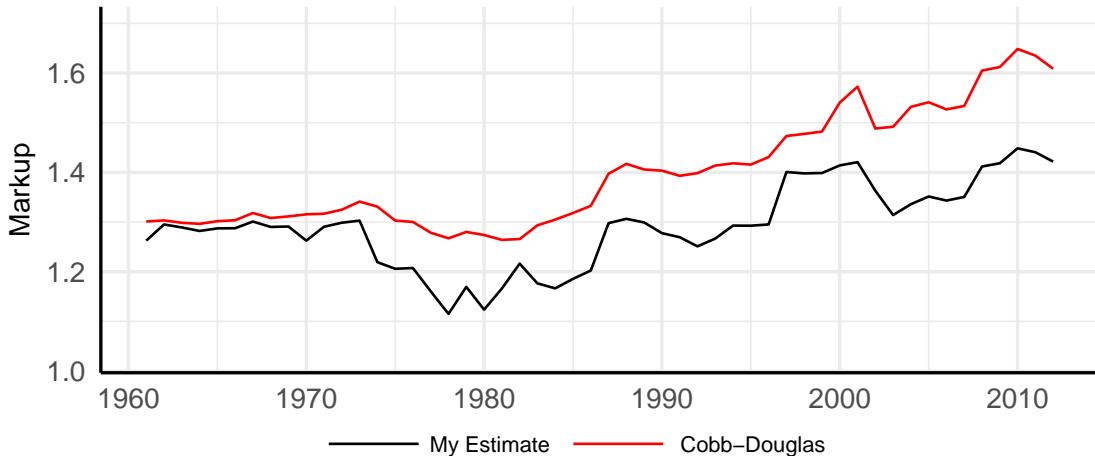
Next, I compare my results with the Cobb-Douglas estimates. Cobb-Douglas estimation is essentially a replication of De Loecker et al. (2018), who estimated a Cobb-Douglas production function with a single flexible input. They find a dramatic rise in markups in the US economy since 1960 and interpret this finding as a large increase in market power. My goal is to understand how a flexible production function affects this conclusion.

Figure 10 reports both markups measures. The Cobb-Douglas estimates suggest that markups rose more than 30 percent between 1960 and 2012. This finding mirrors De Loecker et al. (2018)'s finding and is essentially a replication of their result for the manufacturing industry. The markups estimates from my production function also suggest a rise in markup, albeit a more modest one: 13 percent between 1960 and 2012. This rise is even smaller when markups are weighted by cost shares. The overall change is not the only difference. The series closely follow each other between 1960 and 1970, but they start to diverge after 1970. Also, my estimates have cyclical markup estimates, consistent with the business cycle in the US.⁶²

This result has an important implication for the evolution of market power in the US manufacturing industry. As shown by the variance decomposition, using a restrictive production function does not indicate any change in production technology over time, and markup estimates are driven by the change in revenue share. Viewed in this light, the rise in markup in the US manufacturing industries, according to the Cobb-Douglas specification, is explained by the decline in the labor

⁶²As a robustness check, I estimate a nested CES production function with labor-augmenting productivity in Supplemental Appendix Section 5.4 and compare markups estimates.

Figure 10: 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.

share.

9 Extensions

In this section, I briefly discuss three extensions to my model by showing how to account for (i) heterogeneity in input prices, (ii) unobserved materials prices and (iii) non-random firm exit. I provide details for these extensions in Supplemental Appendix 3.

9.1 Heterogeneous Input Prices

My main 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 and detailed datasets.⁶³ To accommodate this case, I develop an extension in Supplemental Appendix 3.1, which assumes that firms might face different input prices, but they do not have market power in the input markets. This extension requires incorporating heterogeneous input prices into the model and modifying the estimation procedure, but the general framework and identification strategy remains the same.

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.⁶⁴ The key in this extension is to show that unobserved materials-augmenting

⁶³For examples, see De Loecker et al. (2016), Grieco et al. (2016), and Atalay (2014).

⁶⁴In a recent paper, Grieco et al. (2016) study this question under the assumption that the production function is CES. My extension can be seen as a generalization of their framework to a nonparametric production function.

productivity is equivalent to a model with unobserved and heterogeneous 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 3.2.

9.3 Accounting for Firm Selection

In Supplemental Appendix 3.3, I present a way of incorporating non-random firm exit into my estimation framework. I achieve this extension under two simplifying assumptions. First, I assume that the non-separable innovations to productivity shocks defined in Section 3 are independent of each other conditional on previous period's productivity. Second, I assume that 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 observables and firms that draw Hicks-neutral productivity below that cutoff exit. I estimate the propensity of exit conditional on the previous period's inputs and current period's capital level, which allows me to control for selection. The empirical results from implementing this selection correction procedure are provided in Supplemental Appendix 5.3.

10 Conclusions

This paper first proposed an approach to estimate nonparametric production functions with labor-augmenting productivity. Then, it used this new approach to estimate output elasticities and markups using manufacturing data in five countries.

Methodologically, I contribute to the literature by introducing an identification and estimation method for production functions with labor-augmenting and Hicks-neutral productivity. Unlike previous methods, the identification strategy does not rely on parametric restrictions or variation in input prices. The identification is challenging due to two sources of unobserved heterogeneity and absence of parametric restrictions. To address this challenge, I first incorporate labor-augmenting productivity into the standard proxy variable framework from Olley and Pakes (1996). Then, using a novel control variable approach, I show how to overcome the endogeneity of productivity shocks. Finally, after showing that flexible inputs elasticities are not identified, I propose exploiting first-order conditions without parametric assumptions.

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 and overestimate labor elasticity. The estimates also document substantial firm-level heterogeneity in the output elasticities. To what extent these biases and heterogeneity translate into the inferred markups? The estimates suggest that the standard methods generate an upward bias in both the level and growth of markups. I also revisit the recent findings on the rise of US markups. My estimates suggest that markup growth in the US manufacturing sector is 15 percent, in contrast to 30 percent as suggested by recent papers.

A Supplementary Lemmas

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 $f(x | y)$ 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 $f(x, z | y) = f(x | y)f(z | y)$. Using Bayes's rule for continuous random variables I obtain

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

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

Lemma A.2. Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be continuously differentiable functions. If there exists a differentiable function $s : \mathbb{R}_+^2 \rightarrow \mathbb{R}$:

$$f(zh(x)) = s(x, z) \quad (\text{A.1})$$

Then

$$\frac{\log'(h(x))}{z} = \frac{s_1(z, x)}{s_2(z, x)} \quad (\text{A.2})$$

where $s_j(z, x)$ denotes the derivative of $s(z, x)$ with respect to its j -th argument and $\log'(h(x))$ denotes derivative of $\log(h(x))$ with respect to x .

Proof. By the assumptions we can differentiate equation (A.1). Differentiating with respect to z to get

$$f'(zh(x))h(x) = s_1(z, x). \quad (\text{A.3})$$

Differentiating with respect to x yields

$$f'(zh(x))zh'(x) = s_2(z, x). \quad (\text{A.4})$$

Taking the ratio of Equation (A.4) and (A.3) I obtain

$$\frac{zh'(x)}{h(x)} = \frac{s_2(z, x)}{s_1(z, x)}, \quad (\text{A.5})$$

which gives

$$\log'(h(x))z = \frac{s_2(z, x)}{s_1(z, x)}. \quad (\text{A.6})$$

So the ratio of derivatives of $s(z, x)$ does not depend on f . \square

Lemma A.3. Let $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ and $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are differentiable functions. If there exists a differentiable function $s : \mathbb{R}_+^3 \rightarrow \mathbb{R}$ with

$$f(w, zh(x)) = s(w, x, z) \quad (\text{A.7})$$

then

- Ratio of derivative of $s(w, z, x)$ does not depend on w and depends only on z and some function of x

$$\frac{s_2(w, z, x)}{s_3(w, z, x)} = \frac{\log'(h(x))}{z}.$$

- Derivative of $s_1(w, z, x)$ with respect to w depends only on w and $zh(x)$

$$f_2(w, zh(x)) = s_1(w, z, x).$$

Proof. Taking derivative of the both sides of Equation in (A.7) with respect to z we obtain

$$f_2(w, zh(x))h(x) = s_2(w, z, x).$$

Taking derivative of the both sides of Equation in (A.7) with respect to x we obtain

$$f_1(w, zh(x))zh'(x) = s_2(w, z, x)$$

Taking the ratio between the two

$$\frac{zh'(x)}{h(x)} = \frac{s_3(w, z, x)}{s_2(w, z, x)}. \quad (\text{A.8})$$

which gives

$$\frac{z}{\log'(h(x))} = \frac{s_3(w, z, x)}{s_2(w, z, x)}. \quad (\text{A.9})$$

Taking derivative with respect to w

$$f'(w, zh(x)) = s_1(w, z, x).$$

□

Lemma A.4. Consider the following model

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

where (y, x, z) are observed random variables and $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ are unknown functions. Let (f_0, h_0, g_0) denote 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 Ω be the set of functions that obey the model restrictions and assumptions, so $(f_0, h_0, g_0) \in \Omega = \Omega_f \times \Omega_h \times \Omega_g$. Define the set of log-linear functions as $\Omega_{\log} = \{f(x) : f(x) = a \log(x) + b, (a, b) \in \mathbb{R}^2\}$ and assume that they are excluded from Ω_f , i.e., $\Omega_{\log} \cap \Omega_f = \emptyset$.

I next provide some definitions based on Matzkin (2007). $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent if and only if

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

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

Based on these definitions and under my assumptions, g is identified up to a constant, h is identified up to a scale and f is identified up to a constant and a normalization specified below in

the proof. Since identification fails only in special cases we say that the functions, (f, h, g) , are generically identified. The special cases where identification fails are testable.

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

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

Since $\mathbb{E}[y | z, x]$ 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 ϵ .

For contradiction assume $(f, h, g) \in \Omega$ and $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ are observationally equivalent. Using the definition of identification given above, this implies:

$$f(zh(x)) + g(x) = \tilde{f}(z\tilde{h}(x)) + \tilde{g}(x). \quad (\text{A.10})$$

I will show that if Equation (A.10) holds, then (f, h, g) and $(\tilde{f}, \tilde{h}, \tilde{g})$ have to obey the normalization restrictions below

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

for $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}$. To show this, I will take the derivatives of Equation (A.10) with respect to x and z . Taking derivative with respect to z yields

$$f'(zh(x))h(x) = \tilde{f}'(z\tilde{h}(x))\tilde{h}(x). \quad (\text{A.11})$$

This gives me the first restriction. Next, taking derivative with respect to x gives

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

Rearranging this to collect similar terms, I obtain

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

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

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

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

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

By Equation (A.11) we have that $f'(zh(x))h(x) = \tilde{f}'(z\tilde{h}(x))\tilde{h}(x)$. Using this

$$\begin{aligned} z \left(f'(zh(x))h(x) \log'(h(x)) - f'(zh(x))h(x) \log'(\tilde{h}(x)) \right) &= \tilde{g}'(x) - g'(x) \\ z f'(zh(x))h(x) \left(\log'(h(x)) - \log'(\tilde{h}(x)) \right) &= \tilde{g}'(x) - g'(x). \end{aligned} \quad (\text{A.12})$$

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

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

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

$$f'(zh(x)) = \frac{a}{zh(x)} \quad \text{and} \quad \frac{\tilde{g}'(x) - g'(x)}{\left(\frac{h'(x)}{h(x)} - \frac{\tilde{h}'(x)}{\tilde{h}(x)}\right)} = \frac{1}{a},$$

for some constant a . This solution gives

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

which was excluded from Ω_f by my assumption. Therefore, we cannot have $h(x) \neq \frac{\tilde{h}(x)}{\lambda}$, which implies

$$\log'(h(x)) = \log'(\tilde{h}(x)), \quad \tilde{g}'(x) = g'(x) \tag{A.13}$$

Next, using equation (A.12) we also have

$$\tilde{g}'(x) = g'(x) \tag{A.14}$$

Integrating these equations, there exists λ and a such that

$$h(x) = \frac{\tilde{h}(x)}{\lambda} \quad g(x) = \tilde{g}(x) - a$$

Now using these results and Equation (A.11) we solve for $f(zh(x))$ and $\tilde{f}(zh(x))$

$$f(zh(x)) = \tilde{f}(z\tilde{h}(x)) + \tilde{g}(x) - g(x) \tag{A.15}$$

$$= \tilde{f}(z\lambda h(x)) + a \tag{A.16}$$

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

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

In the second part of the proof, I 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

$$y(x, z) := \mathbb{E}[y | z, x] = \lambda \log z + h(x) + g(x). \tag{A.17}$$

which is testable by estimating $\mathbb{E}[y | z, x]$ from data. If part is trivial. To show the only if part, by fundamental theorem of calculus, Equation (A.17) implies that $\frac{\partial t(x, z)}{\partial \log z} = \lambda$. Using this

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

From this, I obtain

$$f'(zh(x))h(x) = \frac{\lambda}{z}. \quad (\text{A.18})$$

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

Lemma A.5. *Under Assumption 3.6 u_{it}^1 and u_{it}^2 are independently distributed conditional on W_{it-1} .*

Proof. We have that

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \quad \omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2)$$

By assumption 3.6, 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

B Proofs

Proof of Proposition 2.1

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

Part (i)

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

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

The firm minimizes the cost of flexible inputs for a given level of planned output, \bar{Y}_{it} . This problem can be written as

$$\begin{aligned} & \min_{L_{it}, M_{it}} \quad p_t^l L_{it} + p_t^m M_{it} \\ & \text{s.t.} \quad \mathbb{E}[F(K_{it}, h(K_{it}, \omega_{it}^L L_{it}, M_{it})) \omega_{it}^H \epsilon_{it} \mid \mathcal{I}_{it}] \geq \bar{Y}_{it}. \end{aligned}$$

Because the firm's information set includes both productivity shocks we can write the firm's problem as follows:

$$\begin{aligned} & \min_{L_{it}, M_{it}} \quad p_t^l L_{it} + p_t^m M_{it} \\ & \text{s.t.} \quad F(K_{it}, h(K_{it}, \omega_{it}^L L_{it}, M_{it})) \omega_{it}^H \mathcal{E}_{it}(\mathcal{I}_{it}) \geq \tilde{Y}_{it}, \end{aligned} \quad (\text{B.1})$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}]$. I use $\bar{L}_{it} := \omega_{it}^L L_{it}$ to denote the effective (quality-adjusted) labor and $\bar{p}_{it}^l := p_t^l / \omega_{it}^L$ to denote the quality-adjusted price of labor. With this notation, I can reformulate the firm's problem as another cost minimization, where the firm chooses the effective labor facing

the quality-adjusted input prices. The two problems are equivalent because the firm takes ω_{it}^L as given. Therefore, the cost minimization problem in Equation (B.1) can be rewritten as

$$\begin{aligned} & \min_{M_{it}, \bar{L}_{it}} \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} \\ \text{s.t. } & F(K_{it}, h(K_{it}, \bar{L}_{it}, M_{it})) \omega_{it}^H \geq \bar{Y}_{it}(\mathcal{I}_{it}), \end{aligned} \quad (\text{B.2})$$

where $\tilde{Y}_{it} := \bar{Y}_{it}/\mathcal{E}_{it}(\mathcal{I}_{it})$. So, for what follows, I suppress keep the argument (\mathcal{I}_{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 can be written as:

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

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 (B.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:

$$\begin{aligned} \bar{L}_{it} &= \frac{\partial C(\tilde{Y}_{it}, K_{it}, \omega_{it}^H, \bar{p}_{it})}{\partial \bar{p}_{it}^l} = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) \frac{\partial C_2(K_{it}, \bar{p}_{it}^l, p_t^m)}{\partial \bar{p}_{it}^l}, \\ M_{it} &= \frac{\partial C(\tilde{Y}_{it}, K_{it}, \omega_{it}^H, \bar{p}_{it})}{\partial p_t^m} = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) \frac{\partial C_2(K_{it}, \bar{p}_{it}^l, p_t^m)}{\partial p_t^m}. \end{aligned}$$

The ratio of materials to the effective labor equals:

$$\frac{M_{it}}{\bar{L}_{it}} = \frac{\partial C_2(K_{it}, \bar{p}_{it}^l, p_t^m)/\partial p_t^m}{\partial C_2(K_{it}, \bar{p}_{it}^l, p_t^m)/\partial \bar{p}_{it}^l} = \frac{C_m(K_{it}, \bar{p}_{it}^l, p_t^m)}{C_l(K_{it}, \bar{p}_{it}^l, p_t^m)},$$

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

$$\frac{M_{it}}{L_{it}} = \frac{C_m(K_{it}, \bar{p}_{it}^l, p_t^m)\omega_{it}^L}{C_l(K_{it}, \bar{p}_{it}^l, p_t^m)}.$$

This function depends only on capital, labor-augmenting productivity and input prices. Hence

$$\tilde{M}_{it} \equiv r(K_{it}, \omega_{it}^L, p_t^m, p_t^l) \equiv r_t(K_{it}, \omega_{it}^L), \quad (\text{B.4})$$

for some function $r_t(K_{it}, \omega_{it}^L)$, as input prices do not vary across firms. This completes the first part of the proof.

Part (ii)

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

$$\frac{\partial r_t(K_{it}, \omega_{it}^L)}{\partial \omega_{it}^L} > 0 \quad \text{for all } (K_{it}, \omega_{it}^L) \quad \text{or} \quad \frac{\partial r_t(K_{it}, \omega_{it}^L)}{\partial \omega_{it}^L} < 0 \quad \text{for all } (K_{it}, \omega_{it}^L).$$

In part (i), I showed that

$$r_t(K_{it}, \omega_{it}^L) = \tilde{M}_{it} = \frac{C_m(K_{it}, \bar{p}_{it}^l, p_t^m) \omega_{it}^L}{C_l(K_{it}, \bar{p}_{it}^l, p_t^m)}.$$

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:

$$\tilde{M}_{it} \equiv \frac{\tilde{C}_m(K_{it}, \tilde{p}_{it}) \omega_{it}^L}{\tilde{C}_l(K_{it}, \tilde{p}_{it})}, \quad (\text{B.5})$$

where $\tilde{p}_{it} := \frac{\bar{p}_{it}^l}{p_t^m}$, $\tilde{C}_m := C_m(K_{it}, \tilde{p}_{it}, 1)$ and $\tilde{C}_l(K_{it}, \tilde{p}_{it}) := C_l(K_{it}, \tilde{p}_{it}, 1)$. Taking the logarithm of Equation (B.5), the logarithm of input is given by

$$\log(\tilde{M}_{it}) = \log \left(\frac{\tilde{C}_l(K_{it}, \tilde{p}_{it})}{\tilde{C}_m(K_{it}, \tilde{p}_{it})} \right) + \log(\omega_{it}^L).$$

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

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

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

Proof of Lemma 3.1

By Assumption 2.2 we have that

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

Substituting ω_{it}^L from Equation (3.1)

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

Since $g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$ is strictly monotone in u_{it}^1 , Equation (B.6) implies independence of u_{it}^1 and \mathcal{I}_{it-1} conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$

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

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 (B.7) implies

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

Using Equations (2.7) and (2.9), $(\omega_{it-1}^L, \omega_{it-1}^H)$ can be expressed as functions of W_{it-1} . Thus

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

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

Proof of Lemma 3.2

By Assumption 2.2, we have

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

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

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

It follows from Equation (B.8), the fact that u_{it}^2 is normally 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

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

Using Equations (2.7) and (2.9), $(\omega_{it-1}^L, \omega_{it-1}^H)$ can be expressed as functions of W_{it-1} . This gives

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

which completes the proof.

Proof of Proposition 4.3

The proof consists of two parts. First, I will show that two different set of structural functions, lead to observationally equivalent $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. 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 be identified. Looking at the elasticities first, θ_{it}^L and θ_{it}^M can be written

as a function of production function in the following way

$$\theta_{it}^L = f_2 h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) \bar{r}(K_{it}, \tilde{M}_{it}) L_{it}, \quad (\text{B.9})$$

$$\theta_{it}^M = f_2 h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) M_{it}, \quad (\text{B.10})$$

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

$$\bar{h}_2(K_{it}, \tilde{M}_{it}) = h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) r_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}), \quad (\text{B.11})$$

$$\bar{h}_1(K_{it}, \tilde{M}_{it}) = h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) r_1(K_{it}, \tilde{M}_{it}). \quad (\text{B.12})$$

So the right-hand side of these equations are identified from \bar{h} and the output elasticities θ_{it}^L and θ_{it}^M . 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 $\bar{r}(K_{it}, \tilde{M}_{it})$ has five arguments 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

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

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 lead to observational equivalence since they give the same $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$.

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

This implies that we cannot distinguish $(h_1, h_2, \bar{r}_1, \bar{r}_2)$ from $(h'_1, h'_2, \bar{r}'_1, \bar{r}'_2)$, however h_3 might be identified. Next I show that labor-augmenting productivity, output elasticity of capital and elasticity of substitution depend on $(h_1, h_2, \bar{r}_1, \bar{r}_2)$, so they cannot be recovered from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. Labor-augmenting productivity is given by

$$\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it}).$$

Hence, non-identification of $\bar{r}(K_{it}, \tilde{M}_{it})$ immediately implies that ω_{it}^L is not identified. The output elasticity of capital is given by

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

Since h_1 is not identified, θ_{it}^K is not identified. Finally, to see that the elasticity of substitution is

not identified note that it is defined as

$$\sigma_{it}^{ML} = \frac{\partial \log(L_{it}/M_{it})}{\partial \log(F_M/F_L)},$$

which depends on the ratio of marginal products. We can write the ratio of marginal products as

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

Using this 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 $\bar{r}(K_{it}, \tilde{M}_{it})$ is not identified. Therefore, I conclude that the elasticity of substitution is not identified. Elasticity of substitution with respect to other inputs can similarly be derived and it can be showed than then 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

$$\begin{aligned}\theta_{it}^L &= f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) r(\tilde{M}_{it}) L_{it} \\ \theta_{it}^M &= f_2 h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) M_{it}.\end{aligned}\tag{B.13}$$

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

$$\bar{h}_1(\tilde{M}_{it}) = h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it}) + h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}),\tag{B.14}$$

where $\bar{r}'(\tilde{M}_{it})$ denotes the derivative of $\bar{r}(\tilde{M}_{it})$. Therefore, the right-hand side of Equation (B.14) is identified from $\bar{h}(\tilde{M}_{it})$. Now, taking the ratio of θ_{it}^L/L_{it} and $f_2 \bar{h}_1(\tilde{M}_{it}) - \theta_{it}^M/M_{it}$ gives

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

This shows that the derivative of $\log(r(\tilde{M}_{it}))$ with respect to \tilde{M}_{it} can be identified from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$ as $b(\tilde{M}_{it})$. Therefore, we can recover $\log(r(\tilde{M}_{it}))$ up to a constant by integrating $b(\tilde{M}_{it})$ with respect to \tilde{M}_{it} .

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

Since $\omega_{it}^L = r(\tilde{M}_{it})$, and $\log(r(\tilde{M}_{it}))$ is identified up to a constant, ω_{it}^L is identified up to a scale. Identification of output elasticity of capital is easy to show because it depends on f and \bar{h} only. We

can recover the output elasticity of capital from f and \bar{h} as:

$$\theta_{it}^K = f_1(K_{it}, L_{it}\tilde{h}(\tilde{M}_{it})).$$

This concludes the proof.

Proof of Proposition 4.5

The elasticity of substitution is given by

$$\sigma_{it}^{ML} = \frac{\partial \log(L_{it}/M_{it})}{\partial \log(F_M/F_L)}$$

If production function takes the form in Equation (4.7), we can derive σ_{it}^{ML} as

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

which depends on h_{22} . Since h_{22} is not identified, the elasticity of substitution is not identified.

C Identification

In this section I show that the homothetic and strong homothetic separable production functions in Section 4.5 are identified except special cases using the moment restriction in Equation (5.5). The identification results follow Roehrig (1988)⁶⁵.

Identification for Homothetic Production Function

Under homotheticity assumption, the function function takes the following form

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

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

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

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

$$y_{it} = v k_{it} + f(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})) + c_2(W_{it-1}, \tilde{M}_{it}, \tilde{s}(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} | k_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}] = 0,$$

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

$$y_{it} = v k_{it} + f(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})) + c_2(W_{it-1}, \tilde{M}_{it}, s(k_{it}, \tilde{L}_{it}, \tilde{M}_{it}, W_{it-1})) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} | k_{it}, \tilde{L}_{it}, \tilde{M}_{it}, W_{it-1}] = 0.$$

where $\tilde{s}(x_1, x_2, x_3, x_4) = s(\log(x_1), x_2/(x_3 x_1), x_3, x_4)$. Note that under the modelling assumptions, none of the random variable in $(k_{it}, \tilde{L}_{it}, \tilde{M}_{it}, W_{it-1})$ is stochastically dependent on others. 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

⁶⁵Benkard and Berry (2006) describes an error in the identification proof of Roehrig (1988) when the system involves multiple equations and multi-dimensional errors. Since my setting involves a single equation, Roehrig (1988)'s result still applies.

from the random variables

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

By the moment restriction in Equation (5.5), we have

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

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

$$\alpha w + f(zh(x)) + g(x, t, s(w, z, x, t)) = \tilde{\alpha} w + \tilde{f}(z\tilde{h}(x)) + \tilde{g}(x, t, s(w, z, x, t)). \quad (\text{C.2})$$

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

Proposition C.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) $h'(x) > 0$ almost everywhere. (iv) $f_0 \notin \Omega_{\log}$, where Ω_{\log} is defined in Lemma A.4. (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 A.4, and α_0 is identified.

Proof. I will show that if there exists observationally equivalent (α, f, h, g) and $(\tilde{\alpha}, \tilde{f}, \tilde{h}, \tilde{g})$, then they equal each other up to normalization described in the proposition. The proof adopts the notation that $r_i()$ denotes the derivative of function r with respect to its i -th argument and r' to denote the derivative if function r takes a single argument. Taking the derivative of Equation (C.2) with respect to w we obtain

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

Rearranging this equation:

$$g_3(x, t, s(w, z, x, t))s_1(w, z, x, t) - \tilde{g}_3(x, t, s(w, z, x, t))s_1(w, z, x, t) = \tilde{\alpha} - \alpha. \quad (\text{C.3})$$

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

$$\bar{g}_3(x, t, s(w, z, x, t))s_1(w, z, x, t) = \tilde{\alpha} - \alpha. \quad (\text{C.4})$$

Taking the derivatives of Equation (C.4) with respect to w and z

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

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

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

$$\alpha = \tilde{\alpha}, \quad g_3(x, t, s(w, z, x, t)) - \tilde{g}_3(x, t, s(w, z, x, t)) = 0. \quad (\text{C.5})$$

This shows that α and g_3 are identified. Next, taking the derivative of Equation (C.4) with respect

to t gives

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

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

$$g_2(x, t, s(w, z, x, t)) = \tilde{g}_2(x, t, s(w, z, x, t)). \quad (\text{C.6})$$

Therefore $g_2(x, t, s(w, z, x, t))$ is also identified. Taking the derivative of Equation (C.4) with respect to z to obtain

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

Using $g_3 = \tilde{g}_3$ obtained in Equation in (C.5) gives

$$f'(zh(x))h(x) = \tilde{f}'(z\tilde{h}(x))\tilde{h}(x). \quad (\text{C.7})$$

Finally, taking the derivative of Equation (C.4) with respect to x

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

Rearranging

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

Using Equation (C.7) we can substitute $f'(zh(x))h'(x)$ and, with some algebra, get

$$z\left(\tilde{f}'(z\tilde{h}(x))\tilde{h}(x)(\log'(h(x)) - \log'(\tilde{h}(x)))\right) = \tilde{g}_1(x, t, s(w, z, x, t)) - g_1(x, t, s(w, z, x, t)) \quad (\text{C.8})$$

Taking the derivative with respect to w

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

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

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

Using $g_{13} = \tilde{g}_{13}$, we have $g_{12}(x, t, s(w, z, x, t)) = \tilde{g}_{12}(x, t, s(w, z, x, t))$. By fundamental theorem of calculus

$$\bar{g}_1(x) \equiv g_1(x, t, s(w, z, x, t)) - g'_1(x, t, s(w, z, x, t)) \quad (\text{C.9})$$

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 (C.8) can be written as

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

Now applying a result in Lemma A.4 we obtain

$$f(x) = \tilde{f}(\lambda x) + a, \quad h(x) = \frac{\tilde{h}(x)}{\lambda}, \quad g(x) = \tilde{g}(x) - a, \quad \alpha = \tilde{\alpha} \quad (\text{C.10})$$

This concludes the proof. \square

Identification for Strong Homothetic Production Function

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

$$y_{it} = f(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})) + \omega_{it}^H + \epsilon_{it}. \quad (\text{C.11})$$

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

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

Under the strong homothetic separable model the control variables are $u_{it}^1 = \tilde{M}_{it}$ and $u_{it}^2 = F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} | K_{it}, W_{it-1}, u_{it}^1)$. Substituting these into Equation (C.11) gives:

$$y_{it} = f(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})) + g(\tilde{M}_{it}, W_{it-1}, \tilde{s}(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} | K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}] = 0.$$

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

$$y_{it} = f(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})) + g(\tilde{M}_{it}, W_{it-1}, s(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})) + \epsilon_{it} \quad \mathbb{E}[\epsilon_{it} | K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1}] = 0$$

where $\tilde{s}(x_1, x_2, x_3, x_4) = s(x_1, x_2/x_3, x_3, x_4)$. Note that under the modelling assumptions, none of the random variable in $(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})$ is stochastically dependent on others. 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

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

By the moment restriction in Equation (5.5), we have

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

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

$$f(w, zh(x)) + g(x, t, s(w, z, x, t)) = \tilde{f}(w, zh(x)) + \tilde{g}(x, t, s(w, z, x, t)). \quad (\text{C.12})$$

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

Proposition C.2. 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) $\mathbb{E}[s_1^2(w, z, x, t) | x, t] > 0$. (vi) Define $q := s_2(w, z, x, t) \log'(h_0(x))z - s_3(w, z, x, t)$. I assume that $\mathbb{E}[q^2 | x, s, t] > 0$ for all (x, s, t) . 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 A.4.

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 described in the proposition. Denote $\mathbb{E}[y | w, x, z, t]$ by $y(w, z, x, t)$. Taking the derivative of $y(w, z, x, t)$ with respect to (w, z, x) we have

$$y_1(w, z, x, t) = f_1(w, zh(x)) + g_2(x, s(w, z, x, t), t)s_1(w, z, x, t), \quad (\text{C.13})$$

$$y_2(w, z, x, t) = f_2(w, zh(x))h(x) + g_2(x, s(w, z, x, t), t)s_2(w, z, x, t), \quad (\text{C.14})$$

$$y_3(w, z, x, t) = f_2(w, zh(x))h'(x) + g_2(x, s(w, z, x, t), t)s_3(w, z, x, t) + g_1(x, s(w, z, x, t), t). \quad (\text{C.15})$$

The same equations hold when we replace (f, h, g) by $(\tilde{f}, \tilde{h}, \tilde{g})$. Multiplying $y_2(w, z, x, t)$ by

$\log'(h(x))z$ and subtracting $y_3(w, z, x, t)$ we obtain

$$y_2(w, z, x, t) \log'(h(x))z - y_3(w, z, x, t) = \quad (\text{C.16})$$

$$g_2(x, s(w, z, x, t), t)(s_2(w, z, x, t) \log'(h(x))z - s_3(w, z, x, t)) - g_1(x, s(w, z, x, t), t). \quad (\text{C.17})$$

We obtain a similar equation for $(\tilde{f}, \tilde{h}, \tilde{g})$.

$$y_2(w, z, x, t) \log'(\tilde{h}(x))z - y_3(w, z, x, t) = \quad (\text{C.18})$$

$$\tilde{g}_2(x, s(w, z, x, t), t)(s_2(w, z, x, t) \log'(\tilde{h}(x))z - s_3(w, z, x, t)) - \tilde{g}_1(x, s(w, z, x, t), t). \quad (\text{C.19})$$

In Equation (C.16), the unknown functions are $h(x)$, $g_1(x, s(w, z, x, t), t)$ and $g_2(x, s(w, z, x, t), t)$, and other functions are known or identified. None of the unknown functions depend on w . By assumption (vi), conditional on (x, s, t) there is variation in $(x, s(w, z, x, t), t)(s_2(w, z, x, t) \log'(\tilde{h}(x))z - s_3(w, z, x, t))$. Therefore, g_2 and g_1 can be identified from Equations (C.16) and (C.18) for a given $h(x)$ and $\tilde{h}(x)$. Therefore, g_1 and \tilde{g}_1 can be written as a function of observed or identified random variables and $h(x)$ and similarly \tilde{g}_2 and \tilde{g}_1 can be written as a function of observed or identified random variables and $\tilde{h}(x)$. So we write

$$\begin{aligned} g_2(x, s, t) &= \bar{g}_2(y_2, y_3, z, h(x), w, t, s_2, s_3), \\ \tilde{g}_2(x, s, t) &= \bar{g}_2(y_2, y_3, z, \tilde{h}(x), w, t, s_2, s_3), \\ g_1(x, s, t) &= \bar{g}_1(y_2, y_3, z, h(x), w, t, s_2, s_3) \\ \tilde{g}_1(x, s, t) &= \bar{g}_1(y_2, y_3, z, \tilde{h}(x), w, t, s_2, s_3) \end{aligned}$$

where \bar{g}_1 and \bar{g}_2 are known functions that can be derived from Equations (C.16) and (C.18). This implies that g_2 and \tilde{g}_2 equal each other up to a transformation of their first argument. And similarly for g_1 and \tilde{g}_1 . So we can write

$$g_2(x, s(w, z, x, t), t) = \tilde{g}_2(h^{-1}(\tilde{h}(x)), s(w, z, x, t), t), \quad (\text{C.20})$$

$$g_1(x, s(w, z, x, t), t) = \tilde{g}_1(h^{-1}(\tilde{h}(x)), s(w, z, x, t), t). \quad (\text{C.21})$$

Let $\tilde{r}(x)$ denote $h^{-1}(\tilde{h}(x))$. Now using $y_1(w, z, x, t)$ we can write

$$f_1(w, zh(x)) + \tilde{g}_2(\tilde{r}(x), s(w, z, x, t), t)s_1(w, z, x, t) = \tilde{f}_1(w, z\tilde{h}(x)) + \tilde{g}_2(x, s(w, z, x, t), t)s_1(w, z, x, t)$$

Once we condition on w , this problem falls into the case given in Lemma (A.4) with a slight modification. Therefore, $h(x)$ is identified up to a scale:

$$h(x) = \frac{\tilde{h}(x)}{\lambda}.$$

This implies identification of $g_1(h(x), s(w, z, x, t), t)$ and $g_2(h(x), s(w, z, x, t), t)$ from Equations (C.20) and (C.21). With these identification results, identification of $f_0(w, zh(x))$ follows by Equation (C.13) and (C.14). \square

References

- Acemoglu, Daron.** 2003. "Labor-and capital-augmenting technical change". *Journal of the European Economic Association* 1 (1): 1–37. (Cit. on p. 7).
- Ackerberg, Daniel A.** 2016. "Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context". *Working Paper*. (Cit. on p. 13).
- Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth.** 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451. (Cit. on pp. 4, 5, 17, 18, 29, 30, 38).
- 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. (Cit. on p. 8).
- Antrás, Pol.** 2004. "Is the US Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution". *Contributions in Macroeconomics* 4 (1). (Cit. on p. 8).
- Atalay, Enghin.** 2014. "Materials Prices and Productivity". *Journal of the European Economic Association* 12 (3): 575–611. (Cit. on p. 44).
- Autor, David, Dorn, David, Katz, Lawrence, Patterson, Christina, and Van Reenen, John.** 2019. "The Fall of the Labor Share and the Rise of Superstar Firms". *MIT Working Paper*. (Cit. on pp. 6, 8, 34).
- 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. (Cit. on p. 14).
- Balat, Jorge, Brambilla, Irene, and Sasaki, Yuya.** 2016. "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports". *Working Paper*. (Cit. on p. 5).
- 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. (Cit. on pp. 3, 34).
- Benkard, C Lanier and Berry, Steven.** 2006. "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on Brown (1983) and Roehrig (1988)". *Econometrica* 74 (5): 1429–1440. (Cit. on p. 56).
- 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. (Cit. on p. 8).
- Bernard, Andrew B., Redding, Stephen J., and Schott, Peter K.** 2009. "Products and Productivity". *Scandinavian Journal of Economics* 111 (4): 681–709. (Cit. on pp. 8, 32).
- 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. (Cit. on pp. 3, 34).
- Blundell, Richard and Bond, Stephen.** 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340. (Cit. on p. 5).
- Bond, Stephen and Söderbom, Mans.** 2005. "Adjustment Costs and the Identification of Cobb Douglas Production Functions". *Working Paper, Institute for Fiscal Studies (IFS)*. (Cit. on p. 17).

- Chambers, Robert G.** 1988. *Applied Production Analysis: A Dual Approach*. Cambridge University Press. (Cit. on pp. 50, 52).
- Chen, Xiaohong and Pouzo, Demian.** 2012. “Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals”. *Econometrica* 80 (1): 277–321. (Cit. on p. 26).
- Chirinko, Robert S.** 2008. “: The Long and Short of It”. *Journal of Macroeconomics* 30 (2): 671–686. (Cit. on p. 1).
- Chirinko, Robert S, Fazzari, Steven M, and Meyer, Andrew P.** 2011. “A New Approach to Estimating Production Function Parameters: The Elusive Capital–Labor Substitution Elasticity”. *Journal of Business & Economic Statistics* 29 (4): 587–594. (Cit. on p. 8).
- Collard-Wexler, Allan and De Loecker, Jan.** 2016. “Production Function Estimation with Measurement Error in Inputs”. *Working Paper*. (Cit. on p. 29).
- De Loecker, Jan, Eeckhout, Jan, and Unger, Gabriel.** 2018. “The Rise of Market Power and the Macroeconomic Implications”. *Working Paper*. (Cit. on pp. 2, 3, 6, 26, 35, 36, 38, 43).
- De Loecker, Jan, Goldberg, Pinelopi K., Khandelwal, Amit K., and Pavcnik, Nina.** 2016. “Prices, Markups, and Trade Reform”. *Econometrica* 84 (2): 445–510. (Cit. on pp. 34, 44).
- De Loecker, Jan and Scott, Paul T.** 2016. “Estimating Market Power: Evidence from the US Brewing Industry”. *Working Paper 22957. National Bureau of Economic Research*. (Cit. on p. 28).
- De Loecker, Jan and Warzynski, Frederic.** 2012. “Markups and Firm-Level Export Status”. *American Economic Review* 102 (6): 2437–2471. (Cit. on pp. 5, 33).
- DeSouza, Sergio.** 2006. “Levinsohn and Petrin’s (2003) Methodology Works under Monopolistic Competition”. *Economics Bulletin* 12 (6): 1–11. (Cit. on p. 12).
- 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. (Cit. on p. 23).
- Diez, Federico, Leigh, Daniel, and Tambunlertchai, Suchanan.** 2018. “Global Market Power and its Macroeconomic Implications”. *IMF Working Papers* 18 (137). (Cit. on pp. 6, 36).
- Doraszelski, Ulrich and Jaumandreu, Jordi.** 2018. “Measuring the Bias of Technological Change”. *Journal of Political Economy* 126 (3): 1027–1084. (Cit. on pp. 3, 5, 11).
- . 2013. “R&D and Productivity: Estimating Endogenous Productivity”. *Review of Economic Studies* 80 (4): 1338–1383. (Cit. on pp. 12, 20).
- . 2019. “Using Cost Minimization to Estimate Markups”. *Working Paper*. (Cit. on pp. 4, 6, 13, 20, 35).
- Edmond, Chris, Midrigan, Virgiliu, and Xu, Daniel Yi.** 2018. “How Costly Are Markups?” *Working Paper 24800 National Bureau of Economic Research*. (Cit. on p. 39).
- Ekeland, Ivar, Heckman, James J., and Nesheim, Lars.** 2004. “Identification and Estimation of Hedonic Models”. *Journal of Political Economy* 112 (1): 60–109. (Cit. on pp. 14, 19).
- Eslava, Marcela, Haltiwanger, John, Kugler, Adriana, and Kugler, Maurice.** 2010. “Factor Adjustments after Deregulation: Panel Evidence from Colombian Plants”. *Review of Economics and Statistics* 92 (2): 378–391. (Cit. on p. 27).

- Fox, Jeremy T., Haddad, Valentin, Hoderlein, Stefan, Petrin, Amil K, and Sherman, Robert P.** 2019. "Heterogenous Production Functions, Panel Data, and Productivity Dispersion". *Slide Deck, Rice University*, https://econ2017.sites.olt.ubc.ca/files/2019/03/pdf_seminar_Productivity-Slides_Fox.pdf. (Cit. on p. 5).
- Fox, Jeremy T. and Smeets, Vallerie.** 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review* 52 (4): 961–989. (Cit. on p. 7).
- Gandhi, Amit, Navarro, Salvador, and Rivers, David.** 2018. "On the Identification of Gross Output Production Functions". *Working Paper*. (Cit. on pp. 4, 5, 8, 12, 17, 20, 23, 27).
- 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. (Cit. on p. 44).
- Grieco, Paul LE and McDevitt, Ryan C.** 2016. "Productivity and Quality in Health Care: Evidence from the Dialysis Industry". *The Review of Economic Studies* 84 (3): 1071–1105. (Cit. on p. 20).
- Griliches, Zvi and Mairesse, Jacques.** 1995. "Production Functions: the Search for Identification". *Working Paper 5067 National Bureau of Economic Research*. (Cit. on p. 5).
- 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. (Cit. on p. 14).
- Hall, Robert E.** 1988. "The Relation between Price and Marginal Cost in U.S. Industry". *Journal of Political Economy* 96 (5): 921–947. (Cit. on pp. 5, 33).
- Herrendorf, Berthold, Herrington, Christopher, and Valentinyi, Akos.** 2015. "Sectoral Technology and Structural Transformation". *American Economic Journal: Macroeconomics* 7 (4): 104–33. (Cit. on p. 8).
- 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. (Cit. on p. 32).
- Holmes, Thomas J. and Schmitz, James A.** 2010. "Competition and Productivity: A Review of Evidence". *Annual Review of Economics* 2 (1): 619–642. (Cit. on p. 8).
- Hong, Han and Shum, Matthew.** 2010. "Pairwise-Difference Estimation of a Dynamic Optimization Model". *The Review of Economic Studies* 77 (1): 273–304. (Cit. on p. 14).
- Hu, Yingyao, Huang, Guofang, and Sasaki, Yuya.** 2011. "Estimating Production Functions with Robustness Against Errors in the Proxy Variables". *Working Paper*. (Cit. on p. 29).
- Imbens, Guido and Newey, Whitney.** 2009. "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity". *Econometrica* 77 (5): 1481–1512. (Cit. on pp. 3, 14).
- Jaumandreu, Jordi.** 2018. "Dangerous Shortcuts: Ignoring Marginal Cost Determinants in Markup Estimation". *Working Paper*. (Cit. on p. 13).
- Karabarbounis, L. and Neiman, B.** 2014. "The Global Decline of the Labor Share". *The Quarterly Journal of Economics* 129 (1): 61–103. (Cit. on p. 8).

- Kasahara, Hiroyuki, Schrimpf, Paul, and Suzuki, Michio.** 2015. "Identification and Estimation of Production Function with Unobserved Heterogeneity". *Working Paper*. (Cit. on p. 5).
- Kasy, Maximilian.** 2011. "Identification in triangular systems using Control Functions". *Econometric Theory* 27 (3): 663–671. (Cit. on p. 14).
- Kehrig, Matthias and Vincent, Nicolas.** 2018. "The Micro-Level Anatomy of the Labor Share Decline". *Working Paper 25275 National Bureau of Economic Research*. (Cit. on p. 8).
- 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. 28).
- Kim, Kyoo il, Petrin, Amil, and Song, Suyong.** 2016. "Estimating Production Functions with Control Functions when Capital is Measured with Error". *Journal of Econometrics* 190 (2): 267–279. (Cit. on p. 29).
- Kim, Kyoo and Petrin, Amil.** 2010. "Control Function Corrections for Unobserved Factors in Differentiated Product Models". (Cit. on p. 14).
- Klump, Rainer, McAdam, Peter, and Willman, Alpo.** 2007. "Factor Substitution and Factor-Augmenting Technical Progress in the United States: A Normalized Supply-Side System Approach". *The Review of Economics and Statistics* 89 (1): 183–192. (Cit. on p. 8).
- Kumar, Krishna B, Rajan, Raghuram G, and Zingales, Luigi.** 1999. "What Determines Firm Size?" *Working Paper 7208 National Bureau of Economic Research*. (Cit. on p. 32).
- Levinsohn, James.** 1993. "Testing the Imports-as-market-discipline Hypothesis". *Journal of International Economics* 35 (1): 1–22. (Cit. on p. 27).
- Levinsohn, James and Petrin, Amil.** 2003. "Estimating Production Functions Using Inputs to Control for Unobservables". *The Review of Economic Studies* 70 (2): 317–341. (Cit. on pp. 5, 12, 17).
- Lewbel, Arthur.** 2016. "The Identification Zoo-Meanings of Identification in Econometrics". *Forthcoming on Journal of Economic Literature* 7. (Cit. on p. 25).
- Li, Tong and Sasaki, Yuya.** 2017. "Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function". *arXiv:1711.10031 [econ]*. (Cit. on p. 5).
- Liu, Lili.** 1993. "Entry-exit, Learning, and Productivity Change Evidence from Chile". *Journal of Development Economics* 42 (2): 217–242. (Cit. on p. 27).
- Marschak, Jacob and Andrews, William H.** 1944. "Random Simultaneous Equations and the Theory of Production". *Econometrica* 12 (3): 143–205. (Cit. on p. 5).
- Matzkin, Rosa L.** 2007. "Nonparametric identification". *Handbook of econometrics* 6:5307–5368. (Cit. on p. 47).
- . 2004. "Unobservable Instruments". *Working Paper*. (Cit. on p. 14).
- Melitz, Marc J.** 2000. "Estimating Firm-level Productivity in Differentiated Product Industries". (Cit. on p. 12).
- Mundlak, Yair.** 1961. "Empirical Production Function Free of Management Bias". *Journal of Farm Economics* 43 (1): 44–56. (Cit. on p. 5).
- Mundlak, Yair and Hoch, Irving.** 1965. "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions". *Econometrica* 33 (4): 814–828. (Cit. on p. 5).

- Nadiri, M. Ishaq.** 1982. "Producers theory". In *Handbook of Mathematical Economics*, 2:431–490. Elsevier. (Cit. on pp. 11, 22).
- Oberfield, Ezra and Raval, Devesh.** 2014. "Micro data and macro technology". *Working Paper 20452, National Bureau of Economic Research*. (Cit. on pp. 1, 8).
- Olley, G. Steven and Pakes, Ariel.** 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64 (6): 1263. (Cit. on pp. 1, 3, 5, 8, 12, 17, 36, 45).
- Pavcnik, Nina.** 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants". *The Review of Economic Studies* 69 (1): 245–276. (Cit. on p. 27).
- Raval, Devesh.** 2019a. "Testing the Production Approach to Markup Estimation". *Working Paper*. (Cit. on pp. 1, 4, 5, 35).
- Raval, Devesh R.** 2019b. "The Micro Elasticity of Substitution and Non-neutral Technology". *The RAND Journal of Economics* 50 (1): 147–167. (Cit. on pp. 3, 5, 8, 11, 35).
- Roehrig, Charles S.** 1988. "Conditions for Identification in Nonparametric and Parametric Models". *Econometrica* 56 (2): 433. (Cit. on p. 56).
- Shephard, Ronald W.** 1953. *Cost and Production Functions*. Princeton University Press. (Cit. on pp. 9, 50).
- 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. 27).
- Traina, James.** 2018. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements". *Working Paper*. (Cit. on pp. 3, 34).
- Van Bieseboeck, Johannes.** 2003. "Productivity Dynamics with Technology Choice: An Application to Automobile Assembly". *Review of Economic Studies* 70 (1): 167–198. (Cit. on p. 34).
- Van Reenen, John.** 2018. "Increasing Differences between Firms: Market Power and the Macroeconomy". *Working Paper*. (Cit. on p. 29).
- Wooldridge, Jeffrey M.** 2009. "On Estimating firm-level Production Functions Using Proxy Variables to Control for Unobservables". *Economics Letters* 104 (3): 112–114. (Cit. on p. 5).
- Zhang, Hongsong.** 2019. "Non-neutral Technology, Firm Heterogeneity, and Labor Demand". *Journal of Development Economics* 140:145–168. (Cit. on pp. 3, 5, 11).

Supplementary Appendix to “Production Function Estimation
with Factor-Augmenting Technology: An Application to
Markups”

Mert Demirer
MIT

September 22, 2020

1 Data and Estimation

1.1 Chile

The data for Chile 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. Appendix Table 1.0 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, there are five industries remaining in the sample, which cover around 30 percent of the manufacturing sector of Chile in terms of sales.

1.2 Colombia

The data for Colombia are from the annual Colombian Manufacturing census provided by the Departamento Administrativo Nacional de Estadística, covering the years 1981 through 1991. This dataset contains all manufacturing plants with 10 or more 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 variable input for each industry to remove outliers. Appendix Table 1.0 provides summary statistics. The number of industries after sample restrictions is nine, relatively higher than the number of industries in other datasets. The sample covers around 55 percent of the entire manufacturing sector in Colombia in terms of sales. We see that for most industries, the number of plants is stable, with little change over the sample period.

1.3 India

The Indian data was collected by the Ministry of Statistics and Programme Implementation, Government of India, through the Annual Survey of Industries (ASI), which covers all factories that have ten or more workers and use electricity, or that do not use electricity but have at least twenty 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 selected each year for the survey.

India uses 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. The ASI data include firm and product-level price information for intermediate inputs and produced goods, but my empirical framework does not use them, as it requires extensive data cleaning and price indexes.

For sample restriction and data cleaning I first follow Alcott et al. (2016).¹ Then, I restrict my sample to the Census sample to be able to follow the firms over time. Therefore, compared to other developing countries, the average firm size is large in the Indian data. My final sample includes 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 variable input for each industry to remove outliers.

Appendix Table 1.0 provides summary statistics. Among all datasets, the Indian sample is the least representative of the country manufacturing sector as five industries in the sample make up only 20 percent of the Indian manufacturing sector in terms of sales. We also see a very large increase in the number of plants over the sample period for all industries. This reflects the extensive growth in Indian manufacturing over the sample period.

1.4 Compustat

Compustat data 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 in my sample, which are at the plant level, Compustat is at the firm-level as it only comprises of public firms. Also, the industry classification is based on NAICS and industries are defined at the 2-digit level. Appendix Table 1.0 provides some summary statistics. Since there are only three 2-digit level NAICS industries, my sample covers the entire population of public manufacturing firms, subject to data cleaning. Differently from other countries there is a large increase in sample size from 1961 to 2012. This reflects the fact that the number of public firms has risen enormously in the US over the sample periods. Differently from other datasets, I drop observations that are at the bottom and top 1 percent, instead of 2 percent, of the distribution for Compustat to preserve the sample size.

1.5 Turkey

The data for Turkey are provided by the Turkish Statistical Institute (TurkStat; formerly known as the State Institute of Statistics, SIS), 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. In non-census years, the new private plants with at least 10 employees are obtained from the chambers of industry.

I use a sample covering a period from 1983 to 2000. Data from a more recent period are available, but due to major changes in the survey methodology, it is not possible to link ASMI to the data from a more

¹The code for data cleaning is available at <https://www.aeaweb.org/articles?id=10.1257/aer.20140389>.

recent period. 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 that have more than 250 firms per year on average and private establishments. 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 variable input for each industry to remove outliers. In the final sample, I have 15437 firms and 104271 year-firm observations. Appendix Table 1.0 provides summary statistics. In 2000, the industries in the sample make up 71 percent of all manufacturing sector of Turkey. An industry's share and the number of firms are proportional to each other except for the vehicle industry, which constitutes the 12 percent of sales but only 5 percent of all firms in manufacturing.

1.6 Variable Construction

1.6.1 Labor

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

1.6.2 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 input is calculated as deflated cost of goods sold plus administrative and selling expenses less depreciation and wage expenditures. For the materials' revenue share I use the sum of materials cost divided by total sales during the year.

1.6.3 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 (2019).

1.6.4 Output

For all countries, the output is calculated as deflated sales. For Chile, Colombia, India and Turkey, total sales are given by total production value, plus the difference between the end year and beginning year value of inventories of finished goods. For Compustat, it is net sales from Compustat’s Industrial data file.

1.7 Estimation Algorithm

This section presents the estimation algorithm. Apply data cleaning and variable construction described in Subsection 1.1 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_{it}^2 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

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

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

$$\log(\hat{h}(\tilde{M}_{it})) = a_1 + a_2 \tilde{m}_{it} + a_3 \tilde{m}_{it}^2, \quad (1.1)$$

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

$$\hat{f}(K_{it}, L_{it} \hat{h}(\tilde{M}_{it})) = b_1 + b_2 k_{it} + b_3 k_{it}^2 + b_3 k_{it} v_{it} + b_4 v_{it} + b_5 v_{it}^2, \quad (1.2)$$

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 $\{a_j\}_{j=1}^3$, $\{b_j\}_{j=1}^5$, $\hat{c}_2(\cdot)$ and $\hat{c}_3(\cdot)$ construct the objective function in Equation (5.7). Minimize this objective function to estimate the production function as follows. For a given $\{a_j\}_{j=1}^3$, estimate $\{b_j\}_{j=1}^5$, $\hat{c}_2(\cdot)$ and $\hat{c}_3(\cdot)$ by minimizing the objective function using least squares regression. The algorithm involves two layers. For a candidate value of the parameter vector $\{a_j\}_{j=1}^3$, in the inner loop, estimate $\{b_j\}_{j=1}^5$, $\hat{c}_2(\cdot)$ and $\hat{c}_3(\cdot)$ using least squares regression. In the outer loop use an optimization routine to estimate $\{a_j\}_{j=1}^3$. Minimizing the objective function requires an optimization routine 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.³ Obtain the estimates of f and \bar{h} from the estimates of the parameters $\{a_j\}_1^3$ and $\{b_j\}_1^5$ in Equations (1.1) and (1.2). 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 θ_{it}^V and revenue shares of materials and labor use Equations (4.6) to estimate output elasticity of labor and materials. Finally estimate markups from $\hat{\theta}_{it}^V$ and the revenue share of flexible input as in Equation (??).

For standard errors, resample firms with replacement from sample A then repeat the estimation procedure above. For estimation of the Nested CES model in Section 5.4, I use the same procedure except that I impose the parametric restrictions given by the Nested CES model.

²One can estimate $s(m, k, w, u)$ for every M_{it} observed in the data with additional computational cost.

³I consider a larger sample for markup estimation than production function estimation because given a production function estimate calculating elasticities and markups does not require observing previous period's inputs.

Table 1.0: Descriptive Statistics - Chile

ISIC	Industry	Share (Sales)			Number of Plants		
		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 1.0: Descriptive Statistics - Colombia

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

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

Table 1.0: Descriptive Statistics - India

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

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

Table 1.0: Descriptive Statistics - US

NAICS	Industry	Share (Sales)			Number of Firms		
		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 1.0: Descriptive Statistics - Turkey

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

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

2 Supplementary Lemmas

Proof of Proposition 3.2

This proof closely follows the same lines as 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 3.5, can be written as:

$$\begin{aligned} \min_{L_{it}, M_{it}} \quad & p_t^l L_{it} + p_{it}^m M_{it} \\ \text{s.t.} \quad & \mathbb{E}[F(K_{it}, h(K_{it}, L_{it}, M_{it})) \omega_{it}^H \epsilon_{it} | \mathcal{I}_{it}] \geq \bar{Y}_{it} \end{aligned}$$

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

$$\begin{aligned} \min_{L_{it}, M_{it}} \quad & p_t^l L_{it} + p_{it}^m M_{it} \\ \text{s.t.} \quad & F(K_{it}, h(K_{it}, L_{it}, M_{it})) \omega_{it}^H \mathcal{E}_{it} \geq \bar{Y}_{it}, \end{aligned} \tag{2.1}$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}]$. The cost minimization problem in Equation (2.1) is given by

$$\begin{aligned} \min_{M_{it}, L_{it}} \quad & p_t^l L_{it} + p_{it}^m M_{it} \\ \text{s.t.} \quad & F(K_{it}, h(K_{it}, L_{it}, M_{it})\omega_{it}^H) \geq \tilde{Y}_{it}, \end{aligned}$$

where $\tilde{Y}_{it} := Y_{it}/\mathcal{E}_{it}(\mathcal{I}_{it})$. So, for what follows I suppress the argument \mathcal{I}_{it} of \tilde{Y}_{it} . Letting $\bar{p}_{it} = (p_t^l, p_{it}^m)$ denote the price vector and following the steps I used to obtain Equation (B.3), the cost function can be expressed as:

$$C(\bar{p}_{it}, \tilde{Y}_{it}, K_{it}, \omega_{it}^H) = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H)C_2(K_{it}, p_t^l, p_{it}^m). \quad (2.2)$$

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

$$\begin{aligned} M_{it} &= \frac{\partial C(\bar{p}_{it}, \tilde{Y}_{it}, K_{it}, \omega_{it})}{\partial p_{it}^m} = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}) \frac{\partial C_2(K_{it}, p_t^l, p_{it}^m)}{\partial p_{it}^m}, \\ L_{it} &= \frac{\partial C(\bar{p}_{it}, \tilde{Y}_{it}, K_{it}, \omega_{it})}{\partial p_t^l} = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}) \frac{\partial C_2(K_{it}, p_t^l, p_{it}^m)}{\partial p_t^l}. \end{aligned}$$

Using optimal labor and materials demand the ratio of labor to materials can be obtained as

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

Since M_{it} is not observed we cannot use L_{it}/M_{it} to control for ω_{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

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

Now, we see that this equation has the same structure as Equation (B.4) in the proof of Proposition 2.1, with $\omega_{it}^L = \omega_{it}^M$, $\bar{p}_{it}^l = 1/\omega_{it}^M$, $p_{it}^m = p_t^l$ 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 for the purpose of estimation. This solves the problem that materials quantity, M_{it} , is unobserved as 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.

Proof of Lemma 3.1

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

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

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 (2.7) and (2.9) we substitute $(\omega_{it}^L, \omega_{it}^H)$ as functions of (W_{it-1}) to obtain

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

which shows the desired result.

Proof of Lemma 3.2

This proof closely follows the proof of Lemma 3.2. By Assumption 3.1 we have

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

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

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

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 (2.7) and (2.9) to substitute $(\omega_{it-1}^L, \omega_{it-1}^H)$ as functions of W_{it-1} , I obtain

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

which shows the desired result.

Proof of Lemma 3.4

By Assumption 3.6 we have that

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

Using the Skorokhod representation of ω_{it}^H in Equation (3.8) we write

$$g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2) \perp \mathcal{I}_{it-1}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \mid \omega_{it-1}^L, \omega_{it-1}^H. \quad (2.3)$$

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

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

It follows from Equation (2.3) and the fact that u_{it}^2 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$ that

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

Since $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$

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

which implies

$$u_{it}^2 | K_{it}, W_{it-1}, u_{it}^1 \sim \text{Uniform}(0, 1). \quad (2.4)$$

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

$$\begin{aligned} M_{it} &= s(K_{it}, \omega_{it}^H, \omega_{it}^L), \\ &= s(K_{it}, g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2), c_1(W_{it-1}, u_{it}^1)), \\ &= s(K_{it}, g_2(\tilde{r}(W_{it-1}), \tilde{s}(W_{it-1}), u_{it}^2), c_1(W_{it-1}, u_{it}^1)), \\ &\equiv \bar{s}(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2). \end{aligned} \quad (2.5)$$

The intuition is similar to that of Lemma 3.1. Employing strict monotonicity of \bar{s} in u_{it}^2 and Equation (2.4), we can use Equation (2.5) to identify u_{it}^2 . In particular, u_{it}^2 equals

$$u_{it}^2 = F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} | K_{it}, W_{it-1}, u_{it}^1), \quad (2.6)$$

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

$$\omega_{it}^H \equiv c_2(W_{it-1}, u_{it}^2).$$

This concludes the proof.

3 Extensions

3.1 Heterogeneous Input Prices

This extension assumes that input prices are heterogeneous, but firms are price-takers in input markets. I denote labor and materials prices by p_{it}^l and p_{it}^m , respectively, and use \bar{p}_{it} to denote the input price vector, so $\bar{p}_{it} := (p_{it}^l, p_{it}^m)$. I also use \tilde{p}_{it} to denote the input price ratio. Moreover, differently from the main model, W_{it} includes also input prices, $W_{it} = (K_{it}, L_{it}, M_{it}, \bar{p}_{it})$. I first modify homothetic separability and monotonicity assumptions to incorporate the input prices into the model. With variation in input prices, Assumption 2.1 is replaced by the following assumption.

Assumption 3.1. *The distribution of productivity shocks and input prices obey:*

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

This assumption states that prices and productivity shocks jointly follow an exogenous first-order Markov process. Importantly, this assumption allows for correlation between productivity shocks and input prices. Since we expect that more productive workers, as represented by higher ω_{it}^L , earn higher wages, correlation between input prices and productivity is important to accommodate. As I discuss later, with some additional structure on the joint distribution, I can obtain stronger identification results. An example is independence between the innovations to productivity shocks and input prices. However, I make minimal assumptions in

this section to develop a general framework.

Assumption 3.2. *Firm's materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^H, \omega_{it}^L, \bar{p}_{it}), \quad (3.1)$$

and $s_t(K_{it}, \omega_{it}^H, \omega_{it}^L, \bar{p}_{it})$ is strictly increasing in ω_{it}^H .

This assumption is a natural extension of Assumption 2.3, as the demand for materials 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 3.1.

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

$$\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^L, \tilde{p}_{it}). \quad (3.2)$$

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

The proof of this proposition is a straightforward extension of the proof of Proposition 2.1, and therefore, is omitted. Compared to Proposition 2.1, the only difference is that the flexible input ratio depends also on the input price ratio. It is worth emphasizing 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 in estimation. With this proposition, ω_{it}^L is invertible once we condition on the input price ratio and capital. The invertibility of Hicks-neutral productivity is given by Assumption 3.2. To summarize, by inverting Equations (3.2) and (3.1), and omitting the time subscripts in functions, I can write productivity shocks as:

$$\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it}, \tilde{p}_{it}), \quad \omega_{it}^H = \bar{s}(K_{it}, M_{it}, \tilde{M}_{it}, \bar{p}_{it}). \quad (3.3)$$

In the presence of heterogeneous input prices, the derivation of the control variables proceed similarly as in Section 3 with minor differences. I first use the Skorokhod's representation of ω_{it}^L to write:

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it}, u_{it}^1), \quad u_{it}^1 | \omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it} \sim \text{Uniform}(0, 1). \quad (3.4)$$

Unlike Equation (3.1), I include the ratio of current and past input prices in the representation of ω_{it}^L , given by $g_1(\cdot)$ function. This is needed because, as stated in Proposition 3.1, the optimal flexible input ratio depends on the ratio of input prices. Using Equations (3.2), (3.3) and (3.4), I obtain

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

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

Lemma 3.1. *Under Assumptions 3.1 - 3.2, u_{it}^1 is jointly independent of $(K_{it}, W_{it-1}, \tilde{p}_{it})$:*

$$u_{it}^1 \perp\!\!\!\perp K_{it}, W_{it-1}, \bar{p}_{it}.$$

Proof. See Appendix B.

Using independence and monotonicity, u_{it}^1 can be identified as:

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

Therefore, we can use Equations (3.3) and (3.4) to write ω_{it}^L as:

$$\omega_{it}^L \equiv c_1(W_{it-1}, \tilde{p}_{it}, u_{it}^1).$$

Note that differently from the main model, the CDF in u_{it}^1 calculation is conditional on the price vector \bar{p}_{it} and control function includes a price ratio \tilde{p}_{it} . Prices are included in the conditioning set since they are endogenous. If we assume that input prices are exogenous, $c_1(\cdot)$ does not take \tilde{p}_{it} as an argument. The procedure for deriving the control function for ω_{it}^H is similar to that of ω_{it}^L . I use

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2), \quad u_{it}^2 | \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, \sim \text{Uniform}(0, 1). \quad (3.5)$$

Following the same steps in Equation (3.2) of Section 3, materials demand function can be written as:

$$M_{it} \equiv \tilde{s}(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2),$$

where $\tilde{s}(\cdot)$ is strictly monotone in u_{it}^2 .

Lemma 3.2. *Under Assumptions 3.1 and 3.2, u_{it}^2 is jointly independent of $(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1)$:*

$$u_{it}^2 \perp\!\!\!\perp K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1.$$

Proof. See Appendix B.

By independence and monotonicity we can recover u_{it}^2 as

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

and the control function is given by

$$\omega_{it}^H \equiv c_2(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2).$$

I conclude that in the presence of input prices control functions becomes

$$\omega_{it}^L = c_1(W_{it-1}, \tilde{p}_{it}, u_{it}^1), \quad \omega_{it}^H = c_2(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2).$$

In contrast to the main model, 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 control variables.

3.2 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 and prices of materials. This would be the case, for example, if there is heterogeneity in materials prices. This model is worth discussing because in standard production datasets, average wages are typically observed, however materials prices are not. Since, usually, only the cost of materials is available in the data, quantity of materials cannot be recovered from its cost when materials prices are heterogeneous. This is especially the case if there are differences in quality of materials used by

firms. To discuss this scenario, I need to restrict the productivity shock to be Hicks-neutral. Therefore, I consider the following production function:

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

I assume that the researcher observes (K_{it}, L_{it}) but she does not observe M_{it} . Instead, materials expenditure, denoted by $R_{it}^m = M_{it}p_{it}^m$, is observed. Due to heterogeneity in materials prices across firms, as indicated by p_{it}^m , we cannot recover M_{it} from R_{it}^m . Therefore, we cannot estimate the production function. However we can replace materials with its expenditure in the following way:

$$Y_{it} = F_t(K_{it}, L_{it}, R_{it}^m \omega_{it}^M) \exp(\omega_{it}^H) \exp(\epsilon_{it}), \quad (3.6)$$

where I define $\omega_{it}^M := 1/(p_{it}^m)$. In Equation (3.6), one can interpret materials cost as an input in the production function and the inverse 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 3.3. *Productivity shock and materials price jointly follow an exogenous joint first-order Markov process*

$$P(\omega_{it}^H, p_{it}^m | \mathcal{I}_{it-1}) = P(\omega_{it}^H, p_{it}^m | \omega_{it-1}^H, p_{it-1}^m).$$

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 incorporates the unobserved materials prices into the firm's materials demand function.

Assumption 3.4. *Firm's materials decision is given by*

$$M_{it} = s_t(K_{it}, p_{it}^m, \omega_{it}^H), \quad (3.7)$$

where $s_t(K_{it}, p_{it}^m, \omega_{it}^H)$ is strictly increasing in ω_{it}^H .

We can write the materials expenditure, R_{it}^m , using Equation (3.7) and materials prices as follows:

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

Since $s_t(K_{it}, p_{it}^m, \omega_{it}^H)$ is strictly monotone in ω_{it}^H conditional on (K_{it}, p_{it}^m) , R_{it}^m is also strictly monotone in ω_{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 3.5. *Suppose that*

(i) Production function is of the following form

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

(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 ω_{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 3.2.

(i) Under Assumptions 3.5(i-iii), the ratio of labor and materials cost, denoted by $\tilde{L}_{it} = L_{it}/R_{it}^M$, depends only on K_{it} and ω_{it}^M :

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

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

(ii) Under Assumptions 3.5(iv) $r_t(K_{it}, \omega_{it}^M)$ is strictly monotone in ω_{it}^M .

Proof. See Appendix B.

With this result, I have two monotonicity conditions that are analogous to those in the main model. The difference is that I replace M_{it} with R_{it}^m and ω_{it}^L with $1/p_{it}^m$. Also, this model involves materials-augmenting productivity instead of labor-augmenting productivity. Therefore, following the same steps in the main model I can write ω_{it}^H and ω_{it}^M as:

$$\omega_{it}^H \equiv s_t(K_{it}, R_{it}^m, \tilde{L}_{it}), \quad \omega_{it}^M \equiv r_t(K_{it}, \tilde{L}_{it}).$$

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 derivation and proofs are omitted.

3.3 Selection

In this section, I present a method of incorporating non-random firm exit, which generates selection problem, into my estimation framework under some simplifying assumptions. In particular, I assume that firms decide whether to exit based only on Hicks-neutral productivity, not labor-augmenting productivity. The second simplifying assumption is that innovations to productivity shocks are independent from each other. Under these simplifying assumptions, I show how to adjust my control variables to account for selection. Accounting for selection relies 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 of the model in Section 2.2, and impose additional restrictions.

Assumption 3.6. *Productivity shocks are independent conditional on last period's productivity:*

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

This assumption implies that innovation to ω_{it}^H and innovation to ω_{it}^L are independently distributed.

Assumption 3.7. *The firm's exit decision depends only on ω_{it}^H and K_{it} . In particular, the firm exits if and only if*

$$\omega_{it}^H \leq \bar{\omega}(K_{it}),$$

where $\bar{\omega}$ is a function that gives the exit threshold in ω_{it}^H . It specifies the firm's exit decision conditional on K_{it} .

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

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2), \quad u_{it}^2 | \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0, 1). \quad (3.8)$$

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 (3.6), which implies that innovations to productivity shocks are independent. 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 3.7, $I_{it} = 1$ if and only if $\omega_{it}^H \leq \bar{\omega}(K_{it})$. So, firm i 's exit decision at time t depends on its capital level and current Hicks-neutral productivity. Using the representation of ω_{it}^H in Equation (3.8) I can write the exit rule as:

$$\begin{aligned} g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2) &\leq \bar{\omega}(K_{it}), \\ u_{it}^2 &\leq g_2^{-1}(\omega_{it-1}^L, \omega_{it-1}^H, \bar{\omega}(K_{it})), \\ u_{it}^2 &\leq g_2^{-1}(\bar{r}(W_{it-1}), \bar{s}(W_{it-1}), \bar{\omega}(K_{it})), \\ u_{it}^2 &\leq \tilde{\omega}(W_{it-1}, K_{it}), \end{aligned} \quad (3.9)$$

where in the second line I use the fact g_2 is invertible in u_{it}^2 , and the third line follows from Equation (3.9) and my assumptions. In the last line, I define a new function $\tilde{\omega}$ to write the exit rule based on a cutoff value in u_{it}^2 . This reformulation of exit suggests that conditional on (W_{it-1}, K_{it}) the firm's exit decision depends only on the realization of u_{it}^2 . Using Lemma 3.2, I have

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

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 (3.9), (W_{it-1}, K_{it}) . Therefore, 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:

$$\tilde{\omega}(W_{it-1}, K_{it}) = \text{Prob}(I_{it} = 1 | W_{it-1}, K_{it}) \equiv p(W_{it-1}, K_{it}). \quad (3.10)$$

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:

$$u_{it}^2 | W_{it-1}, K_{it}, (I_{it} = 1) \sim \text{Uniform}(p(W_{it-1}, K_{it}), 1). \quad (3.11)$$

As shown in Equation (3.7), to control for ω_{it}^H I need the distribution of u_{it}^2 conditional on $(W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1))$. This creates a problem because even though (W_{it-1}, K_{it}) is observed for the firms that exit, u_{it}^1 cannot be estimated from data for the firms that exit; we do not observe \tilde{M}_{it} for $(I_{it} = 1)$. To overcome this problem, I next show that probability of exit remains the same when I condition on $(W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1))$. This result uses Assumption 3.6 and is given by the following lemma.

Lemma 3.3.

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

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

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

where the third line follows because u_{it}^1 and u_{it}^2 are independently distributed conditional on W_{it-1} by Lemma A.5. \square

From this result, I obtain

$$u_{it}^2 \mid W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1) \sim \text{Uniform}(p(W_{it-1}, K_{it}, u_{it}^1), 1), \quad (3.12)$$

$$\sim \text{Uniform}(p(W_{it-1}, K_{it}), 1). \quad (3.13)$$

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, now I derive my control function. The next lemma gives the control variable and control function under these new assumptions.

Lemma 3.4. *We have that*

$$\omega_{it}^H \equiv c_2(W_{it-1}, u_{it}^2), \quad u_{it}^2 = F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1). \quad (3.14)$$

Proof. See Appendix B.

Since I do not observe $F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$ but only observe the distribution conditional on selection $F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0)$, u_{it}^2 cannot be recovered using this lemma. However, I can use Lemme (3.3), which gives the distribution of u_{it}^2 for the firms that stay, to write u_{it}^2 as:

$$\begin{aligned} u_{it}^2 &= p(W_{it-1}, K_{it}) \left(1 - F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0) \right) + \\ &\quad F_{M_{it}|K_{it}, W_{it-1}, u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1, I_{it} = 0) \end{aligned}$$

This result uses the distribution of u_{it}^2 conditional on firms that stay, which is given in Equation (3.12). 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.

4 Additional Results

4.1 Application of Control Variables to Cobb-Douglas Production Function

The control function approach developed in this paper can be applied to Hicks-neutral production function. This section presents this application and discusses the advantages of the control variable method of this paper relative to the proxy variable approach. Since the literature has shown that the gross Cobb-Douglas

production is not identified when there are two flexible inputs, I demonstrate this application using the value-added production function studied in Ackerberg et al. (2015). However, one can use the same control variables for gross production functions, subject to the issues highlighted in the literature.

The (log) production function is given by

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

where ω_{it}^H is unobserved scalar productivity and ϵ_t is ex-post shock to productivity, which is mean independent of capital and labor. I consider the standard assumptions in the proxy variable literature: (i) Productivity shocks follow an exogenous first-order Markov process $P(\omega_{it}^H | \mathcal{I}_{it-1}) = P(\omega_{it}^H | \omega_{it-1}^H)$ (ii) Capital is a dynamic input chosen one period in advance, 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 ω_{it}^H .

We now construct a control variable following a similar procedure in Section 3. Productivity can be represented as:

$$\omega_{it}^H = g(\omega_{it-1}^H, u_{it}) \quad u_{it} | \omega_{it-1}^H \sim \text{Uniform}(0, 1), \quad (4.1)$$

where $g(\omega_{it-1}^H, u_{it})$ is strictly increasing in u_{it} by construction. By Markov Assumption $\omega_{it}^H \perp\!\!\!\perp \mathcal{I}_{it-1} | \omega_{it-1}^H$. Substituting ω_{it}^H using Equation (4.1) we obtain

$$g(\omega_{it-1}^H, u_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1} | \omega_{it-1}^H$$

This implies that $u_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} | \omega_{it-1}^H$.

$$\begin{aligned} m_{it} &= s(k_{it}, \omega_{it}) = s(k_{it}, g(\omega_{it-1}^H, u_{it})) = s(k_{it}, g(s^{-1}(k_{it-1}, m_{it-1}), u_{it})) \\ &\equiv \tilde{s}(k_{it}, k_{it-1}, m_{it-1}, u_{it}) \end{aligned}$$

Note that $s(k_{it}, \cdot)$ is a strictly increasing function and $g(\omega_{it-1}^H, \cdot)$ is also strictly increasing function by construction. So $\equiv \tilde{s}$ is strictly increasing in u_{it} . It follows from Lemma 3.1 that

$$u_{it} | k_{it}, m_{it-1}, k_{it-1} \sim \text{Uniform}(0, 1). \quad (4.2)$$

Therefore, we can recover u_{it} as the conditional CDF of m_{it} : $u_{it} = F_{m_{it}}(m_{it} | k_{it}, m_{it-1}, k_{it-1})$. This suggests that, we can employ $(m_{it-1}, k_{it-1}, u_{it})$ as control variables to proxy ω_{it}^H .

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

With this result, we obtain a partially linear model

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c(m_{it-1}, k_{it-1}, u_{it}) + \epsilon_{it} \quad (4.3)$$

with $E[\epsilon_{it} | I_{it}] = 0$, which gives moments for estimation. However, we can develop other moment restrictions using the first-order Markov property of ω_{it}^H as standard in the literature (Ackerberg et al. (2015)). In particular, we write $\omega_{it}^H = c_2(\omega_{it-1}^H) + \xi_{it}$ with $E[\xi_{it} | I_{it-1}] = 0$. We can obtain second model

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_2(m_{it-1}, k_{it-1}) + \xi_{it} + \epsilon_{it} \quad (4.4)$$

with $E[\xi_{it} | I_{it}] = 0$. Now we can estimate the parameters (β_k, β_l) and unknown functions $c_1(\cdot), c_2(\cdot)$ in

Equation (4.3) and (4.4) using the following moment restrictions.

$$\begin{aligned}\mathbb{E}[\epsilon_{it} | k_{it}, l_{it}, m_{it}, m_{it-1}, k_{it-1}, u_{it}] &= 0 \\ \mathbb{E}[\xi_{it} + \epsilon_{it} | k_{it}, m_{it-1}, k_{it-1}] &= 0\end{aligned}$$

I highlight that my control variable approach does not suffer from the functional dependence problem studied in Ackerberg et al. (2015) even if labor is flexible 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 set. Therefore, my procedure leads to a dimension reduction in the conditioning set. Conditional on the control variable u_{it} there is still variation in k_{it} and l_{it} , which can identify the production function parameters. However, this approach is not robust to other issues studied in Ackerberg et al. (2015).

4.2 Nested CES Production Function

In this section, I study the identification of Nested CES production function:

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

where materials and effective labor are nested. Analyzing this model is also useful to see parametric analog more results in a simpler model. We maintain the assumptions in Section 2.2. Taking the logarithm of this production function we write

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

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

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

where \tilde{K}_{it} and \tilde{L}_{it} denote the ratio of capital to material and ratio of labor to material, respectively and m_{it} denotes the logarithm of materials. Taking the first-order conditions of cost minimization, one can show that

$$\omega_{it}^L = \gamma \tilde{L}^{(1-\sigma_1)/\sigma_1}, \quad \gamma := \left(\frac{(1 - \beta_l) p_t^l}{\beta_l p_t^m} \right)^{1/\sigma_1}$$

where γ 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 γ_1 and γ_2 are constants that depend on the model parameters. Note that ω_{it}^L disappeared from the model. This is the parametric analog of my nonparametric inversion result in Proposition 2.1. The 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). We can estimate the model using the objective function in Equation (4.7). One can show that the sum of the flexible input elasticities 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 β_k are not separately identified in the production function, but the ratio is identified. Since θ_{it}^V depends only on the ratio, it is identified after estimating the parameters. Labor and materials elasticities are identified using θ_{it}^V and the ratio of revenue shares of labor and materials as follows:

$$\theta_{it}^L = \theta_{it}^V \frac{\alpha_{it}^L}{\alpha_{it}^V}, \quad \theta_{it}^M = \theta_{it}^V \frac{\alpha_{it}^M}{\alpha_{it}^V}.$$

And finally, the output elasticity of capital is identified as

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

4.3 CES

In this section, I consider the CES production function:

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

Using homotheticity property we can write

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

where \tilde{K}_{it} and \tilde{L}_{it} denote the ratio of capital to material and ratio of labor to material, respectively and m_{it} denotes the logarithm of materials. By taking the first-order conditions of cost minimization, one can show that

$$\omega_{it}^L = \gamma \tilde{L}^{(1-\sigma)/\sigma}, \quad \gamma := \left(\frac{(1 - \beta_l)p_t^l}{\beta_l p_t^m} \right)^{1/\sigma_1}$$

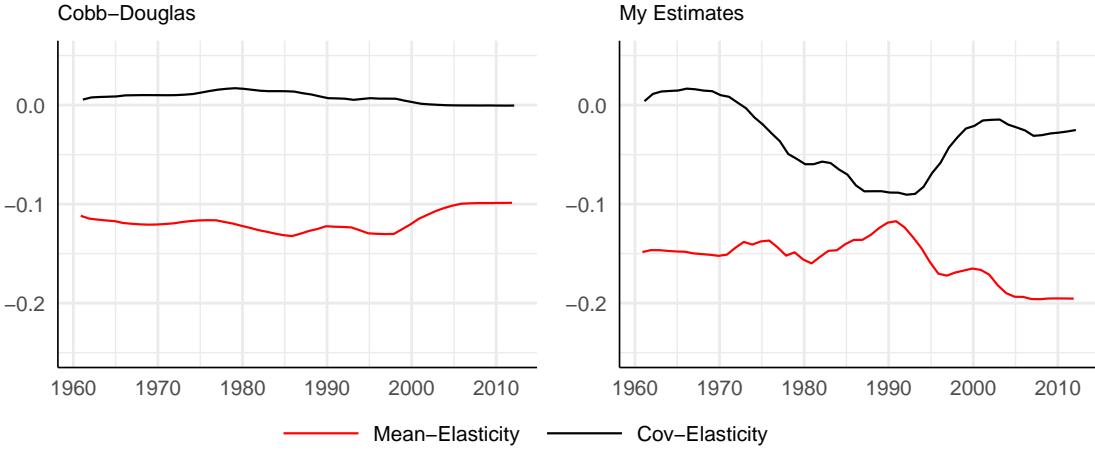
where γ is a constant that depends on input prices and model parameters. Substituting this into the production function we obtain

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

where ω_{it}^L disappeared from the model. This is a parametric analog of my nonparametric inversion result. The model parameters can be estimated using the control functions I develop

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

Figure 4.1: Decomposition of the Change in the Log Aggregate Markups



Notes: Comparisons of the evolution of the mean elasticity (red) and covariance elasticity (black) components of the aggregate log markups produced by my estimates and Cobb-Douglas model using the ACF procedure.

with the same objective function in Section 5. One can again show that the sum of the flexible input elasticities 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 γ_1 are not separately identified from this production function but the ratio is identified. Since θ_{it}^V depends only on the ratio, the sum elasticity is identified. Labor and materials elasticities are identified from θ_{it}^V and the ratio of revenue shares of labor and materials

$$\theta_{it}^L = \theta_{it}^V \frac{\alpha_{it}^L}{\alpha_{it}^V}, \quad \theta_{it}^M = \theta_{it}^V \frac{\alpha_{it}^M}{\alpha_{it}^V}.$$

And finally, the output elasticity of capital is identified as

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

4.4 Decomposing Growth in Markups

One remaining question is where this difference comes from. As I argued in my decomposition exercise in Section 7, two sources can explain the differences in time trends: average flexible input elasticity and the covariance between firm size and flexible input elasticity. To understand the role of these two channels, I plot them according to my production function and Cobb-Douglas in Figure 4.1. According to the Cobb-Douglas production function, the change in both components is limited over the last five decades. This explains why the decline in the share of flexible inputs drives the rise in markups. In contrast, my estimates suggest that both components change over time. The mean elasticity stays the same between 1960-1980, rises slightly for the next ten years and then declines heavily. The decline in the average elasticity can explain why we observe a decline in the revenue share of the flexible input. The covariance term also changes substantially and has a large variation.

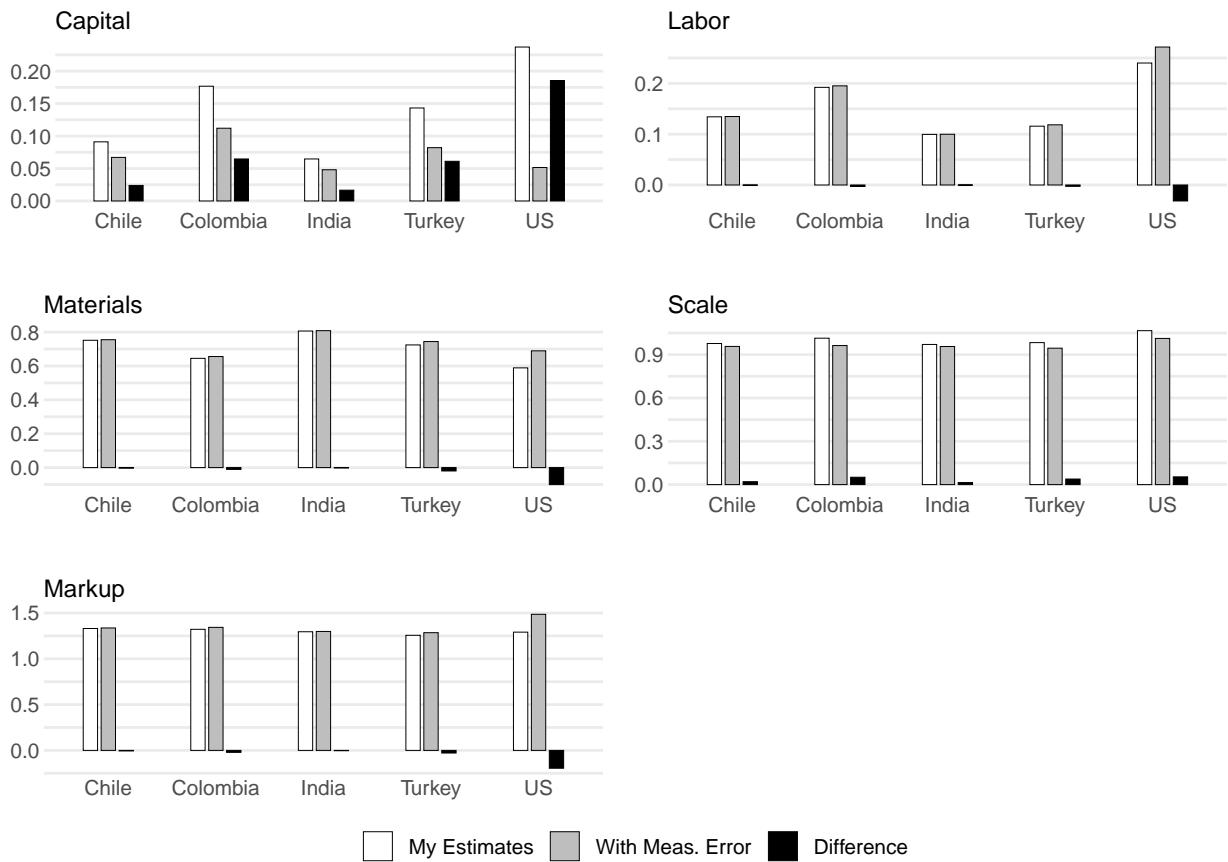
5 Robustness Checks

This section considers four robustness checks. I look at how (i) measurement error in capital, (ii) correction for capacity utilization, (iii) correction for selection and (iv) comparison with a translog production function, affect the empirical results.

5.1 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 the capital input, estimate output elasticities and markups using these dataset and report the average over 100 estimates.

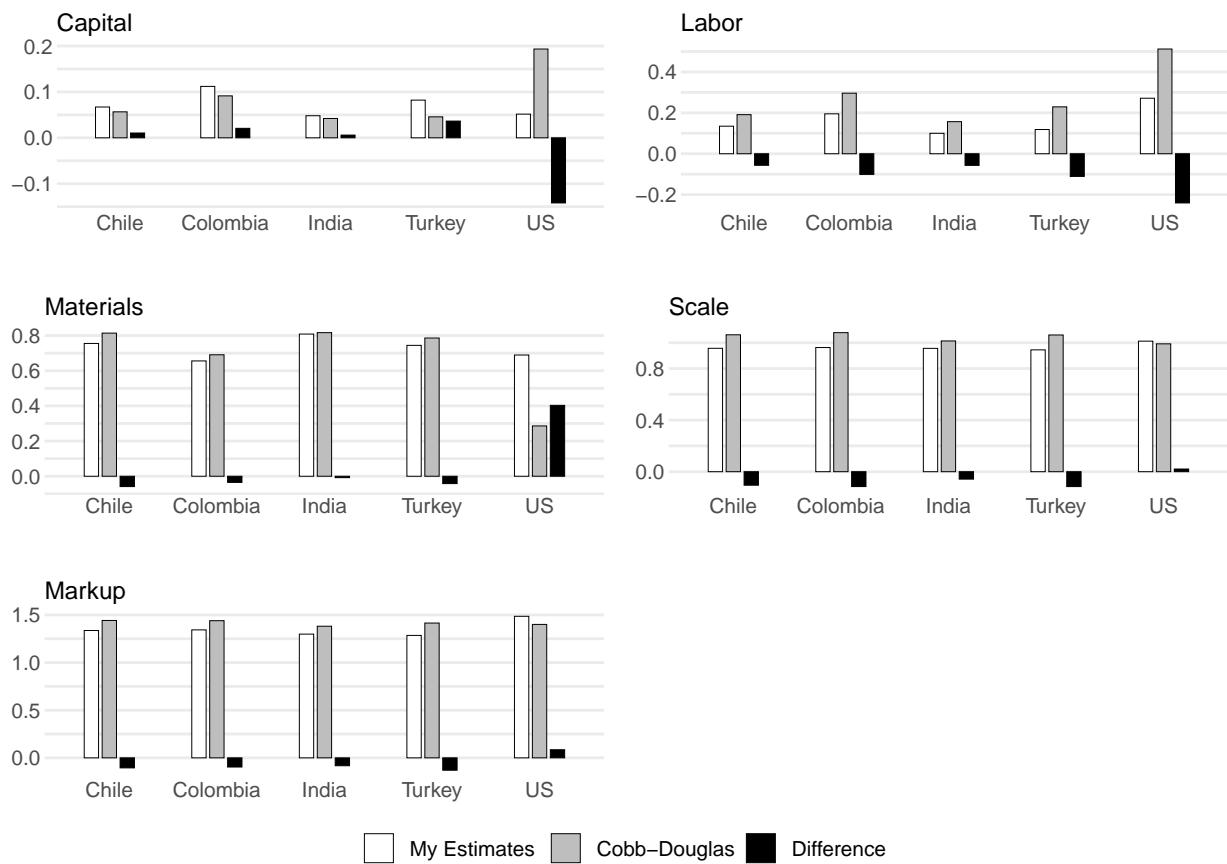
Figure 5.2: Comparison of Estimates with and without Measurement Error



Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without measurement error in capital. White bars report the estimates from the main text and grey bars report the average of 100 estimates obtained from simulated datasets as described in this section.

Figure 5.2 reports the original estimates together with the average of 100 estimates obtained from simulated datasets. As expected, measurement error in capital reduces the magnitude of the output elasticity of capital and increases the magnitude of 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. Figure 5.3, which compares estimates from my method and the Cobb-Douglas model in the presence of measurement error, provides further evidence. We see that the difference of the magnitudes between my model and Cobb-Douglas declines for capital elasticity and markups estimates, and increases for labor elasticity. This suggests that if data contains measurement in capital, my estimates become more conservative.

Figure 5.3: Comparison of Estimates Across Methods with Measurement Errors in Capital



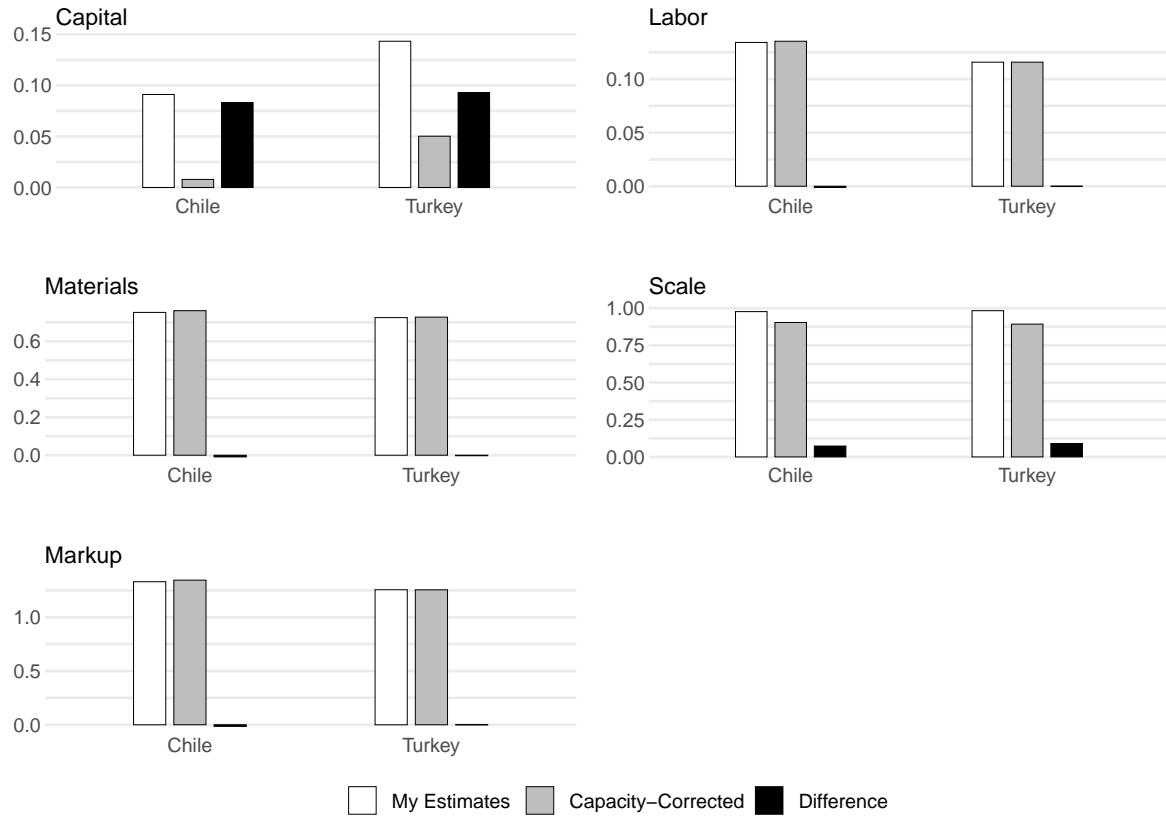
Notes: This figure compares sales-weighted output elasticities and markups estimated using my method and Cobb-Douglas averaged over 100 simulations whose specification is described in this section. Cobb-Douglas specification estimated using the Ackerberg et al. (2015).

5.2 Capital Utilization

In this section, I analyze 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 using the recovered capital.

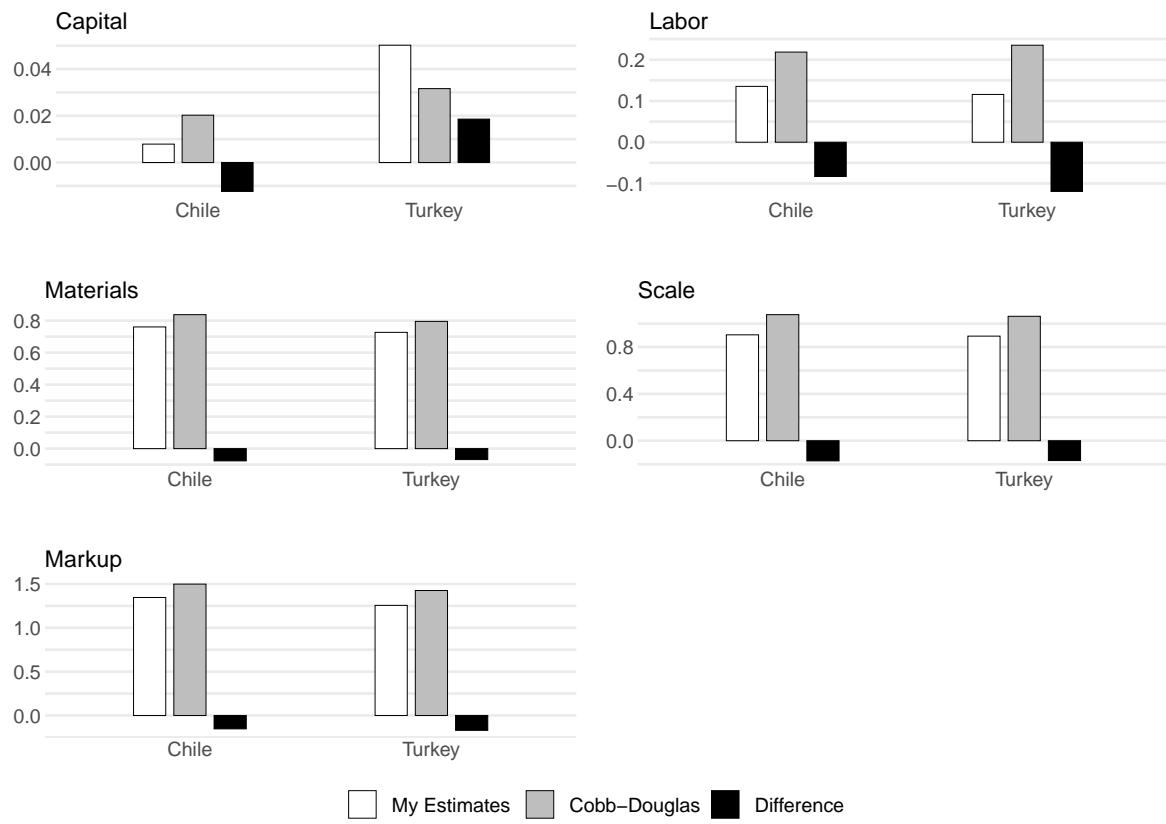
Figure 5.4 reports the original estimates together with the estimates obtained with capacity utilization corrected capital. The results suggest that correcting for capacity utilization affect only capital elasticities, and for other elasticities and markups, the estimates remain the same with negligible differences. For the output elasticity of capital, correcting for capacity utilization changes the estimates in different directions in Chile and Turkey. Figure 5.5 compares the estimates from my method and Cobb-Douglas model with using capacity utilization corrected capital. Comparison between my estimates and Cobb-Douglas estimates lead to the same conclusion as in the main text for all elasticities and markups, with the exception of capital elasticity.

Figure 5.4: Comparison of Estimates with and without Capacity Utilization Correction



Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without capacity utilization in capital. White bars report the estimates from the main text and grey bars report 100 estimates obtained after correcting for capacity utilization as described in this section.

Figure 5.5: Comparison of Estimates Across Methods with Capacity Utilization Correction

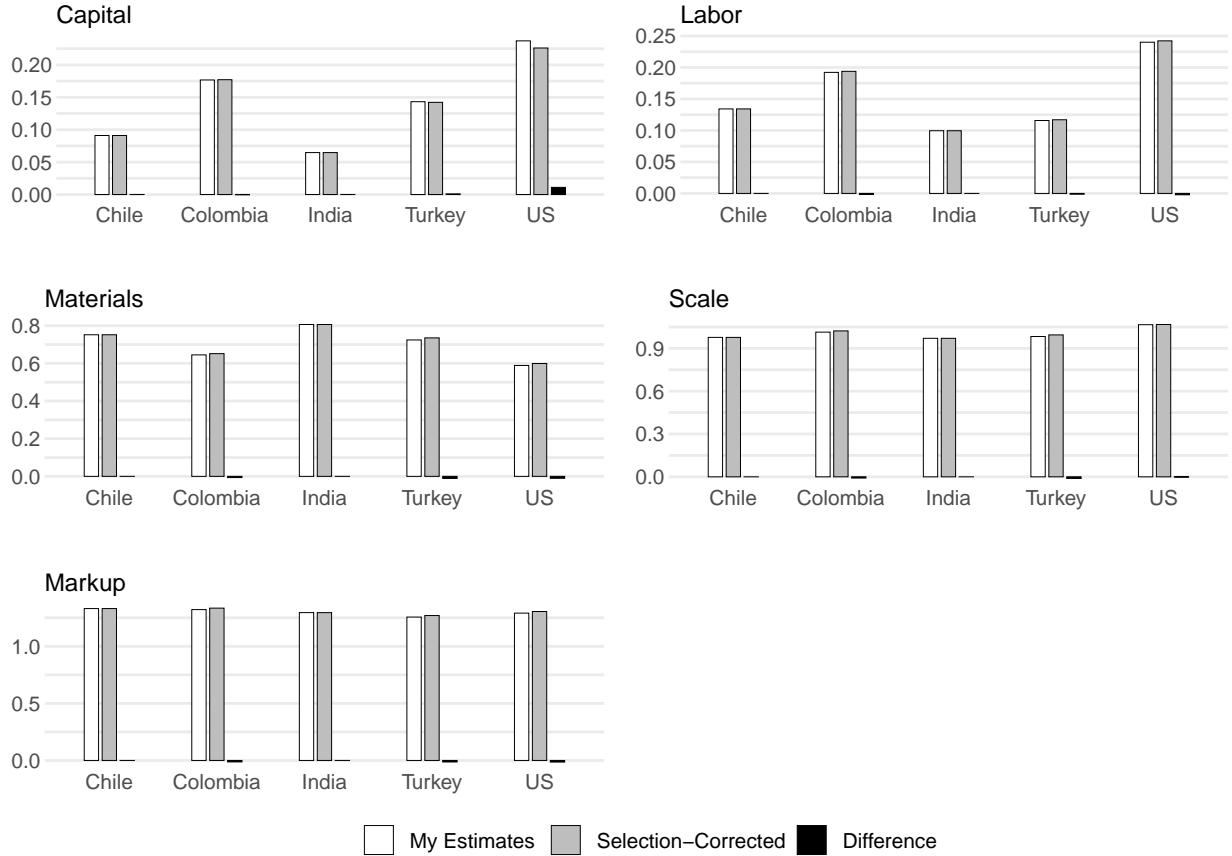


Notes: This figure compares sales-weighted output elasticities and markups estimated using my method and Cobb-Douglas with measurement errors in capital. Cobb-Douglas specification estimated using the Ackerberg et al. (2015).

5.3 Selection

In this section, I estimate output elasticities and markups after accounting for non-random firm exit as described in Subsection 3.3. Figure 5.6 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.

Figure 5.6: Comparison of Estimates with and without Selection Correction



Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without selection correction. White bars report the estimates from the main text and grey bars report the estimates after accounting for selection.

5.4 Markup Comparison to Nested CES with Labor-Augmenting Technology

The model introduced in this paper has two key features: labor-augmenting productivity and absence of parametric restrictions. Analyzing the role of these features theoretically is difficult, but we can disentangle their effects empirically. To this end, I compare my results with a parametric production function with labor-augmenting technology. In particular, I estimate the nested CES production function given in Equation (2.5). The details of the estimation are provided in Appendix 4.2.

Nested CES is a parametric model with labor-augmenting productivity, so a comparison highlights the role of the nonparametric component of my production function. Although this model contains labor-augmenting

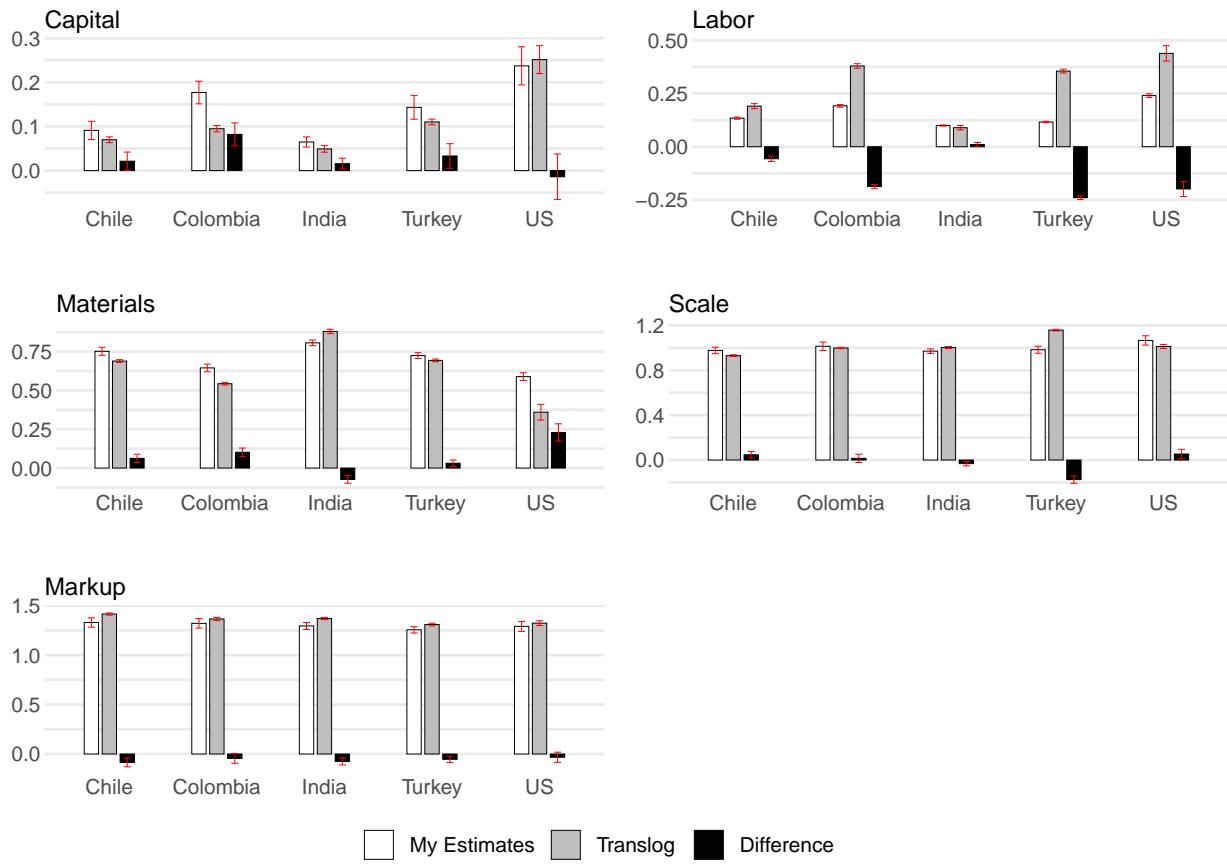
technology, the elasticities of substitutions and returns to scale are restricted to be common across firms.

Appendix Figure 6.17 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 markup comparison in Appendix Figure 6.18. Estimated markups are significantly different for the four developing countries between two models, but the two methods produce similar results for the US. We conclude that differences in the output elasticity estimates affect markups estimates, showing the implications of the parametric restrictions. To understand the source of this difference, I turn to markup decomposition. Appendix Figure 6.19 shows the difference between the first two components in markup decomposition. We see that by allowing for labor-augmenting technology, the difference between the average elasticity estimates vanishes, but the difference in covariances persists. I conclude that the nested CES estimates the average elasticity level correctly, but it does not account for the heterogeneity in the output elasticities.

5.5 Translog Production Function

Figure 5.7: Comparison of my Estimates with Translog Production Function



Notes: This figure compares sales-weighted output elasticities and markups estimated using my method and translog. Translog specification is estimated using the Ackerberg et al. (2015) method.

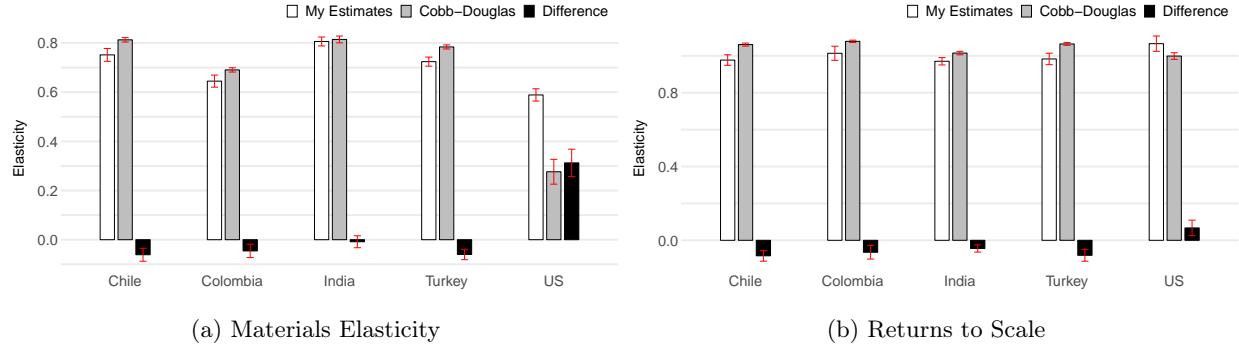
Table 6.8: Unweighted Average Output Elasticities

	Industry 1			Industry 2			Industry 3		
	FA	ACF	OLS	FA	ACF	OLS	FA	ACF	OLS
<i>Chile (311, 381, 321)</i>									
Capital	0.05 (0.01)	0.04 (0.00)	0.05 (0.00)	0.1 (0.02)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)
Labor	0.15 (0.00)	0.14 (0.01)	0.14 (0.01)	0.25 (0.01)	0.31 (0.02)	0.31 (0.02)	0.19 (0.01)	0.23 (0.02)	0.23 (0.02)
Materials	0.84 (0.01)	0.87 (0.01)	0.88 (0.01)	0.68 (0.02)	0.69 (0.01)	0.68 (0.01)	0.71 (0.02)	0.72 (0.01)	0.72 (0.01)
Rts	1.04 (0.02)	1.06 (0.01)	1.06 (0.00)	1.02 (0.03)	1.09 (0.01)	1.09 (0.01)	0.98 (0.02)	1.04 (0.01)	1.04 (0.01)
<i>Colombia(311, 322, 381)</i>									
Capital	0.09 (0.01)	0.07 (0.00)	0.07 (0.00)	0.09 (0.01)	0.07 (0.02)	0.08 (0.01)	0.13 (0.01)	0.13 (0.01)	0.13 (0.01)
Labor	0.18 (0.00)	0.18 (0.01)	0.18 (0.01)	0.36 (0.01)	0.46 (0.02)	0.44 (0.01)	0.33 (0.01)	0.36 (0.01)	0.36 (0.01)
Materials	0.75 (0.01)	0.8 (0.00)	0.8 (0.00)	0.59 (0.01)	0.56 (0.01)	0.54 (0.01)	0.57 (0.01)	0.61 (0.01)	0.61 (0.01)
Rts	1.01 (0.01)	1.05 (0.00)	1.05 (0.00)	1.04 (0.01)	1.09 (0.01)	1.06 (0.01)	1.02 (0.02)	1.1 (0.01)	1.09 (0.01)
<i>India(230, 265, 213)</i>									
Capital	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.08 (0.01)	0.02 (0.01)	0.04 (0.01)	0.04 (0.01)	0.02 (0.02)	0.08 (0.01)
Labor	0.12 (0.00)	0.09 (0.01)	0.09 (0.01)	0.31 (0.00)	0.43 (0.02)	0.35 (0.01)	0.13 (0.00)	0.42 (0.04)	0.35 (0.02)
Materials	0.82 (0.01)	0.84 (0.01)	0.84 (0.01)	0.57 (0.01)	0.54 (0.01)	0.56 (0.01)	0.77 (0.01)	0.64 (0.04)	0.54 (0.02)
Rts	0.99 (0.01)	0.98 (0.00)	0.98 (0.00)	0.96 (0.01)	0.99 (0.01)	0.95 (0.01)	0.95 (0.01)	1.08 (0.03)	0.96 (0.01)
<i>Turkey(321,311,322)</i>									
Capital	0.09 (0.01)	0.04 (0.00)	0.04 (0.00)	0.05 (0.01)	0.03 (0.00)	0.03 (0.00)	0.06 (0.02)	0.03 (0.01)	0.04 (0.01)
Labor	0.16 (0.00)	0.21 (0.01)	0.21 (0.01)	0.1 (0.00)	0.17 (0.01)	0.17 (0.01)	0.16 (0.00)	0.28 (0.01)	0.28 (0.01)
Materials	0.72 (0.02)	0.79 (0.01)	0.79 (0.00)	0.82 (0.01)	0.84 (0.00)	0.84 (0.00)	0.79 (0.02)	0.72 (0.01)	0.72 (0.01)
Rts	0.97 (0.02)	1.04 (0.00)	1.04 (0.00)	0.97 (0.01)	1.04 (0.00)	1.04 (0.00)	1.02 (0.03)	1.04 (0.01)	1.03 (0.01)
<i>US(33,32,31)</i>									
Capital	0.16 (0.01)	0.2 (0.01)	0.19 (0.01)	0.15 (0.02)	0.24 (0.03)	0.23 (0.02)	0.2 (0.04)	0.27 (0.05)	0.28 (0.05)
Labor	0.32 (0.00)	0.53 (0.02)	0.52 (0.02)	0.28 (0.01)	0.47 (0.03)	0.46 (0.03)	0.21 (0.01)	0.45 (0.05)	0.46 (0.04)
Materials	0.55 (0.01)	0.27 (0.02)	0.26 (0.02)	0.55 (0.02)	0.31 (0.06)	0.3 (0.04)	0.58 (0.03)	0.22 (0.06)	0.24 (0.05)
Rts	1.03 (0.01)	0.99 (0.01)	0.98 (0.01)	0.98 (0.02)	1.02 (0.01)	1 (0.01)	0.99 (0.04)	0.94 (0.02)	0.97 (0.02)

Note: Comparison of unweighted 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 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).

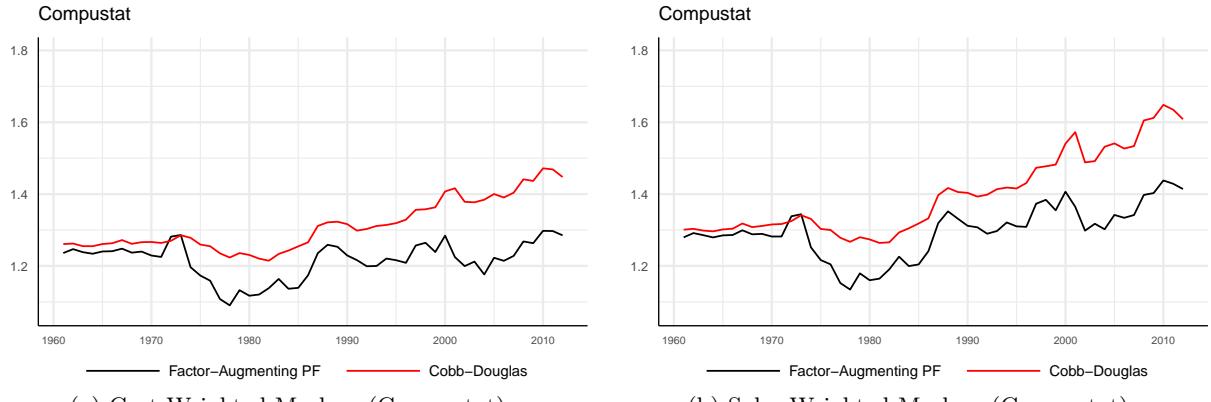
6 Additional Tables and Figures

Figure 6.8: Comparison of Output Elasticities



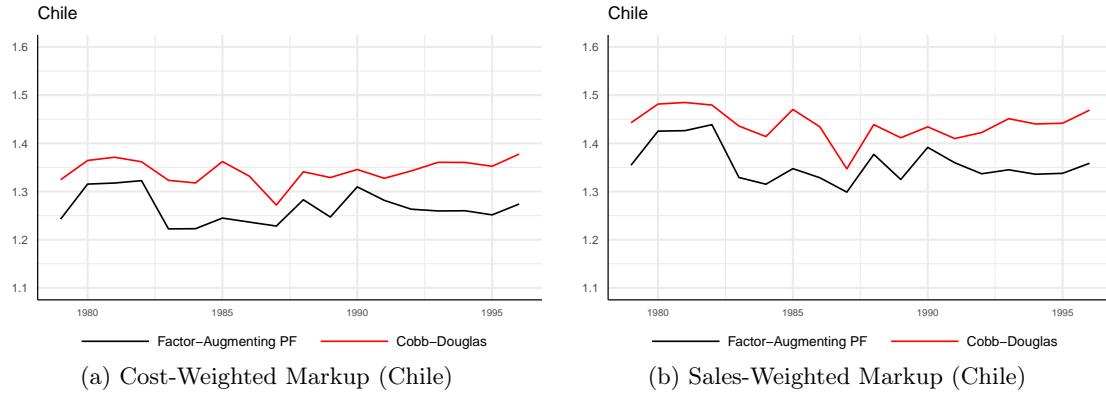
Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95 percent confidence intervals calculated using bootstrap (100 iterations).

Figure 6.9: Evolution of Markups (Compustat)



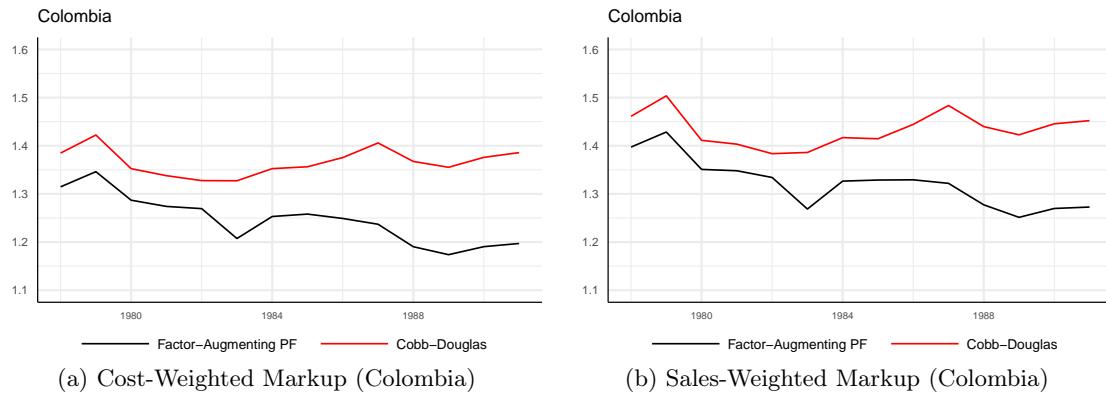
Notes: Comparisons of the evolution of markups in the US manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure 6.10: Evolution of Markups (Chile)



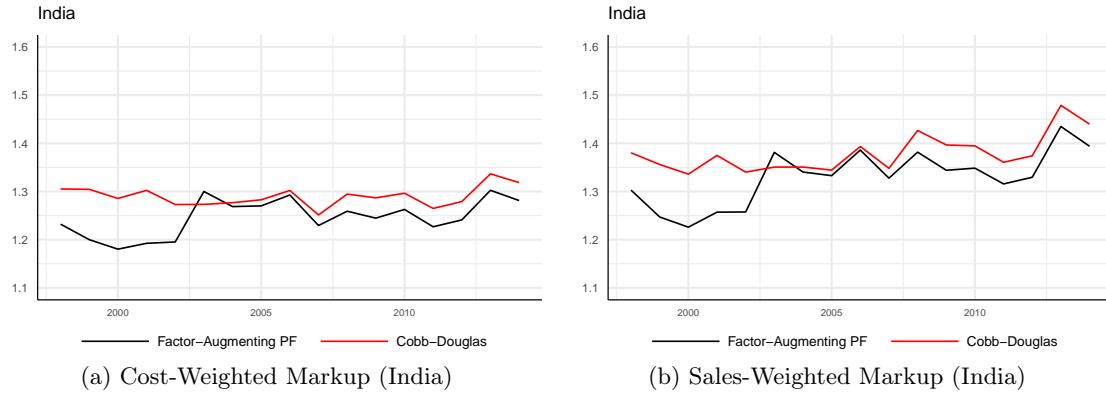
Notes: Comparisons of the evolution of markups in the Chilean manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure 6.11: Evolution of Markups (Colombia)



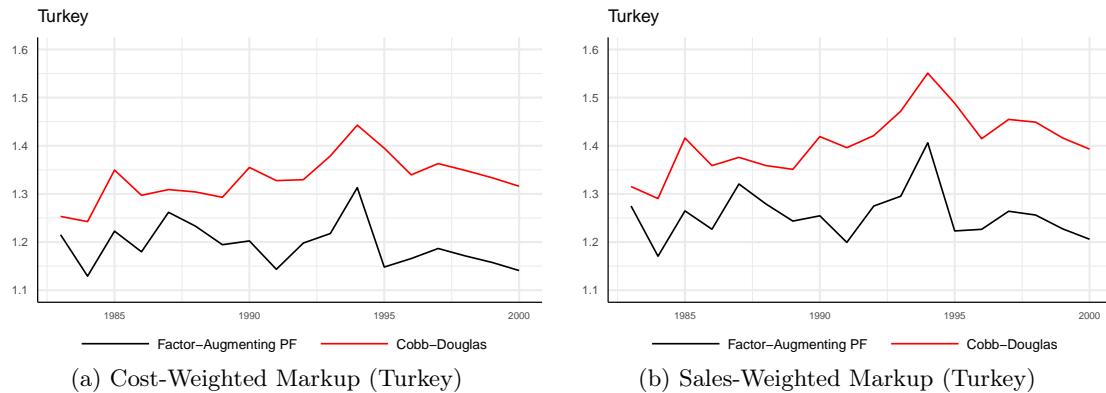
Notes: Comparisons of the evolution of markups in the Colombian manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure 6.12: Evolution of Markups (India)



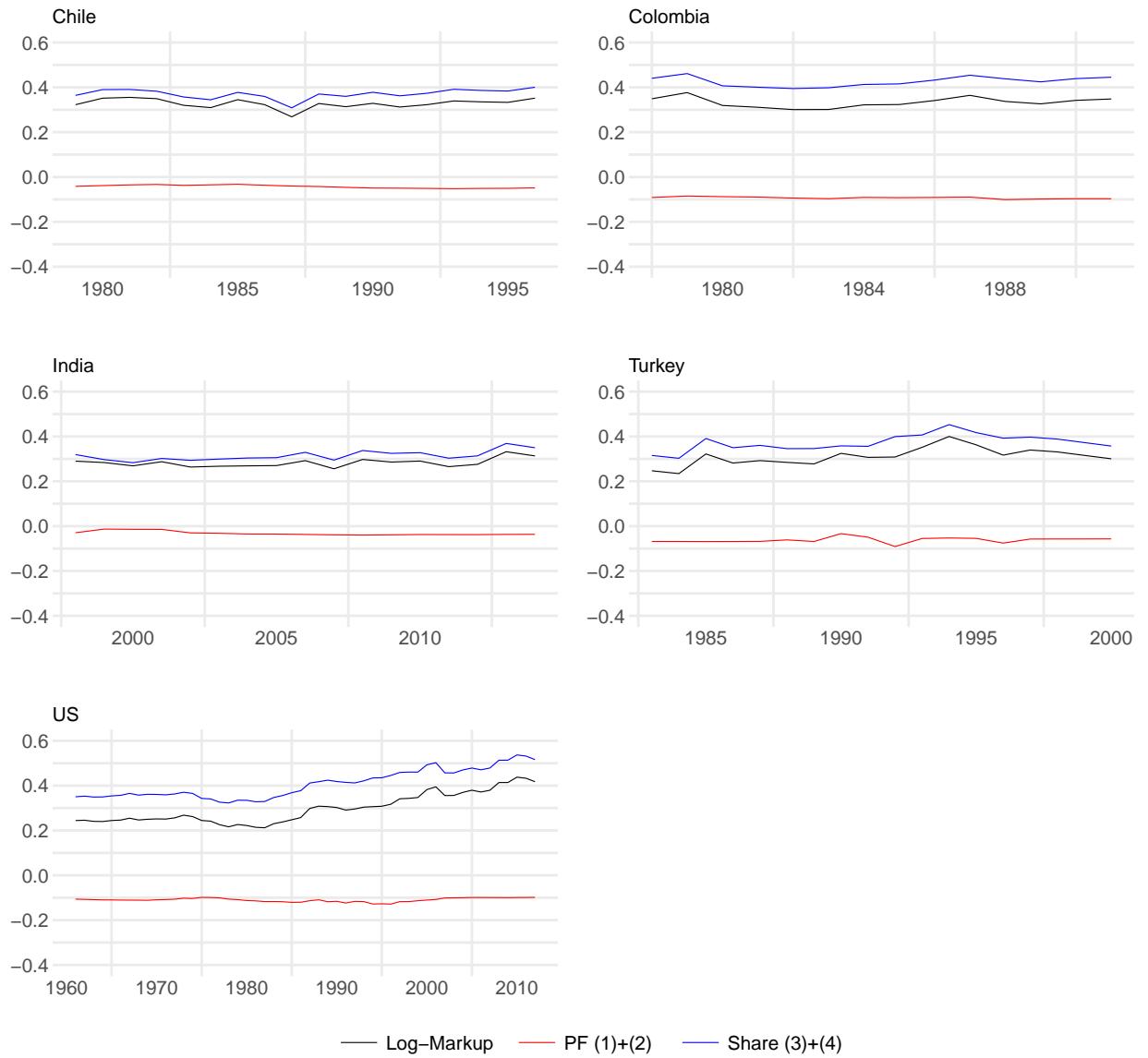
Notes: Comparisons of the evolution of markups in the Indian manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure 6.13: Evolution of Markups (Turkey)



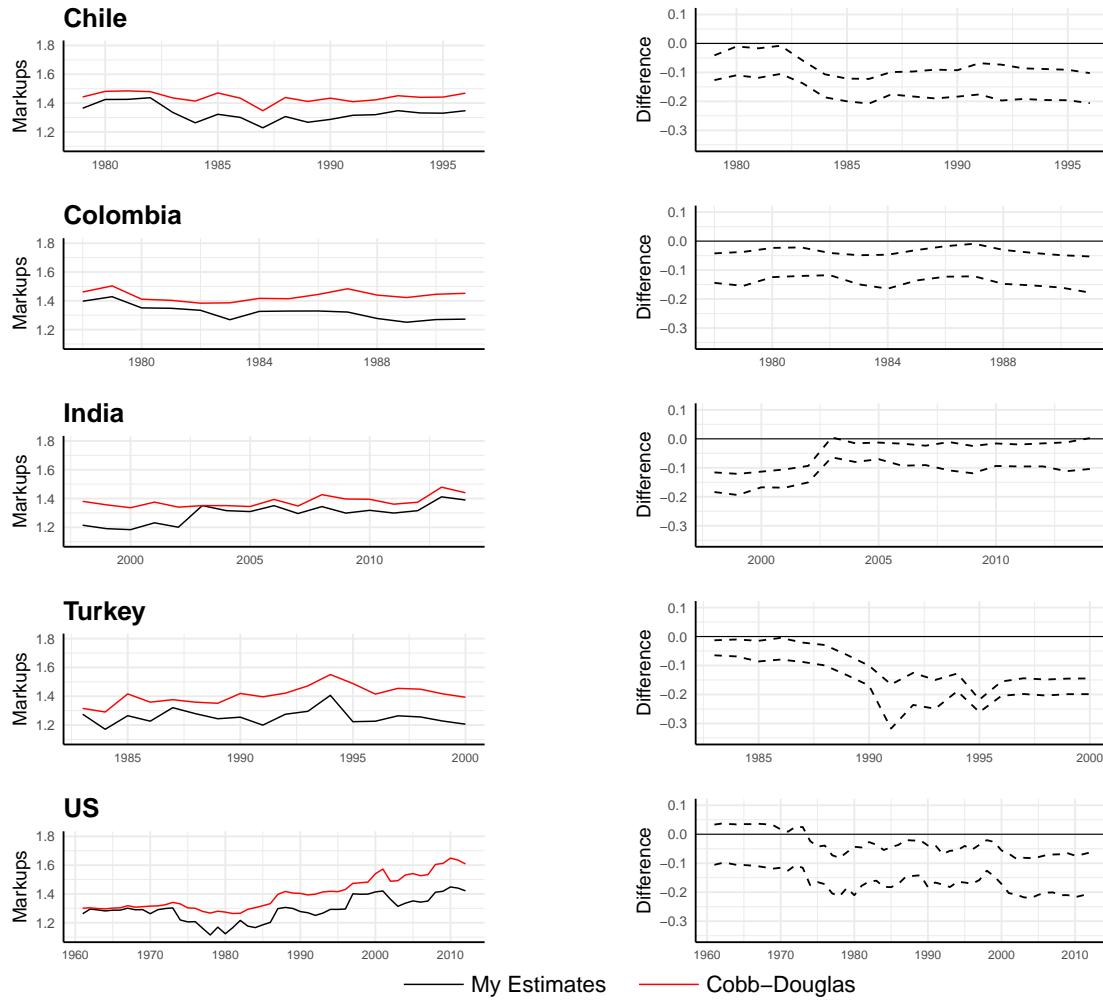
Notes: Comparisons of the evolution of markups in the Turkish manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure 6.14: Decomposition of Markup: Elasticity vs Share



Notes: This figure shows the evolution of the two components of log aggregate markup given in Equation 7.3. Black line displays the log aggregate markups, red line displays the component from production function estimation and, blue line displays the component from revenue share of flexible inputs.

Figure 6.15: 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 6.16: Sales-Weighted

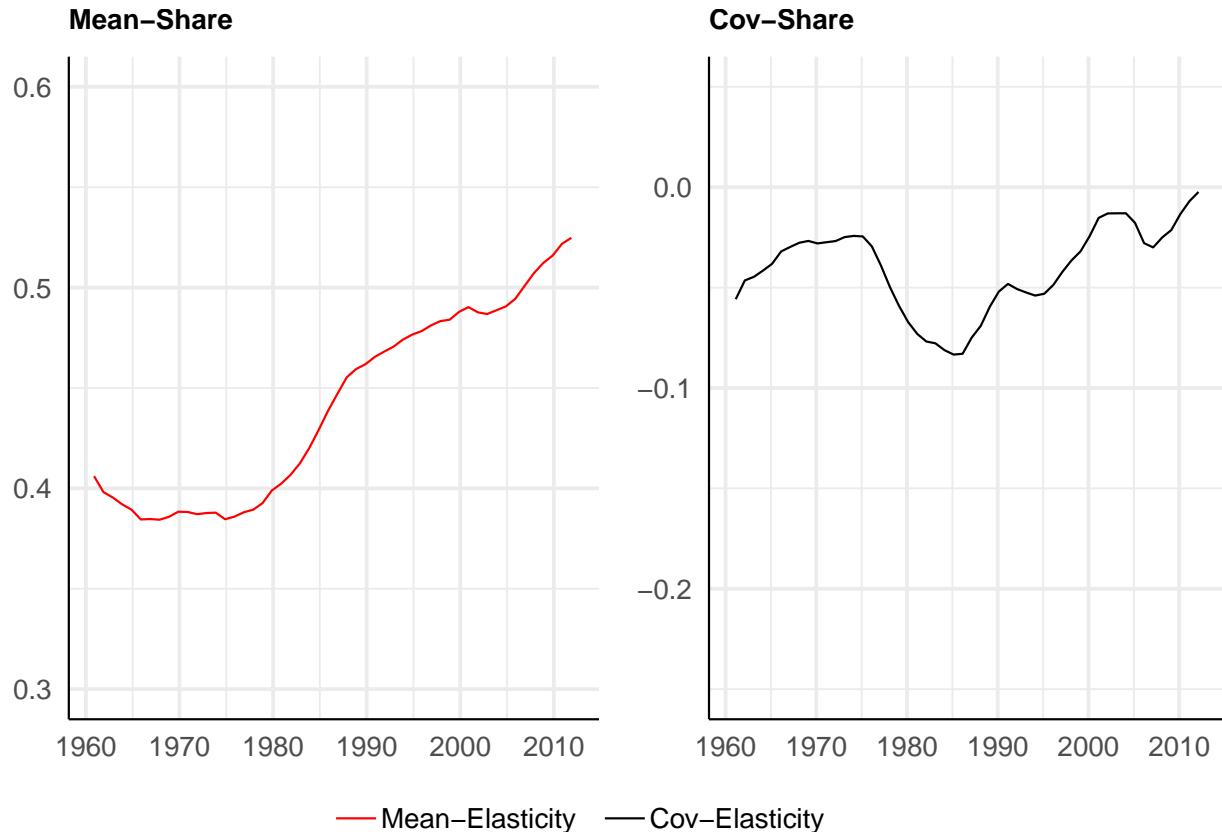
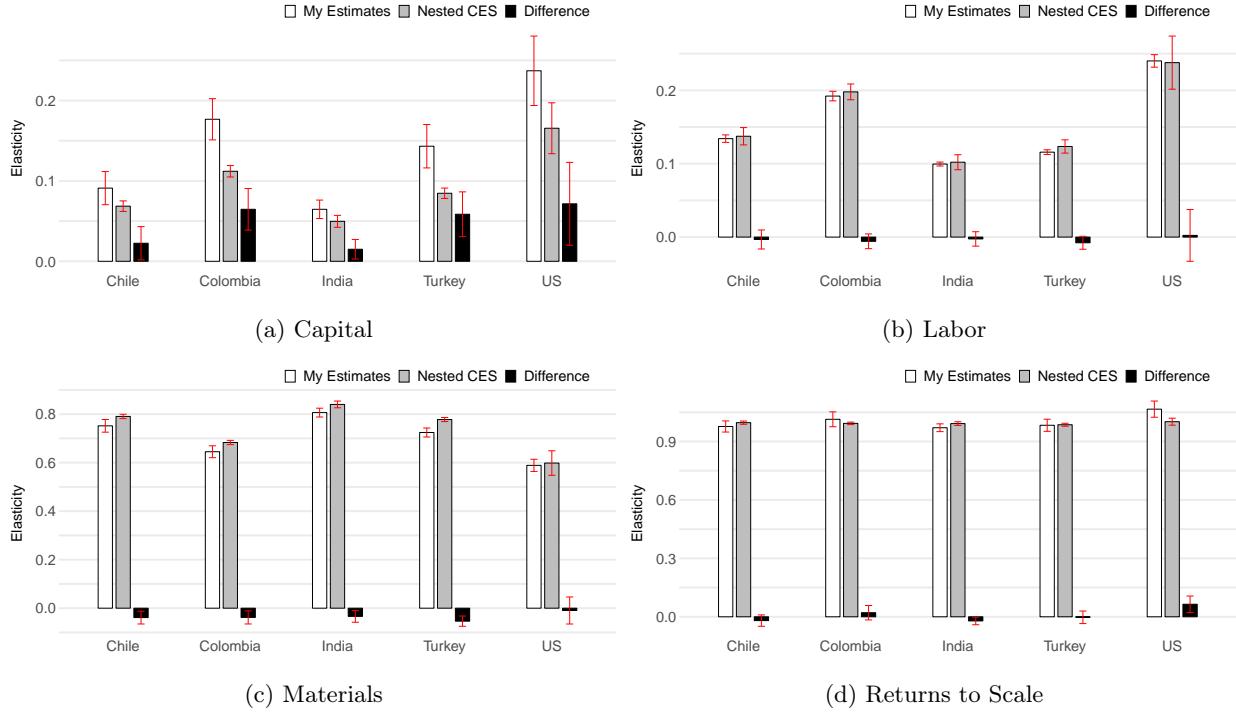
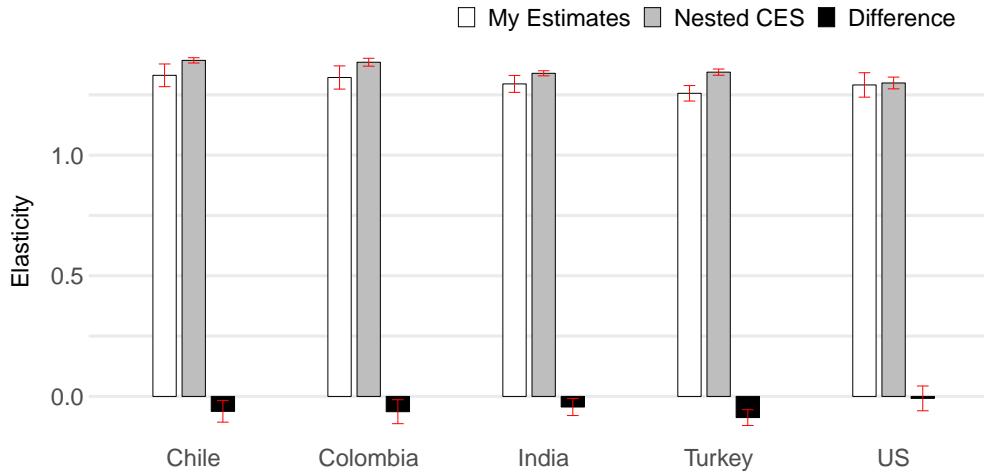


Figure 6.17: Comparison with Nested CES



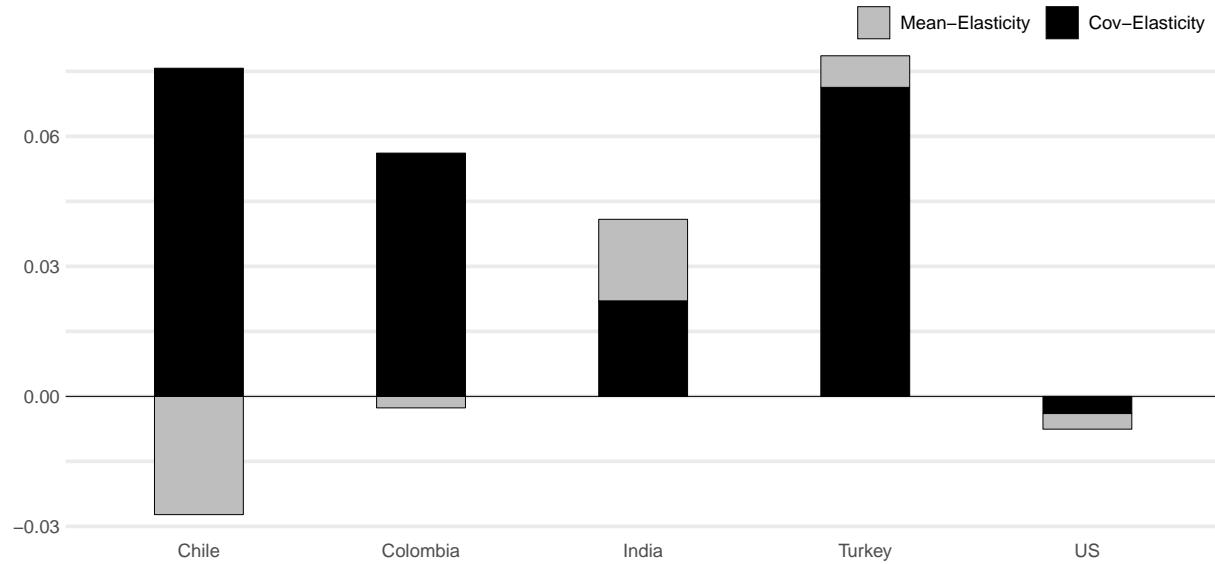
Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Nested CES estimated by procedure given in Subsection 4.2 for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95 percent confidence intervals calculated using bootstrap (100 iterations).

Figure 6.18: Markups Comparison with Nested CES



Note: Comparison of sales-weighted markups produced by my estimates (white) and Nested CES estimated by procedure given in Subsection 4.2 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 6.19: Decomposition of Markup Difference - Nested CES



Notes: This figure decomposes the difference between the aggregate log markups produced by non-parametric labor-augmenting production function and labor-augmenting CES production function estimated using the procedure described in Subsection 4.2. The decomposition is based on Equation 8.1.

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. 16, 17, 21, 24, 27, 28).
- 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 pp. 1, 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).
- Olley, G. Steven and Pakes, Ariel.** 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64 (6): 1263. (Cit. on p. 13).
- Raval, Devesh.** 2019. “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. 3).