Essays on Production Function Estimation

by

Mert Demirer

M.A., Koc Universisty (2013)

B.Sc., Bogazici University (2011)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2020

Author . . . .
Department of Economics
May 14, 2020

Certified by . . . .
Nikhil Agarwal
Associate Professor of Economics
Thesis Supervisor

Certified by . . . .
Daron Acemoglu
Elizabeth and James Killian Professor of Economics
Thesis Supervisor

Accepted by . . . .
Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics
Chairman, Department Committee on Graduate Studies

Essays on Production Function Estimation

by

Mert Demirer

Submitted to the Department of Economics
on May 14, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis contains three chapters.

This first chapter 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.

The second chapter develops partial identification results that are robust to deviations from the commonly used control function approach assumptions and measurement errors in inputs. In particular, the model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates a more flexible timing assumption for capital. I show that under these assumptions production function parameters are partially identified by an 'imperfect proxy' variable via moment inequalities. Using these moment inequalities, I derive bounds on the parameters and propose an estimator. An empirical application is presented to quantify the informativeness of the identified set.

The third chapter develops an approach in which endogenous networks is a source of identification in estimations with network data. In particular, I study a linear model where network data can be used to control for unobserved heterogeneity and partially identify the parameters of the linear model. My method does not rely on a parametric model of network formation. Instead, identification is achieved by assuming that the network satisfies latent homophily — the tendency of individuals to be linked with others who are similar to themselves. I first provide two definitions of homophily: weak and strong homophily. Then, based on these definitions, I characterize the identified sets and show that they are bounded under weak conditions. Finally, to illustrate the method in an empirical setting, I estimate the effects of education on risk preferences and peer effects using social network data from 150 Chinese villages.

Jel Classification: L110

Thesis Supervisor: Nikhil Agarwal
Title: Associate Professor of Economics

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics

Acknowledgments

My journal at MIT and the completion of this thesis would never have been possible without the support of many people.

I am deeply grateful to my advisors Nikhil Agarwal, Daron Acemoglu, Anna Mikusheva, and Victor Chernozhukov, for their guidance, help, and encouragement. Nikhil played the most critical role in shaping my research; without him, this thesis would have been very different. His encouragement and guidance when I was exploring new areas were crucial for me in taking my research in a new direction. He also has had a massive impact in shaping my thinking about research and economics. Daron has always been an inspiration. His patience and advice were instrumental for me to develop the ideas in my thesis. Anna has always been extremely helpful, and her on-point comments were amazingly useful in writing my dissertation. I have always looked forward to my meetings with her. Victor has always been a friendly mentor and advisor. I've benefited immensely from his wisdom and learned a lot from him as his RA and coauthor. In addition to my academic advisors, I am so grateful for the advice I received from Whitney Newey, Tobias Saltz, Glenn Ellison, Sara Fisher Ellison, Nancy Rose, and Mike Whinston.

I have had a great experience in graduate school, and for the most part, I owe this to terrific and wonderful classmates. My interactions with Max Cytrynbaum, Arda Gitmez, Colin Gray, Ryan Hill, Omer Karaduman, Maddie Mckelway, Alex Olssen, Ali Palida, Sophie Sun, Oguzhan Celebi, Nathan Zorzi, among many other, shaped my entire experience as a Ph.D. student. I have learned a lot from them, and they made my MIT journey enjoyable. I owe special thanks to my coauthor and friend Alex Olssen. Our numerous conversations over lunches and coffee not only led to our research projects but also, and more importantly, inspired me. It has been a real joy working with him and being his classmate. I am particularly grateful for the friendship of Colin Gray, without whom I do not know how I could finish the first two years of my Ph.D. His curiosity and passion for learning were contagious. Finally, I would not be able to achieve this journey without Arda Gitmez and Omer Karaduman. Arda has always been an inspiration for me, and without him as an example, doing a Ph.D. at MIT would not have crossed my mind. Omer has always been in my life as a classmate, roommate, officemate, and neighbor.

I am incredibly indebted to Kamil Yilmaz, my mentor, and coauthor. When going to MIT was inconceivable, he had tremendous belief in me, pushed me to believe in myself, and always encouraged me to achieve better. It is no doubt that without his constant support, I would have never been here.

Finally, I am thankful to my parents Nuray and Gazanfer Demirer for their unconditional love and support throughout my life. I feel so grateful and blessed to have parents like them, and I would not have gotten where I am today without them in my life. I dedicate this thesis to the most important person in my life, my wife, Cansu. Without her endless love and support over the past eight years, I have not been able to complete this journey. Doing a Ph.D. is challenging, and my secret to overcoming those challenges was to keep in mind that I will go home to her every night.

The UniCredit Foundation supported my coursework at MIT. I particularly thank Giannantonio De Roni and Annalisa Aleati for making the UniCredit Foundation an excellent organization.

Contents

List of Figures

List of Tables

Chapter 1

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

1.1 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 production function model 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

[1]For 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.

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

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 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

[3] My approach does not rely on variation in input prices. Instead, I use optimal expenditure on flexible inputs.

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 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 nonparametric setup. The rest of the assumptions extend

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

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

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

the standard proxy variable framework of Olley 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 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,

[7] In 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.

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.

1.1.1 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

[8] The 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).

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 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.

1.2 Model

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

[9]Raval (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.

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

1.2.1 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 heterogeneity in production technology across firms.

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

$$Y_{it} = F_t(K_{it}, \omega_{it}^L L_{it}, M_{it}) \exp(\omega_{it}^H) \exp(\epsilon_{it}), \tag{1.2.1}$$

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., $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}] = 0$, 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 (1.2.1). Although the industry-specific production function is restrictive, I introduce firm-level heterogeneity using a nonparametric model with firm-specific productivities. 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 to-

[11] I present this extension in Appendix A.2.1.

tal 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 sources of productivity differences, which typically are not available in standard production datasets. 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 by other production functions. To show an example, 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 (1.2.1) 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

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

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

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

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 nonparametric production function. 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]

1.2.2 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.

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.

than domestic firms (Bernard et al. (2007)).

[15]Although estimates vary, the consensus is that the aggregate elasticity of substitution between capital and labor is less than one. Antras (2004) and Klump et al. (2007) find estimates from 0.5 to 0.9 allowing for biased technical change. Herrendorf et al. (2015) estimate an elasticity of substitution of 0.84 for the US economy and 0.80 for the manufacturing sector. Alvarez-Cuadrado et al. (2018) estimate an elasticity of substitution of 0.78 for the manufacturing sector. 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.

[16]There 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. 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).

Assumption 1.2.1 (Homothetic Separability)**.** *Suppose that*
(i) *The production function is of the form*

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

(ii) $h_t(K_{it}, \cdot, \cdot)$ *is homogeneous of arbitrary degree (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 1.2.1(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 1.2.1(ii) states that $h_t(\cdot)$ is a homothetic function in effective labor and materials for any capital level. Combined with Assumption 1.2.1(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. I will use this property to control for ω_{it}^L. Note that the homotheticity of h_t does not imply that the production function is homothetic.

Assumption 1.2.1(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. Since my production function is nonparametric, cost-minimization does not give rise to parametric first-order conditions. As a result, this assumption is less restrictive in my model than in most models in the literature, which usually make functional form assumptions.

Assumption 1.2.1(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 1.2.1(iv) precludes this possibility.

Next, I state a result establishing the properties of the ratio of flexible inputs using Assumption 1.2.1.

Proposition 1.2.1.
(i) *Under Assumptions 1.2.1(i-iii), the flexible input ratio, denoted by* $\tilde{M}_{it} = M_{it}/L_{it}$*, depends only on* K_{it} *and* ω_{it}^L

$$\tilde{M}_{it} \equiv r_t(K_{it}, \omega_{it}^L), \tag{1.2.2}$$

for some unknown function $r_t(K_{it}, \omega_{it}^L)$.
(ii) *Under Assumption 1.2.1(iv),* $r_t(K_{it}, \omega_{it}^L)$ *is strictly monotone in* ω_{it}^L.

Proof. See Appendix A.4.

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 1.2.1. Under homothetic separability and cost minimization, the cost function can be derived as:

$$C(\bar{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, p_t^m, p_t^l) = C_1(K_{it}, \omega_{it}^L, p_t^m, p_t^l) C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H), \tag{1.2.3}$$

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:

$$M_{it} = \frac{\partial C(\bar{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, p_t^m, p_t^l)}{\partial p_t^m} = \frac{\partial C_1(K_{it}, \omega_{it}^L, p_t^m, p_t^l)}{\partial p_t^m} C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H), \tag{1.2.4}$$

$$L_{it} = \frac{\partial C(\bar{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, p_t^m, p_t^l)}{\partial p_t^l} = \frac{\partial C_1(K_{it}, \omega_{it}^L, p_t^m, p_t^l)}{\partial p_t^l} C_2(K_{it}, \bar{Y}_{it}, \omega_{it}^H). \tag{1.2.5}$$

From these equations, it is clear that the ratio of materials to labor does not change with ω_{it}^H.

The second part of Proposition 1.2.1 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 of above one, Assumption 1.2.1(iv) restricts the elasticity of substitution.[17] 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 1.2.1.

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

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

My framework nests the CES production function with $h\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big) = \beta_l \left[\omega_{it}^L L_{it}\right]^\sigma + (1 - \beta_l - \beta_m) M_{it}^\sigma$. 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

[17] In 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.

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:

$$\log(\tilde{M}_{it}) = \sigma \tilde{p}_t + \log(\omega_{it}^L), \tag{1.2.6}$$

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.[18] 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 1.2.1. I show that ω_{it}^L is invertible under more general conditions.[19]

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

$$Y_{it} = \Big(\beta_k K_{it}^{\sigma} + (1-\beta_k)\left(\beta_l \left[\omega_{it}^L L_{it}\right]^{\sigma_1} + (1-\beta_l) M_{it}^{\sigma_1}\right)^{\sigma/\sigma_1}\Big)^{v/\sigma} \exp(\omega_{it}^H)\exp(\epsilon_{it}). \tag{1.2.7}$$

This is a special case of my model with $h\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big) = \big(\beta_l \left[\omega_{it}^L L_{it}\right]^{\sigma_1} + (1-\beta_l-\beta_m) M_{it}^{\sigma_1}\big)^{1/\sigma_1}$. $h(\cdot)$ 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.[20] Appendix A.6 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.[21] In contrast, I impose a mild restriction on the elasticity of substitution given by Assumption 1.2.1(iv), so it can vary freely subject to this restriction.[22] 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.[23]

---

[18] In order to estimate Equation (1.2.6), one needs to observe heterogeneous input prices at the firm level.

[19] Doraszelski and Jaumandreu (2018) discuss informally using $\tilde{M}_{it}$ to control for ω_{it}^L without parametric assumptions. However, they consider a more restricted version than my model and do not show that $\tilde{M}_{it}$ is invertible in ω_{it}^L.

[20] I am not aware of any empirical work estimating this functional form.

[21] For a discussion, see Nadiri (1982).

[22] Some examples of production functions with non-constant elasticity of substitution are the Liu-Hildenbrand function (Liu and Hildebrand (1965)), transcendental production function (Revankar (1971)), constant marginal shares functions (Bruno (1968)), and Revankar's variable elasticity function (Revankar (1971)).

[23] Although I allow heterogeneity in returns to scale, I show how to impose common returns to scale across firms in Section 1.4.5.

Other Assumptions

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

Assumption 1.2.2 (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.[24,25] 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.[26]

Assumption 1.2.3 (Monotonicity)**.** *Materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^L, \omega_{it}^H), \tag{1.2.8}$$

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. This is an example of an additional complication introduced by labor-augmenting.[27,28]

[24]The model in Olley and Pakes (1996) has only Hicks-neutral productivity shock, which follows an exogenous first-order Markov process.

[25]The 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)). This extension would increase the dimension of the control variables in the estimation.

[26]Note 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.

[27]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-specific demand shocks, violating two-dimensional unobserved heterogeneity. Some examples of demand models satisfying this condition are monopolistic and Cournot competitions. For more discussion; see Jaumandreu (2018) and Doraszelski and Jaumandreu (2019).

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

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.[29] 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. Higher productivity has two effects on input demands: a negative effect since more productive firms need lower amount of inputs for production, and a positive effect due to the increase in output. For this assumption to hold, the positive effect should outweigh the negative effect, meaning that more productive firms produce sufficiently more output than less productive firms.[30]

Assumption 1.2.4 (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 $t-1$.

This assumption is called timing assumption in the literature. It states that current investment becomes productive in the next period, which implies that firms choose capital one period in advance. As a result, K_{it} belongs to the information set at period $t-1$, that is, $K_{it} \in \mathcal{I}_{it-1}$[31].

I use the assumptions in this section for two purposes. First, I use homothetic separability and monotonicity assumptions to express productivity shocks as functions of inputs. Then, I employ first-order Markov and timing assumptions to construct control variables.

1.2.3 Invertibility: Expressing Unobserved Productivity Using Inputs

Proposition 1.2.1 provides the necessary conditions, monotonicity and scalar unobserved heterogeneity, to invert out ω_{it}^L using the flexible input ratio:

$$\omega_{it}^L = r_t^{-1}(K_{it}, \tilde{M}_{it}) \equiv \bar{r}_t(K_{it}, \tilde{M}_{it}). \tag{1.2.9}$$

Similarly, Assumption 1.2.3 provides a monotonicity result for ω_{it}^H using materials demand function in Equation (1.2.8). Inverting that function yields

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \omega_{it}^L). \tag{1.2.10}$$

[29]Note 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.

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

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

This function contains another unobservable, ω_{it}^L, as an argument. Substituting for it from Equation (1.2.9) gives

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \tilde{r}_t(K_{it}, \tilde{M}_{it})) \equiv \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}). \tag{1.2.11}$$

Equations (1.2.9) and (1.2.11) 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 (1.2.11) 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 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.

1.3 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. My control variables differ in two ways. First, I construct control variables under conditional independence rather than independence. I need this extension because the modeling assumptions provide only conditional independence of productivity shocks. Second, my model involves two-dimensional unobserved heterogeneity, for which standard control variable approach does not work (Kasy (2011). I overcome this challenge by forming a triangular structure of productivity shocks using Equations (1.2.9) and (1.2.10).

The control variable approach has a long tradition in industrial organization. It has been used for estimating demand (Bajari and Benkard (2005), Ekeland et al. (2004), Kim and Petrin (2010)), dynamic discrete choice models (Hong and Shum (2010)) and auctions (Guerre et al. (2009)). These models fall into the standard framework with single-dimensional unobserved heterogeneity and independence restriction on unobserved variable. 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.

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.

1.3.1 Derivation of the Control Variable for Factor-Augmenting Technology

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

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1), \qquad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0,1). \tag{1.3.1}$$

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 $\tilde{M}_{it} = r\left(K_{it}, \omega_{it}^L\right)$. Substituting for ω_{it}^L from Equation (1.3.1) and using Equations (1.2.9) and (1.2.11), I obtain

$$\begin{aligned}
\tilde{M}_{it} &= r\big(K_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)\big), \\
\tilde{M}_{it} &= r\big(K_{it}, g_1\big(\bar{r}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}(K_{t-1}, \tilde{M}_{it-1}, M_{it-1}), u_{it}^1\big)\big), \\
\tilde{M}_{it} &\equiv \tilde{r}\left(K_{it}, W_{it-1}, u_{it}^1\right),
\end{aligned} \tag{1.3.2}$$

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 1.2.1, 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 (1.3.2) to derive the control variable for ω_{it}^L.

Lemma 1.3.1. *Under Assumptions 1.2.2 - 1.2.4, u_{it}^1 is jointly independent of (K_{it}, W_{it-1}).*

Proof. See Appendix A.4.

The intuition behind this result is as follows. We condition on $(\omega_{it-1}^L, \omega_{it-1}^H)$ first. By the timing assumption, (K_{it}, W_{it-1}) belongs to the previous period's information set. 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 (1.3.1), we can recover it from data as:

$$u_{it}^1 = F_{\tilde{M}_{it}|K_{it},W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}), \tag{1.3.3}$$

where $F_{\tilde{M}_{it}|K_{it},W_{it-1}}$ denotes the CDF of $\tilde{M}_{it}$ conditional on (K_{it}, W_{it-1}).[32] 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. This result establishes that u_{it}^1 is identified from the data and can be treated as known for identification. Using this result, I can express ω_{it}^L as a function of the control variable and past input:

$$\begin{aligned}
\omega_{it}^L &= g_1(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \\
\omega_{it}^L &= g_1\left(\bar{r}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1\right) \\
&\equiv c_1\left(W_{it-1}, u_{it}^1\right),
\end{aligned} \tag{1.3.4}$$

where $c_1(\cdot)$ is an unknown function. Condition on (W_{it-1}, u_{it}^1) there is no variation in ω_{it}^L, allowing me to control for ω_{it}^L.

1.3.2 Derivation of the Control Variable for Hicks-Neutral Technology

Control variable derivation for ω_{it}^H proceeds similarly, with one important difference: I need to condition on u_{it}^1. This is required because materials' input demand depends on ω_{it}^L in Assumption 1.2.3, and two productivity shocks can be dependent. I first use the Skorohod representation of ω_{it}^H to write:

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \text{Uniform}(0,1). \tag{1.3.5}$$

Unlike 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.[33] Next, I use the monotonicity of materials

[32] For 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 $F_{\tilde{M}_{it}|K_{it},W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}) = 1 - u_{it}^1$, a monotone transformation.

[33] 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 Appendix

in ω_{it}^H given by Assumption 1.2.3, $M_{it} = s\left(K_{it}, \omega_{it}^L, \omega_{it}^H\right)$, to write

$$\begin{aligned} M_{it} &= s\left(K_{it}, c_1\left(W_{it-1}, u_{it}^1\right), g_2\left(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2\right)\right) \\ &= s\left(K_{it}, c_1\left(W_{it-1}, u_{it}^1\right), g_2\left(\bar{r}\left(W_{it-1}\right), \bar{s}\left(W_{it-1}\right), u_{it}^1, u_{it}^2\right)\right) \\ &\equiv \tilde{s}\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right), \end{aligned} \tag{1.3.6}$$

where $\tilde{s}(\cdot)$ is an unknown function. Note that $\tilde{s}\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right)$ is strictly monotone in u_{it}^2 because $s\left(K_{it}, \omega_{it}^L, \omega_{it}^H\right)$ is strictly monotone in ω_{it}^H by Assumption 1.2.3, and $g_2\left(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2\right)$ is strictly monotone in u_{it}^2 by construction.

Lemma 1.3.2. *Under Assumptions 1.2.2 - 1.2.4, u_{it}^2 is jointly independent of $(K_{it}, W_{it-1}, u_{it}^1)$.*

Proof. See Appendix A.4.

The proof and intuition for this result are similar to those for Lemma 1.3.1. Having strict monotonicity and independence, we can use Equation (1.3.6) to identify u_{it}^2. In particular,

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1), \tag{1.3.7}$$

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:

$$\omega_{it}^H \equiv c_2\left(W_{it-1}, u_{it}^1, u_{it}^2\right), \tag{1.3.8}$$

for an unknown function $c_2(\cdot)$ whose derivation is the same as Equation (1.3.4). This result and Equation (1.3.4) 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.[34,35]

Remark 1.3.1 (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. Appendix A.6.1 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 $u_{it} = F_{m_{it}|m_{it-1},k_{it-1}}(m_{it} \mid k_{it}, m_{it-1}, k_{it-1})$. Substituting

A.2.3 for control variable derivation under this assumption.

[34]Using the same procedure and substituting past productivities recursively, we can write productivity shocks as $\omega_{it}^L = c_1\left(W_{it-k}, \{u_{it-l}^1\}_{l=0}^{k-1}\right)$ and $\omega_{it}^H = c_2\left(W_{it-k}, \{u_{it-l}^1\}_{l=0}^{k-1}, \{u_{it-l}^2\}_{l=0}^{k-1}\right)$ for any integer k, where u_{it-l}^1 and u_{it-l}^2 are defined as in Equation (1.3.3) and (1.3.7). This would lead to more identifying variation at the expense of having to estimate more control variables.

[35]I show in Appendix A.2.1 how to extend the control variable approach when there is heterogeneity in input prices.

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 $\mathbb{E}[\epsilon_{it} \mid k_{it}, l_{it}, m_{it-1}, k_{it-1}, u_{it}] = 0$. As I discuss in Appendix A.6.1, 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 $(m_{it-1}, k_{it-1}, u_{it})$[36]. Second, there are efficiency gains, as my approach fully uses the independence condition given by the Markov assumption.

Remark 1.3.2 (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.

Remark 1.3.3 (Source of Variation)**.** In the control variable approach, the identifying variation comes from any remaining variation after conditioning on control variables. In my framework, variables affecting the firm's dynamic problem, which are outside the model, generate this variation. In particular, there are several possible firm-specific unobservables such as adjustment costs, expectations of future demand, borrowing constraints, and investment prices, which affect the firm's dynamic input decisions but not flexible inputs decisions.[37] Therefore, my estimation method does not require modeling these variables, even though they help with identification.

1.3.3 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 1.2.3 to control

[36]To see this, if labor is perfectly flexible, we can write it as $l_{it} = l(k_{it}, \omega_{it}^H) = l\big(k(k_{it-1}, \omega_{it-1}, \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})\big) = l\big(k(k_{it-1}, s^{-1}(k_{it-1}, m_{it-1}), \nu_{it-1})), c(m_{it-1}, k_{it-1}, u_{it})\big) =: \tilde{l}(k_{it-1}, m_{it-1}, u_{it}, \nu_{it-1})$, 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.

[37]There is strong evidence for heterogeneity in adjustment cost. The literature on optimal investment has highlighted firm-level heterogeneity in investment decisions. For example, Goolsbee and Gross (2000) present empirical evidence on heterogeneity in adjustment cost. Cooper and Haltiwanger (2006) argue that there is substantial heterogeneity in capital associated with heterogeneity in adjustment costs. Hamermesh and Pfann (1996), in a review paper, claim that heterogeneity in adjustment cost is a key source of heterogeneity across firms and should be included in models of firm behavior. Theoretical models of optimal investment decisions, such as Caballero and Engel (1999), Abel and Eberly (2002), and Eberly (1997), allow heterogeneity in capital adjustment costs and argue that realistic models should account for this heterogeneity.

for endogeneity,

$$\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it}), \qquad \omega_{it}^H = \bar{s}(K_{it}, M_{it}, \tilde{M}_{it}). \tag{1.3.9}$$

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:

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

which requires past inputs and control variables, u_{it}^1 and u_{it}^2, to control for endogeneity. Consequently, I do not need to condition on any of the current period inputs directly, which reduces the dimension of the conditioning variables. I achieve this result by exploiting the Markov assumption. Papers using the proxy variable framework, such as Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), have also assumed that productivity follows a first-order Markov process, but they have 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 gives me 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.

1.4 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 fully 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. The rest of the section examines identification of the other features of the production function and explores how further economic restrictions can be imposed on the production function.

1.4.1 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

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

Since $h(\cdot)$ is homothetic in its second and third arguments by Assumption 1.2.1, I assume, without loss of generality, that it is homogeneous of degree one.[38] Using this property, I rewrite the production function as

$$y_{it} = f\big(K_{it}, L_{it} h(K_{it}, \omega_{it}^L, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{1.4.1}$$

This reformulation is convenient because ω_{it}^L becomes an argument in $h(\cdot)$. In Subsection 1.2.3, I showed that $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$. Substituting this into Equation (1.4.1) gives

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

This equation reveals an identification problem.

Proposition 1.4.1. *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.[39] Therefore, h is not identified from the relationship between the inputs and output.[40] 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

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

where f_k denotes the derivative of f with respect to its $k-$th component. I also use 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. Therefore, the elasticities cannot be identified from the variation in the inputs and output.

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:

$$y_{it} = f\big(K_{it}, L_{it} \bar{h}(K_{it}, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{1.4.2}$$

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. On the other hand, $h(\cdot)$ and $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$

[38] This is without loss of generality because if h is not homogeneous of degree one, we can take observationally equivalent monotone transformations of f and h to obtain a new h function that is homogeneous of degree 1.

[39] As I show in Appendix A.6.1, with variation in input prices $\bar{r}(K_{it}, \tilde{M}_{it})$ depends also on the price ratio and functional dependence brakes down.

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

are *structural* features of the model. 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 $\bar{h}(\cdot)$.

1.4.2 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.

Identifying the Ratio of Labor and Materials Elasticities

The multicollinearity problem presented in Subsection 1.4.1 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:

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

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

$$\underbrace{\frac{F_V V_{it}}{Y_{it}}}_{\text{Elasticity}(\theta_{it}^V)} \frac{\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}] \lambda_{it}}{\exp(\epsilon_{it}) p_{it}} = \underbrace{\frac{V_{it} p_t^v}{Y_{it} p_{it}}}_{\text{Revenue Share of Input}(\alpha_{it}^V)}, \tag{1.4.3}$$

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 (1.4.3) for $V = M$ and $V = L$ yields

$$\frac{\theta_{it}^M}{\theta_{it}^L} = \frac{\alpha_{it}^M}{\alpha_{it}^L}. \tag{1.4.4}$$

The ratio of the output elasticities of labor and materials is identified as the ratio of revenue shares using the cost-minimization assumption.[41] The revenue shares are often observed in the data so that we can calculate the ratio of elasticities without estimation. In a recent paper, Doraszelski and Jaumandreu (2019) also used the ratio of revenue shares to identify the ratio of elasticities.

The idea underlying identification using the first-order conditions is that firms observe productivity shocks before making input choices. As a result, the input mix used in production reveals information about the production function. This information alone identifies the ratio of elasticities. 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.

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 (1.4.3) in a nonparametric fashion. They assume perfect competition in the output market, which implies $\lambda_{it} = p_{it}$, so 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.

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 (1.4.2).

Proposition 1.4.2. *The sum of labor and materials elasticities is identified from f and $\bar{h}$ as*

$$\theta_{it}^{V} := \theta_{it}^{M} + \theta_{it}^{L} = f_2\big(K_{it}, L_{it}\bar{h}(K_{it}, \tilde{M}_{it})\big) L_{it}\bar{h}(K_{it}, \tilde{M}_{it}), \tag{1.4.5}$$

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

Proof. Using Equation (1.4.1) the output elasticities of materials and labor can be obtained as:

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

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

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

□

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

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.[42] 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

$$\theta_{it}^L = \theta_{it}^V \frac{\alpha_{it}^L}{\alpha_{it}^V}, \qquad \theta_{it}^M = \theta_{it}^V \frac{\alpha_{it}^M}{\alpha_{it}^V}, \tag{1.4.6}$$

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.

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

Proof. See Appendix A.4.

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 1.4.1. As a solution to this problem, I next ask what further restrictions are required to identify the objects in Proposition 1.4.3.

1.4.3 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

[42] Note 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 $f_2\tilde{h} = f_2'\tilde{h}'$. Therefore, the observationally equivalent $(f, \tilde{h})$ and $(f', \tilde{h}')$ give the same sum of flexible input elasticities.

not identified. Consider the following production function:

$$y_{it} = f\big(K_{it}, h(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{1.4.7}$$

This model differs from the main model in that h does not take K_{it} as an argument.[43] Since this is a special case, Proposition 1.2.1 applies to this production function with $\omega_{it}^L = \bar{r}(\tilde{M}_{it})$. Substitution this into Equation (1.4.7), I obtain the *reduced form* for the production function in Equation (1.4.7) as

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{1.4.8}$$

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 1.4.4. *If we replace the production function in Assumption 1.2.1 with Equation (1.4.7), the capital elasticity is identified and labor-augmenting productivity is identified up to scale from* $(f, \bar{h}, \theta_{it}^L, \theta_{it}^M)$ *as:*

$$\theta_{it}^K = f_1(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})), \qquad \log(\omega_{it}^L) = \log(\bar{r}(\tilde{M}_{it})) = \int_{\underline{\tilde{M}}}^{\tilde{M}_{it}} b(\bar{M}_{it}) d\bar{M}_{it} + k. \tag{1.4.9}$$

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 A.4.

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 1.4.5. *Under the conditions of Proposition 1.4.4 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 A.4.

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.

[43] This 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)).

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 Appendix A.2.1, 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.[44]

An important implication of using the first-order conditions for identification is that the output elasticities can only be identified for values of $(L_{it}, \omega^L_{it}, M_{it})$ on the surface $\{(\omega^L_{it}, M_{it}) \mid \omega^L_{it} = \bar{r}(\tilde{M}_{it})\}$. 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^L_{it} constant and asking how change in inputs affects output.[45] 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.

1.4.4 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

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

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

$$y_{it} = vk_{it} + \tilde{f}\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}, \tag{1.4.10}$$

where $\tilde{f} = f\big(1, \tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big)$. 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, $\theta^K_{it} = v - \theta^L_{it} - \theta^M_{it}$.

A researcher might want to restrict the returns to scale for two reasons. First, there might be a priori theoretical reason to assume that the returns to scale are common across firms. The second reason is related to estimation. Restricting returns to scale reduces the dimension of the functions, leading to more precise estimates.

[44] I leave this as future work.

[45] Note 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^L_{it} is non-separable from the production function, so output elasticities depend on an unobserved variable.

1.4.5 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.

$$y_{it} = f\big(K_{it}, h(K_{it}, \omega^L_{it} L_{it}, M_{it})\big) + \omega^H_{it} + \epsilon_{it} \qquad \text{(Weak Homothetic Sep.)}$$

$$y_{it} = f\big(K_{it}, h(\omega^L_{it} L_{it}, M_{it})\big) + \omega^H_{it} + \epsilon_{it} \qquad \text{(Strong Homothetic Sep.)}$$

$$y_{it} = vk_{it} + f\big(\tilde{L}_{it} h(\omega^L_{it}, \tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it} \qquad \text{(Homogeneous)}$$

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

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

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.

1.5 Empirical Model and Data

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

1.5.1 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 (1.4.7), which leads to the following estimating equation:

$$y_{it} = f\big(K_{it}, L_{it} \bar{h}(\tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}. \tag{1.5.1}$$

In Section 1.4, I showed how to identify the output elasticities from f and $\bar{h}$, so the goal is to identify these functions.[46] To control for Hicks-neutral productivity, I use the control variables developed in Equation (1.3.8), $\omega_{it}^H = c_2\left(W_{it-1}, u_{it}^1, u_{it}^2\right)$. Substituting this into Equation (1.5.1), the estimating equation can be written as

$$y_{it} = f\big(K_{it}, L_{it}\tilde{h}(\tilde{M}_{it})\big) + c_2\left(W_{it-1}, u_{it}^1, u_{it}^2\right) + \epsilon_{it}. \tag{1.5.2}$$

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

$$\mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0. \tag{1.5.3}$$

Since all right-hand-side variables are orthogonal to the error term, Equation (1.5.2) can be estimated by minimizing the sum of squared residuals. However, Equation (1.5.3) 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. I next show how to use this condition to augment the moment restriction in Equation (1.5.3). 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 $\mathbb{E}[\xi_{it} \mid \mathcal{I}_{it-1}] = 0$. This innovation term is different from those defined in Section 1.3 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:

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it}\big) + c_3(W_{it-1}) + \xi_{it} + \epsilon_{it}. \tag{1.5.4}$$

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

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

[46]Note 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.

with two conditional moment restrictions:

$$\mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u^1_{it}, u^2_{it}] = 0 \tag{1.5.5}$$
$$\mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}, W_{it-1}] = 0. \tag{1.5.6}$$

Output elasticity estimation requires estimating the unknown functions f, $\bar{h}$, c_2 and c_3 using these moment restrictions. In Appendix A.5, I analyze the identification of f and $\bar{h}$ based on the moment restriction in Equation (1.5.5) and show that it identifies f and $\bar{h}$ except for special cases.[47] These cases include some support conditions on the derivatives of conditional CDF in Equation (1.3.7), so they are testable.[48] Since Equation (1.5.5), by itself, generically identifies the output elasticities, the moment restriction in Equation (1.5.6) provides efficiency gains and overidentifying restrictions. As described in Section 1.3.1, it is possible to construct other moment restrictions to increase efficiency at the expense of a more complicated estimation procedure.

The estimation proceeds in two steps. In the first step, I estimate the control variable u^2_{it} by estimating the conditional CDF in Equation (1.3.7). In the strongly separable model, u^1_{it} 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 (1.5.5) and (1.5.6).

Estimation Procedure

In this section, I provide an overview of the estimation procedure. A more detailed estimation algorithm is given in Appendix A.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).[49]

The estimation involves two stages. In the first stage, I learn conditional distribution function in Equation (1.3.7). 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.[50]

[47] This is sometimes called generic identification; see Lewbel (2016).

[48] I also provide these conditions for homothetic production function in Appendix Proposition A.5.1 and for strong homothetic separable production function in Appendix Proposition A.5.2.

[49] The 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.

[50] The results are robust to using other degrees of polynomials.

Replacing the true functions with the approximations yields

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

I construct an objective function using the moment restrictions in Equations (1.5.5) and (1.5.6). In particular, I use the sum of squared residuals from Equation (1.5.5) and timing moments from Equation (1.5.6) to obtain the following objective function:

$$J(\widehat{f}, \widehat{\widetilde{h}}, \widehat{c}_2, \widehat{c}_3) = \underbrace{\frac{1}{TN}\sum_{i,t}\widehat{\epsilon}^2_{1it}}_{\text{Sum of Squared Residuals}} + \underbrace{\Big(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K_{it}\Big)^2 + \Big(\frac{1}{TN}\sum_{i,t}(\widehat{\xi}_{it} + \widehat{\epsilon}_{2it})K^2_{it}\Big)^2}_{\text{Timing Moments}} \quad (1.5.7)$$

I minimize this objective function to estimate the unknown functions. The estimation of $\widehat{c}_2(W_{it-1})$ and $\widehat{c}_3(W_{it-1})$ are computationally simple because they can be partialed out for a given $(\widehat{f}, \widehat{\widetilde{h}})$. To estimate the $\widehat{f}$ and $\widehat{\widetilde{h}}$, I minimize the objective function after partialing out the control functions. After obtaining the estimates for f and $\widehat{\widetilde{h}}$, I calculate the output elasticities as described in Equations (1.4.5), (1.4.6) and (1.4.9).

Deriving the large sample distribution of the output elasticities and other estimates used in the empirical applications is difficult. First, I need to account for estimation error in the first stage, and then I need to understand how estimation errors in the output elasticities translate into further stages. To avoid these complications, I use the bootstrap to estimate standard errors. The bootstrap procedures treat firms as independent observations and resample firms with replacement. See Appendix A.1.7 for more details on the estimation procedure.

1.5.2 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, which comes from firms' financial statements. For other countries, I use plant-level production datasets. The sample periods are given in Table 1.1, 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.

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.[51] 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.[52] Finally, the Indian data come from the Annual Survey of Industries conducted by the Indian statistical institute for plants with 100 or more employees.[53]

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 Appendix A.1.

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.[54] 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 Appendix A.1.4

Some concerns about Compustat data are worth mentioning. 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. Also, some of firms in Compustat are multinational companies. Second, firms drop out of the sample due to mergers and acquisitions and enter the sample as they become public. Finally, the data is compiled from accounting data, which is low-quality compared to traditional manufacturing censuses. Accounting variables

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

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

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

[54] The total input expenditure in Compustat data bundles labor and materials cost.

do not usually correspond to the inputs used by the firm, and their definition are subject to change over time as reporting practice in Compustat changes.

The concerns on Compustat cast doubt on the suitability of Compustat data for production function and estimation.[55] 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. 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.

1.6 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. Finally, I find that heterogeneity is largely explained by across-firm variation.

1.6.1 Output Elasticities

Table 1.2 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.[56] This finding reflects the advantage of using the firm's first-order conditions in production function estimation. 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.

[55] See Basu (2019) for a discussion on using Compustat data for markup estimation.

[56] I will discuss this point in Section 1.8 when I report markups 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 1-1.[57] The results suggest that I estimate a higher output elasticity of capital and lower elasticity of labor in all countries. The difference is statically 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.[58,59]

After showing significant differences in the average elasticities, I next ask whether elasticities are heterogeneous across firms.

1.6.2 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.

To measure heterogeneity, I estimate the coefficient of variation (CV) of the output elasticities within each industry-year group. Figure 1-2 displays the average and 10-90th percentiles of the CV estimates for all countries.[60] 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. Importantly, the 10-90th percentiles show that this result is not driven by only a small number of industries. Also, I find almost no heterogeneity in returns to scale. This is reasonable because existence of firms with too large or too small returns to scale would not be consistent with the economic theory and existing empirical evidence. Finally, note that the coefficient of variation equals zero under the Cobb-Douglas assumption as output elasticities are constant. Therefore, the commonly used Cobb-Douglas specification cannot capture heterogeneity in output elasticities.

The presence of heterogeneity in production technology is an important finding. Yet, it leads to a more interesting question: what explains this heterogeneity? Although the

[57]Other elasticity estimates are reported in Appendix Figure A-3.

[58]My estimates also suggest significantly lower returns to scale than the Cobb-Douglas estimates.

[59]A common concern in production function estimation is measurement error in capital, which is more important in a nonparametric model. If capital is measured with error, the capital elasticity estimates will be biased towards zero, and other elasticities will be biased upwards since they are usually positively correlated with capital. I verify this prediction using a simulation exercise in Section A.7.1. 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.

[60]The coefficient of variation is defined as the ratio of the standard deviation to the mean. It is robust to the scale of the variable and facilitates comparison.

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 test these two hypotheses.

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

$$d_{ijt} = \alpha_0 + \gamma \times \text{Firm Size}_{ijt} + \delta_{jt} + \epsilon_{it}, \tag{1.6.1}$$

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 as log capital elasticity divided by labor elasticity. I use log-sales to proxy for firm size. Note that even though production function estimation is at the 3-digit level, I use 4-digit industry-year interaction fixed effects to control for industry-specific variations.

Table 1.3 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 is similar when I use the capital elasticity as the outcome variable, which removes the effects of labor elasticity estimates. 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 are more capital intensive, use more capital, and use less labor and flexible input.

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.[61] The coefficient of interest reflects the average difference of the outcome variable between exporters and non-exporters. Table 1.4 presents the coefficient estimates, which 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 a validation exercise for my output elasticity estimates 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 other empirical evidence.

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

1.6.3 Persistence in Output Elasticities

The heterogeneity documented in the previous subsection can come from two sources: within-firm or cross-firm heterogeneity. We should expect a lower contribution from within-firm heterogeneity as firm-specific technology is likely to persist over time. In this section, I study within-firm heterogeneity by estimating the firm-level CV and running AR(1) regressions.

The first analysis estimates the CV of the flexible input elasticity for each firm and compares its distribution with the unconditional CV. Figure 1-3 displays the distribution of flexible input elasticity CV for all countries and the blue line shows the unconditional CV.[62] I find that in all countries, the CV at the firm-level is small and less than 0.05 for most firms, suggesting that the firm's flexible input elasticity does not change considerably over time. Also, the CV is higher than the unconditional CV for only a small number of firms. These results confirm that within-firm variation is small relative to cross-firm variation.

The second analysis estimates AR(1) regressions separately for each output elasticity with industry-year interaction fixed effects. Results reported in Table 1.5 show that all output elasticities are persistent, as indicated by the coefficient estimates that are close to one. These results provide further evidence that technology at the firm-level is persistent.

1.7 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. 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 markups from production data under the assumption that firms are cost-minimizers with respect to at least one flexible input. To illustrate this approach, consider a generic production function $F(K_{it}, X_{it}^1, \ldots, X_{it}^V)$, where X_{it}^V is a flexible input. If the firm is a price-taker in the input markets, first order conditions of cost minimization gives $\lambda_{it} F_V = p_{it}^V$. Here, λ_{it} denotes the Lagrange-multiplier on the output constraint, which equals the marginal cost, and p_{it}^V denotes the price of X_{it}^V. Therefore, the firm sets the marginal product of X_{it}^V, multiplied by marginal cost, equal to input price. Rearranging terms and some algebra yields

$$\underbrace{\frac{\partial F_t}{\partial X_{it}^V} \frac{X_{it}^V}{F_t}}_{\text{Elasticity}} = \underbrace{\frac{p_{it}}{\lambda_{it}}}_{\text{Markup}} \underbrace{\frac{p_{it}^V X_{it}^F}{p_{it} F_{it}}}_{\text{Revenue Share}},$$

where p_{it} denotes the output price. This expression specifies a simple relationship between the markup, output elasticity of X_{it}^V, and its revenue share:

$$\mu_{it} := \frac{\theta_{it}^V}{\alpha_{it}^V}, \tag{1.7.1}$$

[62] The results for the other elasticities are reported in Appendix A.8.

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 (1.7.1) holds for all flexible inputs, we only need to estimate the output elasticity of one flexible input.

In recent years, estimating markups from production data has become popular for a number of reasons.[63] First, it does not require a model of competition. Second, the data requirement is low compared to estimating markups from demand data as firm-level production data is sufficient. Third, firm-level markups can be used to investigate a range of research questions. For example, one can aggregate firm-level markups to study market power, or estimate a cross-sectional regression to understand the link between markups and other economic variables.[64]

1.7.1 How Does the Form of the Production Function Affect Markup Estimates?

Output elasticity is the only estimated component of markup in Equation (1.7.1). Therefore, when the production approach is used for markup estimation, the bias in output elasticity estimates has first-order effects on markups. This fact makes the markup estimate sensitive to the form of the production function.[65] In this section, I provide some examples of how the commonly used functional forms of production functions affect markups.

Heterogeneity in Markups. Much of the empirical research estimating markups assumes a Cobb-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. For a demonstration, consider the logarithm of markups under the Cobb-Douglas specification:

$$\log(\mu_{it}) = \log(\theta_t) - \log(\alpha_{it}),$$

where the industry-specific output elasticity, θ_t, shifts only the location of the log markup distribution, with no effect on its shape. 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 in production technology. In this case, production function estimation is redundant.

[63]See, for example, De Loecker et al. (2016), Autor et al. (2019), Traina (2018), and De Loecker and Scott (2016).

[64]For the former see De Loecker et al. (2018), Diez et al. (2018) and for the latter see De Loecker and Warzynski (2012).

[65]Van 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.

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.[66]

To see why the standard models are inadequate, note that identical markups from labor and materials under the Cobb-Douglas assumption imply that $\theta_t^L/\theta_t^M = \alpha_{it}^L/\alpha_{it}^M$. That is, all firms in the same industry have the same ratio of revenue shares of labor and materials, a prediction that is strongly rejected in all commonly used production datasets. Furthermore, using a more flexible Hicks-neutral production function, such as translog, does not address the problem.[67] In that case, the output elasticities depend only on observables, which implies that observables explain all the variation in the ratio of revenue shares.

1.7.2 Labor-augmenting Productivity Gives Identical Markup Estimates

The discussion in the previous section suggests 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 1.4.2. The latter immediately implies that the two markups estimates are the same:

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

where μ_{it}^L and μ_{it}^M denote markup estimates from labor and materials, respectively.[68]

[66]To 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.

[67]Raval (2019a) compares markups estimated from the translog production function and finds similar results obtained from the Cobb-Douglas estimates.

[68]Doraszelski and Jaumandreu (2019) also make this observations.

The role of labor-augmenting technology is subtle. In a Hicks-neutral production function, the ratio of the flexible input elasticities is over-identified. One source of identification is production function estimation, and the other source is first-order conditions. Getting different estimates from these sources rejects the model. By introducing unobserved labor-augmenting productivity, I make the model exactly-identified. The elasticities are not identified from the production function, which allows me to use first-order conditions. Therefore, the model becomes exactly-identified. 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.

1.7.3 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. Aggregate markup is useful because it can be considered as a measure of market power in the economy.[69] 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 be able to implement the Olley-Pakes decomposition, I look at the logarithm of markup, which equals the difference between the log output elasticity and log revenue share, $\log(\mu_{it}) = \log(\theta_{it}) - \log(\alpha_{it})$. 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-

[69] See Berry et al. (2019) for a discussion of market power interpretation of markups.

Pakes decomposition to both terms to obtain:

$$\tilde{\mu}_t = \underbrace{\underbrace{\bar{\theta}_t}_{\text{Average Elasticity (1)}} + \underbrace{\text{Cov}(w_{it}, \log(\theta_{it}))}_{\text{Heterogeneity in Technology (2)}}}_{\text{Estimation}} - \underbrace{\underbrace{\bar{\alpha}_t}_{\text{Average Share (3)}} - \underbrace{\text{Cov}(w_{it}, \log(\alpha_{it}))}_{\text{Heterogeneity in Shares (4)}}}_{\text{Data}} \tag{1.7.2}$$

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.

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.[70] 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.

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

[70] It 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.

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.

1.8 Empirical Results: Markups

I estimate markups using the output elasticities reported in Section 1.6. 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.[71] My aggregate markup estimates are lower than the Cobb-Douglas estimates in all countries. To explain the discrepancy, I use the markup decomposition and find that two factors drive this different: (1) Cobb-Douglas overestimates the average output elasticity heterogeneity, 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 in markups in the US. I find that the markup growth is lower according to my estimates.

1.8.1 Testing the Cobb-Douglas Specification using Markups

As discussed in Section 1.7, 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 1-4 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):

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

[71]This section mainly focuses on the comparison with Cobb-Douglas since it is the most commonly used specification. In Appendix A.7, I compare my results with the translog production function with Hicks-neutral productivity and in Subsection 1.8.4) with Nested CES production function with labor-augmenting productivity.

Here, v_{it} is the combined flexible input of labor and materials, defined as the deflated sum of labor and materials cost, and β_v corresponds to the flexible input elasticity. Having 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. I estimate this model using the ACF method and calculate markups as

$$\mu_{it}^{CD} = \frac{\beta_v}{\alpha_{it}^V}.$$

Since labor and materials generate identical markup estimates in my model, I use the flexible input elasticity divided by flexible input's revenue share as my markup measure,

$$\mu_{it} = \frac{\theta_{it}^L + \theta_{it}^M}{\alpha_{it}^V}. \tag{1.8.1}$$

This markup measure equals $\theta_{it}^L/\alpha_{it}^L$ and $\theta_{it}^M/\alpha_{it}^M$, the markups obtained from labor and materials elasticities.

1.8.2 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 1.7.3 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.[72] Figure 1-5 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.[73]

What drives these differences in markups estimates? I answer this question by decomposing markups into its four components, as presented in Section 1.7.3. These components, averaged over time, are presented in Figure 1-6. The red and white bars come directly from the data, and their magnitudes are the same for both estimation methods. Therefore, markup estimates between the two methods differ only through the first and second

[72] Edmond 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 Appendix A.8 and find qualitatively similar results.

[73] Appendix Figure A-14 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.

components (grey and black bars in Figure 1-7). The largest difference is in the second component, the covariance between firm size and elasticity. While this component is negligible under the Cobb-Douglas assumption, my estimates suggest that it is negative, and therefore, contributes negatively to the aggregate markups. This is not surprising because both the literature and my analysis in Section 1.6 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:

$$\tilde{\mu}_t^{CD} - \tilde{\mu}_t = \underbrace{\bar{\theta}_t^{CD} - \bar{\theta}_t}_{\text{Mean-Elasticity}} + \underbrace{\mathrm{Cov}(w_{it}, \log(\theta_{it}^{CD})) - \mathrm{Cov}(w_{it}, \log(\theta_{it})),}_{\text{Cov-Elasticity}} \tag{1.8.2}$$

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 1-7. 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 Cobb-Douglas markup estimates.

Within Industry and Between Industry Decomposition

In this section, I further decompose the covariance between firm size and output elasticity into within- and across-industry covariance. If across-industry covariance is important, estimating industry-specific production functions can account for heterogeneity in technology. However, if within-industry covariance is important, one needs a flexible production function that models within-industry heterogeneity. I use the law of total covariance for this decomposition and write $\mathrm{Cov}(w_{it}, \theta_{it})$ as

$$\mathrm{Cov}(w_{it}, \theta_{it}) = \underbrace{\mathbb{E}\big[\mathrm{Cov}(w_{it}, \theta_{it}) \mid \text{Industry}\big]}_{\text{Within Industry}} + \underbrace{\mathrm{Cov}\Big(\mathbb{E}\big[w_{it} \mid \text{Industry}\big], \mathbb{E}\big[\theta_{it} \mid \text{Industry}\big]\Big)}_{\text{Between Industry}}.$$

Figure 1-8 shows the two components of the covariance estimated using my method.[74] The results indicate that a large fraction of heterogeneity comes from within-industry variation rather than from cross-industry variation. A key implication of this result is that the standard practice of estimating industry-specific production functions is not enough to capture the effects of heterogeneous production technology on markups.

[74]By construction, the within-industry component equals zero under the Cobb-Douglas assumption. Therefore I do not report the results from the Cobb-Douglas specification.

1.8.3 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.

Variance Decomposition of the Aggregate Markups

I implement the standard variance decomposition analysis. In particular, I decompose the variance of the aggregate log-markup into the variance of (1)+(2) and variance of (3)+(4) in Equation (1.7.2), ignoring the covariance between the two. Figure 1-9 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; production function estimation is not needed.[75] Is the role of change in technology really minimal? For the rest of this section, I seek to answer this question.

Change in Markups in the US Manufacturing Sector

This section investigates the evolution of the aggregate markup in the US manufacturing sector. Figure 1-10 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.

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 1-10 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 sample size of Compustat changes with mergers and

[75] This is also evident in Appendix Figure A-13, which displays the evolution of markups along with its two components. The aggregate markups closely track the revenue share in all countries.

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.

To explore the importance of weighing in aggregation, Appendix Figure A-8 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. This could be explained by firms with higher markups having higher market shares when measured by sales, keeping everything else the same. Thus, weighting by sales artificially puts more weight on firms with high markups, causing a double-counting problem. This issue affects both the level and the change in markup.

Comparison to De Loecker et al. (2018)

Finally, 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. Although the sample does not cover the entire US economy, the manufacturing industry is large enough that the evidence is informative about the US economy.

Figure 1-11 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.

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 share.

One remaining question is where this difference comes from. As I argued in my decomposition exercise in Section 1.7, 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 1-12. 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.

To compare these results with the evolution of shares, I report the average flexible input's share and the covariance between flexible input's share and firm size in Appendix Figure A-15. These time series directly come from the data and are not affected by estimation. The figure shows a significant variation in both series over time. The average share of the flexible input has been increasing since the 1980s. The covariance is always negative, but we see some fluctuation. From this comparison, I conclude that it would be unreasonable to observe a large fluctuation in the revenue shares but to observe a small and insignificant change in the elasticities. These findings support my hypothesis that the flexible production function captures changes in production technology and heterogeneity in production technology.

1.8.4 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 (1.2.7). The details of the estimation are provided in Appendix A.6.2.

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 A-16 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 A-17. 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 A-18 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.

1.9 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 Appendix A.2.

1.9.1 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.[76] To accommodate this case, I develop an extension in Appendix A.2.1, 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 are the same. I incorporate two important modifications to the model. First, with heterogeneous input prices, input demand functions depend on input prices. So demand functions in Equation (1.2.2) and Equation (1.2.8) take input prices as arguments. Second, I modify Assumption 1.2.2 so that productivity shocks and input prices jointly follow a first-order Markov process. The main difference in the implementation is accounting for input prices when constructing the control variables. In particular, the control variables become

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

where $\bar{p}_{it}$ denotes the ratio of input prices and $\tilde{p}_{it}$ denotes the input price vector. These control variables lead to the following representations of productivity shocks

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

Note that incorporating input prices into the model breaks the multicollinearity problem discussed in Subsection 1.4.1. Therefore, even if input prices are endogenous, they can aid identification.

1.9.2 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

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

quality inputs at different prices. In 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. The key in this extension is to show that unobserved materials-augmenting 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 Appendix A.2.2.

1.9.3 Accounting for Firm Selection

In Appendix A.2.3, 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 1.3 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.[77] Using these assumptions, I rely on Olley and Pakes (1996)'s insight that there is a cutoff in Hicks-neutral productivity and firms that draw Hicks-neutral productivity below that cutoff exit.

Incorporating firm exit requires estimating the control variables conditional on firms that stay in operation. To achieve this, I estimate the propensity of exit conditional on the previous period's inputs and current period's capital level. Using this propensity score, I estimate the distribution of the innovations to productivity shocks conditional on firms that stay in operation, which allows me to control for selection. The empirical results from implementing this selection correction procedure are provided in Appendix A.7.3.

1.10 Conclusion

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.

[77] Modeling exit based on two productivity shocks is a difficult problem, which requires accounting for the firm's dynamic problem and understanding how labor-augmenting productivity affects the firm's dynamic problem. I leave this as future work.

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.

Several follow-up extensions emerge from this paper to understand the issues left open. For instance, a critical assumption for identification is that labor is a flexible input, an assumption that might be difficult to justify in some cases. Therefore, allowing dynamic labor is a particularly interesting direction. Another interesting extension is incorporating the demand side of the market, about which this paper is agnostic. This extension would allow us to accommodate a richer set demand model and makes the welfare implications of markups more transparent. Finally, on the empirical side, future research could use a better quality of data for the US, such as the US Census of Manufacturers, which would address the caveats with the Compustat dataset. I leave these extensions for future work.

1.11 Tables and Figures

Table 1.1: 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.

Table 1.2: Sales-Weighted Average Output Elasticities for Three Largest Industries

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

Note: Comparison of sales-weighted average output elasticities produced by different methods. FA refers to my estimates, ACF refers to Ackerberg et al. (2015) estimates and OLS is Cobb-Douglas estimated by OLS. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. Numbers in each panel correspond to the SIC code of the largest, second largest and third largest industries, respectively, in each country. Industry codes are provided in parentheses in each panel. Corresponding industry names are Food Manufacturing (311), Equipment Manufacturing (381), Paper Manufacturing (322), Glass Manufacturing (311), Cotton ginning (230), Textile (265). Bootstrapped standard errors in parentheses (100 iterations).

Table 1.3: Regressions of the Output Elasticities on Firm Size

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

Notes: Regressions of firm size on the output elasticities and capital intensity controlling for 4 digit industry-year fixed effects based on Equation (1.6.1). 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.

Table 1.4: Regression of Capital Intensity on Export Status

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

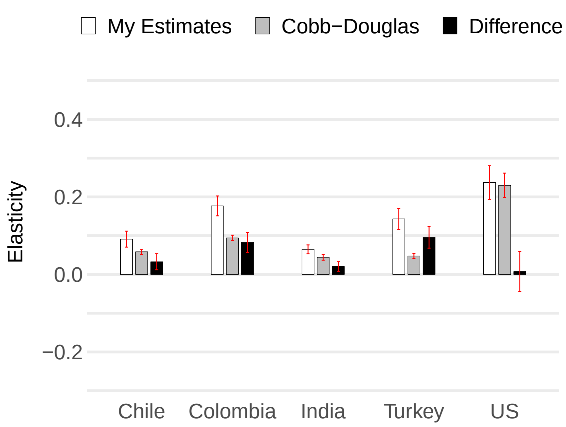
Notes: Regressions of capital elasticity on a dummy of whether the firm exports, controlling for 4 digit industry-year fixed effects. Each row corresponds to a separate regression where the left-hand side variable is given in the first column. Standard errors are clustered at the firm level and reported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

Table 1.5: AR(1) Regression of Output Elasticity

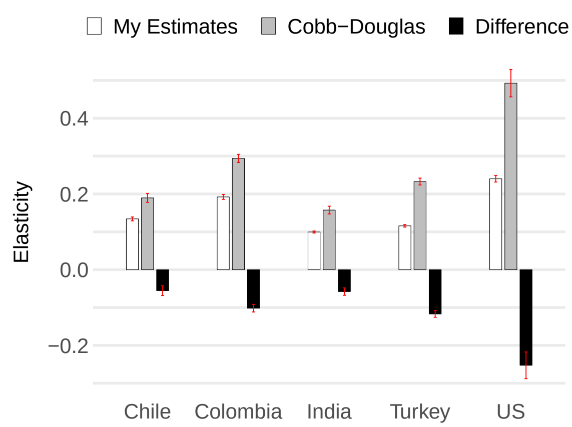
	Chile	Colombia	India	Turkey	US
Capital Elasticity	0.91	0.98	0.88	0.95	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Labor Elasticity	0.75	0.77	0.88	0.77	0.85
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Materials Elasticity	0.73	0.74	0.88	0.78	0.85
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Returns to Scale	0.78	0.82	0.71	0.85	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: Estimates from AR(1) regression for each output elasticity and returns to scale controlling for 4 digit industry-year fixed effects. Each row corresponds to a separate regression. Standard errors are clustered at firm level and reported in parentheses.

Figure 1-1: Average Capital and Labor Elasticities Comparison



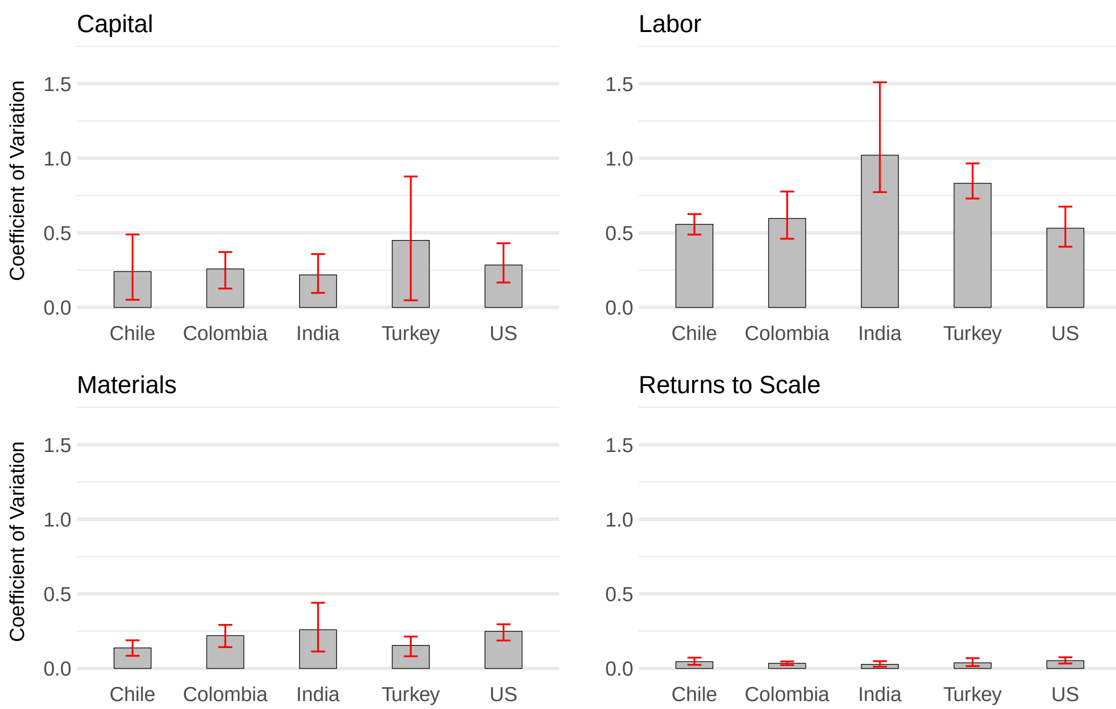
(a) Capital Elasticity



(b) Labor Elasticity

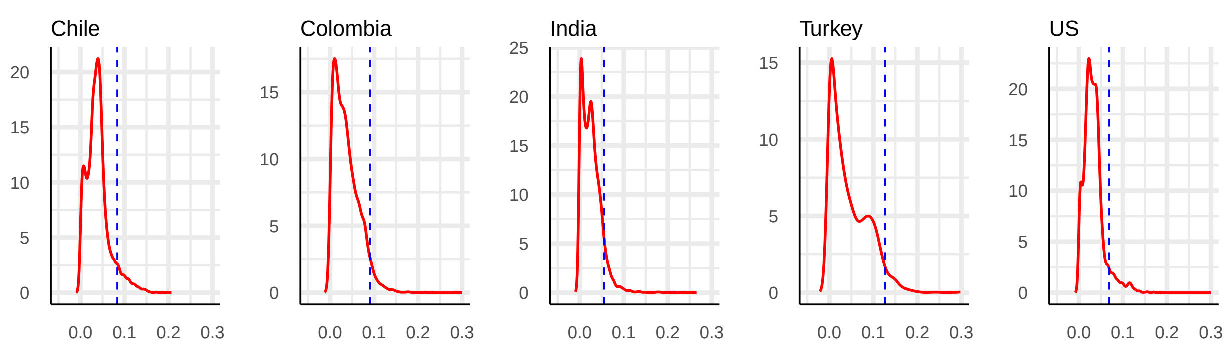
Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages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

Figure 1-2: Average Coefficient of Variation



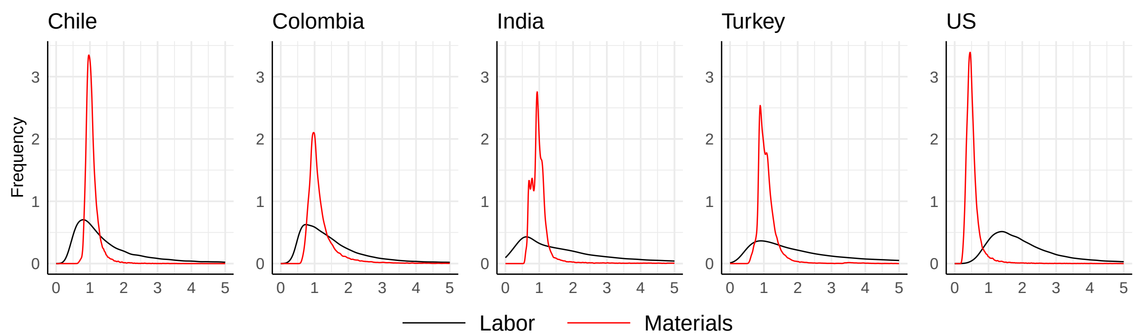
Note: This figure shows the average coefficient of variation for the output elasticities averaged across industries over years. In each panel, each bars reports the average CoV of the output elasticity of the corresponding input for all countries. The error bars indicate the 10th and 90th percentile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

Figure 1-3: Distribution of Coefficient of Variation of Sum Elasticity



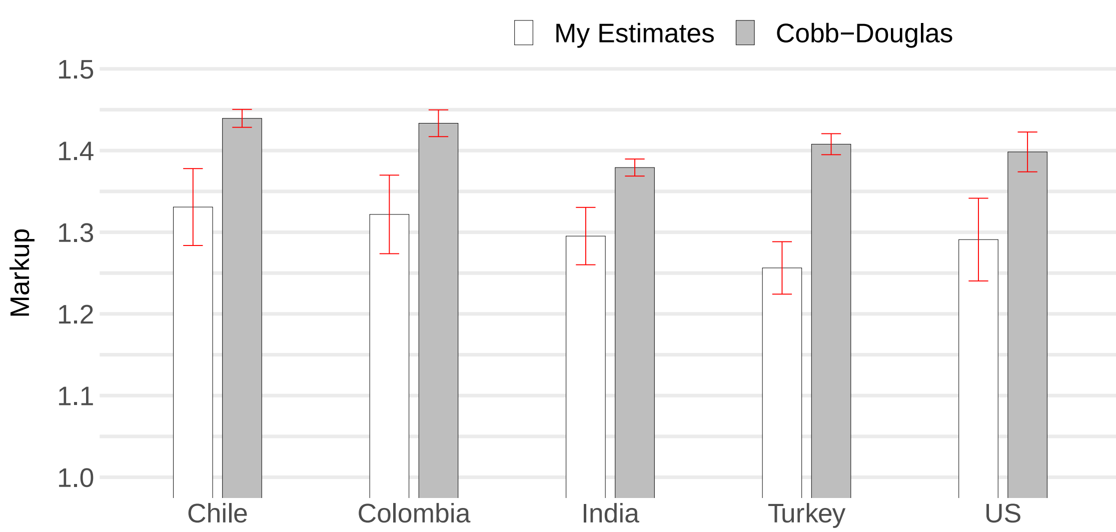
Notes: This figure compares the distribution of coefficient of variation of the sum elasticity within firm (red) with the estimates of unconditional coefficient of variation in the entire sample (blue) for each country.

Figure 1-4: Distribution of Markups Implied by Labor and Materials (Cobb-Douglas)



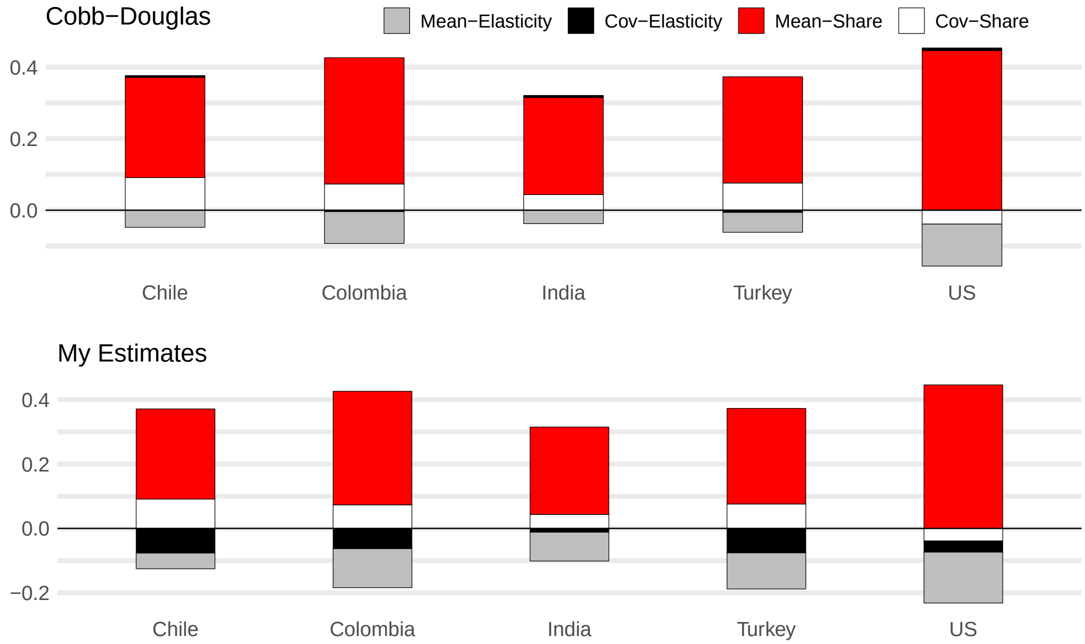
Notes: This figure compares the distribution of markups implied by labor (black) and materials (red) elasticities from the Cobb-Douglas specification estimated using the Ackerberg et al. (2015) procedure for each country. Each plot shows the result for the country given in the top-left corner.

Figure 1-5: Average Markups Comparison



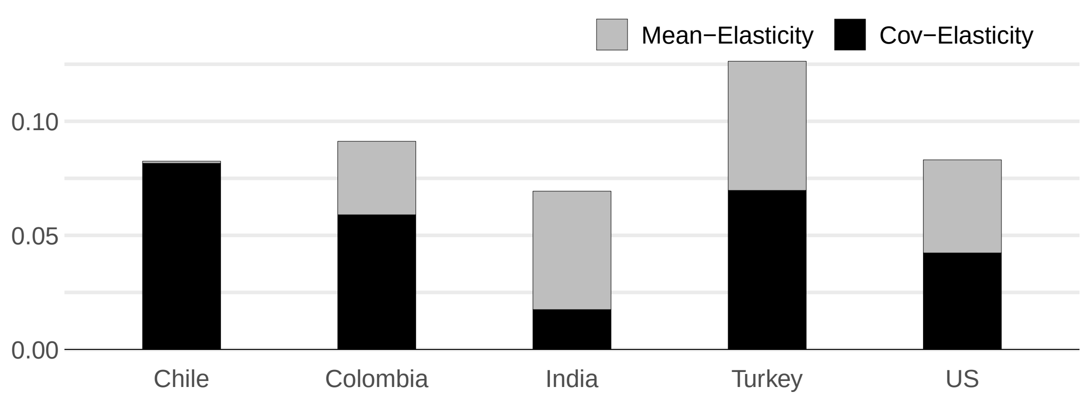
Notes: Comparison of sales-weighted average markups produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages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

Figure 1-6: Markup Decomposition



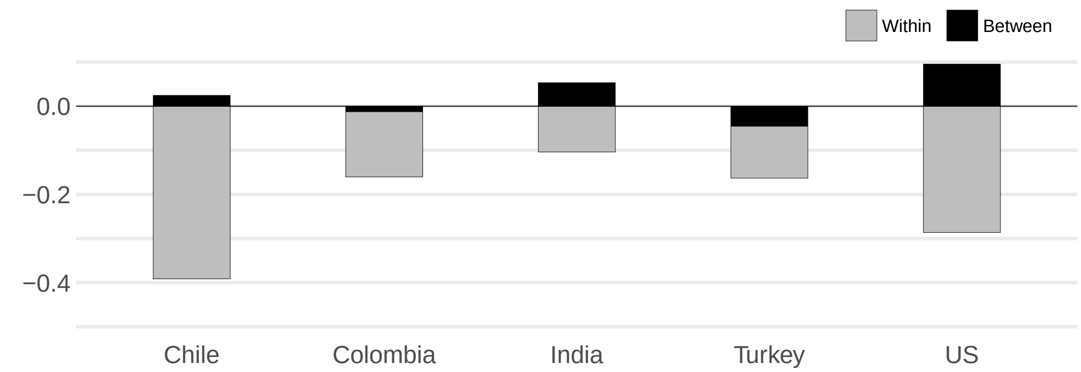
Notes: This figure compares the four components of the aggregate log-markups given in Equation (1.7.2) produced by my method and Cobb-Douglas estimated with ACF procedure. For each country, each component is averaged over years and indicated by a different color.

Figure 1-7: Decomposition of the Difference between Aggregate Markups



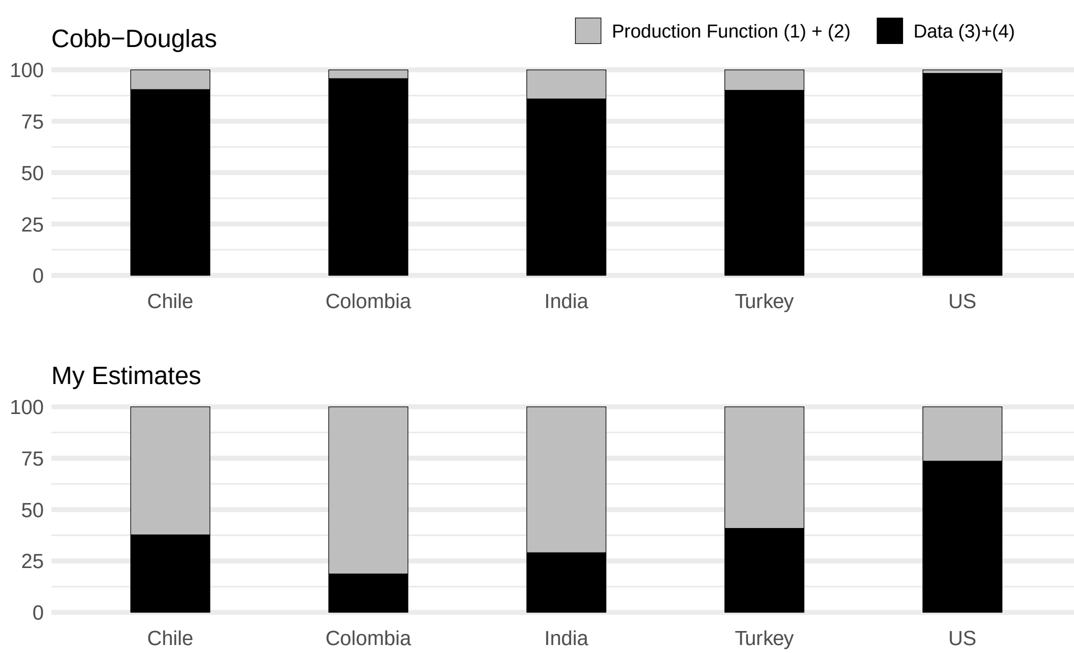
Notes: This figure decomposes the difference between the aggregate log markups produced by my method and the Cobb-Douglas model estimated using the ACF procedure (Equation 1.8.2).

Figure 1-8: Within and Between Industry Decomposition



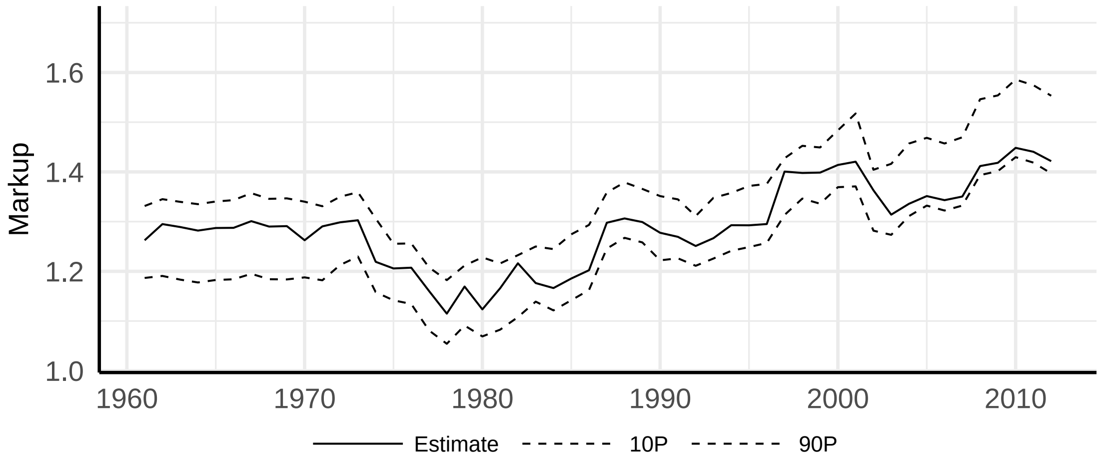
Notes: This figure decomposes the covariance between log elasticity and firm size into within- and between-industry covariance.

Figure 1-9: Variance Decomposition of the Change in Markups



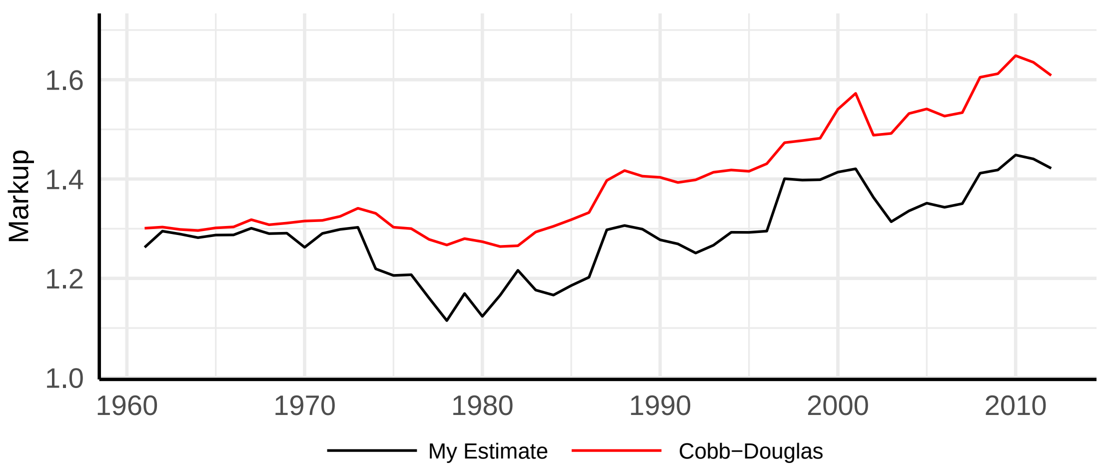
Notes: This figure shows the results by decomposing the annual aggregate log markups time series into the components obtained from elasticities (gray) and revenue shares (black). The covariance between the two components are subtracted from the total variance so that the two components sum to 100.

Figure 1-10: Evolution of the Aggregate Markup



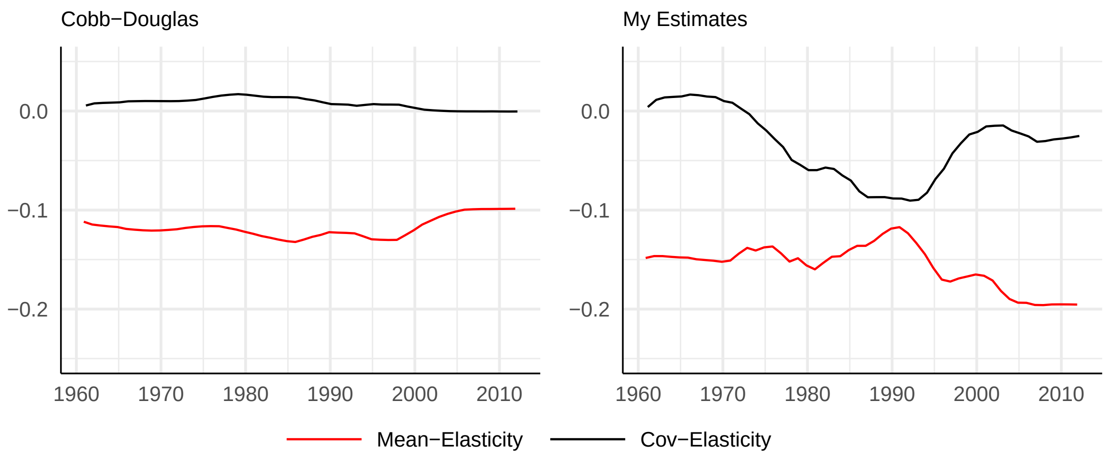
Notes: The evolution of markups in the US manufacturing industry. The dotted lines report the 10-90the percentile of the bootstrap distribution (100 iterations).

Figure 1-11: 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.

Figure 1-12: 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.

Chapter 2

Production Function Estimation with Imperfect Proxies

2.1 Introduction

Production functions are a critical input in many economic studies. These studies typically require estimating a production function using firm-level data. A major challenge in production function estimation is the endogeneity of inputs. Firms observe their productivity before choosing production inputs; however, productivity is unobservable to the researcher. This results in inputs being correlated with productivity, a standard endogeneity problem.

A commonly used method to address the endogeneity problem is the proxy variable approach. Introduced by Olley and Pakes (1996) (henceforth OP), this approach relies on using a variable, called proxy, to control for unobserved productivity. OP use investment as a proxy variable, which is assumed to be a strictly increasing function of productivity conditional on capital. By inverting this unknown function, they essentially recover the productivity shock, and control for it in the estimation. The proxy variable approach has become the workhorse for estimating production functions and has been extended by several papers. Levinsohn and Petrin (2003) (LP) have proposed using materials as a proxy, and Ackerberg et al. (2015) (ACF) have introduced a unified framework of proxy variable approach that deals with some practical concerns.

A limitation of proxy variable approach is that it relies on strong assumptions, such as single-dimensional unobserved heterogeneity and strict monotonicity. These assumptions have important economic implications, as observed by others (Ackerberg et al. (2007), Ackerberg et al. (2015)). First, firms are differentiated only by a single productivity shock, which restricts firm-level heterogeneity. Second, there is no heterogeneity in adjustment costs and investment prices, as investment depends only on productivity. Third, estimation requires restricting competition in the output market. Moreover, the proxy variable approach is not robust to measurement errors in inputs, an important concern, especially for capital.

In this paper, I develop a partial identification approach that is robust to certain deviations from proxy variable assumptions and measurement errors in inputs. In particular, my model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates a more general timing assumption, and (iv) is robust to measurement errors in all inputs. With these changes, the standard proxy variable becomes an 'imperfect proxy,' which can be used to derive moment inequalities for identification. Using these moment inequalities, I characterize the identified set for the parameters and propose an estimator.

An 'imperfect proxy' variable contains information about productivity, but it cannot be directly used to control for productivity in estimation. Instead, an imperfect proxy gives a stochastic ordering of productivity distributions, which can be used for identification. To show this result, I first group firms into 'high' investment firms, firms that invest more than a cutoff value, and 'low' investment firms, firms that invest less than the cutoff value. Then, I show that the productivity distribution of 'high' investment firms first-order stochastically dominates the productivity distribution of 'low' investment firms. The main idea for identification is to use this stochastic ordering in the form of moment inequalities to obtain bounds for production function parameters.

I derive moment inequalities and study identification under a wide range of assumptions. The first identification result relies only on the assumption that productivity shocks follow an exogenous Markov process. This is the least restrictive specification, and therefore, gives the widest bounds. The other identification results exploit modeling assumptions fully to derive moment inequalities via the imperfect proxy. These moment inequalities give an identified set for the main specification of the paper. I also show how to tighten the identified set under additional distributional assumptions and shape restrictions. Analyzing identification under a wide range of assumptions makes the role of each assumption in identification transparent. For example, one can start with the most general model to impose as few restrictions as possible. If the estimated set is not informative, then a nested model can be considered to shrink the identified set. Also, comparing the results from a nested model and a general model tests the restrictions imposed by the nested model.

The partial identification approach allows me to have a model with rich heterogeneity. My model includes two productivity shocks, one persistent and the other transitory. The firm can observe both of these shocks, so both can create endogeneity. Moreover, my model includes unobserved variables that affect the firm's choice of investment. Consequently, it allows for heterogeneity in input prices and adjustment costs as well as demand shock in the output market. Finally, the identification approach is robust to measurement errors in all inputs. This robustness is particularly crucial for capital, which is most prone to measurement error.

My method is generic in that it applies to production functions under different specifications. First, one can use my method to partially identify the parameters of both value-added or gross Cobb-Douglas production functions. Second, the model is agnostic about which proxy variable to use, so both investment and materials can serve as an imperfect proxy for estimation. Third, the model can accommodate different timing assumptions about capital. One can assume that capital is chosen one period in advance, as in prior approaches, or that firms choose capital after (partially) observing productivity shocks. Finally, the model is not specific to the Cobb-Douglas production function. A nonlinear production function that is known up to a finite-dimensional parameter vector can be considered.

This paper contributes to the large literature on production function estimation using proxy variables (OP, LP, ACF; Gandhi et al. (2018) (GNR)).[1] OP find the conditions under which investment can be used as a 'proxy' to control for unobserved productivity. Motivated by 'zero' and 'lumpy' investment problem, LP propose using materials as a proxy variable. ACF point out a collinearity issue in these papers and propose an alternative proxy variable approach that avoids the collinearity problem. My paper extends these approaches by showing how to make inferences when the standard proxy variable approach assumptions are violated.

A few recent papers study production function estimation with measurement errors in capital (Hu et al. (2011), Collard-Wexler and De Loecker (2016), Kim et al. (2016)). These papers require either an instrumental variable or another proxy variable to address measurement errors. In contrast, my method does not require an additional variable, but it gives a bound rather than a point estimate.

This paper is related to the literature on monotone instrumental variables (Manski (1997), Manski and Pepper (2000)). This literature assumes that the means of potential outcomes can be ordered conditional on an observed variable, which is called the monotone instrument. In my model, the monotone instrument corresponds to the indicator variable that specifies whether the proxy variable is greater than a cutoff. My approach differs from the monotone instrument variable approach in that the monotone instrument costructed from inside the model.

Notation. I use the notation $F_a(t)$ and $F_a(t \mid b)$ to donate the distribution of variable a and the distribution of a conditional on b, respectively. Similarly, I use $f_a(t)$ and $f_a(t \mid b)$ to denote the probability density function of random variable a and the probability density function of a conditional on b, respectively.

2.2 Model

In this section, I describe the production function model and assumptions. The model builds on the proxy variable framework introduced by OP, but allows for deviations from some of OP's assumptions. I discuss how my model differs from the proxy variable framework and the implications of the differences for identification.

[1] The 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 and Hoch (1965), Mundlak (1961)). However, in practice, these methods do not give satisfactory answers, as summarized by Griliches and Mairesse (1995). See also Blundell and Bond (2000).

2.2.1 Production Function

I consider a value-added Cobb-Douglas production function to demonstrate the main results of the paper.[2] The production function is given by

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \omega_{it} + \epsilon_{it}, \tag{2.2.1}$$

where y_{it} denotes log-output, k_{it} denotes log-capital, and l_{it} denotes log-labor input.[3] The model includes two unobserved productivity shocks, ω_{it} and ϵ_{it}. ω_{it} represents the persistent component of productivity; it is correlated over time. On the other hand, ϵ_{it} represents the transitory component of productivity; it is independently and identically distributed over time, and it does not provide information about future productivity. Firms observe ω_{it} before choosing inputs, whereas ϵ_{it} can be partially or fully observed. Therefore, both productivity shocks can be correlated with inputs and can generate endogeneity.

The data consists of a panel of firms observed over periods $t = 1, \ldots, T$. To simplify the exposition I assume that $T = 3$, which is sufficient for the identification results. Observations are independently and identically distributed across firms, but they can be serially correlated within the firm. The objective is to estimate the production function parameters, (θ_l, θ_k). Since inputs are endogenous, OLS estimation would give biased coefficient estimates.

I assume that capital is a dynamic input, meaning that the firm's current period capital level affects the firm's future production.[4] As a result, capital is a state variable in the firm's dynamic optimization problem. However, unlike the standard proxy variable framework, I do not assume that capital is a predetermined input. That is, capital may be chosen after persistent productivity, ω_{it}, is (partially) observed by the firm.[5] The model is agnostic about labor input, so it can be a dynamic or static input.

2.2.2 Assumptions

My assumptions follow the structure of the proxy variable approach assumptions but relax them in several ways. This section presents the assumptions and describes how they lead to a less restrictive model than the proxy variable model, in terms of their economic implications. The first assumption defines the firm's information set.

Assumption 2.2.1 (Information Set)**.** *Let $\mathcal{I}_{it}$ denote the firm i's information set at period t. I assume that past and current persistent productivity shocks are in firm's information set, that is, $\{\omega_{i\tau}\}_{\tau=-\infty}^{t} \in \mathcal{I}_{it}$. The transitory shocks satisfy $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$.*

---

[2] The identification strategy applies to other forms of production functions. I show how the model can be extended to other commonly used production functions in Section 2.5.

[3] Following the tradition, lowercase letters correspond to the logarithm of uppercase variables.

[4] This can happen, for example, due to adjustment cost in investment.

[5] It is important to note that my results do not rely on this assumption since the model can accommodate predetermined capital. This requires a minor modification in the estimation procedure, as discussed in Subsection 2.5.3.

This assumption distinguishes the roles of two productivity shocks. The persistent productivity, ω_{it}, is observed by the firm. The transitory productivity, ϵ_{it}, can be observed, partially observed, or not observed. I also assume that the transitory shock cannot be predicted by the firm in the sense that $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$. Note that since this includes mean independence, not full independence, the firm's dynamic decision can still be affected by ϵ_{it}, as the distribution of ϵ_{it} can give information about future production. The next assumption restricts the distribution of persistent productivity shock.

Assumption 2.2.2 (Markov Property)**.** *Persistent productivity shocks follow an exogenous first-order Markov process*

$$P(\omega_{it+1} \mid \mathcal{I}_{it}) = P(\omega_{it+1} \mid \omega_{it}),$$

and the distribution is stochastically increasing in ω_{it}*.*

This assumption is standard in the literature and states that the only information about future productivity (in the firm's information set) is given by current productivity. An implication of this assumption is that the transitory productivity shock ϵ_{it} is not informative about the distribution of future productivity shocks, once we condition on ω_{it}. However, this assumption does not restrict the correlation between two productivity shocks, as we might expect a positive correlation between them.

The second part of the assumption says that the distribution of future persistent productivity is stochastically increasing in ω_{it}. This assumption was first made by OP and indicates that more productive firms at the current period are likely to be more productive next period.[6] This assumption is critical for the moment inequality approach developed in this paper. Even though the literature makes this assumption, this is the first paper to use it for identification.

Assumption 2.2.3 (Capital Accumulation)**.** *Capital accumulates according to*

$$k_{it} = \delta k_{it-1} + i_{it}.$$

Capital is depreciated at the rate of δ, and firms make investments to accumulate capital.[7] An important feature of this assumption is that investment made at time t is productive immediately. Therefore, it makes capital a dynamic but not necessarily a predetermined input. This assumption relaxes the standard timing assumption, which assumes that capital is lagged by one period, that is, $k_{it} = \delta k_{it-1} + i_{it-1}$. This timing assumption is critical for identification in standard methods.

I am able to relax the timing assumption because my method uses an imperfect proxy instead of the perfect proxy. Also, my goal is partial identification, rather than point identification. I also emphasize that Assumption 2.2.3 is not required for my identification results,

[6] A distribution is stochastically increasing if $P(\omega_{it+1} \mid \bar{\omega}_{it})$ first-order stochastically dominates $P(\omega_{it+1} \mid \tilde{\omega}_{it})$ if and only of $\bar{\omega}_{it} \geqslant \tilde{\omega}_{it}$.

[7] My framework allows for a more complicated capital accumulation function; however, I do not consider it for simplicity.

and therefore, my method can accommodate the standard timing assumption. I show this as an extension in Section 2.5.

Assumption 2.2.4 (Investment Function)**.** *The firm's investment decision is given by,*

$$i_{it} = f_t(k_{it-1}, \omega_{it}, \xi_{it}),$$

where $\xi_{it} \in \mathbb{R}^L$ is a vector of unobserved random variables that affects firm's investment and it is assumed to be jointly independent of ω_{it} conditional on k_{it-1}.

According to this assumption, investment depends on the two standard state variables, capital, and persistent productivity, as well as other unobserved variables denoted by ξ_{it}. The unobserved vector ξ_{it} can include variables that affect the firm's dynamic problem, such as heterogeneity in adjustment cost, investment prices, or demand shocks.[8] As noted by ACF, OP do not allow heterogeneity in these variables, as the only unobservable that affects the firm's investment decision is assumed to be ω_{it}. Note that ξ_{it} can also include ϵ_{it}, the transitory productivity. According to Assumption 2.2.3 the firm chooses k_{it} at time t after observing the persistent productivity shock ω_{it}.

Since ξ_{it} can include demand shocks to produced goods, I do not restrict market power and competitive conduct. This is important because, in prior approaches, identification is possible only under particular competition assumptionsx'z' in the output market. For example, OP, LP, and ACF implicitly assume a perfect competition or monopolistic competition with identical demand curves. GNR consider a perfectly competitive market. This shows the importance of allowing multidimensional heterogeneity in a model of a firm to capture a richer competition structure. This is especially relevant when production function estimates are used for calculating markups (De Loecker et al. (2018)).

Another condition in Assumption 2.2.4 is that ξ_{it} is independent of persistent productivity conditional on the last period's capital. Therefore, given the firm's capital level, the variables that affect the firm's investment decision are not informative about the persistent productivity shock, ω_{it}. Although this assumption is restrictive, it allows for multi-dimensional heterogeneity in investment function and productivity. Some of the identification results presented later do not require this assumption, so it is still possible to make inferences on the parameters without this assumption. However, this assumption gives additional moment inequalities, which are likely to make the identified set tighter.

Finally, Assumption 2.2.4 also accommodates measurement error in capital, as one interpretation of ξ_{it} could be measurement error in investment. I discuss this point in Section 2.5.5, since measurement error in capital is an important concern in production function estimation.

[8] There is strong evidence for heterogeneity in adjustment cost. For example, Goolsbee and Gross (2000) present empirical evidence on heterogeneity in adjustment cost. Cooper and Haltiwanger (2006) argue that there is substantial heterogeneity in capital associated with heterogeneity in adjustment costs. Hamermesh and Pfann (1996), in a review paper, claim that heterogeneity in adjustment cost is a key source of heterogeneity across firms and should be included in models of firm behavior.

Assumption 2.2.5 (Imperfect Proxy)**.** *$f_t(k_{it-1}, \omega_{it}, \xi_{it})$ is weakly increasing in ω_{it} conditional on (k_{it-1}, ξ_{it}).*

This assumption relaxes the standard condition that investment is strictly monotonict in productivity. It instead assumes a weak monotonicity. Strict monotonicity is the key restriction in the proxy variable approach, which makes it possible to invert and essentially 'observe' the productivity using the proxy variable. Under my assumption, investment is no longer a 'perfect proxy' because the one-to-one relationship between investment and productivity disappears. However, by weak monotonicity, investment is still informative about productivity, so it becomes an imperfect proxy. My identification approach relies on capturing the information in an imperfect proxy variable via moment inequalities.

Relaxing strict monotonicity has important practical implications. As observed by LP and other papers, investment is often lumpy in the data. Moreover, several firms do not invest in some periods. This suggests that investment is not continuous in productivity. OP drop firms with zero investment to overcome this problem. LP propose using materials instead of investment as a proxy. My approach is robust to both zero and lumpy investments in the data.

Assumption 2.2.6 (Measurement Error in Labor)**.** *Labor is measured with error*

$$l_{it} = l^*_{it} + \eta^l_{it},$$

where l^*_{it} denotes the true labor input and η^l_{it} is measurement error in labor. I assume that measurement error, η^l_{it}, is orthogonal to the information set at $t-1$, $\mathbb{E}[\eta^l_{it} \mid I_{it-1}] = 0$.

This assumption addresses an important concern in production function estimation literature as production inputs are prone to measurement errors. My estimation method can accommodate measurement error in labor as long as it is orthogonal to the firm's information set $t-1$. Note that η^l_{it} is more general than the classical measurement error because it can be correlated with right-hand side variables. Also, this assumption nests the situation where labor does not have measurement error when we set $\eta^l_{it} = 0$.

2.2.3 Discussion

Overall, Assumptions 2.2.4 and 2.2.5 are the key differences of this paper from the standard assumptions, which assume that $i_{it} = f_t(k_{it}, \omega_{it})$ and $f_t(k_{it}, \omega_{it})$ is strongly increasing in ω_{it}. These assumptions limit the dimension of unobserved heterogeneity that impacts firm behavior. I relax these two strong assumptions on the functional form of investment by assuming that (i) investment is weakly increasing in productivity, and (ii) there are other unobservables affecting the investment decision. Under these assumptions, f_t is not invertible, which is the key step in OP to control for unobserved productivity. I deal with controlling for ω_{it} using moment inequalities. This allows me to have multi-dimensional unobserved heterogeneity.

2.3 Identification

This section derives a set of moment inequalities from the assumptions presented in the previous section. I derive moment inequalities and study identification under a wide range of assumptions. The first identification result relies only on the exogenous Markov assumption, so it is the least restrictive and gives the largest bound. Other identification results make use of other modeling assumptions and tighten the bounds.

2.3.1 Identification with Markov Assumption

In this section, I show that the Markov property of ω_{it} in Assumption 2.2.1 provides moment inequalities and set identifies the production function. This result relies on the following proposition.

Proposition 2.3.1. *Under Assumption 2.2.2 we have*

$$\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big]. \tag{2.3.1}$$

Proof. See Appendix A.4.

This proposition states that the difference between productivity shocks across two periods is smaller for the same firm than for two different firms. The key assumption to obtain this result is that conditional distribution of ω_{it} is stochastically increasing in ω_{it-1}. Therefore, firm i's current period productivity, $\omega_{it} + \epsilon_{it}$, is closer to firm i's previous period productivity than firm j's previous period productivity in Euclidean distance. Now, using the Cobb-Douglas functional form, I write the productivity shocks in Proposition 2.3.1 as

$$\Delta\omega_{it} + \Delta\epsilon_{it} = \Delta y_{it} - \theta_k \Delta k_{it} - \theta_l \Delta l_{it},$$
$$\Delta\omega_{ijt} + \Delta\epsilon_{ijt} = \Delta y_{ijt} - \theta_k \Delta k_{ijt} - \theta_l \Delta l_{ijt},$$

where I use $\Delta z_{it} := z_{it} - z_{it-1}$ and $\Delta z_{ijt} := z_{it} - z_{jt-1}$. Combining this with Proposition 2.3.1, I construct a moment inequality

$$\mathbb{E}\Big[\Delta y_{it} - \theta_k \Delta k_{it} - \theta_l \Delta l_{it}\Big] \leqslant \mathbb{E}\Big[\Delta y_{ijt} - \theta_k \Delta k_{ijt} - \theta_l \Delta l_{ijt})\Big], \tag{2.3.2}$$

which consists only of data and parameters, so it can be used for estimating bounds for the parameters. The result uses two assumptions: (i) Persistent productivity shock follows an exogenous Markov process, and (ii) Productivity shocks are additively separable in the production function. Therefore, moment inequalities are obtained under very general conditions. First, inputs could be dynamic or static. Second, we do not need to observe a proxy variable to control for productivity shocks. Finally, the variables that affect the firm's dynamic or static decisions are unrestricted. Of course, this flexibility might come with a cost, as the identified set might not be very informative.

In Proposition 2.3.1, I use the Euclidean distance to derive moment inequalities. However, one can consider other distance measures and obtain different moment inequalities. In that case, different distances would give different identified sets, which can be intersected to obtain tighter bounds.

2.3.2 Identification with Other Assumptions

This section derives a set of conditional moment inequalities based on all the modeling assumptions. The derivation relies on using investment as an imperfect proxy variable. The first step in constructing moment inequalities is stochastically ordering the productivity distributions that involve firms with 'high' and 'low' investment. Let $\tilde{\mathcal{X}} = \{(k, z) : 0 < \text{Prob}(i_{it} < z \mid k_{it-2} = k) < 1\}$. So $\tilde{\mathcal{X}}$ denotes the 'common support', which is commonly used in treatment effect estimation. To simplify the exposition, I also assume that i_{it} has a continuous distribution function. However, the results extend to the case where i_{it} has a discrete support.

Proposition 2.3.2. *Assumptions 2.2.1-2.2.6, along with some regulatory conditions, imply that*
(i) *For* $(k_{it-2}, z) \in \tilde{\mathcal{X}}$

$$\frac{f_{\omega_{it}}(t \mid k_{it-1}, i_{it} > z)}{f_{\omega_{it}}(t \mid k_{it-1}, i_{it} < z)}$$

is increasing in t, that is, it satisfies the Monotone Likelihood Ratio Property (MLRP).
(ii) *The distribution function of ω_{it} conditional on k_{it-1} and $\{i_{it} > z\}$ first order stochastically dominates (FOSD) the distribution function of ω_{it} conditional on k_{it-1} and $\{i_{it} < z\}$*

$$F_{\omega_{it}}(t \mid k_{it-1}, i_{it} > z) \geqslant F_{\omega_{it}}(t \mid k_{it-1}, i_{it} < z),$$

for $t \geqslant 0$ and $(k_{it-2}, z) \in \tilde{\mathcal{X}}$.
(iii) *The mean of ω_{it} conditional on k_{it-1} and $\{i_{it} > z\}$ is greater than the mean of ω_{it} conditional on k_{it-1} and $\{i_{it} < z\}$:*

$$\mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} > z] \geqslant \mathbb{E}[\omega_{it} \mid k_{it-1}, i_{it} < z], \tag{2.3.3}$$

for $(k_{it-2}, z) \in \tilde{\mathcal{X}}$.

Proof. See Appendix A.4.
This proposition shows that weak monotonicity of investment in ω_{it}, along with other assumptions, gives three stochastic orderings: (i) monotone likelihood ratio, (ii) first-order stochastic dominance, and (iii) mean ordering. The proof of this proposition shows that the weak monotonicity encompasses all the information given by the proxy variable. That is, MLRP holds if and only if investment is weakly monotone in productivity.

It is useful to compare the statements of this proposition with the invertibility condition in the proxy variable approach. When investment is invertible, and therefore is a perfect proxy, the ranking of the firm in investment equals to the ranking in productivity. This makes it possible to infer productivity using investment. In my model, investment is not a perfect proxy, so it is not possible to recover productivity from investment. However, by weak monotonicity, investment still provides information about productivity, so it becomes an imperfect proxy.

This proposition shows that an imperfect proxy can be used to order productivity stochastically, rather than deterministically. In particular, Proposition 2.3.2 says that when firms are grouped based on how much they invest, we can infer that high investment firms will be more productive than low investment firms, on average. The main idea for identification is to use these stochastic orderings in the form of moment inequalities to set identify the production function.

My first moment inequality derivation exploits the condition Proposition 2.3.2(iii). To see how to obtain a moment inequality, first note that the Markov assumption implies

$$\omega_{it} = g(\omega_{it-1}) + \zeta_{it},$$

where $\zeta_{it} = \omega_{it} - \mathbb{E}[\omega_{it} \mid \omega_{it-1}]$ and $\mathbb{E}[\zeta_{it} \mid I_{it-1}] = 0$ by construction. Also, the assumption that $P(\omega_{it} \mid \omega_{it-1})$ is stochastically increasing implies that $g(\omega_{it-1})$ is a monotone function. This representation of ω_{it} has been commonly used in the proxy variable approach for constructing moments. Substituting productivity into the production function yields:

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}. \tag{2.3.4}$$

This representation of the production function involves three error terms: innovation to productivity ζ_{it}, measurement error in labor η^l_{it}, and the transitory productivity shock ϵ_{it}. Let me define a function which takes data and parameter:

$$m(w_{it}, \tilde{\theta}) := y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} \tag{2.3.5}$$

with $w_{it} = (k_{it}, l_{it})$ and $\tilde{\theta} = (\tilde{\theta}_l, \tilde{\theta}_k)$. Also let θ denote the vector of true parameter values. The next proposition presents a conditional moment inequality using Equation (2.3.4) and Proposition 2.3.2.

Proposition 2.3.3. *For* $(k_{it-2}, z) \in \tilde{\mathcal{X}}$

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} < z] \geqslant 0. \tag{2.3.6}$$

Proof. Omitted.

This proposition is the main identification result of the paper. Conditional on k_{it-2} if we compare two groups of firms, one with investment greater than z and one with investment lower than z, Equation (2.3.6) is satisfied at the true parameter values. The key conditions needed for this proposition are monotonicity of $g(\omega_{it-1})$ and the weak monotonicity of investment in productivity.

A necessary condition for this proposition to hold is that $(\zeta_{it}, \eta^l_{it}, \epsilon_{it})$ are orthogonal to the firm's information set at $t-1$. Recall that Proposition 2.3.2(iii) provides moment inequalities in terms of ω_{it}. However, we can only recover $g(w_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}$ from the observed variables and parameters. Therefore, we need to account for $(\zeta_{it}, \eta^l_{it}, \epsilon_{it})$. The orthogonality condition allows me to achieve this, as $\zeta_{it} + \eta^l_{it} + \epsilon_{it}$ drop from the moment inequality in Equation (2.3.6) when we take conditional expectations.

Remark 2.3.1 (Comparison to Proxy Variable Approach)**.** Single dimensional unobserved heterogeneity and strong monotonicity of investment in productivity allow OP to invert the investment function and recover productivity shock as

$$\omega_{it} = f_t^{-1}(i_{it}, k_{it}).$$

An invertible investment function means that one can control for unobserved productivity by conditioning on observables. My model relaxes the two necessary conditions for invertibility. First, I allow the investment function to be weakly monotone in productivity. Second, there are additional unobserved variables affecting investment. Therefore, we can no longer compare the productivity levels of two firms by comparing their investments under my assumptions.

Remark 2.3.2 (Conditioning on Two Investment Levels)**.** One might think that a moment inequality similar to Equation (2.3.3) holds, conditional on two different investment levels:

$$\mathbb{E}[\omega_{it} \mid k_{it-2}, i_{it-1} = z_1] \geqslant \mathbb{E}[\omega_{it} \mid k_{it-2}, i_{it-1} = z_2],$$

where $z_1 < z_2$. However, this inequality does not hold, as it is easy to find counterexamples. Therefore, it is crucial to group firms using a cutoff value in investment.

Remark 2.3.3 (Relation to Monotone Instrument and Imperfect Instrument Literature)**.** This paper is related to the literature on monotone instrumental variables (Manski (1997), Manski and Pepper (2000)). This literature assumes that the mean potential outcomes are ordered based on an observed variable, which is called a monotone instrument. In my model, investment can be considered a monotone instrument for productivity. The main difference of my model from the standard monotone instrumental variable approach is that the monotone instrument comes from within the model in this paper. My approach is also related to the 'imperfect instrument approach,' which assumes that the researcher has some prior information about the correlation between the endogenous variable and unobserved heterogeneity. This information is then used to construct moment inequalities. See, for example; Nevo and Rosen (2012) and Conley et al. (2012).

2.3.3 Identified Set

In this section, I characterize the identified set using the derived moment inequalities. Since I have conditional moment inequalities, the identified set is given by intersection bounds.

Recall that the true parameter satisfies:

$$\mathbb{E}[y_{it} - \theta_k k_{it} - \theta_l l_{it} \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[y_{it} - \theta_k k_{it} - \theta_l l_{it} \mid k_{it-2}, i_{it-1} < z] \geqslant 0. \tag{2.3.7}$$

In what follows, I assume that all expectations are conditional on $(k_{it-2} = k)$, so I drop it from notation. Define the following variables indexed by z

$$I_{it}^h(z) := \mathbb{1}\{i_{it-1} > z\}, \qquad I_{it}^l(z) := \mathbb{1}\{i_{it-1} < z\}.$$

$I_{it}^h(z)$ equals 1 if a firm invests more than z at period t and zero otherwise. The opposite is true for $I_{it}^l(z)$. I call firms with $I_{it}^h(z) = 1$ as 'high' investment and $I_{it}^l(z) = 1$ as 'low' investment firms. With some abuse of notation, I treat these variables as events when they are in the conditioning set. I can write the moment inequality in Equation (2.3.7) as:

$$\theta_k\big(\mathbb{E}[k_{it} \mid I_{it}^h(z)] - \mathbb{E}[k_{it} \mid I_{it}^l(z)]\big) + \theta_l\big(\mathbb{E}[l_{it} \mid I_{it}^h(z)] - \mathbb{E}[l_{it} \mid I_{it}^l(z)]\big) \leqslant \big(\mathbb{E}[y_{it} \mid I_{it}^h(z)] - \mathbb{E}[y_{it} \mid I_{it}^l(z)]\big).$$

This expression shows that the identified set depends on the mean capital and labor of 'high' and 'low' investment firms. For example, if average capital and labor do not vary with investment, then the identified set would not be informative. To make this more concrete let me characterize the identified set. Define

$$a_l(z,k) := \mathbb{E}[l_{it} \mid I_{it}^h(z)] - \mathbb{E}[l_{it} \mid I_{it}^l(z)], \tag{2.3.8}$$
$$a_k(z,k) := \mathbb{E}[k_{it} \mid I_{it}^h(z)] - \mathbb{E}[k_{it} \mid I_{it}^l(z)], \tag{2.3.9}$$
$$a_y(z,k) := \mathbb{E}[y_{it} \mid I_{it}^h(z)] - \mathbb{E}[y_{it} \mid I_{it}^l(z)]. \tag{2.3.10}$$

Using these definitions, the moment inequality can be expressed as

$$a_k(z,k)\theta_k + a_l(z,k)\theta_l \leqslant y_l(z,k), \qquad (k,z) \in \tilde{\mathcal{X}}.$$

The identified set, conditional on k and z, is a region defined by a half-plane. Therefore, the identified set is the intersection of these half-planes.

Proposition 2.3.4 (Identified Set)**.** *Assume $\theta \in \tilde{\Theta}$, a compact parameter space. The identified set Θ is defined as the set of parameters that satisfy the conditional moment inequalities*

$$\Theta := \left\{\tilde{\theta} \in \tilde{\Theta} : \bigcap_{(k,z)\in\tilde{\mathcal{X}}} a_y(z,k) - \tilde{\theta}_k a_k(z,k) - \tilde{\theta}_l a_l(z,k) \geqslant 0 \quad a.s.\right\}, \tag{2.3.11}$$

and the identified set contains true parameter value $\theta \in \Theta$.

2.3.4 Moment Inequalities Using FOSD and MLRP

This subsection shows how to construct moment inequalities using MLRP and FOSD under additional assumptions. These assumptions include restrictions on the distribution of unobservables.

Proposition 2.3.2 establishes that the distributions of productivity conditional on high and low investment satisfy MLRP and FOSD. However, when characterizing the identified set, I only used the mean ordering, a weaker implication of MLRP and FOSD. This is because even though MLRP and FOSD hold for ω_{it} conditional on high and low investment, I can only recover $g(w_{it-1}) + \zeta_{it} + \eta_{it}^{l} + \epsilon_{it}$ using the data and parameters. Therefore, I need additional conditions that ensure that MLRP and FOSD are preserved when there are additive errors. The following two theorems present those conditions.

Theorem 2.3.1 (Shaked and Shanthikumar (2007)). *Let X_1 and X_2 be two independent random variables, and Y_1 and Y_2 be another two independent random variables. If $X_i \leqslant_{FOSD} Y_i$ for $i = 1, 2$ then*

$$X_1 + X_2 \leqslant_{FOSD} Y_1 + Y_2.$$

Theorem 2.3.2 (Shaked and Shanthikumar (2007)). *Let X_1, X_2 and Z be random variables such that X_1 and Z are independent and X_2 and Z are independent. If $X_1 \leqslant_{MLRP} X_2$ and Z has a log-concave probability density functions, then*

$$X_1 + Z \leqslant_{MLRP} X_2 + Z.$$

These two theorems suggest that in order for MLRP and FOSD to be preserved under convolutions I need: (i) an independence condition for FOSD, and (ii) independence and log-concavity conditions for MLRP. Therefore, I next impose these conditions on the unobservables to derive moment inequalities using FOSD and MLRP.

Identified Set Using FOSD

As Theorem 2.3.1 suggests, I need to impose independence restrictions on unobservables to preserve FOSD ordering.

Assumption 2.3.1. *$(\eta_{it}^{l}, \zeta_{it}, \epsilon_{it})$ are jointly independent from $\mathcal{I}_{it-1}$.*

With this assumption $g(\omega_{it-1})$ becomes jointly independent from the rest of the unobservables, $\zeta_{it} + \eta_{it}^{l} + \epsilon_{it}$, conditional on $\mathcal{I}_{it-1}$. Therefore, MLRP for $g(\omega_{it-1})$ conditional on high and low investment is preserved for $g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it}$ conditional on high and low investment. Under this assumption, I can strengthen the mean ordering in Equation (2.3.3) to the first-order stochastic dominance ordering, and characterize the identified set accordingly.

Proposition 2.3.5 (Identified Set-FOSD). *Under Assumptions 2.2.1-2.2.6 and Assumption 2.3.1, the true parameter $\theta \in \tilde{\Theta}$ satisfies the following condition*

$$F_{m(w_{it},\theta)}(\cdot \mid k_{it-2} = k, I_{it}^{h}(z) = 1) - F_{m(w_{it},\theta)}(\cdot \mid k_{it-2} = k, I_{it}^{l}(z) = 1) \geqslant 0, \qquad (2.3.12)$$

for all $(k,z) \in \tilde{\mathcal{X}}$. *The identified set (conditional on* k *and* z*) is*

$$\Theta_t^{FOSD} = \{\widehat{\theta} \in \Theta : (2.3.12) \text{ holds with } \theta \text{ in place of } \widehat{\theta}\}.$$

The independence assumption is non-standard in production function models, but it is difficult to imagine situations where mean independence holds and independence does not hold. Next, I specify the assumption that allows me to use the MLRP result in Proposition 2.3.2(i) for identification.

Identified Set Using MLRP

As Theorem 2.3.2 suggests, I need to impose independence and shape restrictions on the distributions of unobservables to be able to use MLRP.

Assumption 2.3.2. $(\eta_{it}^l, \zeta_{it}, \epsilon_{it})$ *are jointly independent from* $\mathcal{I}_{it-1}$*, and each variable in* $\eta_{it}^l, \zeta_{it}, \epsilon_{it}$ *has a log-concave probability distribution function.*

Under this assumption, I can strengthen the mean inequality in Proposition (2.3.3) to MLRP and characterize the identified set accordingly.

Proposition 2.3.6 (Identified Set-MLRP)**.** *Under Assumptions 2.2.1-2.2.6 and Assumption 2.3.2, the true parameter* $\theta \in \tilde{\Theta}$ *satisfies the following inequality:*

$$\begin{aligned} &F_{m(w_{it},\theta)}(\cdot \mid a \leqslant m(w_{it},\theta) \leqslant b, k_{it-2} = k, I_{it}^h(z) = 1) - \\ &F_{m(w_{it},\theta)}(\cdot \mid a \leqslant m(w_{it},\theta) \leqslant b, k_{it-2} = k, I_{it}^l(z) = 1) \geqslant 0 \end{aligned} \tag{2.3.13}$$

for all $(k,z) \in \tilde{\mathcal{X}}$ *and for all* (a,b) *such that* $a < b$. *The identified set (conditional on* (k,z)*) is*

$$\Theta_t^{MLRP} = \{\widehat{\theta} \in \Theta : (2.3.13) \text{ holds with } \widehat{\theta} \text{ in place of } \theta\}.$$

Most well known distributions, such as those in the exponential family, satisfy the assumptions required for this proposition.

2.3.5 Discussion

An advantage of identification analysis under a wide range of assumptions is that we can see the role of each assumption in identification. For example, we can start with the most general model to impose as few restrictions as possible. If the estimated set is not informative, then a nested model can be considered to shrink the identified set. Also, comparing estimates from a nested and a nesting model would test the restrictions imposed by the nested model.

Note also that proxy variable specification is a special case of my framework. So, if the estimates set is not information, one can use the proxy variable approach to point identify the parameters. It is also worth noting that, the identified set under my assumptions does not have to include the point estimates obtained from proxy variable method. The reason is that under the proxy variable assumptions the model is over-identified. If my partial identification method uses overidentification restrictions, the point estimates might be excluded from the identified set. This would mean rejecting the proxy variable assumptions.

One may think that set identifying the production function parameters is not useful unless the set is tight. As in most set identification results, the informativeness of the identified set depends on the data and empirical setting. However, as discussed above, there are other advantages of using my framework. Most importantly, since the standard OP approach is nested under my assumptions, there is no harm in starting with a more general model.

2.4 Estimation

The identification analysis generates conditional moment inequalities. It is convenient to turn conditional moment inequalities into unconditional ones for estimation. To achieve this, I integrate out k_{it-2} and define the propensity of high and low investment conditional on k_{it-2}.

First, I define the propensity score, which equals the probability that investment is greater than a cutoff z as

$$m(k_{it-2}, z) = \mathbb{E}[I_{it}^h(z) \mid k_{it-2}].$$

I define the moments using the propensity scores in the following ways.

$$\mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2}, I_{it}^h(z)] = \mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) I_{it}^h(z)}{m(k_{it-2}, z)} \mid k_{it-2}\right],$$
$$\mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2}, I_{it}^h(z)] = \mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big)\big(1 - I_{it}^h(z)\big)}{\big(1 - m(k_{it-2}, z)\big)} \mid k_{it-2}\right].$$

Integrating out k_{it-2}, the unconditional moment inequality can be written as:

$$\mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) I_{it}^h(z)}{m(k_{it-2}, z)}\right] \geqslant \mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big)\big(1 - I_{it}^h(z)\big)}{1 - m(k_{it-2}, z)}\right].$$

Define

$$s^h(\theta, z) = \mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) I_{it}^h(z)}{m(k_{it-2}, z)}\right], \qquad s^l(\theta, z) = \mathbb{E}\left[\frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big)\big(1 - I_{it}^h(z)\big)}{1 - m(k_{it-2}, z)}\right].$$

Estimation can be carried out by testing the hypothesis $s^h(\theta, z) \geqslant s^l(\theta, z)$ and inverting the test. Specifically, for a given θ, test $s^h(\theta, z) \geqslant \theta^l(\theta, z)$ for all z and include θ in the identified set if the the null hypothesis fails to be rejected. The natural estimators for $s^h(\theta, z)$ and

$s^l(\theta, z)$ are

$$\widehat{s}^h(\theta, z) = \sum_{i=1}^{N} \frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) I_{it}^h(z)}{\widehat{m}(k_{it-2}, z)}, \qquad s^l(\theta, z) = \sum_{i=1}^{N} \frac{\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) I_{it}^l(z)}{1 - \widehat{m}(k_{it-2}, z)}, \quad (2.4.1)$$

where $\widehat{m}(k_{it-2}, z)$ is an estimate of $m(k_{it-2}, z)$. These functions involve the propensity score function, which is a nuisance function and needs to be estimated in the first stage. To make the estimation procedure more robust to an estimation error in the nuisance function, I can define the doubly robust moment functions. This requires other nuisance functions in the moment functions. Define

$$\begin{aligned} g^h(k, z, \theta) &= \mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2} = k, I_{it}^h(z)], \\ g^l(k, z, \theta) &= \mathbb{E}[\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) \mid k_{it-2} = k, I_{it}^l(z)]. \end{aligned}$$

Using these functions, the doubly robust moments are

$$\begin{aligned} s_{db}^h(\theta, z) &= g^h(k_{it-2}, z, \theta) + \frac{I_{it}^h(z)\Big(\big(y_{it} - \theta_k k_{it} - \theta_l l_{it}\big) - g^h(k_{it-2}, z, \theta)\Big)}{m(k_{it-2}, z)}, \\ s_{db}^l(\theta, z) &= g^l(k_{it-2}, z, \theta) + \frac{\big(1 - I_{it}^h(z)\big)\big((y_{it} - \theta_k k_{it} - \theta_l l_{it}) - g^l(k_{it-2}, z, \theta)\big)}{1 - m(k_{it-2}, z)}. \end{aligned}$$

The sample analog of these moments can be obtained similar to Equation (2.4.1). Expectations of the doubly robust moments and the original moments equal to each other:

$$\mathbb{E}\big[s_{db}^h(\theta, z)\big] = \mathbb{E}\big[s^h(\theta, z)\big], \qquad \mathbb{E}\big[s_{db}^l(\theta, z)\big] = \mathbb{E}\big[s^l(\theta, z)\big].$$

at the true nuisance functions. Doubly robust moments have the property that if one of the nuisance functions is correct, then the moment is correct. Semenova (2017) studies moment inequality estimation with nuisance functions shows how to do inference when the nuisance functions are estimated using machine learning methods.

One can also consider using conditional moment inequalities to tighten the identified set instead of integrating out capital. For that estimation problem one can use many moment inequalities framework of Chernozhukov et al. (2018b) or conditional moment inequality estimation framework of Andrews and Shi (2013).

2.5 Extensions

The approach developed in this paper can be extended to other forms of production functions. To give some examples, I discuss the application to gross production function and using materials as the proxy variable instead of investment. I also show how my model accommodates measurement error in capital.

2.5.1 Gross Production Function

In this subsection, I show how to extend my model to a gross production function. The estimation procedure remains the same, with an increase in the number of parameters.
A gross production function is given by

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \theta_m m_{it} + \omega_{it} + \eta_{it}^l + \epsilon_{it}. \tag{2.5.1}$$

Similar to the main model, using the Markov assumption we can express the production function as

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + \theta_m m_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta_{it}^l + \epsilon_{it}.$$

All the propositions presented for the main model apply to this model because they do not depend on the functional form of the production function. Therefore, we can construct a moment function as

$$m(w_{it}, \tilde{\theta}) := y_{it} - \tilde{\theta}_k k_{it} - \tilde{\theta}_l l_{it} - \tilde{\theta}_m m_{it},$$

with $w_{it} = (k_{it}, l_{it}, m_{it})$, $\tilde{\theta} = (\tilde{\theta}_l, \tilde{\theta}_k, \tilde{\theta}_m)$. Let θ denote the vector of true parameter values.

Proposition 2.5.1. *For all $(k_{it-2}, z) \in \tilde{\mathcal{X}}$*

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-2}, i_{it-1} < z] \geqslant 0.$$

Proof. Omitted.

This moment inequality can be used to estimate the model parameters. The only difference is that there are more parameters to estimate, so the estimated bounds might be less informative.

2.5.2 Using Materials as an Imperfect Proxy

Levinsohn and Petrin (2003) propose using materials instead of investment as a proxy for productivity shock motivated by the fact that investment is often lumpy in the data. Even though my framework is robust to this problem, this paper allows for using materials as a proxy instead of investment. To show this, I need to replace Assumption 2.2.4 with the following assumption.

Assumption 2.5.1. *Firms' materials decision is given by*

$$m_{it} = m_t(k_{it-1}, \omega_{it}, \xi_{it}),$$

where ξ_{it} is a vector of unobserved random variables that affect the firm's materials decision and it is assumed to be independent of ω_{it} conditional on k_{it-1}.

When materials is used as an imperfect proxy, the moment inequalities take the form in the following proposition.

Proposition 2.5.2. *For all $(k_{it-2}, z) \in \tilde{\mathcal{X}}$, the true parameter value, θ, satisfies*

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, m_{it} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, m_{it} < z] \geqslant 0.$$

The moment function, $m(w_{it}, \theta)$, is defined in Equation (2.3.5). This moment inequality can be used to estimate the model parameters.

2.5.3 Identification with Predetermined Capital

To accommodate predetermined capital, the standard timing assumption in proxy variable framework, I replace Assumption 2.2.3 with the following assumption.

Assumption 2.5.2. *Capital accumulates according to*

$$k_{it} = \delta k_{it-1} + i_{it-1}.$$

This assumption implies that the amount of capital used for time t production is determined at time $t-1$. I also need to replace Assumption (2.2.4) with the following assumption.

Assumption 2.5.3. *Firms' investment decision is given by*

$$i_{it} = f_t(k_{it}, \omega_{it}, \xi_{it}),$$

where ξ_{it} is a vector of unobserved variables that affect firm's investment decision and it is assumed to be independent of ω_{it} conditional on k_{it}.

These changes in the assumptions affect only the conditioning set in moment inequalities. In particular, I need to condition on (k_{it-1}, i_{it-1}) instead of (k_{it-2}, i_{it-1}). So the moment inequality becomes:

$$\mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, i_{it-1} > z] - \mathbb{E}[m(w_{it}, \theta) \mid k_{it-1}, i_{it-1} < z] \geqslant 0,$$

where the moment function, $m(w_{it}, \theta)$, is defined in Equation (2.3.5). This moment inequality can be used to estimate the parameters. This modification does not affect the remained of the estimation.

2.5.4 Deviations from Cobb-Douglas Functional Form

In my identification strategy, the Cobb-Douglas functional form is necessary only to account for measurement errors. Thus, if I rule out measurement errors, I can use a nonlinear production function known up to a finite dimensional parameter vector. To demonstrate this extension, consider the following model:

$$y_{it} = r(\theta, w_{it}) + \omega_{it} + \epsilon_{it}, \tag{2.5.2}$$

where $r(\theta, w_{it})$ is a known function but the parameter vector, θ, is unknown. For this model, the moment function becomes:

$$m^*(w_{it}, \tilde{\theta}) := y_{it} - r(\tilde{\theta}, w_{it}). \tag{2.5.3}$$

With this moment function, the results for the main model can be applied to estimate Equation (2.5.2).

2.5.5 Measurement Error in Capital

My framework can also accommodate measurement error in capital, which is important because among all inputs, capital is most prone to measurement error.

Let η^k_{it} denote measurement error in investment. It is natural to model measurement error in capital using measurement error in investment because capital is accumulated through investment. Also, capital is often constructed from investment series in the data.[9] Let i^*_{it} denote true investment. The observed investment is given by:

$$\begin{aligned} i_{it} &= i^*_{it} + \eta^k_{it} \\ &= f_t(k_{it-1}, \omega_{it}, \xi_{it}) + \eta^k_{it}. \end{aligned}$$

Measurement error, η^k_{it}, can be included into f_t function as a part of ξ_{it} vector. Define $\xi^*_{it} = (\xi_{it}, \zeta^k_{it})$ and rewrite the investment function as

$$i_{it} = f_t(k_{it-1}, \omega_{it}, \xi^*_{it}).$$

With measurement error in investment, observed capital takes the form

$$k_{it} = k^*_{it} + \sum_{j=0}^{t} (1-\delta)^j \eta^k_{it},$$

where k^*_{it} is the true capital observed by the firm. Substituting this into the production function yields

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} - \Big(\sum_{j=0}^{t} (1-\delta)^j \eta^k_{it} \Big) \theta_k + \eta^l_{it} + \epsilon_{it}. \tag{2.5.4}$$

I can combine the measurement errors in capital and labor as $\eta_{it} = -\Big(\sum_{j=0}^{t} (1-\delta)^j \eta^k_{it} \Big) \theta_k + \eta^l_{it}$. With the combined measurement error, production function becomes:

$$y_{it} = \theta_k k_{it} + \theta_l l_{it} + g(\omega_{it-1}) + \zeta_{it} + \eta_{it} + \epsilon_{it}. \tag{2.5.5}$$

[9]For example, the perpetual inventory method.

This equation takes the form I analyzed for my partial identification results. So if we assume that η_{it}^k be is independent of ω_{it} conditional on k_{it-1} and $\mathbb{E}[\eta_{it} \mid \mathcal{I}_{it-1}] = 0$, we can use the same moment inequalities derived in Equation (2.3.6).

Since measurement error in capital is an important problem, the literature has paid particular attention to measurement error in capital. Some examples are Hu et al. (2011), Collard-Wexler and De Loecker (2016), and Kim et al. (2016). These papers require an instrumental variable or another proxy variable to control for measurement error in capital and show point identification. In contrast, my method does not require another variable, but it gives a bound instead of a point.

2.6 Conclusion

This paper extends the production function estimation literature by relaxing the restrictive assumptions of the proxy variable approach and showing that the parameters remain partially identified. My model (i) allows for multi-dimensional unobserved heterogeneity, (ii) relaxes strict monotonicity to weak monotonicity, (iii) accommodates a more general timing assumption. Also, the method is robust to measurement errors in inputs, an important problem in production function estimation.

I accomplish this by using an 'imperfect proxy' variable for identification. An 'imperfect proxy' variable contains information about productivity, but it cannot directly be used to control for productivity as in the proxy variable approach. Instead, an imperfect proxy variables generates stochastic orderings of productivity distributions, which can be exploited for estimation. I show how to use this stochastic ordering in the form of moment inequalities to obtain bounds for production function parameters.

Chapter 3

Partial Identification of Linear Models Using Homophily in Network Data

3.1 Introduction

Homophily, the tendency to interact with others with similar traits, is a prominent feature of networks. It has been widely studied and documented across a wide array of characteristics, including ethnicity, age, profession, and religion (McPherson et al. (2001)). The ubiquity of homophily in networks led economists to study its implications for economic phenomena, such as learning and information diffusion.

An implication of homophily is that networks are endogenous; individuals choose whom to associate based on observed and unobserved preferences. This endogeneity is considered a threat for identification in empirical models that involves network data. An important example is the estimation of peer effects, where peers are defined based on friendships in the network. If network endogeneity is not accounted for, peer effect estimates will be biased. The majority of papers in the literature ignore this problem by assuming that the network is formed exogenously.

This paper proposes using homophily as a solution to network endogeneity. In particular, I develop a method that uses network data to control for endogeneity in a linear model and partially identify the parameters. The method relies on constructing moment inequalities using network data under homophily assumption. Unlike prior approaches, my approach does not impose a specific network formation model. Instead, it relies on homophily, a feature of a broad family of network formation models. In this sense, my method is robust to a misspecified network formation model as long as homophily is satisfied. My model can also be used to estimate peer effects. I provide both simulation and empirical evidence that the estimated bounds are informative.

The object of interest is the parameters of a linear model with endogenous regressors. The researcher observes a network that includes observations of the linear model. Some examples are returns to schooling estimation, where the researcher observes a social network, and cross-country regressions, where the researcher observes trade or spatial networks between countries. My approach relies on two assumptions: (i) The same unobserved variable affects both network formation, and the outcome of the linear model (ii) The network exhibits unobserved homophily conditional on the unobserved variable. Homophily implies that linked nodes in the network are closer in terms of unobserved heterogeneity than non-linked nodes. Therefore, network data provides information about the unobserved variable. This paper shows how to use this information to control for endogeneity in the linear model and set identify the parameters.

The first contribution of the paper is to provide several definitions of homophily. Unlike previous definitions in the literature, these definitions do not rely on a network formation model, but they are satisfied by commonly-used network formation models. My framework, instead, uses restrictions on the shape of the conditional link probability function or stochastic orderings of conditional latent distances. This framework gives a range of homophily definitions, which can be ordered from strongest to weakest. These model-free homophily definitions are of independent interest, even though they are not the focus of the paper. For example, researchers can use these definitions to quantify homophily in observed networks or incorporate these definitions into network formation models.

I use two of the homophily definitions, weak and strong homophily, to identify the parameters of the linear model. Weak homophily is satisfied if the expected difference between unobserved heterogeneity is smaller if two agents are linked than if they are non-linked. Strong homophily orders (in the first-order stochastic dominance sense) the distributions of distance in the unobserved space based on the presence of links. Weak homophily is implied by strong homophily.

I show that homophily in networks can be used to construct moment inequalities. Weak homophily gives a conditional moment inequality, while strong homophily gives a continuum of conditional moment inequalities. To analyze how informative these moment inequalities are, I first derive the identified sets implied by these moments inequities. Then, I characterize the conditions under which the identified sets are bounded, and the conditions under which the signs of the parameters are identified. These conditions are weak and testable. Finally, for estimation, I develop doubly robust moment inequalities and show that standard methods of moment inequality estimation can be used to estimate my model.

I use Monte Carlo simulations to analyze the identified set under different scenarios. I investigate how the extent of homophily and endogeneity affects the size of the identified set. The simulations suggest that the identified set becomes more informative as the degree of homophily in the network increases. Thus, my model is more effective for networks with large homophily, such as social networks. The simulation also suggests that the identified set gets larger as the extent of endogeneity increases. However, this effect is small in my simulations.

I demonstrate my method in an empirical application using social network data from 185 villages in rural China. I consider a peer effect model (Manski (1993)), which estimates the effects of own literacy and peer literacy on perceptions about future disasters. The estimated set excludes the OLS estimates and its confidence band, suggesting that the model has endogeneity, and my approach controls for that endogeneity using network data. The OLS estimates are close to zero, whereas my estimated set implies both own and peer literacy have statistically significant effects. However, the estimated set does not pin down the signs of these effects but suggests that both own and peer effects have the same signs. Finally, I compare the estimated sets under strong and weak homophily assumptions. As expected, the identified set is more informative under strong homophily than under weak homophily.

This paper is related to a large literature on estimation using network data. This literature overwhelmingly focuses on peer effect estimation, employing a common specification called the linear-in-means model (Manski (1993)). However, most of the literature ignores network endogeneity and assumes that the network is exogenous (Bramoullé et al. (2009), Blume et al. (2015)). Another strand of the literature tries to address network endogeneity by randomly assigning links in networks either through intervention or a quasi-experiment. However, this solution is applicable only in very controlled environments.

A small but growing literature addresses network endogeneity (Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Johnsson and Moon (2017), Arduini et al. (2015)). The common theme in these papers is a parametric network formation model. They jointly estimate the linear model and network formation model to overcome endogeneity. My approach is related to this literature in that I also assume that the linear model and network share the same unobserved heterogeneity. However, I do not impose a specific network formation model, but use an important feature of a family of network formation models. Importantly, my assumptions are satisfied in the models commonly-used in this literature. Therefore, my model is robust to deviations from a specified network formation model. This flexibility comes at a cost: I identify bounds on the parameters, whereas this literature considers point identification.

The closest paper to my paper is Auerbach (2019), which uses network data to overcome endogeneity in a partially linear model. His method relies on identifying individuals with the same unobserved heterogeneity by finding pairs with similar link distributions. He does not impose a parametric network formation model but assumes that links are formed independently in a non-strategic fashion. My paper differs from his paper in three ways. First, I consider a more general family of network formation models, with possibly dependent and strategic pairwise network formation models. Second, I use a structural assumption, homophily, to achieve identification, so my results rely on different features of networks to achieve identification. Finally, I only consider partial identification, whereas he studies point identification.

I also contribute to the network literature in economics by studying homophily in networks. The literature studies homophily either using simple statistics based on group membership (Currarini et al. (2009)), or in a network formation model (Boucher (2015), Graham (2016)). In contrast, I provide a unified and flexible framework to analyze homophily based on the statistical properties of network data rather than a fully-specified model. My framework gives rise to several definitions of homophily that can be used to investigate observed networks in economics.

In the next section, I provide a formal definition of homophily. In section 3.3, I investigate the identification of a linear model with an endogenous variable using weak and strong homophily and characterize the identified set. Section 3.4 shows how my model accommodates peer effects. Section 3.5 discusses estimation, and section 3.6 discusses several extensions. Section 3.7 reports the result from Monte Carlo simulations. I provide an empirical application to demonstrate my method in Section 3.8.

Notation. I use the notation $F_a(t)$ and $F_a(t \mid b)$ to donate the distribution of variable a and the distribution of a conditional on b, respectively. Similarly, I use $f_a(t)$ and $f_a(t \mid b)$ to denote the probability density function of random variable a and the probability density function of a conditional on b, respectively. I also use $F_{|a|}(t)$ to denote the distribution of the absolute value of random variable a. Δ is the pairwise difference operator, so $\Delta x_{ij} := x_i - x_j$.

3.2 Defining Homophily in Networks

This section provides a discussion on homophily, a widely observed feature of networks, and then presents a series of formal definitions of homophily in a unified framework.

Homophily is one of the fundamental patterns of networks. It is defined as the tendency of individuals to associate with others similar to themselves. Homophily has been widely studied and documented in sociology across a wide array of characteristics, including ethnicity, age, profession, and religion (McPherson et al. (2001), Shrum et al. (1988), Moody (2001), Baerveldt et al. (2004)). The ubiquity of homophily in networks also led economists to study it from different perspectives. Several papers have investigated the implications of homophily for important economic phenomena, such as learning and information diffusion (Golub and Jackson (2012), Jackson and López-Pintado (2013), Bramoullé et al. (2012)). Also, the network formation literature in economics has studied how to incorporate homophily into network formation models.[1]

There are two types of homophily: observed and unobserved. The literature mostly focuses on observed homophily as it is easy to measure and document. However, there might also be homophilic preferences for characteristics that are unobservable to researchers. This would generate unobserved (latent) homophily in networks. Although it is hard to quantify, unobserved homophily has been considered to be an important feature of networks. For example, identifying an unobserved community structure from network data often relies on the latent homophily assumption. (Copic et al. (2009)).

Although homophily has been documented across a wide range of characteristics, to the best of my knowledge, there is no consensus on how to formally define homophily in networks. The literature often relies on simple statistics to measure homophily. For categorical variables, this is often the ratio of connections within the same category, divided by the total number of connections (Jackson (2008), Currarini et al. (2009)). However, these statistics do not extend to continuous variables, and their properties are not well-known.

The literature also analyzes homophily via network formation models. This approach specifies the link formation probability as a function of the similarity between nodes. In particular, the probability of forming a link between two nodes decreases in the distance

[1] See Jackson et al. (2017) for a survey on the economic consequences of the structure of networks.

between the observed and unobserved characteristics. To give an example, consider a model where links are formed according to

$$G_{ij} = \mathbb{1}\{c_0 + c_1|\alpha_i - \alpha_j| + c_2|x_i - x_j| \leqslant u_{ij}\}, \tag{3.2.1}$$

where G_{ij} is an indicator variable that denotes whether i and j are connected, and u_{it} is a random variable with known distribution, typically Extreme Value Type I. The model parameters are (c_0, c_1, c_2). In this model, α_i denotes the unobserved characteristic and x_i denotes the observed characteristics of node i.[2] Thus, both observed and unobserved homophily are captured. This model, or one of its variants, has commonly been used in the literature (Goldsmith-Pinkham and Imbens (2013), Johnsson and Moon (2017), Qu and Lee (2015), Hsieh and Lee (2017), Boucher and Mourifié (2017)). When defined this way, homophily corresponds to the relation between link formation probability and the distance between nodes' characteristics.

A drawback of defining homophily using a link formation function is that it requires a fully-specified network formation model. These models are useful for fully characterizing and analyzing the network; however, they are not ideal if the goal is to analyze and measure homophily. For example, a misspecified network formation model would lead to incorrect conclusions about homophily.

The central objective of this paper is to use homophily for identification without relying on a specific network formation model. For this reason, I present alternative definitions of homophily that do not rely on a network formation model. Instead, my definitions employ stochastic ordering of random variables. I will consider only unobserved homophily; however, the same definitions can be written in terms of observed homophily. I start with the definition of strong homophily. Define $\tilde{\mathcal{X}}$ as the support of (x_i, x_j) such that such that $\mathbb{E}[G_{ij} \mid x_i, x_j]$ is bounded below zero and above one.

Definition 3.2.1 ((Conditional) Strong Homophily)**.** *A network formation model satisfies strong homophily if*

$$F_{(\Delta\alpha_{ij})^2}\big(t \mid x_i, x_j, G_{ij} = 1\big) \geqslant F_{(\Delta\alpha_{ij})^2}\big(t \mid x_i, x_j, G_{ij} = 0\big),$$

for $t \geqslant 0$ *and* $(x_i, x_j) \in \tilde{\mathcal{X}}$.

This definition specifies a stochastic ordering for the distances between latent characteristics of linked and non-linked nodes. Conditional on observed characteristics x_i and x_j, the squared distance in latent space for non-linked nodes first-order stochastically dominates the squared distance for linked nodes. An interpretation of this property is that nodes that are 'close' in the latent space are more likely to be connected than nodes that are 'away' from each other. Strong homophily is satisfied by the network formation model in Equation (3.2.1); therefore, it imposes less restriction than the commonly-used network formation models. Next, I provide a weaker notion of homophily.

[2] I assume that all parameters are scalar to simplify the analysis.

Definition 3.2.2 ((Conditional) Weak Homophily). *A network formation model satisfies weak homophily if*

$$\mathbb{E}\big[(\Delta\alpha_{ij})^2 \mid x_i, x_j, G_{ij} = 0\big] \geqslant \mathbb{E}\big[(\Delta\alpha_{ij})^2 \mid x_i, x_j, G_{ij} = 1\big],$$

for $(x_i, x_j) \in \tilde{\mathcal{X}}$.

This definition restricts only the second moment of the distances instead of their distributions. It is nested by strong homophily. Both definitions of homophily stochastically order distances in latent space based on the existence of a link. Weak homophily orders the moments; strong homophily orders the distributions in the first-order stochastic dominance sense.

Even though this section provides two definitions, there are other ways of defining homophily using different stochastic orderings. In Appendix C.1, I present a range of homophily definitions that use both stochastic ordering and properties of link formation probability. I also examine the relationship between these definitions: I order them from strongest to weakest, and I show which definition nests the others. As a result, one contribution of this paper is to provide a toolkit for understanding and measuring homophily in networks. I focus on weak and strong homophily assumptions because I will only use those two to demonstrate how to address endogeneity in an empirical model.

There are several reasons why a broad range of homophily definitions is of independent interest. First, researchers can use these definitions to quantify homophily in observed networks. Since networks are present in a number of empirical contexts, different homophily definitions may be better suited for different types of networks. Also, these definitions make it possible to compare networks in terms of homophily. Second, network formation models that aim to capture homophily can incorporate these definitions.

The rest of the paper investigates how to use homophily to address endogeneity in a linear regression model.

3.3 Model and Assumptions

This section first presents a linear model and assumptions. Then, I show how to construct moment inequalities using network data under weak and strong homophily assumptions. Finally, I characterize the identified sets implied by moment inequalities.

The goal is to estimate the parameters of a linear model with endogeneity. The outcome, y_i, is given by the following model:

$$y_i = \beta_0 x_i + \alpha_i + \epsilon_i, \tag{3.3.1}$$

where x_i is an observed and α_i is an unobserved scalar random variable, and ϵ_i is orthogonal to both x_i and α_i. Thus, α_i represents unobserved heterogeneity and creates endogeneity in this model. β_0 is the parameter of interest, which gives the causal effect of x_i on the outcome.

I consider a model with a scalar endogenous variable to simplify the exposition. The model can accommodate multi-dimensional endogenous variables and exogenous control variables.[3]

I also assume that the researcher observes whether each pair (i, j) is linked in a network. This situation might arise in different empirical contexts. For example, Equation (3.3.1) can estimate the returns to schooling, and the researcher observes a social network of individuals. Another example is a cross-country regression. In this case, the network observed by the researcher could be geography or trade networks.

To introduce the network into the model, let G_{ij} be an indicator variable that equals one if i and j are linked and zero otherwise. The model has three stages. In the first stage, variables $\{x_i, \alpha_i, \epsilon_i\}$ are drawn independently from the same distribution. In the second stage, links are formed according to an unknown, and possibly complicated, network formation model, so G_{ij} is realized.[4] In the last stage, outcomes y_i are realized according to Equation (3.3.1). This timing assumption implies that the outcome variable y_i does not affect link formation. This sequential structure, where the network is formed, and then outcomes are realized, is standard in the literature.[5,6]

Without further assumptions, β_0 cannot be identified, as x_i is endogenous. To overcome this non-identification problem, I will impose restrictions on the observed network and exploit these restrictions to (partially) identify β_0. The overall idea is that if unobserved heterogeneity α_i in Equation (3.3.1) also affects link formation, then network data are informative about α_i. This information can be used to control for α_i when estimating Equation (3.3.1). The particular assumption I impose is unobserved homophily, which relates link formation probability to the distance between units in latent space. To see how homophily can be useful for estimation, I take the pairwise difference of Equation (3.3.1) to write:

$$\Delta y_{ij} = \beta_0 \Delta x_{ij} + \Delta\alpha_{ij} + \Delta\epsilon_{ij}. \tag{3.3.2}$$

This model relates the differences in outcomes to the differences in endogenous variables, Δx_{ij}, and unobserved heterogeneity, $\Delta\alpha_{ij}$. The pairwise difference model is useful because it is written in terms of the difference of $\Delta\alpha_{ij}$, for which homophily provides information.[7] In particular, homophily says that $\Delta\alpha_{ij}$ should be smaller in absolute value for linked pairs than for non-linked pairs. So, one can use network data, along with Equation (3.3.2), to make inference about β_0.

To be able to use the homophily assumption for estimation, I first need to formally define

[3]I discuss some of these extensions in Section 3.6.

[4]Link formation may be affected by variables that are outside the model.

[5]Alternatively, one can assume that the second and third stages occur simultaneously and independently. However, for the peer effects model developed later, I need the three-stage timing assumption. If the second and third stages occur simultaneously, the outcome variable y_i depends on variables affected by network formation, so the second and third stage are not independent with the presence of peer effects.

[6]Another important issue in estimations with network data is whether asymptotics is based on a single large network or many networks. This is beyond the scope of this paper. The empirical application involves many networks, which allows me to do inference using the standard methods.

[7]Pairwise difference models have a long history in econometrics; see Aradillas-Lopez et al. (2007), Honoré and Powell (1994).

homophily. I use the weak and strong homophily definitions given in the previous section. To simplify notation, I drop the ij subscript from difference variables, so any variable with Δ is implicitly indexed by ij. I also drop ij from the link variable G_{ij}.

3.3.1 Identification under Weak Homophily

In this section, I show that a network with weak homophily provides moment inequalities for estimating β_0. Then, I characterize the identified set given by the derived moment inequalities. To be able to use weak homophily for identification, I need to impose additional restrictions on the distribution of ϵ_i. The next assumption states these restrictions.

Assumption 3.3.1. *Suppose that*
(i) $\mathbb{E}\big[\epsilon_i \mid \alpha_i, x_i\big] = 0$.
(ii) $\mathbb{E}\big[\epsilon_i^2 \mid \alpha_i, x_i\big] = \mathbb{E}\big[\epsilon_i^2 \mid x_i\big]$.
(iii) (ϵ_i, ϵ_j) *is jointly independent of* G *conditional on* $(\alpha_i, \alpha_j, x_i, x_j)$.
(iv) *The network satisfies weak homophily defined in Definition (3.2.1).*

Part (i) assumes that ϵ_i is mean independent of (x_i, α_i), so it is not a source of endogeneity. Part (ii) states that the variance of ϵ_i does not change with α_i conditional on x_i. This assumption is non-standard in linear models. I need this assumption because homophily imposes restrictions on the second moment of distance, $\Delta\alpha^2$, instead of α_i. For this reason, I need to restrict the second moment of ϵ_i. Part (iii) implies that (ϵ_i, ϵ_j) do not affect link formation once we condition on other unobservables and observables in the model. Part (iv) states that the network satisfies the weak homophily assumption. Using these assumptions, I obtain the following proposition.

Proposition 3.3.1. *Under the conditions of Assumption 3.3.1,*

$$\mathbb{E}\big[(\Delta\alpha + \Delta\epsilon)^2 \mid x_i, x_j, G = 0\big] \geqslant \mathbb{E}\big[(\Delta\alpha + \Delta\epsilon)^2 \mid x_i, x_j, G = 1\big],$$

for $(x_i, x_j) \in \tilde{\mathcal{X}}$.

Proof. See Appendix A.4.

This proposition establishes that weak homophily is robust to additive noise, ϵ_i, under Assumption 3.3.1. I need this proposition to ensure that weak homophily is preserved when α_i is contaminated by ϵ_i. By assumption (iv), weak homophily is satisfied in terms of α_i. However, the linear model in Equation (3.3.1) includes another unobserved variable, ϵ_i. And from data and parameters, I can recover only the sum of the unobserved variables, $\alpha_i + \epsilon_i$. So I need weak homophily for $\Delta\alpha + \Delta\epsilon$.

Identified Set

This section derives the moment inequalities implied by the weak homophily assumption and characterizes the identified set. The first step is to write unobservables as a function of parameters and data. The pairwise difference model in Equation (3.3.2) yields

$$\Delta\alpha_{ij} + \Delta\epsilon_{ij} = \Delta y_{ij} - \beta_0 \Delta x_{ij}.$$

Substituting this expression into Proposition 3.3.1, I obtain a conditional moment inequality:

$$\mathbb{E}[(\Delta y - \beta_0 \Delta x)^2 \mid x_i, x_j, G = 1] \leqslant \mathbb{E}[(\Delta y - \beta_0 \Delta x)^2 \mid x_i, x_j, G = 0]. \tag{3.3.3}$$

This inequality orders two means conditional on the existence and absence of a link. Note that it involves observed variables and parameters only, so I can use this moment inequality to partially identify β_0. I first characterize the identified set given by the conditional moment inequality in Equation (3.3.3). The identified set is the intersection of conditionally identified sets. Let $x_{ij} := (x_i, x_j)$. Equation (3.3.3) can be written as

$$\beta_0 \Delta x \big(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0]\big) \geqslant \mathbb{E}[(\Delta y)^2 \mid x_{ij}, G = 1] - \mathbb{E}[(\Delta y)^2 \mid x_{ij}, G = 0]. \tag{3.3.4}$$

This inequality suggests that the identified set (conditional on x_{ij}) is a half-space. Also, it shows that the direction of half-spaces depends on the sign of the conditional mean differences on the left-hand side of Equation (3.3.4). Using this, I characterize the identified set conditional on x_{ij} as follows:

$$\mathcal{B}(x_{ij}) = \begin{cases} \left(-\infty, \dfrac{\mathbb{E}[\Delta y^2 \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y^2 \mid x_{ij}, G = 0]}{2\Delta x \left(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0]\right)} \right], & \text{if } \Delta x \left(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0 \right. \\ \Big(-\infty, \infty \Big), & \text{if } \Delta x \left(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0 \right. \\ \left[\dfrac{\mathbb{E}[\Delta y^2 \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y^2 \mid x_{ij}, G = 0]}{2\Delta x \left(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0]\right)}, \infty \right), & \text{if } \Delta x \left(\mathbb{E}[\Delta y \mid x_{ij}, G = 1] - \mathbb{E}[\Delta y \mid x_{ij}, G = 0 \right. \end{cases}$$

The identified set is given by the intersection of these sets:

$$\tilde{\mathcal{B}} = \{\cap_{x_{ij} \in \tilde{\mathcal{X}}} \ \mathcal{B}(x_{ij})\},$$

Next, I investigate the conditions under which the identified set is informative.

Characterizing the Identified Set

I focus on the conditions under which the identified set is bounded, and the sign of β_0 is identified.

Since the identified set is given by the intersection of half-spaces, its informativeness depends on the shape of these half-spaces. To obtain a bounded identified set, there should exist two x_{ij}'s such that identified sets (conditional on the x_{ij}'s) should extend in different directions. For the sign of β_0 to be identified, we need at least one x_{ij} such that conditional on that, the identified set excludes zero. The next proposition establishes these results.

Proposition 3.3.2.

(i) *The identified set is bounded if there exists* $(\bar{x}_{ij}, \tilde{x}_{ij}) \in \tilde{\mathcal{X}}^2$ *such that*

$$\Delta\bar{x}\big(\mathbb{E}[\Delta y \mid \bar{x}_{ij}, G=1] - \mathbb{E}[\Delta y \mid \bar{x}_{ij}, G=0]\big) \geqslant 0 \geqslant \Delta\tilde{x}\left(\mathbb{E}[\Delta y \mid \tilde{x}_{ij}, G=1] - \mathbb{E}[\Delta y \mid \tilde{x}_{ij}, G=0]\right).$$

(ii) *The sign of* β_0 *is identified if there exists* $\tilde{x}_{ij} \in \tilde{\mathcal{X}}$ *such that*

$$\big(\mathbb{E}[(\Delta y)^2 \mid \tilde{x}_{ij}, G=1] - \mathbb{E}[(\Delta y)^2 \mid \tilde{x}_{ij}, G=0]\big) \geqslant 0. \tag{3.3.5}$$

The proof of this proposition is trivial, and therefore omitted. The first part of the proposition specifies the condition under which the identified set is bounded. This condition is testable because all moments in the proposition can be identified from data. However, it is also useful to interpret this condition in terms of model primitives. For this purpose, I write the condition to obtain a bounded identified set as:

$$\begin{aligned} &\Delta\bar{x}\left(\mathbb{E}[\Delta\alpha \mid \bar{x}_{ij}, G=1] - \mathbb{E}[\Delta\alpha \mid \bar{x}_{ij}, G=0]\right) + \beta_0(\Delta\bar{x})^2 \geqslant 0, \\ &\Delta\tilde{x}\left(\mathbb{E}[\Delta\alpha \mid \tilde{x}_{ij}, G=1] - \mathbb{E}[\Delta\alpha \mid \tilde{x}_{ij}, G=0]\right) + \beta_0(\Delta\tilde{x})^2 \leqslant 0. \end{aligned} \tag{3.3.6}$$

This reformulation is written in terms of unobserved heterogeneity $\Delta\alpha$ and does not depend on Δy. The difference of expectations in Equation (3.3.6) is similar to the condition given in the definition of weak homophily, except that it involves the first moment of $\Delta\alpha$, instead of its second moment. Therefore, whether the identified set is bounded depends on the expected difference in $\Delta\alpha$ of linked and non-linked nodes and how that difference varies with x_{ij}.

The second part of Proposition (3.3.2) concerns the identification of the sign of β_0. The sign is identified if expected $(\Delta y)^2$ for the linked nodes is greater than expected $(\Delta y)^2$ for the non-linked nodes. Note that this condition is closely related to the definition of weak homophily: to identify the sign of β_0, the network should not exhibit homophily in terms of y_i. This can happen in two ways. To explain them, I rewrite Equation (3.3.5) as

$$\underbrace{\big(\mathbb{E}[(\Delta\alpha)^2 \mid \tilde{x}_{ij}, G=1] - \mathbb{E}[(\Delta\alpha)^2 \mid \tilde{x}_{ij}, G=0]\big)}_{I} + 2\beta_0\Delta\tilde{x}\underbrace{\big(\mathbb{E}[\Delta\alpha \mid \tilde{x}_{ij}, G=1] - \mathbb{E}[\Delta\alpha \mid \tilde{x}_{ij}, G=0]\big)}_{II} \geqslant 0, \tag{3.3.7}$$

which involves two parts. Part I is negative by the weak homophily assumption. Thus, in order for this inequality to hold, the second summand should be positive and larger than the first summand in absolute value. Three quantities affect the second summand: (i) β_0, (ii) $\Delta\tilde{x}$ and (iii) the expression labeled as II. If $\beta_0 < 0$ then $\Delta\tilde{x}$ and II should have different signs, otherwise, they should have the same signs. Since the sign of II depends on the underlying network formation, which is not specified, it is not possible to further analyze the identification of the sign of β_0.

From Equation (3.3.7) we can also see that in two cases it is not possible to identify the sign of β_0: (i) $\beta_0 = 0$ and (ii) II equals zero. The first case is obvious. To rule out the second case, link formation should be related not only to $(\Delta\alpha)^2$ but also to the expected difference in unobserved heterogeneity, $\Delta\alpha$.

After analyzing identification under weak homophily, I next turn to what we can learn from networks with strong homophily.

3.3.2 Identification under Strong Homophily

In this section, I show how to use the strong homophily assumption to derive moment inequalities. Then, I characterize the identified set based on these moment inequalities.

Since strong homophily implies weak homophily, the identified set obtained from strong homophily is tighter than the identified set obtained from weak homophily. However, I need more conditions on the distribution of ϵ_i to exploit strong homophily for identification. In particular, I need to strengthen the restrictions on the distribution of ϵ given in Assumption 3.3.1.

Assumption 3.3.2. *Suppose that*
(i) *ϵ_i is independently distributed from α_i conditional on (x_i, G_{ij}).*
(ii) *(ϵ_i, ϵ_j) is jointly independent of G_{ij} conditional on $(\alpha_i, \alpha_j, x_i, x_j)$.*
(iii) *$f_{\Delta\epsilon}(\Delta\epsilon \mid x_i, x_j)$ is unimodal around zero for all $(x_i, x_j) \in \tilde{\mathcal{X}}$.*
(iv) *The network satisfies the strong homophily assumption.*

Part (i) of Assumption 3.3.2 states that two unobserved variables are independently distributed conditional on observables. This assumption strengthens the mean independence in Assumption 3.3.1(ii) to independence. I need this restriction on the distribution of ϵ_i because strong homophily involves distribution functions, not just moments. In Part (ii), I maintain the assumption that ϵ_i does not affect network formation after conditioning on observed and unobserved characteristics.

Part (iii) imposes a shape restriction on the distribution of $\Delta\epsilon$. I need this assumption to show that strong homophily in terms of $\Delta\alpha$ implies strong homophily in terms of $\Delta\alpha + \Delta\epsilon$. This assumption is not as strong as it seems because it requires unimodality only for $\Delta\epsilon$; the distribution of ϵ_i does not have to be unimodal. Unimodality for the distribution of the difference between two independently and identically distribution variables is a substantially weaker condition that unimodality for the distribution of both of these variables.[8] For most well-known distributions, such as those in the exponential family, the difference of two identically and independently distributed random variables has a unimodal distribution. One primitive condition ensuring unimodality is that ϵ_i has a log-concave distribution function.

The next proposition shows that strong homophily continues to hold when ϵ_i is added to α_i.

Proposition 3.3.3. *Under the conditions given in Assumptions 3.3.2*

$$F_{|\Delta\alpha+\Delta\epsilon|}\big(t \mid x_i, x_j, G = 1\big) \geqslant F_{|\Delta\alpha+\Delta\epsilon|}\big(t \mid x_i, x_j, G = 0\big),$$

[8] Note that, by construction, the probability density function of $\Delta\epsilon$ is symmetric, so we need to rule out distributions that are symmetric but not unimodal.

for $(t, x_i, x_j) \in \mathbb{R}^+ \times \tilde{\mathcal{X}}$.

Proof. See Appendix A.4.

According to this proposition, strong homophily is robust to additive noise.[9] Unimodality of the distribution of $\Delta\epsilon$ and independence of ϵ_i and α_i are sufficient to achieve this. The proof of this proposition uses the fact that the distribution of the sum of two independent random variables is given by a convolution. So, under the condition of Assumption 3.3.2(i), I can use convolution analysis to derive the distribution of $\alpha_i + \epsilon_i$ conditional on $(G = 1)$ and $(G = 0)$, and then establish first-order stochastic dominance. The unimodality of the distribution of $\Delta\epsilon$ ensures that the convolution is well behaved. Although these conditions are strong, they allow me to use strong homophily, which can lead to a substantially tighter identified set than weak homophily.

Identified Set

Proposition 3.3.3 gives a first-order stochastic dominance relation between two distribution functions. To use this for identification, I turn the stochastic dominance relation into moment inequalities. One way to achieve this is to write first-order stochastic dominance in terms of a continuum of moment inequalities:

$$\mathbb{E}[\mathbb{1}\{|\Delta\alpha + \Delta\epsilon| \leqslant \gamma\} \mid x_{ij}, G = 1] \geqslant \mathbb{E}[\mathbb{1}\{|\Delta\alpha + \Delta\epsilon| \leqslant \gamma\} \mid x_{ij}, G = 0] \tag{3.3.8}$$

for $(\gamma, x_i, x_j) \in \mathbb{R}^+ \times \mathbb{X}^c$. Substituting $\Delta\alpha + \Delta\epsilon$ as a function of data and parameters I obtain:

$$\mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 1] \geqslant \mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 0]. \tag{3.3.9}$$

This inequality says that the true parameter value satisfies a continuum of conditional moment inequalities. To see the difference between strong and weak homophily, note that FOSD implies that if A dominates B, then the expectation of any monotone transformation B is smaller than the expectation of the same monotone transformation of A. Therefore, strong homophily gives moment inequalities for any monotone transformation of $|\Delta\alpha|$. This is in contrast to weak homophily, which gives moment inequalities only for the second moment of $|\Delta\alpha|$. Thus, strong homophily should give a more informative identified set.

To state the identified set, I turn Equation (3.3.9), which involves two moments, into a inequality with a single moment. Note that one can write the expectations in Equation

[9] FOSD is preserved under Fourier transformation for the random variables. However, we need FOSD for the difference between two random variables. This requires additional conditions on the distribution of the random variables.

(3.3.9) as

$$\mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 1] = \mathbb{E}\left[\frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} G}{\mathbb{E}[G \mid x_{ij}]} \;\middle|\; x_{ij}\right],$$

$$\mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 0] = \mathbb{E}\left[\frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\}(1 - G)}{1 - \mathbb{E}[G \mid x_{ij}]} \;\middle|\; x_{ij}\right],$$

where $\mathbb{E}[G \mid x_{ij}]$ corresponds to the propensity of link formation conditional on observed characteristics. This transformation has been widely-used in the treatment effect estimation literature (Rosenbaum and Rubin (1983)). Now, define a function g that takes parameters and data:

$$g(x_i, x_j, \Delta y, G, \gamma, \beta) := \frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\}(1 - G)}{1 - \mathbb{E}[G \mid x_{ij}]} - \frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} G}{\mathbb{E}[G \mid x_{ij}]}.$$

Using this function, I define the identified set.

Proposition 3.3.4 (Identified Set)**.** *Assume $\beta_0 \in \tilde{B}$, a compact parameter space. The identified set B is defined as the set of parameter values that satisfy the conditional moment inequalities*

$$B := \left\{\beta \in \tilde{B} : \bigcap_{(x_i, x_j) \in \tilde{\mathcal{X}}} \bigcap_{\gamma > 0} \mathbb{E}[g(x_i, x_i, \Delta y, G, \gamma, \beta) \mid x_{ij}] \leqslant 0 \quad a.s.\right\},$$

and the identified set contains the true parameter value $\beta_0 \in B$.

Next, I analyze the identified set given in this proposition.

Characterization of the Identified Set

Under strong homophily, it is difficult to characterize the identified set because moment inequalities involve non-linear functions. In this section, I derive some properties of the identified set under additional regulatory conditions, which are given by the next assumption.

Assumption 3.3.3. *Suppose that there exists $x_{ij} \in \tilde{\mathcal{X}}$ such that*
(i) *Δy has a bounded support conditional on $x_{ij} \in \tilde{\mathcal{X}}$.*
(ii) *There exists $M \in \mathbb{R}$ such that $\mathbb{E}\big[G = 1 \mid x_{ij}, \Delta\alpha = \Delta\big] < \mathbb{E}[G = 1]$ for $\Delta z \in (-\infty, -M) \cup (M, \infty)$.*

Part (i) imposes a mild support restriction. According to Part (ii), if the distance in latent space is too large, then the conditional link formation probability is less than the unconditional link formation probability. This condition imposes a mild restriction on the shape of conditional link probability, and is consistent with, albeit not implied by, strong homophily. This is because, by definition, observations that are too far in the latent space should have a very low probability of link formation. I use the following lemma to show that the identified set is bounded under Assumption 3.3.3.

Lemma 3.3.1. *Suppose that there exists* $(t, x_i, x_j), (t', x_i, x_j) \in \mathcal{R}^+ \times \mathbb{X}^c$ *and* $t \neq t'$ *such that*

$$F_{\Delta\alpha}(t \mid x_{ij}, G = 1) - F_{\Delta\alpha}(t' \mid x_{ij}, G = 0) < 0,$$
$$F_{\Delta\alpha}(t \mid x_{ij}, G = 1) - F_{\Delta\alpha}(t' \mid x_{ij}, G = 0) > 0.$$

Then, the identified set is bounded.

Proof. See Appendix A.4.

This lemma rules out a stochastic ordering in $\Delta\alpha$ conditional on $(G = 1)$ and conditional on $(G = 0)$. Note that this is different from the strong homophily condition, which involves the absolute value of the difference. Since the model does impose restrictions on $(\Delta\alpha)^2$, not $(\Delta\alpha)$, it is plausible to expect that Lemma 3.3.1 holds. If this condition is satisfied, then it is possible to show that the strong homophily condition is violated for large and small candidate parameter values. Therefore, we can reject parameter values that are larger than $\bar{\beta}$ and smaller than $-\bar{\beta}$, where $\bar{\beta}$ is positive. This implies that the identified set is bounded.

As for the identified set under weak homophily, the difference in expectation of $\Delta\alpha$ between linked and non-linked nodes plays a critical role in the identified set. As a result, the informativeness of the identified set depends on the underlying network formation model. Next, I show that Assumption 3.3.3 is sufficient for the conditions of this lemma to hold.

Proposition 3.3.5. *Under Assumptions 3.3.2–3.3.3, the conditions of Lemma 3.3.1 are satisfied, and therefore, the identified set given in Proposition 3.3.4 is bounded.*

Proof. See Appendix A.4.

This proposition uses the Bayes rule for continuous variables to show that the conditions of Lemma 3.3.1 are satisfied.

3.3.3 Discussion and Relation to Literature

Exploiting network data to address endogeneity has been proposed in the literature, especially in peer effects estimation. The standard implementation of this idea is to assume a network formation model and jointly estimate a linear model of peer effects and a network formation model. To give an example, Goldsmith-Pinkham and Imbens (2013) consider the following model:[10]

$$y_i = \beta_0 x_i + \alpha_i + \epsilon_i, \qquad \epsilon_i \mid \alpha_i, x_i, G_{ij} \sim \mathcal{N}(0, \sigma_1^2). \tag{3.3.10}$$

To address endogeneity in Equation (3.3.10) they consider the following network formation model:

$$G_{ij} = \mathbb{1}\{\alpha_0 + c_1|\alpha_i - \alpha_j| + c_2|x_i - x_j|\}, \quad u_{ij} \sim \text{EV II}(0, \sigma_2). \tag{3.3.11}$$

This network formation model incorporates both latent and observed homophily. Also, the unobserved heterogeneity, α_i, is the same in Equations (3.3.10) and (3.3.11). Therefore,

[10] Their model considers also a peer effects variable. I omit that variable to facilitate the comparison.

one can estimate these models jointly to account for the correlation between α_i and x_i. Specifically, Goldsmith-Pinkham and Imbens (2013) take a Bayesian approach to estimate the parameters under distributional assumptions.[11] Other papers in the literature take similar approaches with slightly different network formation and estimation models (Hsieh and Lee (2016), Johnsson and Moon (2017), Arduini et al. (2015)).

My approach is related to this literature in that I also assume that the linear model and network formation model share the same unobserved variable. However, I do not impose a specific network formation model, but use an important feature of a family of network formation models. Importantly, the network formation model in Equation (3.3.11) satisfies both weak and strong homophily. Therefore, my model is robust to deviations from a specified network formation model as long as the true network formation model exhibits homophily. Yet this flexibility comes at a cost: I identify a bound on the parameters, whereas the existing literature considers point identification.

This paper is also related to the literature on monotone instrumental variables (Manski (1997), Manski and Pepper (2000)). This literature assumes that the expected potential outcomes are ordered based on an observed variable, which is called a monotone instrument. In my model, G_{ij} can be considered a monotone instrument for the pairwise difference model in Equation (3.3.2) under weak homophily. That is because, in that equation, the second moment of potential outcomes can be ordered based on G. The main difference between my model and the standard monotone instrumental variable approach is that monotonicity holds for the second moment of unobserved heterogeneity in Equation (3.3.2), which requires stronger assumptions than the monotone instrumental variables approach.

My paper is also related to the 'imperfect instrument approach', which assumes that the researcher has some prior information about the correlation between the endogenous variable and unobserved heterogeneity. This information is used to construct moment inequalities. See, for example; Nevo and Rosen (2012) and Conley et al. (2012).

3.4 A Model with Peer Effects

The most common empirical setting with network data is peer effects estimation. In this section, I show that my model accommodates peer effects. The estimation strategy is similar to the main model, but due to complications that come with peer effects, the characterization of the identified set is limited compared to the main model.

Let $\bar{x}_i$ be a scalar variable that is a function of peer characteristics. Following a common specification called the linear-in-means model, I assume that $\bar{x}_i$ equals the average characteristics of person i's peers.[12] With the addition of $\bar{x}_i$, the model in Equation (3.3.1)

[11]Goldsmith-Pinkham and Imbens (2013) study other important questions such as testing the endogeneity of the network and estimating a dynamic network formation model. This paper is not related to their other contributions.

[12]In principle, $\bar{x}_i$ can be any function of peer characteristics.

becomes

$$y_i = \beta_0 x_i + \rho_0 \bar{x}_i + \alpha_i + \epsilon_i.$$

In this model, β_0 corresponds to own effect, and ρ_0 corresponds to peer effects. Since Manski (1993), the linear-in-means model has become the workhorse of peer effects estimation and has been used to estimate peer effects in many different settings. My model, unlike Manski (1993), considers only contextual peer effects, not endogenous peer effects. That is, only peer *characteristics* influence the outcome, not peer *outcomes*. I make this restriction because, with endogenous peer effects, the model becomes a game, which requires solving for the equilibrium and deriving the implied reduced form equation (Blume et al. (2015)).[13]

This model is more difficult to analyze than the model in Equation (3.3.1) due to two endogenous variables: both own characteristics x_i and peer characteristics $\bar{x}_i$ are endogenous.[14] For identification, I consider the same assumptions stated for the main model, but construct the identified set only under strong homophily.[15]

3.4.1 Identified Set with Peer Effects

Recall that Propositions 3.3.1 and 3.3.3 state the moment inequalities under the weak and strong homophily assumptions. These propositions continue to hold with the addition of peer effects because they involve only α_i and ϵ_i, and they are affected by the linear model (by timing assumption). Thus, I can use the same moment inequalities with the addition of peer characteristics, $\bar{x}_i$, in the estimation. Specifically, under strong homophily, the true parameter vector $\theta_0 = (\beta_0, \gamma_0)$ satisfies

$$\mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x - \rho_0 \Delta \bar{x}| \leqslant \gamma\} \mid x_{ij}, G = 1] \leqslant \mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x - \rho_0 \Delta \bar{x}| \leqslant \gamma\} \mid x_{ij}, G = 0],$$

for $\{(t, x_i, x_j) : t \geqslant 0, (x_{ij}) \in \tilde{\mathcal{X}}\}$, which gives a continuum of conditional moment inequalities. Define a function that takes data and model parameters as

$$g^p(x_{ij}, \Delta y, G, \Delta \bar{x}, \theta) := \frac{\mathbb{1}\{|\Delta y - \beta \Delta x - \gamma \Delta \bar{x}| \leqslant \gamma\}(1 - G)}{1 - \mathbb{E}[G \mid x_{ij}]} - \frac{\mathbb{1}\{|\Delta y - \beta \Delta x - \gamma \Delta \bar{x}| \leqslant \gamma\} G}{\mathbb{E}[G \mid x_{ij}]}.$$

Using this function, I define the identified set in the presence of peer effects.

Proposition 3.4.1 (Identified Set with Peer Effects)**.** *Assume $\theta_0 \in \tilde{\Theta}$, a compact parameter space. The identified set Θ is defined as the set of parameters that satisfy the conditional moment inequalities,*

[13] Extending my framework to endogenous peer effects is a subject of future work.

[14] Peer characteristics are endogenous because they are realized based on the network, which is endogenous due to the presence of α_i

[15] Identification under weak homophily is similar, and I do not repeat it here.

$$\Theta := \left\{ \theta \in \tilde{\Theta} : \bigcap_{(x_i,x_j)\in\tilde{\mathcal{X}}} \bigcap_{\gamma>0} \mathbb{E}[g^p(x_i, x_j, \Delta y, G, \Delta\bar{x}, \theta) \mid x_{ij}] \geqslant 0 \quad a.s. \right\},$$

and the identified set contains the true parameter value $\theta_0 \in \Theta$.

With peer effects, analyzing the identified set is complicated for two reasons. First, the model involves two parameters. Second, without peer effects, homophily holds conditional on endogenous variables in the linear model. This is not the case with peer effects, as homophily does not necessarily hold conditional on $\bar{x}$. Thus, I analyze the informativeness of the identified sets under the weak homophily assumption and in some special cases.

3.4.2 Identification Conditional on $(x_i = x_j)$

If the object of interest is the peer effects, not the own effect, one way to simplify estimation and identification analysis is to condition on $(x_i = x_j)$. In this case, β_0 drops from the pairwise model in Equation (3.3.2), so only the peer effects coefficient ρ_0 can be estimated. Conditional on $(x_i = x_j)$, the moment inequality becomes

$$\mathbb{E}[(\Delta y - \rho_0\Delta\bar{x})^2 \mid x_i = x_j, G = 1] \leqslant \mathbb{E}[(\Delta y - \rho_0\Delta\bar{x})^2 \mid x_i = x_j, G = 0].$$

This is similar to the moment inequality I derived in Subsection 3.3.1. This moment inequality can be written as

$$\begin{aligned}\rho_0^2\big(\mathbb{E}[\Delta\bar{x}^2 \mid x_i = x_j, G = 1] - \mathbb{E}[\Delta\bar{x}^2 \mid x_i = x_j, G = 1] \leqslant \\ \mathbb{E}[\Delta y\Delta\bar{x} \mid x_i = x_j, G = 0]\big) + \big(\mathbb{E}[(\Delta y)^2 \mid x_i = x_j, G = 1] - \mathbb{E}[(\Delta y)^2 \mid x_i = x_j, G = 0]\big).\end{aligned} \tag{3.4.1}$$

The following proposition considers whether the identified set is bounded and whether the sign of the coefficient ρ_0 is identified.

Proposition 3.4.2.

(i) *The identified set is bounded if there exists $\tilde{x}_i = \tilde{x}_j$ such that*

$$\big(\mathbb{E}[\Delta\bar{x}^2 \mid \tilde{x}_i, \tilde{x}_j, G = 1] - \mathbb{E}[\Delta\bar{x}^2 \mid \tilde{x}_i, \tilde{x}_j, G = 0]\big) \geqslant 0. \tag{3.4.2}$$

(ii) *The identification of the sign of β_0 depends on the solution to inequality in Equation (3.4.1) that is quadratic in parameters.*

The proof of this proposition is trivial and therefore omitted. The conditions stated in this proposition are different from those in Equation (2.5.3). Notably, moment inequality (3.4.2) is quadratic in the parameter. The quadratic structure arises because in deriving the moment inequality (2.5.3) for the main model, I condition on the endogenous variable, whereas in this proposition, it is not possible to condition on $\bar{x}_i$.

Part (i) of this proposition requires an interesting condition. The expectation of the difference in peer characteristics of linked nodes must be greater than the difference in peer characteristics of non-linked nodes. In other words, the identified set is bounded if there is no homophily in terms of peer characteristics. This is a strong restriction, but it is only a sufficient condition for obtaining a bounded identified set. Part (ii) requires solving the quadratic equation in Equation (3.4.2). If the solution to that quadratic equation is an interval, then the identified set is bounded.

3.4.3 Identification Conditional on $(x_i \neq x_j)$

If the researcher is also interested in estimating β_0, we need to condition on x_{ij} such that $x_i \neq x_j$. Let $\mathbb{E}_k[Y] = \mathbb{E}[Y \mid x_{ij}, G = k]$, for $k \in \{0, 1\}$. Using this notation, I obtain

$$\begin{aligned}&\big(\mathbb{E}_1[\Delta y^2] - \mathbb{E}_0[\Delta y^2]\big) + \rho_0^2\big(\mathbb{E}_1[\Delta \bar{x}^2] - \mathbb{E}_0[\Delta \bar{x}^2]\big) - 2\beta_0 \Delta x\big(\mathbb{E}_1[\Delta y] - \\ &\qquad \mathbb{E}_0[\Delta y]\big) - 2\beta_0\rho_0\Delta x\big(\mathbb{E}_1[\Delta \bar{x}] - \mathbb{E}_0[\Delta \bar{x}]\big) - 2\rho_0\big(\mathbb{E}_1[\Delta y \Delta \bar{x}] - \mathbb{E}_0[\Delta y \Delta \bar{x}]\big) \leqslant 0.\end{aligned}$$

The parameters β_0 and ρ_0 can be partially identified from this moment inequality.

3.5 Estimation

This section considers estimation of a bound for the parameters. In particular, I show how to apply the standard estimation methods of moment inequalities to my empirical model.

The previous section derived moment inequalities from homophily assumptions. For estimation, it is convenient to turn these conditional moment inequalities into unconditional moment inequalities. To do this, I first define the propensity of link formation conditional on observables:

$$m(x_{ij}) = \mathbb{E}[G = 1 \mid x_{ij}].$$

I can write the moments in Equation (3.3.8) using $m(x_{ij})$

$$\begin{aligned}\mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 1] &= \mathbb{E}\left[\frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} G}{m(x_{ij})} \;\Big|\; x_{ij}\right], \\ \mathbb{E}[\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} \mid x_{ij}, G = 0] &= \mathbb{E}\left[\frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\}(1 - G)}{1 - m(x_{ij})} \;\Big|\; x_{ij}\right].\end{aligned}$$

Using these expectations, I write the conditional moment inequality as

$$\mathbb{E}\left[\frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\}(1 - G)}{1 - m(x_{ij})} - \frac{\mathbb{1}\{|\Delta y - \beta_0 \Delta x| \leqslant \gamma\} G}{m(x_{ij})} \;\Big|\; x_{ij}\right] \leqslant 0, \qquad (3.5.1)$$

which contains a single conditional moment. As a result, one can use the standard estimation methods of conditional moment inequalities after accounting for $m(x_{ij})$. To avoid the indicator function in the moment inequality and ease implementation, I write the moment inequalities as

$$\mathbb{E}\left[\frac{t\big(|\Delta y-\beta_0\Delta x|\big)(1-G)}{1-m(x_{ij})}-\frac{t\big(|\Delta y-\beta_0\Delta x|\big)G}{m(x_{ij})}\;\Big|\; x_{ij}\right]\geqslant 0, \tag{3.5.2}$$

where $t(x)$ is a monotone function. For example, $t(x)=x^2$ corresponds to the moment inequality obtained under the weak homophily assumption in Equation (2.5.3). For the rest of this section, I present an estimation procedure for $t(x)=x^2$. However, the procedure can be applied to other monotone functions. The estimation relies on a doubly robust version of the moment inequality in Equation (3.5.2). Rearranging Equation (3.5.2) and using $t(x)=x^2$, I obtain

$$\mathbb{E}\left[\frac{\big((\Delta y)^2-2\beta\Delta x\Delta y\big)(1-G)}{1-m(x_{ij})}-\frac{\big((\Delta y)^2-2\beta\Delta x\Delta y\big)G}{m(x_{ij})}\;\Big|\; x_{ij}\right]\geqslant 0.$$

Next, I define some nuisance functions:

$$\begin{aligned} f^k(x_{ij}) &:= \mathbb{E}[(\Delta y)^2 \mid x_{ij}, G=k] \\ h^k(x_{ij}) &:= \mathbb{E}[\Delta y \mid x_{ij}, G=k] \end{aligned}$$

for $k\in\{0,1\}$. The moment inequality can be written in a doubly robust form as:

$$\begin{aligned} \mathbb{E}\Bigg[\Big(f^0(x_{ij})-2\beta\Delta x h^0(x_{ij})\Big)+\frac{\Big(\big((\Delta y)^2-f^0(x_{ij})\big)-2\beta\Delta x\big(\Delta y-h^0(x_{ij})\big)\Big)(1-G)}{1-m(x_{ij})}- \\ \Big(f^1(x_{ij})-2\beta\Delta x h^1(x_{ij})\Big)-\frac{\Big(\big((\Delta y)^2-f^1(x_{ij})\big)-2\beta\Delta x\big(\Delta y-h^1(x_{ij})\big)\Big)G}{m(x_{ij})}\;\Big|\; x_{ij}\Bigg]\geqslant 0. \end{aligned}$$

For doubly robust moments, the inequality holds if either of the nuisance functions is correct, so this estimation strategy is robust to misspecification in one of the nuisance functions. It also has other appealing properties, as studied by Chernozhukov et al. (2018a). Estimation of moment inequalities with doubly robust moments has been studied by Semenova (2017). She shows that using doubly robust moments, one can estimate nuisance functions with machine learning methods and obtain uniformly consistent estimates.

If x_i is a continuous variable, a straightforward way of estimation would be choosing a finite number of monotone $t(x)$ functions and integrating over x_{ij} to obtain unconditional moment inequalities. Then, one can use the method proposed by Semenova (2017). An alternative approach would be to use estimation methods for many moment inequalities. This approach could capture more information from the continuum of moment inequalities given by first-order stochastic dominance. For estimating many moment inequalities, see Chernozhukov et al. (2018b) and Andrews and Shi (2013).

If x_i is a discrete variable, it is also possible to estimate the set by inverting a hypothesis test. For weak homophily, one can test mean inequality using the standard two-sample t-test and invert the test to obtain the estimated set. For strong homophily, one can use the available first-order stochastic dominance test and invert the test to obtain the estimated set (Linton et al. (2005), Barrett and Donald (2003), Horváth et al. (2006)). I use this approach in my empirical application since my endogenous variables are binary.

3.6 Extensions

In this section, I briefly discuss two extensions of my main model by showing how to account for (i) exogenous control variables and (ii) nonlinear models.

3.6.1 Model with Controls

My main model does not include control variables. In this section, I incorporate the control variables into the model as:

$$y_i = \beta_0 x_i + \lambda_0 z_i + \alpha_i + \epsilon_i,$$

where z_i is a vector of exogenous control variables. To accommodate controls in identification, I modify the strong homophily assumption.

Definition 3.6.1 ((Conditional) Strong Homophily with Controls)**.** *Network formation satisfies*

$$F_{|\Delta\alpha|}\big(t \mid x_i, x_j, z_i, z_j, G = 1\big) \leqslant F_{|\Delta\alpha|}\big(t \mid x_i, x_j, z_i, z_j, G = 0\big),$$

for $(t, z_i, z_j, x_i, x_j) \in \mathbb{R}^+ \times \tilde{\mathcal{X}}^c$*.*

$\tilde{\mathcal{X}}^c$ is defined as the support of (z_i, z_j, x_i, x_j) such that $\mathbb{E}[G_{ij} \mid z_i, z_j, x_i, x_j]$ is bounded below zero and above one. In this definition, latent homophily holds conditional on all control variables, which accounts for the possibility that z_i and z_j might affect network formation. I also need to alter other assumptions to accommodate control variables.

Assumption 3.6.1. *Suppose that*
(i) $\epsilon_i \perp\!\!\!\perp \alpha_i, x_i, z_i, G_{ij}$*.*
(ii) $\Delta\epsilon$ *has a unimodal distribution.*
(iii) $\mathbb{E}[z_i(\alpha_i + \epsilon_i)] = 0$*.*
(iv) *The network satisfies strong homophily with controls.*

This assumption differs from Assumption (3.3.2) in that the conditions hold conditional on control variables, and also it assumes that z_i is exogenous. Since control variables are endogenous, I can partial out z_i and rewrite the model as

$$\tilde{y}_i = \beta_0 \tilde{x}_i + \alpha_i + \epsilon_i,$$

where $\tilde{x}_i := x_i - z_i \mathbb{E}\big[z_i' z_i\big]^{-1} \mathbb{E}\big[z_i' x_i\big]^{-1}$ and $\tilde{y}_i := y_i - z_i \mathbb{E}\big[z_i' z_i\big]^{-1} \mathbb{E}\big[z_i' y_i\big]^{-1}$.

After partialing out, α_i and ϵ_i are still in the model. This suggests that having exogenous controls in the model does not affect my identification results significantly. It is easy to show that Propositions 3.3.1 and 3.3.3 hold when x_i and x_j are replaced with $\tilde{x}_i$ and $\tilde{x}_j$. After satisfying these conditions, I can follow the same procedure from Section (3.5). Moreover, since the homophily assumption holds conditional on z_i, it is possible to obtain a tighter identified set. Also, including the control variables is important for homophily, as homophily might hold only conditional on some observables.

3.6.2 Deviations from the Linear Model

My identification strategy does not rely on a linear model. The linear function in Equation (3.3.1) can be replaced with a known nonlinear function up to a finite dimensional parameter vector. To demonstrate this extension, consider the model

$$y_i = r(\theta_0, x_i) + \alpha_i + \epsilon_i,$$

where $r(\theta_0, x_i)$ is a known function up to the parameter vector θ_0. For this model, I can obtain

$$\mathbb{E}[(\Delta y - r(\theta_0, x_i) + r(\theta_0, x_j) \mid x_i, x_j, G = 1] \leqslant \mathbb{E}[(\Delta y - r(\theta_0, x_i) + r(\theta_0, x_j) \mid x_i, x_j, G = 0].$$

This moment inequality can be used to partially identify θ_0.

3.7 Monte Carlo Simulations

I use Monte Carlo simulations to analyze the identified set under different scenarios.

3.7.1 Simulation Design I

The first simulation examines how the model features affect the size of the identified set. I consider two simulation designs. The first design looks at how the identified set changes with the degree of homophily. Intuitively, stronger homophily should lead to a more informative identified set. In the limit where there is no homophily, the identified set is not informative at all. The purpose of the second design is to understand the impact of endogeneity on the identified set. This is important because the characterization of the identified set does not provide guidance about how endogeneity affects the identified set. For these two exercises, I consider the following outcome and link formation models:

$$y_i = \beta_0 x_i + \alpha_i + \epsilon_i, \tag{3.7.1}$$

$$G_{ij} = 1\{\gamma|\alpha_i - \alpha_j| \leqslant \eta\}. \tag{3.7.2}$$

The data generating process is given by

$$\begin{pmatrix} \alpha_i \\ x_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix}\right], \tag{3.7.3}$$

where $x_i = 1\{\tilde{x}_i > 0\}$, $\eta \sim U[-1, 5]$, $\epsilon \sim N[0, 1]$ and $\beta_0 = 0$. The network formation model in Equation (3.7.2) is simple and does not include observables. It incorporates homophily since the link formation probability decreases with the distance in latent space. Two parameters in this model control the features I aim to study. p equals the correlation between x_i and α_i, so it controls the degree of endogeneity in the model. The second parameter, γ, controls for the degree of homophily in the network formation model. As γ increases, the impact of the distance in latent space on link formation increases. My goal is to understand how these two parameters affect the length of the identified set. For this purpose, I investigate the identified sets for different values of p and γ.

I set the number of observations to ten million to reduce the impact of estimation error and focus on the identified set. Since x_i is an indicator variable, the model provides a single conditional moment inequality. I derive the identified set under the strong homophily assumption. Since my focus is on the identified set, I simulate this model only once.

Figure 3-2 reports the identified sets from Monte Carlo simulations. Panel (a) displays the relationship between the size of the identified set and the degree of homophily as controlled by γ. As expected, the size of the identified set becomes smaller as the degree of homophily increases. At a very large value of γ, the identified set is very informative and converges to a point. This result suggests that in networks with a high degree of homophily, the identified set should be more informative. Social networks are a good example of networks with a high degree of homophily (McPherson et al. (2001), Shrum et al. (1988), Moody (2001)).

In Panel (b), I report the identified set for different values of p, which controls the extent of endogeneity. The size of the identified set is negatively related to the degree of endogeneity: The identified set is narrower when there is no endogeneity, and it gets wider as endogeneity increases. However, the effect of endogeneity is small, as we see an only slight change in the size of the identified set.

3.7.2 Simulation Design II

My second simulation design considers a data generating process with two endogenous variables, and a network formation model with both observed and unobserved homophily. I assume the following model:

$$\begin{aligned} y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \alpha_i + \epsilon_i \\ G_{ij} &= 1\{|\alpha_i - \alpha_j| + |x_{1i} - x_{1j}| + |x_{2i} - x_{2j}| \leqslant \eta\}. \end{aligned}$$

This simulation design is more realistic in the sense that the network formation model includes both observed and unobserved homophily. To generate the variables, I first simulate $(\alpha, \tilde{x}_1, \tilde{x}_2)$ from a multivariate normal distribution:

$$\begin{pmatrix} \alpha \\ \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \right],$$

and generate (x_1, x_2) as $x_1 = 1\{\tilde{x}_1 > 0\}, \quad x_2 = 1\{\tilde{x}_2 > 0\}$. The other random variables are distributed $\eta \sim U[-1, 5]$, $\epsilon \sim N(0, 1)$ with true parameter values $\beta_1 = -1$, $\beta_2 = 1$. In this simulation, I report the estimated set with the number of observations equal to 500. Also, this simulation allows me to examine the identified set with two endogenous variables, which I have not analyzed in Section 3.3.1.

I report the results in Figure 3-1. The reported set shown in red covers the true parameter with a 95% probability, and the black point shows the true parameter values. The results suggest that the identified set is bounded and quite informative, even though the number of observations is small. For example, the identified set excludes $(0, 0)$, rejecting the null hypothesis of no effects. Another interesting point worth noting is the shape of the identified set. This shape arises because the two variables are positively correlated.

3.8 Empirical Application

I demonstrate my method in an empirical setting using social network data from villages in rural China. This dataset was collected and used by Cai et al. (2015) to study the influence of social networks on weather insurance adoption. The authors conduct a randomized control trial where households are randomly assigned to information sessions about insurance products. Their paper combines this randomized control trial with network data to understand how information diffusion through the network affects insurance purchases. In my empirical application, I do not use insurance adoption or treatment variables, so my empirical setting is not affected by the experiment.

The data are collected from two surveys: a social network survey carried out before the experiment, and a household survey completed after households had made their insurance purchase decisions. The household survey includes questions on demographics, income, experience in purchasing any insurance, risk attitudes, and perception about future disasters. The dataset provides a census of the population of 185 villages with a total number of 5,335 households surveyed.

In my empirical application, I consider a peer effects model. The model includes both own effect and peer effects:

$$y_i = \theta x_i + \gamma \bar{x}_i + \lambda z_i + \alpha_i + \epsilon_i. \tag{3.8.1}$$

The outcome variable, y_i, is perceptions about future disasters, x_i is an indicator variable for literacy, and the controls, z_i, include age, gender, and village fixed effects.[16] I choose these variables because they are the most suitable variables in the dataset for my empirical application.[17,18] Since the endogenous variable is an indicator variable, the estimation gives a single moment inequality. I estimate the model under both the weak homophily and strong homophily assumptions.

We should expect that disaster perception and literacy are related to each other. A literate individual should acquire information from sources that an illiterate individual cannot. Also, we should also expect that unobservables affecting risk perception have an effect on whether an individual is literate. My objective is not to draw a definite conclusion about the relationship between perception of disaster and literacy, but to understand how my approach helps solve the endogeneity problem.

The goal of this empirical application is twofold. First and most important, I look at whether the identified set is informative. Since the data is a social network, there should be some degree of homophily, but whether this homophily is informative is an empirical question. My second goal is to compare my results with the OLS estimates. Excluding the OLS estimates would suggest that the model has endogeneity, and my approach controls for that endogeneity using network data.

I present the estimation results in Figure (3-3). The black region shows the estimated set (with 95% coverage) from my method. The blue points show the OLS estimates, and the red regions are 95% confidence bands for the OLS estimates. The panel on the left reports estimates under the weak homophily assumption and the panel on the right reports results for the strong homophily assumption. We observe that under both assumptions, the identified excludes the OLS estimates and their confidence bands. The OLS estimates suggest no effect of education and peer education on risk perception. In contrast, my estimated set excludes null effects.

Even though my estimates exclude the OLS estimates from the identified set, they do not give a definite answer on the sign of the effects. Under both the strong and weak homophily assumptions, my estimates suggest either positive or negative own and peer effects. Their signs are not identified, but we learn that the own effect and peer effects cannot have different signs. Finally, I compare the estimated sets from strong and weak homophily assumptions. As expected, the identified set is tighter under strong homophily than under weak homophily. Also, the identified set given by strong homophily is less well-behaved than the one given by weak homophily. This might be due to non-linearities in the moment inequalities under the

[16]The perceived probability of future disasters was elicited by asking, 'What do you think is the probability of a disaster that leads to a more than 30 percent loss in yield next year?'.

[17]Most outcome variables in the dataset are binary, which is not covered in my model.

[18]In a future iteration of this project, I will use data from AddHealth, which provides a richer set of variables and a larger number of observations.

strong homophily assumption.

3.9 Conclusion

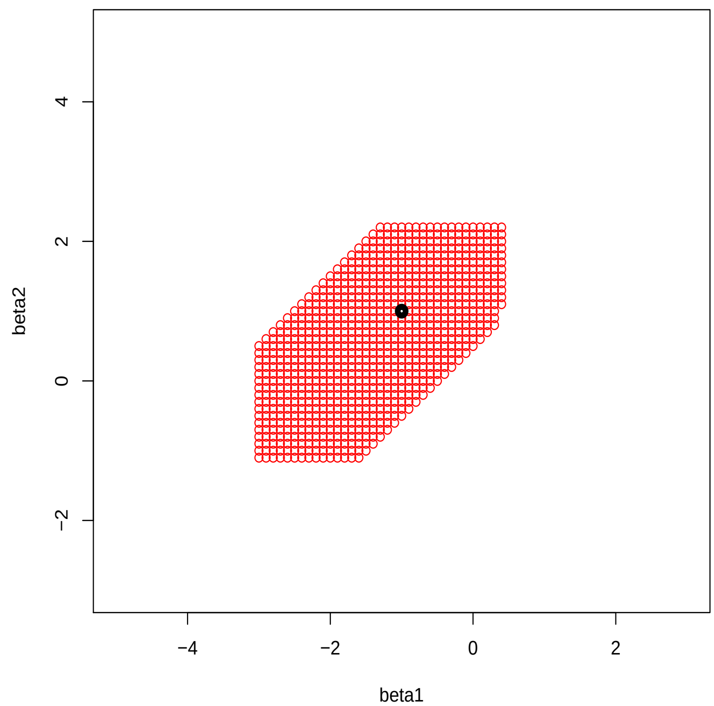
This paper provides set identification results for the parameters of a linear model using homophily in networks. The model assumes that unobserved heterogeneity in the linear model also affects network formation. Using this assumption, I show that the homophily in unobservables, which is recognized as an essential feature of networks, provides information about unobserved heterogeneity in the linear model. I show how to use this information in the form of moment inequalities and make inferences about the parameters.

I analyze the identified set under two weak and strong homophily assumptions. Weak homophily gives a conditional moment inequality, while strong homophily gives a continuum of conditional moment inequalities. To analyze the informativeness of these moment inequalities, I derive the identified sets implied by these moments inequities and characterize the conditions under which the identified sets are bounded, and the conditions under which the signs of the parameters are identified. These conditions are weak and testable.

Importantly, my identification strategy is agnostic about the network formation model. I exploit a feature, homophily, that exists in a broad class of network formation models, rather than relying on a correctly specified network formation model. This makes my approach robust to misspecifications in the network formation model.

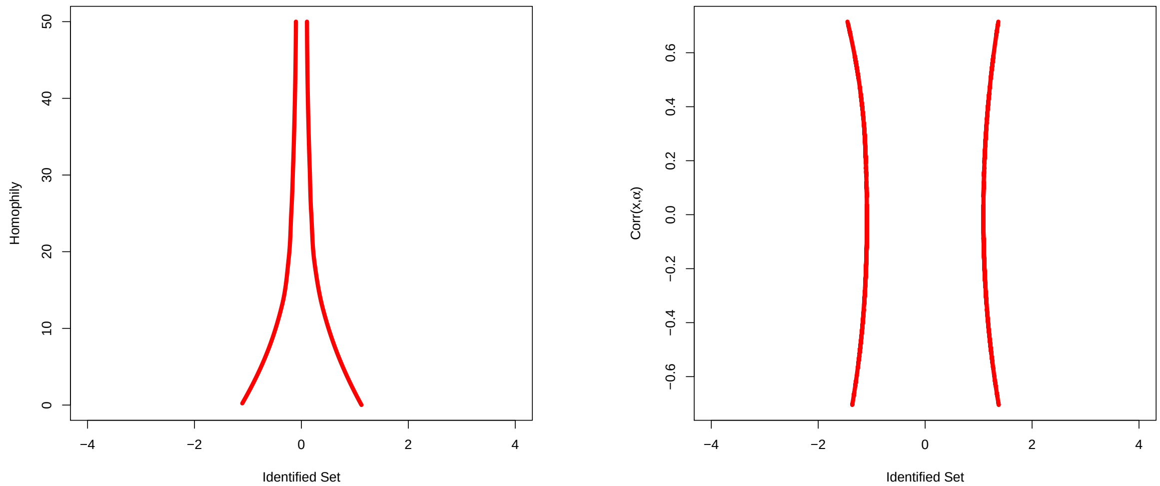
Finally, an empirical application and Monte Carlo simulations show that the identified set is informative about the parameters. Estimating own effect and peer effects of education on risk perception using social network data from villages in rural China, I find that my identified set excludes the OLS estimate and its confidence band, pointing to an endogeneity problem. The simulation exercise suggests that the identified set becomes more informative as the degree of homophily in the network increases. Therefore, my method is more effective to address endogeneity in networks with high degree of homophily, such as in social networks.

Figure 3-1: Simulation with Two Endogenous Variables - Estimated Set



Note: The estimated set obtained from the simulation design in section (3.7.2). The reported set shown in red covers the true parameter with a 95% probability, and the black point shows the true parameter values.

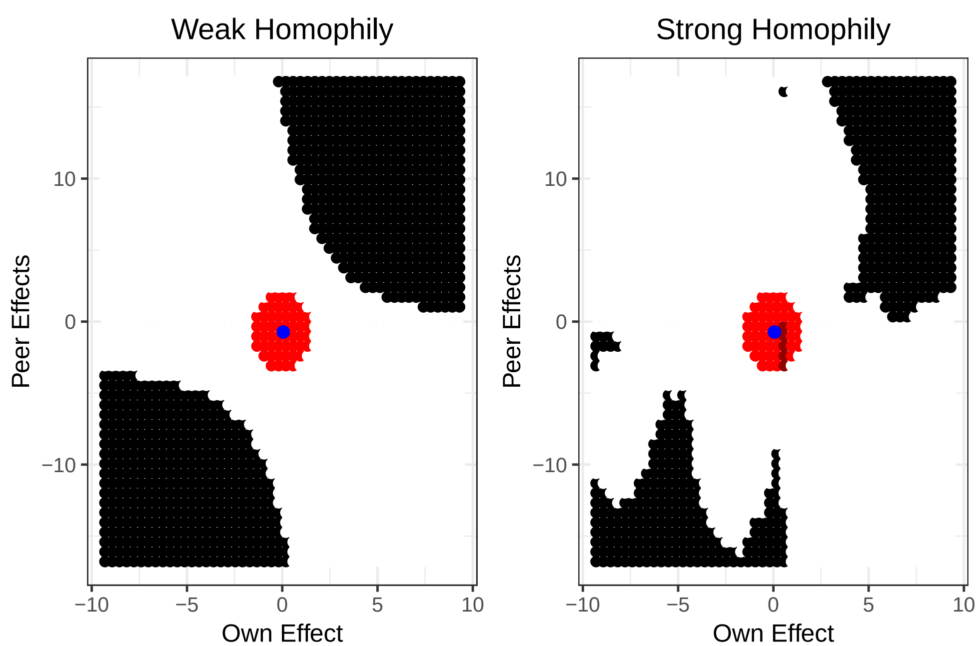
Figure 3-2: Identified Set-Simulation Design I



(a) Degree of Homophily and Identified Set (b) Degree of Homophily and Identified Set

Note: The left panel shows the width of the identified set as the degree of homophily, governed by γ in Equation (3.7.2), changes. The left panel shows the width of the identified set as the degree of endogeneity, governed by γ in Equation (3.7.3), changes.

Figure 3-3: Own and Peer Effects



Note: Estimated set for the parameters θ and γ in Equation (3.8.1) under weak and strong homophily assumptions. The regions in black cover the true parameter value with 95% probability. The blue points show the OLS estimates, and the red regions are 95% confidence bands for the OLS estimates.

Appendix A

Appendix to Chapter 1

A.1 Data and Estimation Appendix

A.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 A.1 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.

A.1.2 Colombia

The data for Colombia are from the annual Colombian Manufacturing census provided by the Departamento Administrativo Nacional de Estadistica, covering the years 1981 through 1991. This dataset contains all manufacturing plants with 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 A.2 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.

A.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 Allcott 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 A.3 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.

A.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

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

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 A.4 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.

A.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 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 A.5 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.

A.1.6 Variable Construction

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.

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.

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 (2019a).

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.

A.1.7 Estimation Algorithm

This section presents the estimation algorithm. Apply data cleaning and variable construction described in Subsection A.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

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

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

$$\log(\widehat{h}(\tilde{M}_{it})) = a_1 + a_2\tilde{m}_{it} + a_3\tilde{m}^2_{it}, \tag{A.1.1}$$

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 $\widehat{h}(\tilde{M}_{it})$ described in Section 1.4. Let $V_{it} := L_{it}\widehat{h}(\tilde{M}_{it})$. Approximate the production function as

$$\widehat{f}(K_{it}, L_{it}\widehat{h}(\tilde{M}_{it})) = b_1 + b_2k_{it} + b_3k^2_{it} + b_3k_{it}v_{it} + b_4v_{it} + b_5v^2_{it}, \tag{A.1.2}$$

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 $\{a_j\}^3_{j=1}$, $\{b_j\}^5_{j=1}$, $\widehat{c}_2(\cdot)$ and $\widehat{c}_3(\cdot)$ construct the objective function in Equation (1.5.7). Minimize this objective function to estimate the production function as follows. For a given $\{a_j\}^3_{j=1}$, estimate $\{b_j\}^5_{j=1}$, $\widehat{c}_2(\cdot)$ and $\widehat{c}_3(\cdot)$ by minimizing the objective function using least squares regression. The algorithm involves two layers. For a candidate value of the parameter vector $\{a_j\}^3_{j=1}$, in the inner loop, estimate $\{b_j\}^5_{j=1}$, $\widehat{c}_2(\cdot)$ and $\widehat{c}_3(\cdot)$ using least squares regression. In the outer loop use an optimization routine to estimate $\{a_j\}^3_{j=1}$. 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 $\{a_j\}^3_1$ and $\{b_j\}^5_1$ in Equations (A.1.1) and (A.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 (1.4.5) and (1.4.9) by taking numerical derivatives. Then given an estimate of θ^V_{it} and revenue shares of materials and labor use Equations (1.4.6) to estimate output elasticity of labor and materials. Finally estimate markups from $\widehat{\theta}^V_{it}$ and the revenue share of flexible input as in Equation (1.8.1).

For standard errors, resample firms with replacement from sample A then repeat the

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

[3]I 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.

estimation procedure above. For estimation of the Nested CES model in Section 1.8.4, I use the same procedure except that I impose the parametric restrictions given by the Nested CES model.

A.2 Extensions

A.2.1 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^l_{it} and p^m_{it}, respectively, and use $\bar{p}_{it}$ to denote the input price vector, so $\bar{p}_{it} := (p^l_{it}, p^m_{it})$. I also use $\tilde{p}_{it}$ 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, Assumptions 1.2.1 is replaced by the following assumption.

Assumption A.2.1. *The distribution of productivity shocks and input prices obey:*

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

This assumption states that prices and productivity shocks jointly follow an exogenous first-order Markov process. Importantly, this assumption allows for correlation between productivity shocks and input prices. Since we expect that more productive workers, as represented by higher ω^L_{it}, 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 A.2.2. *Firm's materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega^H_{it}, \omega^L_{it}, \bar{p}_{it}), \tag{A.2.1}$$

and $s_t(K_{it}, \omega^H_{it}, \omega^L_{it}, \bar{p}_{it})$ is strictly increasing in ω^H_{it}.

This assumption is a natural extension of Assumption 1.2.3, as the demand for materials should depend on both input prices. In this section, I maintain the other assumptions in the model, namely Assumptions 1.2.2 and 1.2.4, and state the following proposition.

Proposition A.2.1.
(i) *Under Assumptions 1.2.2(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^L_{it} and $\tilde{p}_{it}$*

$$\tilde{M}_{it} = r_t(K_{it}, \omega^L_{it}, \tilde{p}_{it}). \tag{A.2.2}$$

(ii) *Under Assumptions 1.2.2(v), $r_t(K_{it}, \omega^L_{it}, \tilde{p}_{it})$ is strictly monotone in ω^L_{it}.*

The proof of this proposition is a straightforward extension of the proof of Proposition 1.2.1, and therefore, is omitted. Compared to Proposition 1.2.1, 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 A.2.2. To summarize, by inverting Equations (A.2.2) and (A.2.1), and omitting the time subscripts in functions, I can write productivity shocks as:

$$\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it}, \tilde{p}_{it}), \qquad \omega_{it}^H = \bar{s}(K_{it}, M_{it}, \tilde{M}_{it}, \bar{p}_{it}). \tag{A.2.3}$$

In the presence of heterogeneous input prices, the derivation of the control variables proceed similarly as in Section 1.3 with minor differences. I first use the Skorokhod's representation of ω_{it}^L to write:

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it}, u_{it}^1), \qquad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H, \tilde{p}_{it-1}, \tilde{p}_{it} \sim \text{Uniform}(0,1). \tag{A.2.4}$$

Unlike Equation (1.3.1), 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 A.2.1, the optimal flexible input ratio depends on the ratio of input prices. Using Equations (A.2.2), (A.2.3) and (A.2.4), I obtain

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

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

Lemma A.2.1. *Under Assumptions A.2.1 - A.2.2, 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 A.4.

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

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

Therefore, we can use Equations (A.2.3) and (A.2.4) 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

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, \sim \text{Uniform}(0,1). \tag{A.2.5}$$

Following the same steps in Equation (1.3.2) of Section 1.3, materials demand function can be written as:

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

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

Lemma A.2.2. *Under Assumptions A.2.1 and A.2.2, 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 A.4.*

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

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

and the control function is given by

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

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

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

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.

A.2.2 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:

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

I assume that the resercher 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:

$$Y_{it} = F_t\left(K_{it}, L_{it}, R_{it}^m\omega_{it}^M\right)\exp(\omega_{it}^H)\exp(\epsilon_{it}), \tag{A.2.6}$$

where I define $\omega_{it}^M := 1/(p_{it}^m)$. In Equation (A.2.6), 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 A.2.3. *Productivity shock and materials price jointly follow an exogenous joint first-order Markov process*

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

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

Assumption A.2.4. *Firm's materials decision is given by*

$$M_{it} = s_t(K_{it}, p_{it}^m, \omega_{it}^H), \tag{A.2.7}$$

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 (A.2.7) 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 1.2.1 to accommodate heterogeneous materials prices.

Assumption A.2.5. *Suppose that*
(i) *Production function is of the following form*

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

(ii) $h_t\big(K_{it}, \cdot, \cdot\big)$ *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 A.2.2.
(i) *Under Assumptions A.2.5(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 A.2.5(iv)* $r_t(K_{it}, \omega_{it}^M)$ *is strictly monotone in* ω_{it}^M.

Proof. See Appendix A.4.
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 productiocity 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:

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

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.

A.2.3 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 1.2.2, and impose additional restrictions.

Assumption A.2.6. *Productivity shocks are independent conditional on last period's productivity:*

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

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

Assumption A.2.7. *The firm's exit decision depends only on ω_{it}^H and K_{it}. In particular, the firm exits if and only if*

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

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 1.3.1. However, for ω_{it}^H, differently from Equation (1.3.5), I use the following representation:

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H \sim \text{Uniform}(0,1). \tag{A.2.8}$$

In contrast to Equation (1.3.5), $g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2)$ does not include u_{it}^1. This follows by Assumption (A.2.6), 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 A.2.7, $I_{it} = 1$ if and only if $\omega_{it}^H \leqslant \bar{\omega}(K_{it})$. So, firm i's exit decision at time t depends on its capital level and current Hicks-neutral productivity. Using the representation of ω_{it}^H in Equation (A.2.8) I can write the exit rule as:

$$\begin{aligned} g_2(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2) &\leqslant \bar{\omega}(K_{it}), \\ u_{it}^2 &\leqslant g_2^{-1}\big(\omega_{it-1}^L, \omega_{it-1}^H, \bar{\omega}(K_{it})\big), \\ u_{it}^2 &\leqslant g_2^{-1}\big(\bar{r}(W_{it-1}), \bar{s}(W_{it-1}), \bar{\omega}(K_{it})\big), \\ u_{it}^2 &\leqslant \tilde{\omega}(W_{it-1}, K_{it}), \end{aligned} \tag{A.2.9}$$

where in the second line I use the fact g_2 is invertible in u_{it}^2, and the third line follows from Equation (1.3.9) 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 rformulation 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 1.3.2, I have

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

This is useful because the variable that determines whether a firm exits, (u_{it}^2), is uniform and independent from the variables I need to condition on in Equation (A.2.9), (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:

$$\tilde{\omega}(W_{it-1}, K_{it}) = \text{Prob}(I_{it} = 1 \mid W_{it-1}, K_{it}) \equiv p(W_{it-1}, K_{it}). \tag{A.2.10}$$

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:

$$u_{it}^2 \mid W_{it-1}, K_{it}, (I_{it} = 1) \sim \text{Uniform}\big(p(W_{it-1}, K_{it}), 1\big). \tag{A.2.11}$$

As shown in Equation (1.3.7), 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 $\big(W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1)\big)$. This result uses Assumption A.2.6 and is given by the following lemma.

Lemma A.2.3.

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

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

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

where the third line follows because u_{it}^1 and u_{it}^2 are independently distributed conditional on W_{it-1} by Lemma A.3.5. □

From this result, I obtain

$$\begin{aligned}
u_{it}^2 \mid W_{it-1}, K_{it}, u_{it}^1, (I_{it} = 1) &\sim \text{Uniform}(p(W_{it-1}, K_{it}, u_{it}^1), 1), &\text{(A.2.12)} \\
&\sim \text{Uniform}(p(W_{it-1}, K_{it}), 1). &\text{(A.2.13)}
\end{aligned}$$

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 A.2.4. *We have that*

$$\omega_{it}^{H} \equiv c_2\left(W_{it-1}, u_{it}^2\right), \qquad u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1). \tag{A.2.14}$$

Proof. See Appendix A.4.
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 use Lemme (A.2.3), which gives the distribution of u_{it}^2 for the firms that stay, to write u_{it}^2 as:.

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

This result uses the distribution of u_{it}^2 conditional on firms that stay, which is given in Equation (A.2.11). 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.

A.3 Supplementary Lemmas

Lemma A.3.1. *Suppose x, y and z are scalar and continuous random variables with a joint probability density function given by $f(x, y, z)$. Assume that (x, y) are jointly independent from z. Then x and z are independent conditional on y.*

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

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

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

Lemma A.3.2. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ be continuously differentiable functions. If there exists a differentiable function $s : \mathbb{R}_+^2 \to \mathbb{R}$:*

$$f(zh(x)) = s(x, z) \tag{A.3.1}$$

Then

$$\frac{\log'\big(h(x)\big)}{z} = \frac{s_1(z, x)}{s_2(z, x)} \tag{A.3.2}$$

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

Proof. By the assumptions we can differentiate equation (A.3.1). Differentiating with respect to z to get

$$f'(zh(x))h(x) = s_1(z,x). \tag{A.3.3}$$

Differentiating with respect to x yields

$$f'(zh(x))zh'(x) = s_2(z,x). \tag{A.3.4}$$

Taking the ratio of Equation (A.3.4) and (A.3.3) I obtain

$$\frac{zh'(x)}{h(x)} = \frac{s_2(z,x)}{s_1(z,x)}, \tag{A.3.5}$$

which gives

$$\log'\big(h(x)\big)z = \frac{s_2(z,x)}{s_1(z,x)}. \tag{A.3.6}$$

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

Lemma A.3.3. *Let $f : \mathbb{R}_+^2 \to \mathbb{R}$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ are differentiable functions. If there exists a differentiable function $s : \mathbb{R}_+^3 \to \mathbb{R}$ with*

$$f(w, zh(x)) = s(w,x,z) \tag{A.3.7}$$

then

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

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

- *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.3.7) 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.3.7) with respect to x we obtain

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

Taking the ratio between the two

$$\frac{zh'(x)}{h(x)} = \frac{s_3(w,z,x)}{s_2(w,z,x)}. \tag{A.3.8}$$

which gives

$$\frac{z}{\log'\big(h(x)\big)} = \frac{s_3(w,z,x)}{s_2(w,z,x)}. \tag{A.3.9}$$

Taking derivative with respect to w

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

$\square$

Lemma A.3.4. *Consider the following model*

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

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

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

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

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

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

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

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

$$f\big(zh(x)\big) + g(x) = \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}(x). \tag{A.3.10}$$

I will show that if Equation (A.3.10) holds, then (f,h,g) and $(\tilde{f}, \tilde{h}, \tilde{g})$ have to obey the normalization restrictions below

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

for $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}$. To show this, I will take the derivatives of Equation (A.3.10) with respect to x and z. Taking derivative with respect to z yields

$$f'\big(zh(x)\big)h(x) = \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x). \tag{A.3.11}$$

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

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

Rearranging this to collect similar terms, I obtain

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

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

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

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

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

By Equation (A.3.11) we have that $f'\big(zh(x)\big)h(x) = \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)$. Using this

$$\begin{aligned} z\Big(f'\big(zh(x)\big)h(x)\log'(h(x)) - f'\big(zh(x)\big)h(x)\log'(\tilde{h}(x))\Big) &= \tilde{g}'(x) - g'(x) \\ zf'\big(zh(x)\big)h(x)\Big(\log'(h(x)) - \log'(\tilde{h}(x))\Big) &= \tilde{g}'(x) - g'(x). \end{aligned} \tag{A.3.12}$$

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

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

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

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

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

$$\log'(h(x)) = \log'(\tilde{h}(x)), \qquad \tilde{g}'(x) = g'(x) \tag{A.3.13}$$

Next, using equation (A.3.12) we also have

$$\tilde{g}'(x) = g'(x) \tag{A.3.14}$$

Integrating these equations, there exists λ and a such that

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

Now using these results and Equation (A.3.11) we solve for $f\big(zh(x)\big)$ and $\tilde{f}\big(zh(x)\big)$

$$f\big(zh(x)\big) = \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}(x) - g(x) \tag{A.3.15}$$

$$= \tilde{f}\big(z\lambda h(x)\big) + a \tag{A.3.16}$$

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

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

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

$$y(x, z) := \mathbb{E}[y \mid z, x] = \lambda \log z + h(x) + g(x). \tag{A.3.17}$$

which is testable by estimating $\mathbb{E}[y \mid z, x]$ from data. If part is trivial. To show the only if part, by fundamental theorem of calculus, Equation (A.3.17) implies that $\frac{\partial t(x,z)}{\partial \log z} = \lambda$. Using this

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

From this, I obtain

$$f'(zh(x))h(x) = \frac{\lambda}{z}. \tag{A.3.18}$$

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 $\mathbb{E}[y \mid z, x]$ with respect to $\log(z)$ is constant. □

Lemma A.3.5. *Under Assumption A.2.6 u^1_{it} and u^2_{it} are independently distributed conditional on W_{it-1}.*

Proof. We have that

$$\omega^L_{it} = g_1(\omega^L_{it-1}, \omega^H_{it-1}, u^1_{it}) \qquad \omega^H_{it} = g_2(\omega^L_{it-1}, \omega^H_{it-1}, u^1_{it}, u^2_{it})$$

By assumption A.2.6, 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}. □

A.4 Proofs

Proof of Proposition 1.2.1

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:

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

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

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

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

$$\begin{aligned} \min_{L_{it}, M_{it}} \quad & p_t^l L_{it} + p_t^m M_{it} \\ \text{s.t.} \quad & F\left(K_{it}, h(K_{it}, \omega_{it}^L L_{it}, M_{it})\right) \omega_{it}^H \mathcal{E}_{it}(\mathcal{I}_{it}) \geqslant \tilde{Y}_{it}, \end{aligned} \tag{A.4.1}$$

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 (A.4.1) can be rewritten as

$$\begin{aligned} \min_{M_{it}, \bar{L}_{it}} \quad & \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_t \\ \text{s.t.} \quad & F\big(K_{it}, h(K_{it}, \bar{L}_{it}, M_{it})\big) \omega_{it}^H \geqslant \bar{Y}_{it}(\mathcal{I}_{it}), \end{aligned} \tag{A.4.2}$$

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:

$$
\begin{aligned}
C(\tilde{Y}_{it}, K_{it}, \omega_{it}^H, \bar{p}_{it}) &= \min_{\bar{L}_{it}, M_{it}} \Big\{ \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : \tilde{Y}_{it} \leqslant F\big(K_{it}, h(K_{it}, \bar{L}_{it}, M_{it})\big)\omega_{it}^H \Big\}, \\
&= \min_{\bar{L}_{it}, M_{it}} \Big\{ \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}) \leqslant h(K_{it}, \bar{L}_{it}, M_{it}) \Big\}, \\
&= \min_{\bar{L}_{it}, M_{it}} \Big\{ \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : 1 \leqslant h\Big(K_{it}, \bar{L}_{it}/F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}), M_{it}/F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it})\Big) \Big\}, \\
&= \min_{\bar{L}_{it}, M_{it}} \Big\{ F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}) \left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h\big(K_{it}, \bar{L}_{it}, M_{it}\big) \Big\}, \\
&= F^{-1}(\tilde{Y}_{it}/\omega_{it}^H, K_{it}) \min_{\bar{L}_{it}, M_{it}} \Big\{ \left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h\big(K_{it}, \bar{L}_{it}, M_{it}\big) \Big\}, \\
&\equiv C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_2(K_{it}, \bar{p}_{it}^l, p_t^m). \qquad \text{(A.4.3)}
\end{aligned}
$$

The second line follows by the assumption that $F(\cdot,\cdot)$ is strictly monotone in its second argument. The third and fourth lines are due to homotheticity property of $h(K_{it},\cdot,\cdot)$. In the last line I define two new functions that characterize the cost function. Equation (A.4.3) 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:

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

The ratio of materials to the effective labor equals:

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

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

$$
\tilde{M}_{it} \equiv r(K_{it}, \omega_{it}^L, p_t^m, p_t^l) \equiv r_t(K_{it}, \omega_{it}^L), \qquad \text{(A.4.4)}
$$

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:

$$\tilde{M}_{it} \equiv \frac{\tilde{C}_m(K_{it},\tilde{p}_{it})\omega_{it}^L}{\tilde{C}_l(K_{it},\tilde{p}_{it})}, \tag{A.4.5}$$

where $\tilde{p}_{it} := \dfrac{\bar{p}_{it}^l}{p_t^m}$, $\tilde{C}_m := C_m(K_{it},\tilde{p}_{it},1)$ and $\tilde{C}_l(K_{it},\tilde{p}_{it}) := C_l(K_{it},\tilde{p}_{it},1)$. Taking the logarithm of Equation (A.4.5), 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

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

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 subtitution between effective labor and materials. By Assumption 1.2.1(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 1.3.1

By Assumption 1.2.2 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 (1.3.1)

$$g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{A.4.6}$$

Since $g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$ is strictly monotone in u_{it}^1, Equation (A.4.6) implies independence of u_{it}^1 and $\mathcal{I}_{it-1}$ conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$

$$u_{it}^1 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{A.4.7}$$

Note that by normalization u_{it}^1 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$ and by timing assumption $(K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H) \in \mathcal{I}_{it-1}$. Therefore, Equation (A.4.7) 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 (1.2.9) and (1.2.11), $(\omega_{it-1}^L, \omega_{it-1}^H)$ can be expressed as functions of W_{it-1}. Thus

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

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

Proof of Lemma 1.3.2

By Assumption 1.2.2, 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 (1.3.1) and (1.3.5) 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.3.1 imply

$$u_{it}^2 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1. \tag{A.4.8}$$

It follows from Equation (A.4.8), 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

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

Using Equations (1.2.9) and (1.2.11), $(\omega_{it-1}^L, \omega_{it-1}^H)$ 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 1.4.3

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 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

$$\theta_{it}^L = f_2 h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}(K_{it}, \tilde{M}_{it})L_{it}, \tag{A.4.9}$$

$$\theta_{it}^M = f_2 h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it}, \tag{A.4.10}$$

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

$$\bar{h}_2(K_{it}, \tilde{M}_{it}) = h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}), \tag{A.4.11}$$

$$\bar{h}_1(K_{it}, \tilde{M}_{it}) = h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})r_1(K_{it}, \tilde{M}_{it}). \tag{A.4.12}$$

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 $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. Labor-augmenting productivity is given by

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

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

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

Since h_1 is not identified, θ^K_{it} is not identified. Finally, to see that the elasticity of substitution is not identified note that it is defined as

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

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

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

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 1.4.4

If production function takes the form given Equation (1.4.7) the output elasticities with respect to labor and materials, as a function of f and h, can be written as

$$\begin{aligned} \theta^L_{it} &= f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) r(\tilde{M}_{it}) L_{it} \\ \theta^M_{it} &= f_2 h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) M_{it}. \end{aligned} \tag{A.4.13}$$

Since I already showed in Equation (1.4.6) that θ^L_{it} and θ^M_{it} are identified, the right-hand sides of these equations are identified. The identification of θ^M_{it} immediately implies that $h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$ is identified from $(f_2, \theta^M_{it}$. Taking the derivative of the reduced form function $\bar{h}$ and using $\bar{h}(\tilde{M}_{it}) = h(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$ I obtain

$$\bar{h}_1(\tilde{M}_{it}) = h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it}) + h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}), \tag{A.4.14}$$

where $\bar{r}'(\tilde{M}_{it})$ denotes the derivative of $\bar{r}(\tilde{M}_{it})$. Therefore, the right-hand side of Equation (A.4.14) is identified from $\bar{h}(\tilde{M}_{it})$. Now, taking the ratio of θ^L_{it}/L_{it} and $f_2\bar{h}_1(\tilde{M}_{it}) - \theta^M_{it}/M_{it}$

gives

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

This shows that the derivative of $\log(r(\tilde{M}_{it})$ with respect to $\tilde{M}_{it}$ can be identified from $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$ 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 $\omega^L_{it} = r(\tilde{M}_{it})$, and $\log(r(\tilde{M}_{it}))$ is identified up to a constant, ω^L_{it} 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:

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

This concludes the proof.

Proof of Proposition 1.4.5

The elasticity of substitution is given by

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

If production function takes the form in Equation (1.4.7), we can derive σ^{ML}_{it} as

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

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

Proof of Proposition A.2.2

This proof closely follows the same lines as the proof of Proposition 1.2.1. 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 A.2.5, can be written as:

$$\begin{aligned} \min_{L_{it}, M_{it}} \quad & p^l_t L_{it} + p^m_{it} M_{it} \\ \text{s.t.} \quad & \mathbb{E}\big[F\left(K_{it}, h(K_{it}, L_{it}, M_{it})\right) \omega^H_{it} \epsilon_{it} \mid \mathcal{I}_{it}\big] \geqslant \bar{Y}_{it} \end{aligned}$$

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

$$\begin{aligned} \min_{L_{it},M_{it}} \quad & p_t^l L_{it} + p_{it}^m M_{it} \\ \text{s.t.} \quad & F\left(K_{it}, h(K_{it}, L_{it}, M_{it})\right) \omega_{it}^H \mathcal{E}_{it} \geqslant \bar{Y}_{it}, \end{aligned} \tag{A.4.15}$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it}]$. The cost minimization problem in Equation (A.4.15) is given by

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

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 (A.4.3), the cost function can be expressed as:

$$C(\bar{p}_{it}, \tilde{Y}_{it}, K_{it}, \omega_{it}^H) = C_1(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_2(K_{it}, p_t^l, p_{it}^m). \tag{A.4.16}$$

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

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

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

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

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 (A.4.4) in the proof of Proposition 1.2.1, with $\omega_{it}^L = \omega_{it}^M$, $\bar{p}_{it}^l = 1/\omega_{it}^M$, $p_t^m = p_t^l$ and $\tilde{M}_{it} = L_{it}/R_{it}^m$. Therefore, we can treat R_{it}^m as materials input and treat $\omega_{it}^M = 1/p_{it}^M$ as the materials-augmenting productivity 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 1.2.1 and, therefore, is omitted.

Proof of Lemma A.2.1

This proof closely follows the proof of Lemma 1.3.1. By Assumption A.2.1 we have

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

Monotonicity of g_1 with respect to its last argument and Lemma A.3.1 imply

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

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

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

Using Equations (1.2.9) and (1.2.11) we substitute $(\omega^L_{it}, \omega^H_{it})$ as functions of (W_{it-1}) to obtain

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

which shows the desired result.

Proof of Lemma A.2.2

This proof closely follows the proof of Lemma 1.3.2. By Assumption A.2.1 we have

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

Monotonicity of g_1 and g_2 with respect to their last arguments and Lemma A.3.1 imply that

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

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

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

Using Equations (1.2.9) and (1.2.11) to substitute $(\omega^L_{it-1}, \omega^H_{it-1})$ as functions of W_{it-1}, I obtain

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

which shows the desired result.

Proof of Lemma A.2.4

By Assumption A.2.6 we have that

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

Using the Skorokhod representation of ω^H_{it} in Equation (A.2.8) we write

$$g_2(\omega^L_{it-1}, \omega^H_{it-1}, u^2_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1}, g_1(\omega^L_{it-1}, \omega^H_{it-1}, u^1_{it}) \mid \omega^L_{t-1}, \omega^H_{t-1}. \tag{A.4.17}$$

By monotonicity of g_1 and g_2 in their last arguments, u^2_{it} is (conditionally) independent of

$(\mathcal{I}_{it-1}, u_{it}^1)$

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

It follows from Equation (A.4.17) 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 $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$

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

which implies

$$u_{it}^2 \mid K_{it}, W_{it-1}, u_{it}^1 \sim \text{Uniform}(0,1). \tag{A.4.18}$$

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

$$\begin{aligned} M_{it} &= s\left(K_{it}, \omega_{it}^H, \omega_{it}^L\right), \\ &= s\left(K_{it}, g_2\left(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^2\right), c_1\left(W_{it-1}, u_{it}^1\right)\right), \\ &= s\left(K_{it}, g_2\left(\tilde{r}\left(W_{it-1}\right), \tilde{s}\left(W_{it-1}\right), u_t^2\right), c_1\left(W_{it-1}, u_{it}^1\right)\right), \\ &\equiv \bar{s}\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right). \end{aligned} \tag{A.4.19}$$

The intuition is similar to that of Lemma 1.3.1. Employing strict monotonicity of $\bar{s}$ in u_{it}^2 and Equation (A.4.18), we can use Equation (A.4.19) to identify u_{it}^2. In particular, u_{it}^2 equals

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1), \tag{A.4.20}$$

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

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

This concludes the proof.

A.5 Identification

In this section I show that the homothetic and strong homothetic separable production functions in Section 1.4.5 are identified except special cases using the moment restriction in Equation (1.5.5). The identification results follow Roehrig (1988)[4].

[4]Benkard 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.

A.5.1 Identification for Homothetic Production Function

Under homotheticity assumption, the function function takes the following form

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

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

$$y_{it} = vk_{it} + f\big(\tilde{L}_{it}\bar{h}(\tilde{M}_{it})\big) + g\big(W_{it-1}, u_{it}^1, u_{it}^2)\big) + \epsilon_{it}, \qquad \mathbb{E}[\epsilon_{it} \mid k_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}] = 0. \tag{A.5.1}$$

Under homothetic model the control variables are $u_{it}^1 = \tilde{M}_{it}$ and $u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. Substituting these into Equation (A.5.1), I obtain

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

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:

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

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 from the random variables

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

By the moment restriction in Equation (1.5.5), we have

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

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

$$\alpha w + f\big(zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) = \tilde{\alpha} w + \tilde{f}\big(z\tilde{h}(x)\big) + \tilde{g}\big(x, t, s(w, z, x, t)\big). \tag{A.5.2}$$

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 A.5.1. *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 $\Omega_{\log}$ is defined in Lemma A.3.4. (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.3.4, 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 (A.5.2) with respect to w we obtain

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

Rearranging this equation:

$$g_3\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t) - \tilde{g}_3\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t) = \tilde{\alpha} - \alpha. \tag{A.5.3}$$

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

$$\bar{g}_3\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t) = \tilde{\alpha} - \alpha. \tag{A.5.4}$$

Taking the derivatives of Equation (A.5.4) with respect to w and z

$$\bar{g}_{33}\big(x,t,s(w,z,x,t)\big)s_1^2(w,z,x,t) + \bar{g}_3\big(x,t,s(w,z,x,t)\big)s_{11}(w,z,x,t) = 0.$$
$$\bar{g}_{33}\big(x,t,s(w,z,x,t)\big)s_1(w,z,x,t)s_2(w,z,x,t) + \bar{g}_3\big(x,t,s(w,z,x,t)\big)s_{12}(w,z,x,t) = 0.$$

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

$$\alpha = \tilde{\alpha}, \qquad g_3\big(x,t,s(w,z,x,t)\big) - \tilde{g}_3\big(x,t,s(w,z,x,t)\big) = 0. \tag{A.5.5}$$

This shows that α and g_3 are identified. Next, taking the derivative of Equation $(A.5.4)$ with respect to t gives

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

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

$$g_2\big(x,t,s(w,z,x,t)\big) = \tilde{g}_2\big(x,t,s(w,z,x,t)\big). \tag{A.5.6}$$

Therefore $g_2\big(x,t,s(w,z,x,t)\big)$ is also identified. Taking the derivative of Equation $(A.5.4)$ with respect to z to obtain

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

Using $g_3 = \tilde{g}_3$ obtained in Equation in (A.5.5) gives

$$f'\big(zh(x)\big)h(x) = \tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x). \tag{A.5.7}$$

Finally, taking the derivative of Equation $(A.5.4)$ with respect to x

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

Rearranging

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

Using Equation (A.5.7) we can substitute $f'\big(zh(x)\big)h'(x)$ and, with some algebra, get

$$z\Big(\tilde{f}'\big(z\tilde{h}(x)\big)\tilde{h}(x)(\log'(h(x)) - \log'(\tilde{h}(x)))\Big) = \tilde{g}_1\big(x,t,s(w,z,x,t)\big) - g_1\big(x,t,s(w,z,x,t)\big) \tag{A.5.8}$$

Taking the derivative with respect to w

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

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

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

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

$$\bar{g}_1(x) \equiv g_1\big(x,t,s(w,z,x,t)\big) - g_1'\big(x,t,s(w,z,x,t)\big) \tag{A.5.9}$$

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 (A.5.8) can be written as

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

Now applying a result in Lemma A.3.4 we obtain

$$f(x) = \tilde{f}(\lambda x) + a, \qquad h(x) = \frac{\tilde{h}(x)}{\lambda}, \qquad g(x) = \tilde{g}(x) - a, \qquad \alpha = \tilde{\alpha} \tag{A.5.10}$$

This concludes the proof. □

Identification for Strong Homothetic Production Function

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

$$y_{it} = f\big(K_{it}, L_{it}\bar{h}(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{A.5.11}$$

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

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

Under the strong homothetic separable model the control variables are $u_{it}^1 = \tilde{M}_{it}$ and $u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. Substituting these into Equation (A.5.11) gives:

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

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

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

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

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

By the moment restriction in Equation (1.5.5), we have

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

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

$$f\big(w, zh(x)\big) + g\big(x, t, s(w, z, x, t)\big) = \tilde{f}\big(w, z\tilde{h}(x)\big) + \tilde{g}\big(x, t, s(w, z, x, t)\big). \tag{A.5.12}$$

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

Proposition A.5.2. *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) $\mathbb{E}[s_1^2(w, z, x, t) \mid x, t] > 0$. (vi) Define $q := s_2(w, z, x, t) \log'\big(h_0(x)\big)z - s_3(w, z, x, t)$. I assume that $\mathbb{E}[q^2 \mid x, s, t] > 0$ 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.3.4.*

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

$$y_1(w, z, x, t) = f_1\big(w, zh(x)\big) + g_2\big(x, s(w, z, x, t), t\big)s_1(w, z, x, t), \tag{A.5.13}$$

$$y_2(w, z, x, t) = f_2\big(w, zh(x)\big)h(x) + g_2\big(x, s(w, z, x, t), t\big)s_2(w, z, x, t), \tag{A.5.14}$$

$$y_3(w, z, x, t) = f_2\big(w, zh(x)\big)h'(x) + g_2\big(x, s(w, z, x, t), t\big)s_3(w, z, x, t) + g_1\big(x, s(w, z, x, t), t\big). \tag{A.5.15}$$

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

$$y_2(w, z, x, t) \log'\big(h(x)\big)z - y_3(w, z, x, t) = \tag{A.5.16}$$

$$g_2\big(x, s(w, z, x, t), t\big)\big(s_2(w, z, x, t) \log'\big(h(x)\big)z - s_3(w, z, x, t)\big) - g_1\big(x, s(w, z, x, t), t\big). \tag{A.5.17}$$

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

$$y_2(w,z,x,t)\log'\left(\tilde{h}(x)\right)z - y_3(w,z,x,t) = \tag{A.5.18}$$

$$\tilde{g}_2\big(x, s(w,z,x,t), t\big)\big(s_2(w,z,x,t)\log'\left(\tilde{h}(x)\right)z - s_3(w,z,x,t)\big) - \tilde{g}_1\big(x, s(w,z,x,t), t\big). \tag{A.5.19}$$

In Equation (A.5.16), the unknown functions are $h(x)$, $g_1\big(x, s(w,z,x,t), t\big)$ and $g_2\big(x, s(w,z,x,t), t\big)$, 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 $\big(x, s(w,z,x,t), t\big)\big(s_2(w,z,x,t)\log'\left(\tilde{h}(x)\right)z - s_3(w,z,x,t)\big)$. Therefore, g_2 and g_1 can be identified from Equations (A.5.16) and (A.5.18) 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

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

where $\bar{g}_1$ and $\bar{g}_2$ are known functions that can be derived from Equations (A.5.16) and (A.5.18). 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

$$g_2\big(x, s(w,z,x,t), t\big) = \tilde{g}_2\big(h^{-1}(\tilde{h}(x)), s(w,z,x,t), t\big), \tag{A.5.20}$$

$$g_1\big(x, s(w,z,x,t), t\big) = \tilde{g}_1\big(h^{-1}(\tilde{h}(x)), s(w,z,x,t), t\big). \tag{A.5.21}$$

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

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

Once we condition on w, this problem falls into the case given in Lemma (A.3.4) with a slight modification. Therefore, $h(x)$ is identified up to a scale:

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

This implies identification of $g_1\big(h(x), s(w,z,x,t), t\big)$ and $g_2\big(h(x), s(w,z,x,t), t\big)$ from Equations (A.5.20) and (A.5.21). With these identification results, identification of $f_0\big(w, zh(x)\big)$ follows by Equation (A.5.13) and (A.5.14). □

A.6 Additional Results

A.6.1 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 $P(\omega_{it}^H \mid \mathcal{I}_{it-1}) = P(\omega_{it}^H \mid \omega_{it-1}^H)$ (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 1.3. Productivity can be represented as:

$$\omega_{it}^H = g(\omega_{it-1}^H, u_{it}) \qquad u_{it} \mid \omega_{it-1}^H \sim \text{Uniform}(0,1), \tag{A.6.1}$$

where $g(\omega_{it-1}^H, u_{it})$ is strictly increasing in u_{it} by construction. By Markov Assumption $\omega_{it}^H \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^H$. Substituting ω_{it}^H using Equation (A.6.1) we obtain

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

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

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

Note that $s(k_{it}, \cdot)$ is a strictly increasing function and $g(\omega_{it-1}, \cdot)$ is also strictly increasing function by construction. So $\equiv \tilde{s}$ is strictly increasing in u_{it}. It follows from Lemma 1.3.1 that

$$u_{it} \mid k_{it}, m_{it-1}, k_{it-1} \sim \text{Uniform}(0,1). \tag{A.6.2}$$

Therefore, we can recover u_{it} as the conditional CDF of m_{it}: $u_{it} = F_{m_{it}}(m_{it} \mid k_{it}, m_{it-1}, k_{it-1})$. This suggests that, we can 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

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c(m_{it-1}, k_{it-1}, u_{it}) + \epsilon_{it} \tag{A.6.3}$$

with $E[\epsilon_{it} \mid I_{it}] = 0$, 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 $E[\xi_{it} \mid I_{it-1}] = 0$. We can obtain second model

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_2(m_{it-1}, k_{it-1}) + \xi_{it} + \epsilon_{it} \tag{A.6.4}$$

with $E[\xi_{it} \mid I_{it}] = 0$. Now we can estimate the parameters (β_k, β_l) and unknown functions $c_1(\cdot), c_2(\cdot)$ in Equation (A.6.3) and (A.6.4) using the following moment restictions.

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

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).

A.6.2 Nested CES Production Function

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

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

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 1.2.2. Taking the logarithm of this production function we write

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

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

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

where $\tilde{K}_{it}$ 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

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

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

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

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 1.2.1. The model parameters can be estimated using the control functions I develop with the following estimating equation

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

where $u_{it}^{1} = \tilde{L}_{it}$ in the Nested CES model because it falls into the model in Equation (1.4.7). We can estimate the model using the objective function in Equation (1.4.7). One can show that the sum of the flexible input elasticites are identified from the model parameters as:

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

where $x = M(\tilde{L}_{it} + \gamma_2)^{1/\sigma_1}$. Note that $(1-\beta_k)\gamma_1$ and 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 revenu shares of labor and materials as follows:

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

And finally, the output elasticity of capital is identified as

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

A.6.3 CES

In this section, I consider the CES production function:

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

Using homotheticity property we can write

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

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

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

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

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

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

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

with the same objective function in Section 1.5. One can again show that the sum of the flexible input elasticites is identified from the model parameters as:

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

where $x = M_{it}(\tilde{L}_{it} + \gamma_2)$. Note that $(1-\beta_l-\beta_m)$ and 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 revenu shares of labor and materials

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

And finally, the output elasticity of capital is identified as

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

A.7 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.

A.7.1 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 A-1 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 1.6.1 cannot be explained by potential measurement error in capital. Figure A-2, 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.

A.7.2 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 A-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 A-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.

A.7.3 Selection

In this section, I estimate output elasticities and markups after accounting for non-random firm exit as described in Subsection A.2.3. Figure A-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.

A.8 Tables and Figures

Table A.1: Descriptive Statistics - Chile

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

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

Table A.2: Descriptive Statistics - Colombia

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

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

Table A.3: Descriptive Statistics - India

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

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

Table A.4: Descriptive Statistics - US

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

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

Table A.5: Descriptive Statistics - Turkey

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

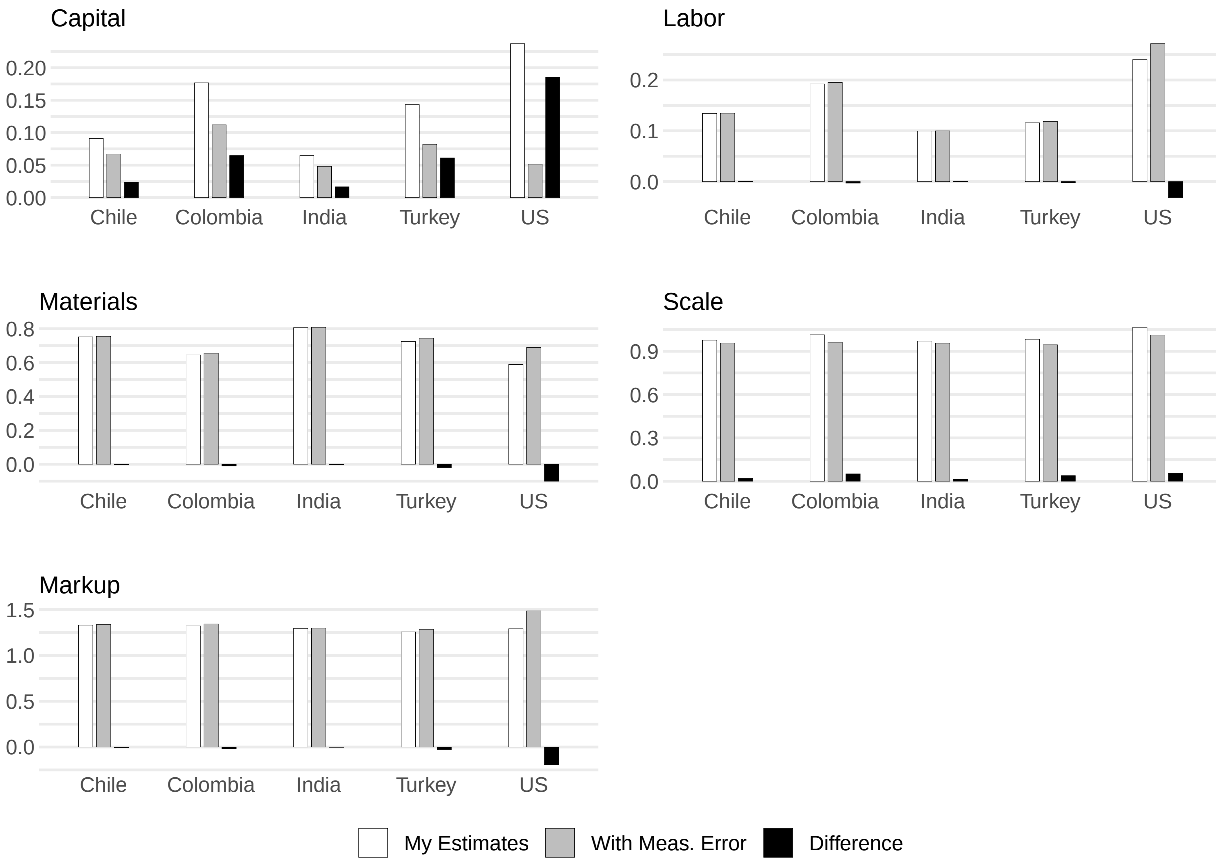
Note: Descriptive Statistics for Turkey. Column 3-5 shows each industry share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Column 6-8 reports the number of active plants. The last row provides information about the industries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

Table A.6: Unweighted Average Output Elasticities

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

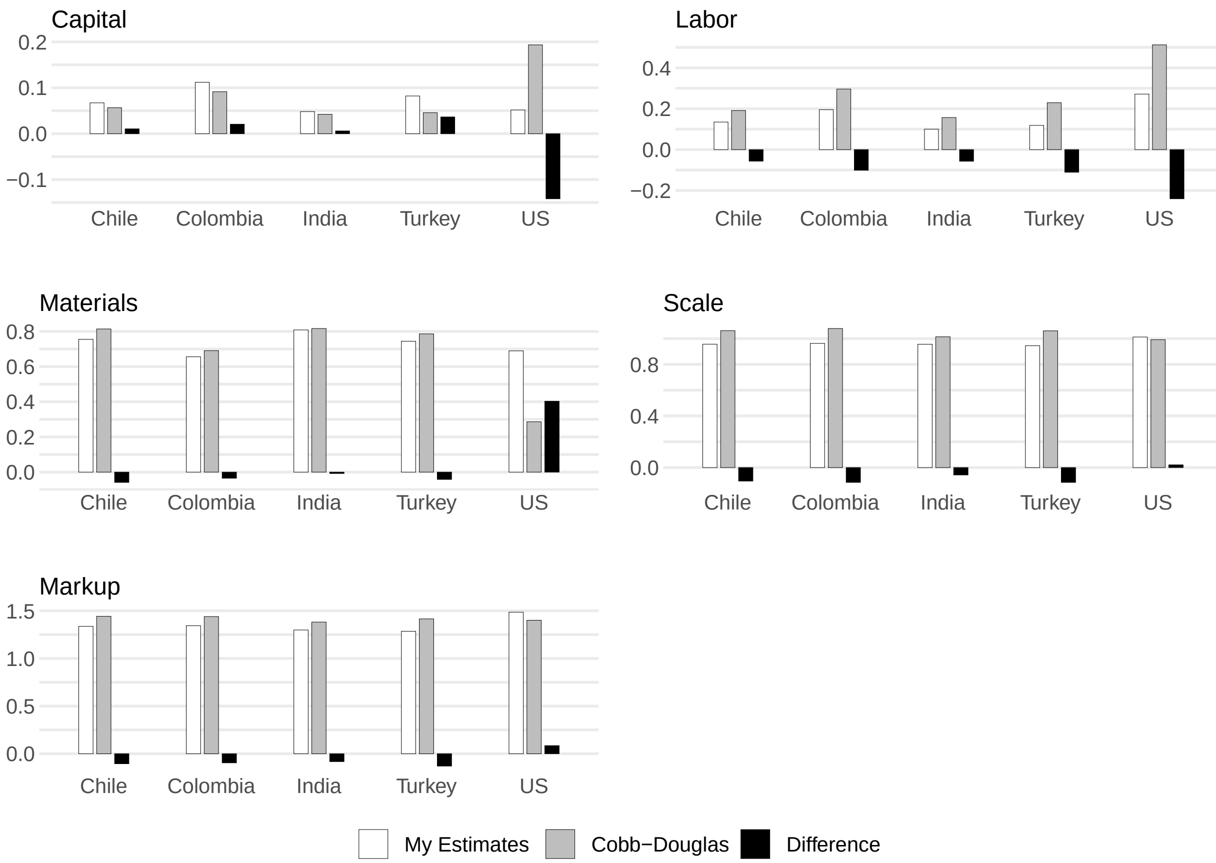
Note: Comparison of unweighted average output elasticities produced by different methods. FA refers to my estimates, ACF refers to Ackerberg et al. (2015) estimates and OLS is Cobb-Douglas estimated by OLS. For each year and industry, sales-weighted averages are calculated, and then simple averages are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

Figure A-1: Comparison of Estimates with and without Measurement Error



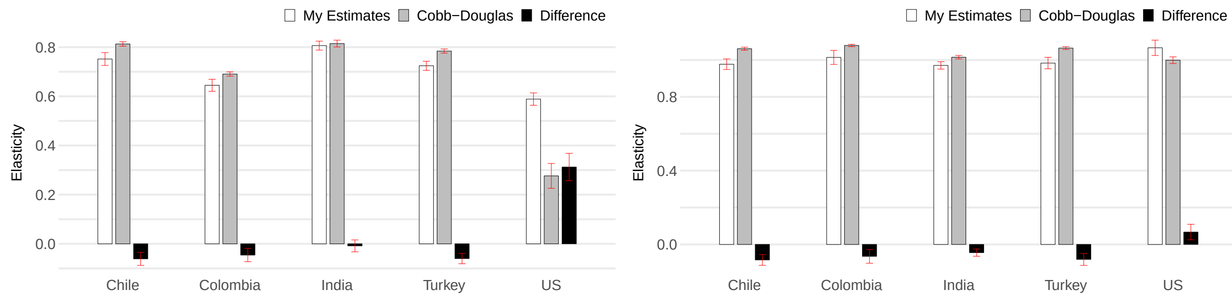
Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without measurement error in capital. White bars report the estimates from the main text and grey bars report the average of 100 estimates obtained from simulated datasets as described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

Figure A-2: 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).

Figure A-3: Comparison of Output Elasticities

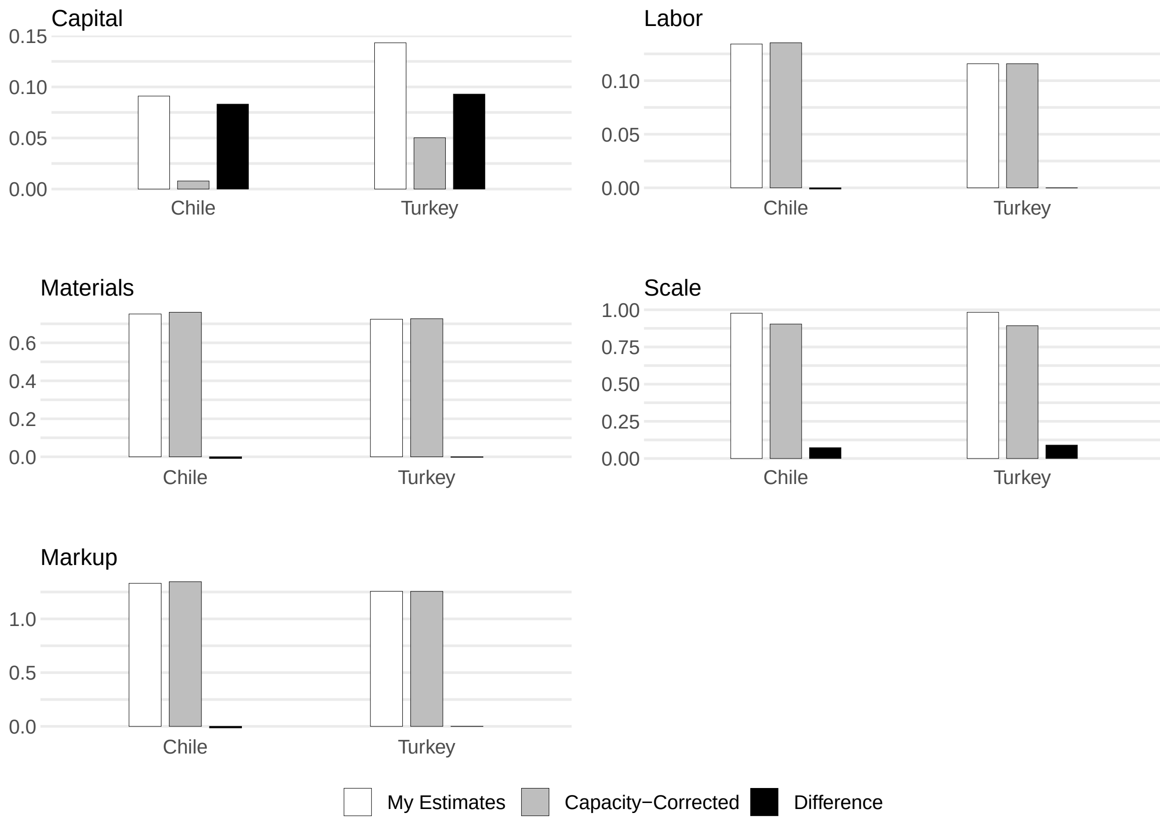


(a) Materials Elasticity

(b) Returns to Scale

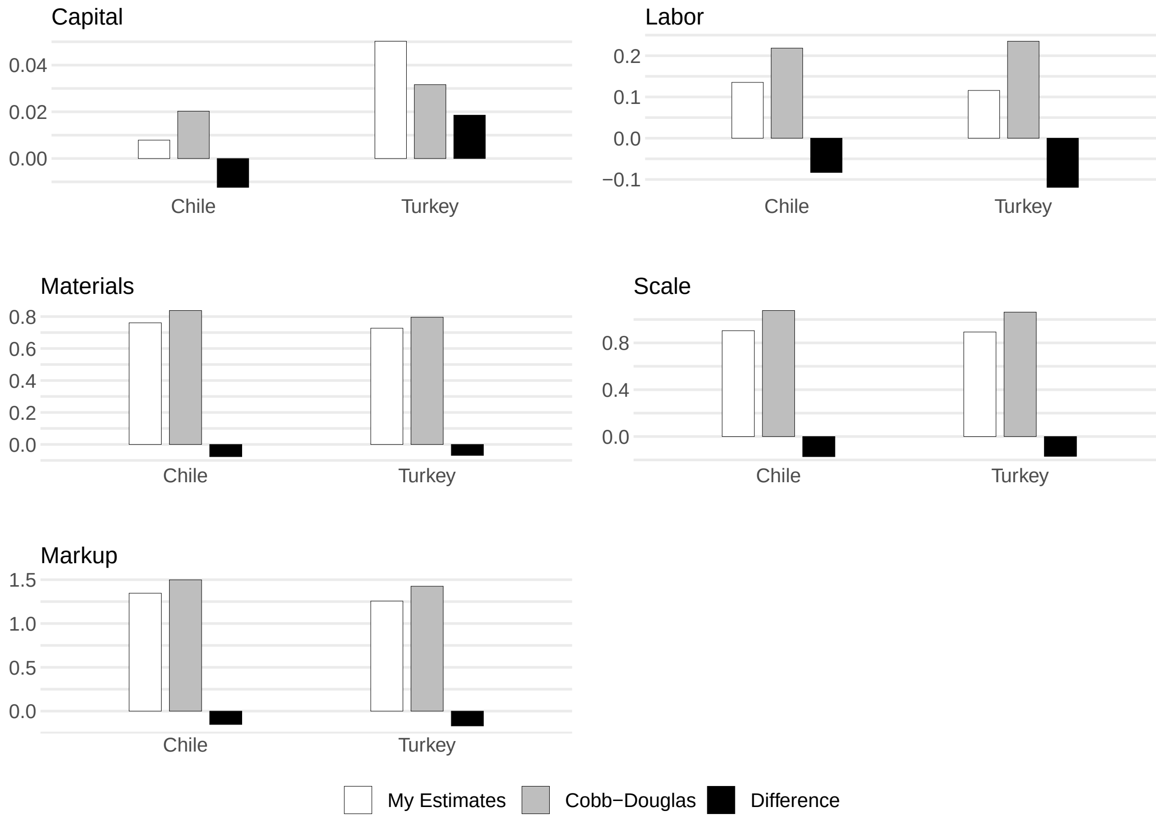
Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Cobb-Douglas estimated by ACF (grey) for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages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

Figure A-4: Comparison of Estimates with and without Capacity Utilization Correction



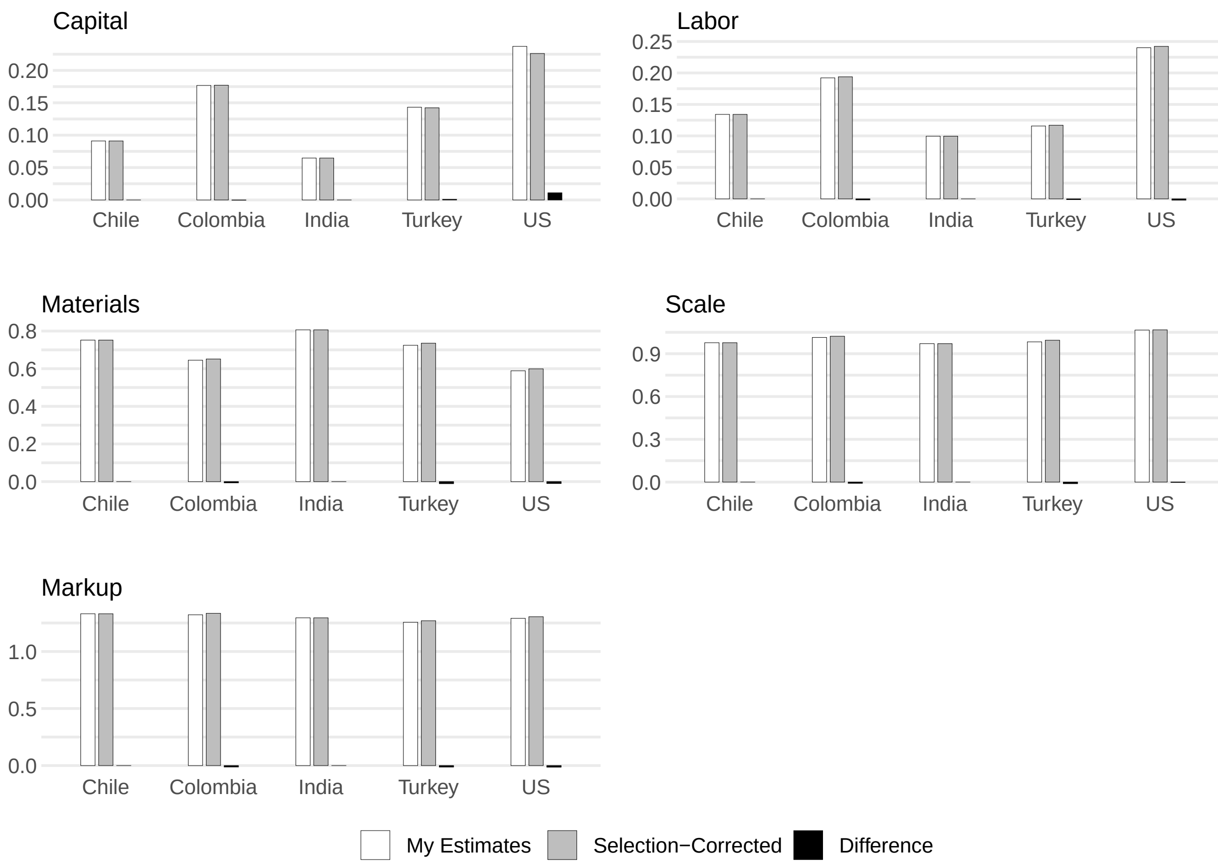
Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without capacity utilization in capital. White bars report the estimates from the main text and grey bars report 100 estimates obtained after correcting for capacity utilization as described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

Figure A-5: Comparison of Estimates Across Methods with Capacity Utilization Correction



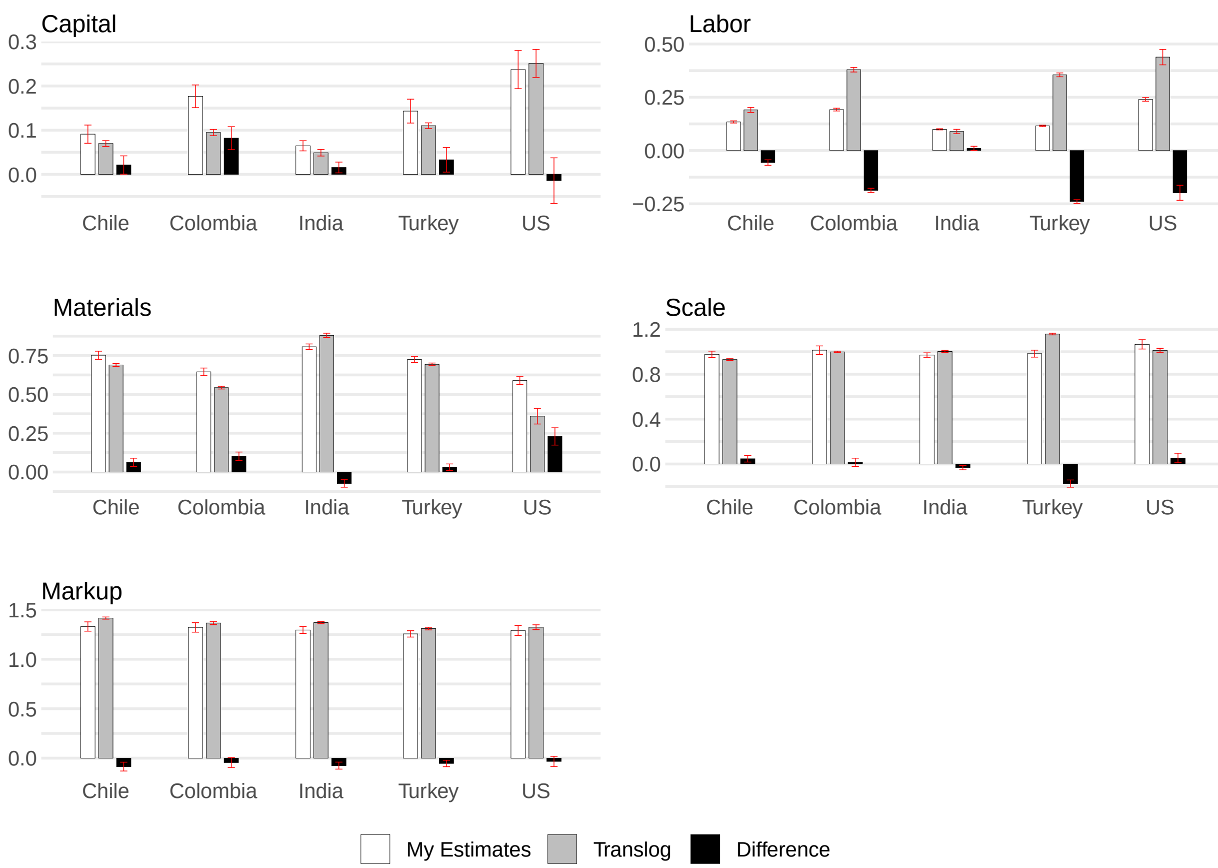
Notes: This figure compares sales-weighted output elasticities and markups estimated using my method and Cobb-Douglas with measurement errors in capital. Cobb-Douglas specification estimated using the Ackerberg et al. (2015).

Figure A-6: Comparison of Estimates with and without Selection Correction



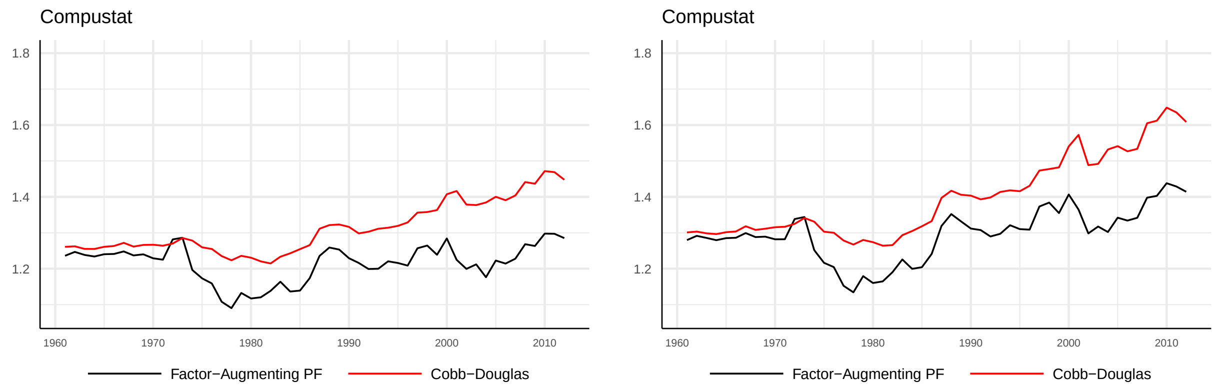
Notes: This figure compares sales-weighted output elasticities and markups estimates obtained using my method with and without selection correction. White bars report the estimates from the main text and grey bars report the estimates after accounting for selection.

Figure A-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.

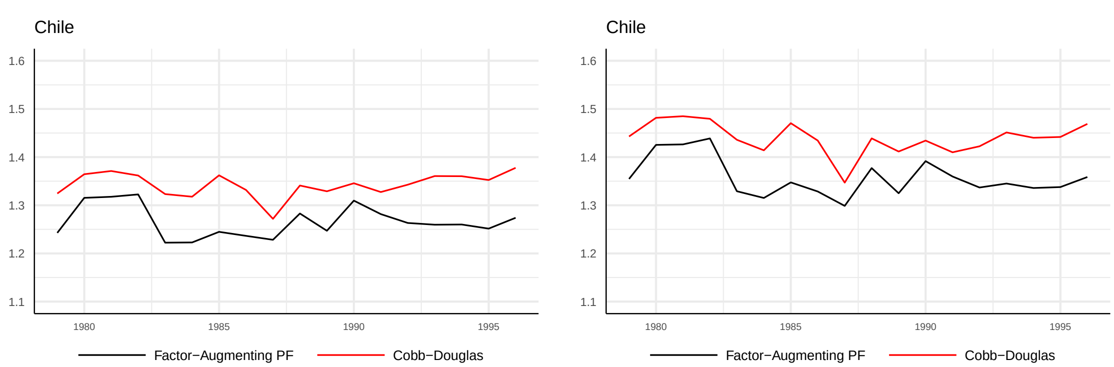
Figure A-8: Evolution of Markups (Compustat)



(a) Cost-Weighted Markup (Compustat) (b) 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. The two panels show results with two different weighting method used when aggregating firm-level markups.

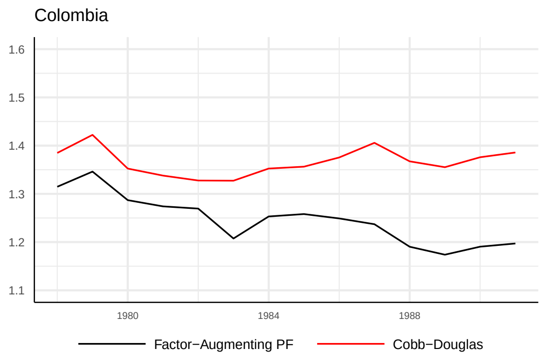
Figure A-9: Evolution of Markups (Chile)



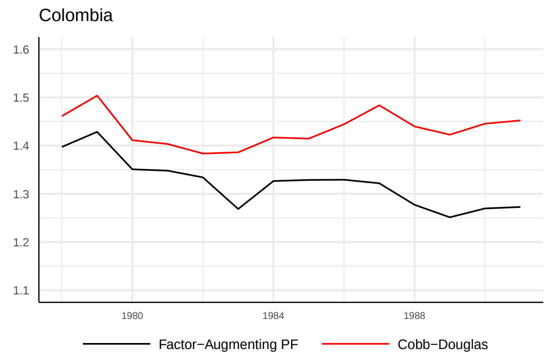
(a) Cost-Weighted Markup (Chile) (b) Sales-Weighted Markup (Chile)

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 A-10: Evolution of Markups (Colombia)



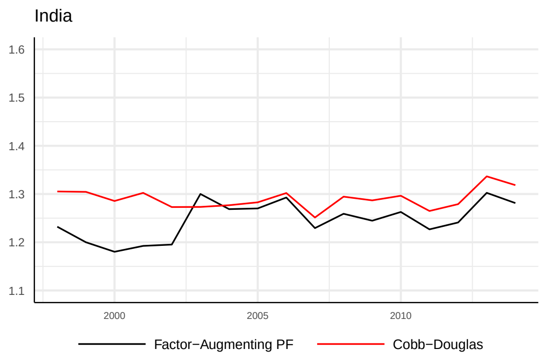
(a) Cost-Weighted Markup (Colombia)



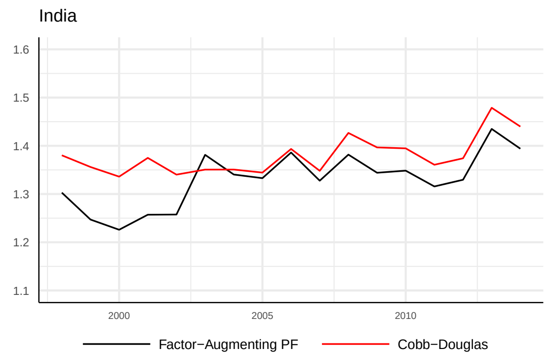
(b) Sales-Weighted Markup (Colombia)

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 A-11: Evolution of Markups (India)



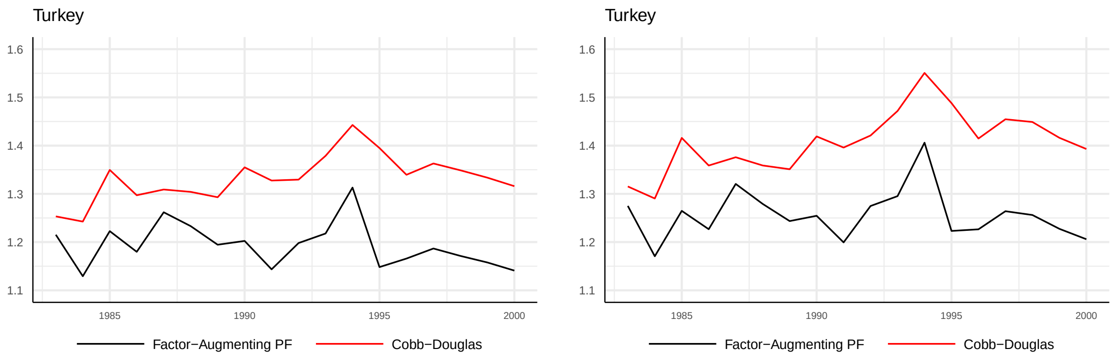
(a) Cost-Weighted Markup (India)



(b) Sales-Weighted Markup (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.

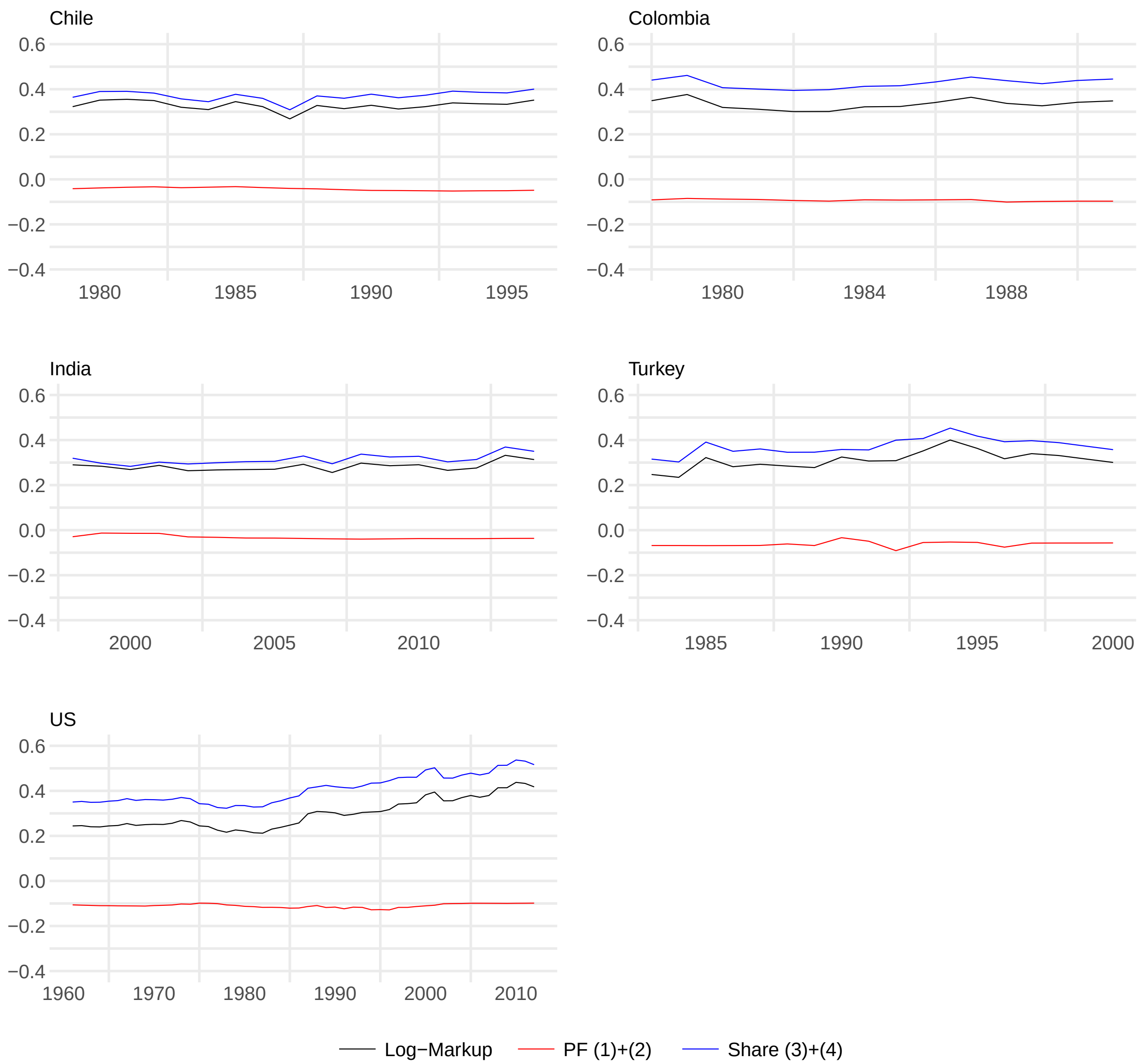
Figure A-12: Evolution of Markups (Turkey)



(a) Cost-Weighted Markup (Turkey) (b) Sales-Weighted Markup (Turkey)

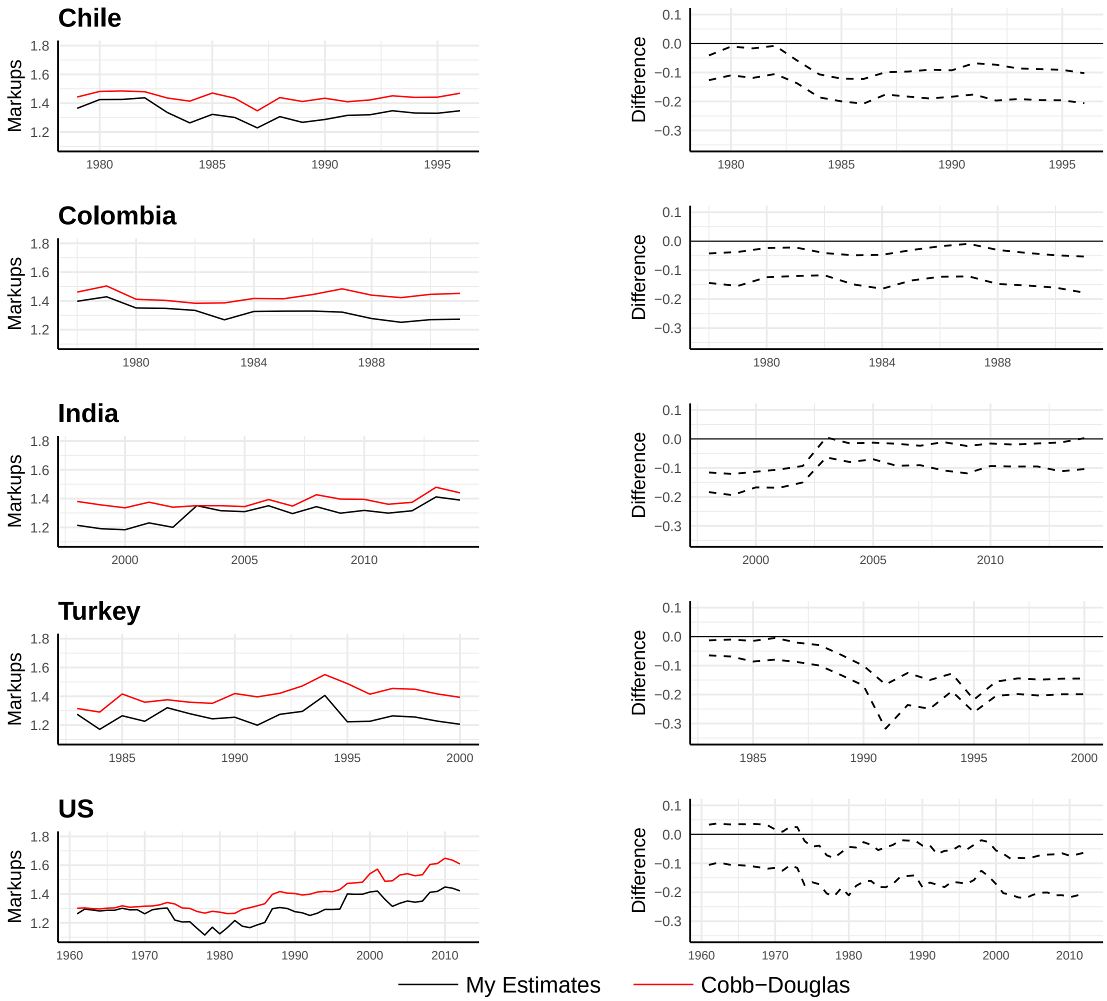
Notes: Comparisons of the evolution of markups in the Turkish manufacturing industry produced by my method and the Cobb-Douglas model estimated using the ACF procedure. The two panels show results with two different weighting method used when aggregating firm-level markups.

Figure A-13: Decomposition of Markup: Elasticity vs Share



Notes: This figure shows the evolution of the two components of log aggregate markup given in Equation 1.7.2. Black line displays the log aggregate markups, red line displays the component from production function estimation and, blue line displays the component from revenue share of flexible inputs.

Figure A-14: 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 10-90the percentile of the bootstrap distribution (100 iterations) for the difference between the two estimates.

Figure A-15: Sales-Weighted

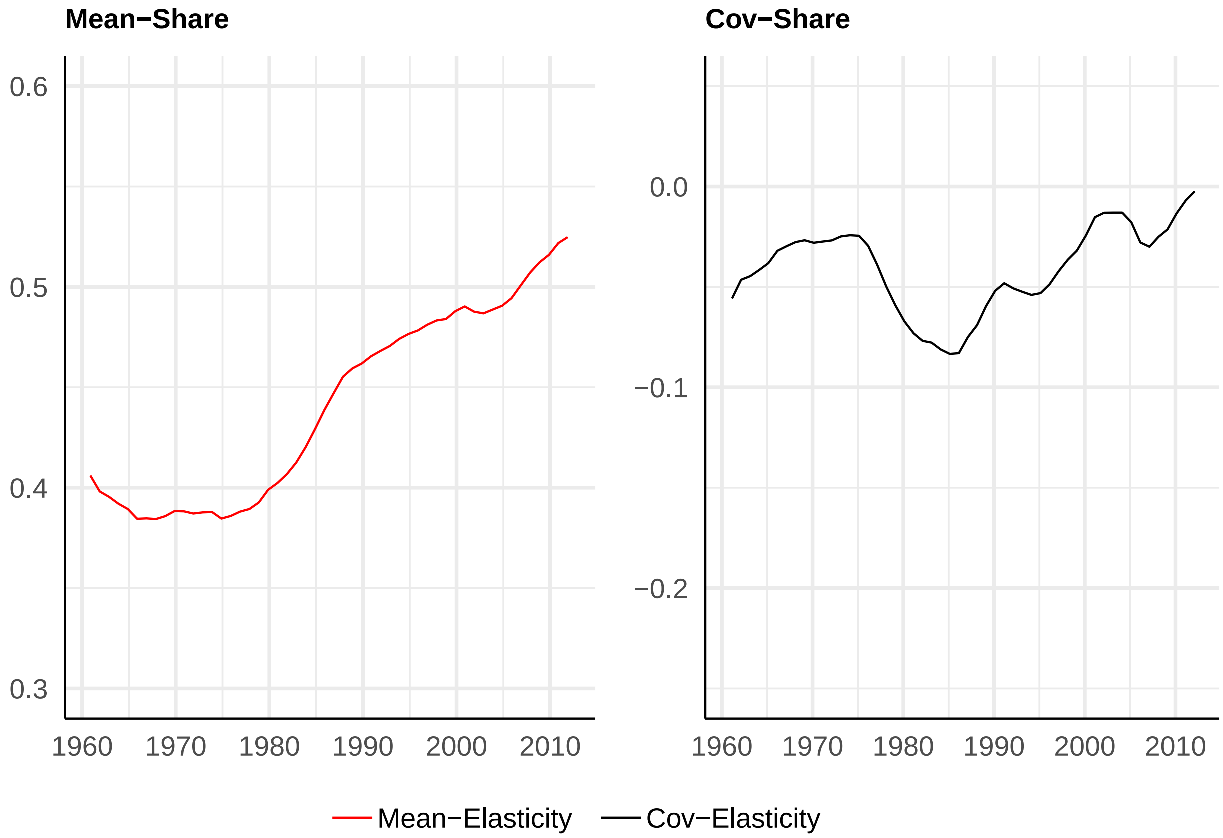
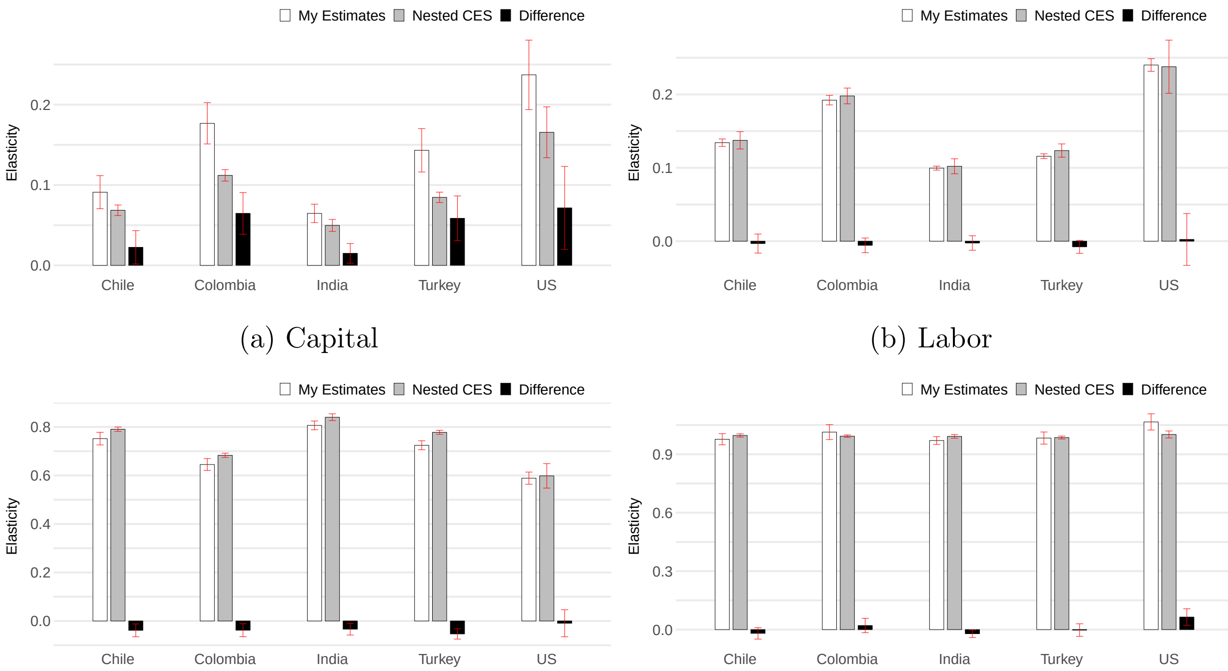


Figure A-16: Comparison with Nested CES

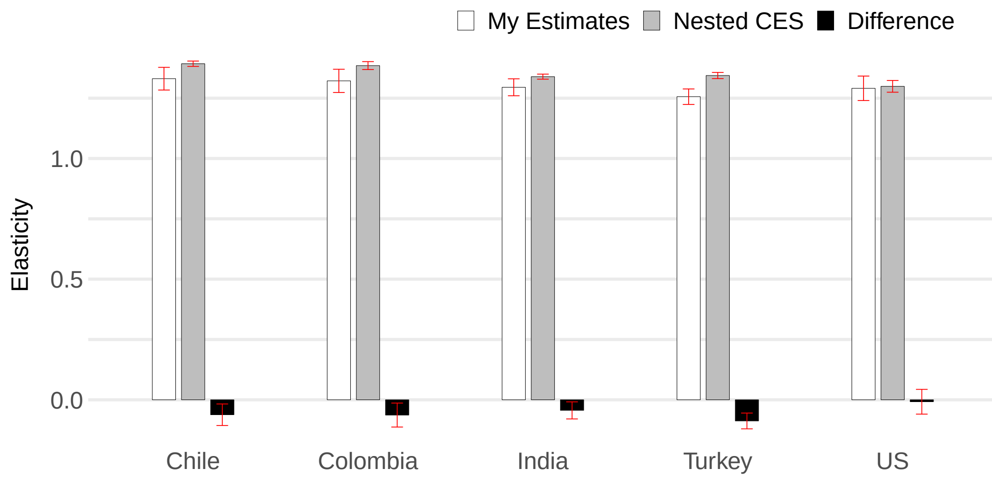


(a) Capital (b) Labor

(c) Materials (d) Returns to Scale

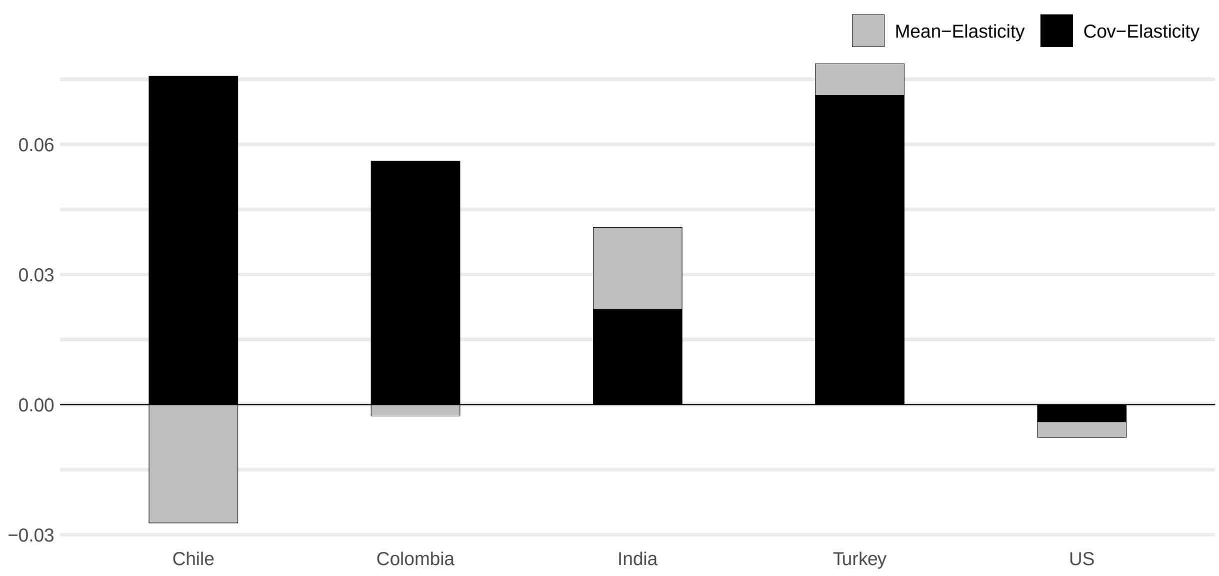
Note: Comparison of sales-weighted average elasticities produced by my estimates (white) and Nested CES estimated by procedure given in Subsection A.6.2 for each country. The difference between the two averages is shown by the black bar. For each year and industry, sales-weighted averages are calculated, and then simple averages are taken over years. The error bars indicate 95 percent confidence intervals calculated using bootstrap (100 iterations).

Figure A-17: 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 A.6.2 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 A-18: 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 A.6.2. The decomposition is based on Equation 1.8.2.

Appendix B

Appendix to Chapter 2

B.1 Proofs

B.1.1 Proof of Proposition 2.3.1

Under Assumption 1.2.2 productivity shock can be written as

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it},$$

where $g(\cdot)$ is a monotone function because the distribution is stochastically increasing in ω_{it} by Assumption 1.2.2. Substituting this into Equation (2.3.1) and expanding the left-hand side yield

$$\begin{aligned}
&\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \\
&\quad = \mathbb{E}\Big[\big(g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big], \\
&\quad = \mathbb{E}\Big[\big((g(\omega_{it-1}) - \omega_{it-1}) + \xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big], \\
&\quad = \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] + \mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big] + 2\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)\Big].
\end{aligned}$$

Expanding the right-hand side similarly

$$\begin{aligned}
&\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) + \xi_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1} + \xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)^2\Big] \\
&= \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big] + \mathbb{E}\Big[(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)^2\Big] + 2\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)\Big].
\end{aligned}$$

Since observations are independently and identically distributed across firms and $\mathbb{E}[\epsilon_{it} \mid \mathcal{I}_{it-1}] = 0$, the second expectations are equal to each other

$$\mathbb{E}\Big[\big(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)^2\Big] = \mathbb{E}\Big[\big(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)^2\Big]$$

Since $(\epsilon_{it}, \epsilon_{it-1}, \xi_{it})$ is orthogonal to ω_{it-1} and observations and independent and identically distributed, the third expectations equal zero:

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)\big(\xi_{it} + \epsilon_{it} - \epsilon_{jt-1}\big)\Big] = \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)\big(\xi_{it} + \epsilon_{it} - \epsilon_{it-1}\big)\Big] = 0$$

Therefore, to prove the proposition, I need to show that

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big].$$

Expanding both sides

$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{it-1}\big)^2\Big] = \mathbb{E}[g(\omega_{it-1})^2] + \mathbb{E}[\omega_{it-1}^2] - 2\mathbb{E}[g(\omega_{it-1})\omega_{it-1}]$$
$$\mathbb{E}\Big[\big(g(\omega_{it-1}) - \omega_{jt-1}\big)^2\Big] = \mathbb{E}[g(\omega_{it-1})^2] + \mathbb{E}[\omega_{jt-1}^2] - 2\mathbb{E}[g(\omega_{it-1})\omega_{jt-1}]$$

The second and third moments are equal to each other by iid assumption

$$\mathbb{E}[g(\omega_{it-1})^2] = \mathbb{E}[g(\omega_{it-1})^2], \qquad \mathbb{E}[\omega_{it-1}^2] = \mathbb{E}[\omega_{jt-1}^2].$$

So I need show that

$$\mathbb{E}[g(\omega_{it-1})\omega_{it-1}] \geqslant \mathbb{E}[g(\omega_{it-1})\omega_{jt-1}].$$

Observe that $\mathbb{E}[g(\omega_{it-1})\omega_{jt-1}] = 0$. We also have $\mathbb{E}[g(\omega_{it-1})\omega_{it-1}] \geqslant 0$, because for a random variable X, $\mathbb{E}[f(X)X] \geqslant 0$ for an increasing function f. This gives the inequality in the proposition.

$$\mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{it-1} - \epsilon_{it-1}\big)^2\Big] \leqslant \mathbb{E}\Big[\big(\omega_{it} + \epsilon_{it} - \omega_{jt-1} - \epsilon_{jt-1}\big)^2\Big].$$

B.1.2 Proof of Proposition 2.3.2

Using the Bayes rule for continuous random variables, I write the conditional probability distribution function of ω_{it} as (by changing the notation slightly)

$$f(\omega_{it} \mid k_{it-1}, i_{it} > z) = \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) f(\omega_{it} \mid k_{it-1}) f(k_{it-1})}{\Pr(i_{it} > z \mid k_{it-1}) f(k_{it-1})} = \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) f(\omega_{it} \mid k_{it-1})}{\Pr(i_{it} > z \mid k_{it-1})},$$

and similarly for $f(\omega_{it} \mid k_{it}, i_{it} < z)$. By taking the ratio of the two

$$\begin{aligned}\frac{f(\omega_{it} \mid k_{it-1}, i_{it} > z)}{f(\omega_{it} \mid k_{it-1}, i_{it} < z)} &= \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) \Pr(i_{it} < z \mid k_{it-1})}{\Pr(i_{it} < z \mid k_{it-1}, \omega_{it}) \Pr(i_{it} > z \mid k_{it-1})} \\ &= \frac{\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) \Pr(i_{it} < z \mid k_{it-1})}{\big(1 - \Pr(i_{it} > z \mid k_{it-1}, \omega_{it})\big)\big(1 - \Pr(i_{it} < z \mid k_{it-1})\big)}\end{aligned}$$

This function is increasing in $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$ because the numerator is increasing and denominator is decreasing in $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$. So if I can show that $\Pr(i_{it} > z \mid$

$k_{it-1}, \omega_{it})$ is weakly increasing in ω_{it} that imply that

$$\frac{f(\omega_{it} \mid k_{it-1}, i_{it} > z)}{f(\omega_{it} \mid k_{it-1}, i_{it} < z)}$$

is a weakly increasing function of ω_{it}, and there the monotone likelihood ratio property holds. The next step is to show that $\Pr(i_{it} > z \mid k_{it}, \omega_{it})$ is weakly increasing in ω_{it}. I write $\Pr(i_{it} > z \mid k_{it}, \omega_{it})$ as

$$\begin{aligned}\Pr(i_{it} > z \mid k_{it-1}, \omega_{it}) &= \int \mathbb{1}\{f(k_{it-1}, \omega_{it}, \xi_{it}) > z\} f(\xi_{it}|\omega_{it}, k_{it-1}) d\xi_{it}, \\ &= \int \mathbb{1}\{f(k_{it-1}, \omega_{it}, \xi_{it}) > z\} f(\xi_{it}|k_{it-1}) d\xi_{it},\end{aligned}$$

where the second line follows from Assumption 1.2.4. Since by assumption Assumption 2.2.5 $f(k_{i-1}, \omega_{it}, \xi_{it})$ is weakly increasing in ω_{it}, I conclude that $\Pr(i_{it} > z \mid k_{it-1}, \omega_{it})$ is weakly increasing in ω_{it}. It is well known that MLRP implies the first order stochastic dominance and ordering of expectations, so other results follow.

B.1.3 Proof of Proposition 2.3.3

Using Equation (2.3.4), I can write the moment function at the true parameter values as

$$m(w_{it}, \theta) = y_{it} - \theta_k k_{it} - \theta_l l_{it} = g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}.$$

Substituting $m(w_{it}, \theta) = g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it}$ into Equation (2.3.6) we need to show that the following inequality holds

$$\mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it} \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \eta^l_{it} + \epsilon_{it} \mid k_{it-2}, i_{it-1} < z] \geqslant 0.$$

I proceed in two steps. First note that, by assumptions ζ_{it}, ϵ_{it} and ηit are orthogonal to information set at $t-1$. Therefore we have

$$\begin{aligned}\mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} > z] &= \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z], \\ \mathbb{E}[g(\omega_{it-1}) + \zeta_{it} + \epsilon_{it} + \eta_{it} \mid k_{it-2}, i_{it-1} < z] &= \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} < z].\end{aligned}$$

To finish the proof, I need to show that

$$\mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z] \geqslant \mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} < z].$$

To see this, by Proposition 1.2.1, I have a first order stochastic dominance for the distribution of ω_{it-1} conditional on 'high' and 'low' investment (conditional on k_{it-1}). Using this

$$F_{\omega_{it-1}}(t \mid k_{it-2}, i_{it-1} > z) \geqslant F_{\omega_{it-1}}(t \mid k_{it-2}, i_{it-1} < z), \qquad \text{for all } t > 0.$$

Since $g(\cdot)$ is a monotone function and stochastic order is preserved under the monotone transformation, we have

$$F_{g(\omega_{it-1})}(t \mid k_{it-2}, i_{it-1} > z) \geqslant F_{g(\omega_{it-1})}(t \mid k_{it-2}, i_{it-1} < z) \qquad \text{for all} \qquad t > 0.$$

This leads to the desired inequality

$$\mathbb{E}[g(\omega_{it-1}) \mid k_{it-2}, i_{it-1} > z] - \mathbb{E}[g(\omega_{it-1} \mid k_{it-2}, i_{it-1} < z] \geqslant 0.$$

Appendix C

Appendix to Chapter 3

C.1 Alternative Definitions of Homophily

In this section, I provide several definitions of homophily and discuss the implication of these definitions. The definition relies on two features: (i) the shape of conditional link probability, and (ii) stochastic ordering of distance measures. I show that my definitions provide a range of homophily structures that might be of independent interest. I analyze these structures by showing that they are often nested with each other, so they can be ordered from strongest to weakest. I also argue that using stochastic ordering, which has not been used before, provides a flexible framework to characterize and study homophily in networks. Finally, this section also comments on how model-based homophily is related to my definitions.

Consistent with the empirical framework, this section considers unobserved homophily conditional on observed variables. However, these definitions can easily be extended to observed homophily. Let α_i denote an unobserved and scalar characteristic, and x_i denote a vector of observed features. G_{ij} is an indicator variable that equals one if i and j are connected. First, I define a function that characterizes the link formation probability,

$$h(\Delta, x_i, x_j) = \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \Delta, x_i, x_j). \tag{C.1.1}$$

This function specifies the link formation probability conditional on the distance in the latent space and observed characteristics. Next, I define a representative network formation model, which nests the commonly used network formation models. Assume that there exists $g(\cdot)$ such that links are formed according to:

$$G_{ij} = \mathbb{1}(g(\Delta\alpha_{ij}, x_i, x_j) \geqslant u_{ij}), \tag{C.1.2}$$

where u_{ij} is independent and identically distributed random variable with strictly increasing cumulative distribution function. The main difference between this model and Equation (C.1.1) is that this model imposes a structural model, whereas Equation (C.1.1) just a function of the unknown data generating process. For example, according to Equation (C.1.2), the links are independently formed, whereas, $h(\Delta, x_i, x_j)$ might exist when network is generated from a more complicated model. I start with the strongest definition of homophily.

C.1.1 Symmetric and Unimodal Network Formation - (SUF)

According to this definition, the network is formed based on Equation (C.1.2) and

$$g(\Delta\alpha_{ij}, x_i, x_j)$$

is symmetric and unimodal around zero with respect to its first argument for all x_i and x_j. The standard models in the literature assumes that g has a known form (possibly up to a parameter vector). This homophily definition is strong because it imposes a particular network formation structure and functional form restriction.

C.1.2 Symmetric and Unimodal Link Formation Probability- (SUP)

This definition states that

$$h(\cdot, x_i, x_j)$$

is symmetric and unimodal around zero. This differs from SUF in that it does not impose a particular network formation structure. As a result, it covers a broader network formation model and data generating processes. For example, this definition does not rule out interdependent link formation and multiple equilibria. I only require that $h(\cdot, x_i, x_j)$ exists.

C.1.3 Unimodal Link Formation Probability - (UP)

According to this definition

$$h(\cdot, x_i, x_j)$$

is unimodal around zero. This is similar to the previous definition but does not impose symmetry. This can happen, for example, when the links are not symmetric, and the nodes in the network are not exchangeable.

C.1.4 Decreasing Link Formation Probability - (DP)

This definition states that

$$\Pr\big(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta, x_i, x_j\big)$$

is decreasing in Δ for all (x_i, x_j). This definition concerns how link formation probability changes with the squared distance. It is slightly weaker than the previous definitions in that it does restrict the entire shape link formation function.

C.1.5 Monotone Likelihood Ratio based Homophily Definition - (MLRP)

This definition states that

$$\frac{f_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 0)}{f_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 1)}$$

is increasing in Δ for all x_i and x_j. This definition is based on stochastic ordering of the distribution of $\Delta\alpha^2$ conditional on a link and no link. Using stochastic ordering is important because it provides a flexible framework to define homophily based on other stochastic orderings implied by MLRP.

C.1.6 Monotone Hazard Rate based Homophily Definition - (MHR)

This definition states that

$$\frac{f_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 0)}{1 - F_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 0)} \geqslant \frac{f_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 1)}{1 - F_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 1)}$$

for all Δ, x_i and x_j. Monotone Hazard Rate is an implication of MLRP and based on a weaker notion of stochastic ordering.

C.1.7 First Order Stochastic Dominance based Homophily Definition - (FOSD)

This definition imposes a first order stochastic ordering. In particular, it says that

$$F_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 0) \geqslant F_{(\Delta\alpha_{ij})^2}(\Delta \mid x_i, x_j, G_{ij} = 1)$$

for all Δ, x_i and x_j. This corresponds to the strong homophily definition I use to identify the linear model in the main text. First order stochastic dominance is an implication of MHR and based on a weaker notion of stochastic ordering.

C.1.8 Moment Inequality based Homophily Definition - (MI)

This definition states that

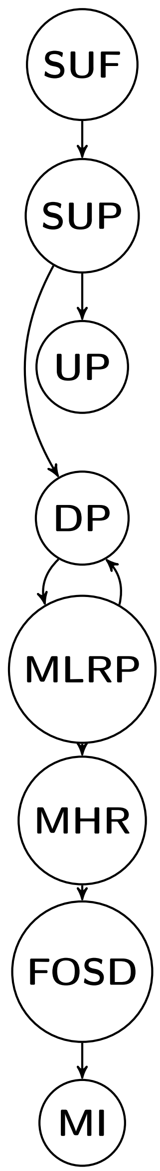
$$\mathbb{E}[(\Delta\alpha_{ij})^2 \mid x_i, x_j, G_{ij} = 0] \geqslant \mathbb{E}[(\Delta\alpha_{ij})^2 \mid x_i, x_j, G_{ij} = 1]$$

for all x_i and x_j. This definition corresponds to the weak homophily definition I use to identify the linear model in the main text. This is again an implication of first order stochastic dominance. This is the weakest homophily definition.

Figure C-1 compares the homophily definitions given above. An arrow from definition A to definition B shows that if a network satisfies definition A, then it also satisfies the definition B. We see that SUF, the model-based definition, is the strongest notion of homophily. In contrast, MI is the weakest. Among all definitions, only DP and MLRP are equivalent. Also, definitions that impose restrictions on the shape of the conditional probability of link (SUF, SUP, UP and DP) in general impose stronger restrictions than the ones that are based on stochastic ordering (MLRP, MHR, FOSD, MI). The equivalence of MLRP and DP relates the two concepts of homophily.

The proofs that show that the relationships between homophily definitions are given in the next section.

Figure C-1: Relationship between Hompohily Definitions



C.1.9 Proofs

In this section, I present the proofs for the relationship between homophily definitions presented in this section. I omit the trivial proofs that SUF implies SUP, SUP implies UP, MLRP implies MHR, MHR implies FOSD, FOSD implies MI.

SUP implies DP

Assume that I condition on x_i and x_j in all expectations and probability distribution functions

$$\begin{aligned}
\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta) &= \frac{\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta}) f_{\Delta\alpha_{ij}}(-\sqrt{\Delta}) + \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta}) f_{\Delta\alpha_{ij}}(\sqrt{\Delta})}{f_{\Delta\alpha_{ij}}(\sqrt{\Delta}) + f_{\Delta\alpha_{ij}}(-\sqrt{\Delta})}, \\
&= \frac{\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta})\big(\Delta f_{\alpha_{ij}}(-\sqrt{\Delta}) + f_{\Delta\alpha_{ij}}(\sqrt{\Delta})\big)}{\big(f_{\Delta\alpha_{ij}}(-\sqrt{\Delta}) + f_{\Delta\alpha_{ij}}(\sqrt{\Delta})\big)}, \\
&= \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta}),
\end{aligned}$$

which is decreasing in Δ by unimodality of $\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \Delta)$ and monotonicity of $\sqrt{\Delta}$ in Δ.

UP Does not Imply DP

I condition on x_i and x_j throughout. Take $\Delta_1 > \Delta_2 \geqslant 0$ Network formation according DP should satisfy

$$\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_2) \geqslant \Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_1).$$

I will show a counterexample where this does not hold when UP holds. The probabilities above can be computed as

$$\begin{aligned}
\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_1) &= \frac{\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_1}) f_{\alpha_{ij}}(-\sqrt{\Delta_1}) + \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_1}) f_{\alpha_{ij}}(\sqrt{\Delta_1})}{f_{\alpha_{ij}}(\sqrt{\Delta_1}) + f_{\alpha_{ij}}(-\sqrt{\Delta_1})}, \\
\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_2) &= \frac{\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_2}) f_{\alpha_{ij}}(-\sqrt{\Delta_2}) + \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_2}) f_{\alpha_{ij}}(\sqrt{\Delta_2})}{f_{\alpha_{ij}}(\sqrt{\Delta_2}) + f_{\alpha_{ij}}(-\sqrt{\Delta_2})}
\end{aligned}$$

Unimodal link formation probability according to UP implies

$$\begin{aligned}
\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_2}) &\geqslant \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_1}), \\
\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_2}) &\geqslant \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_1}).
\end{aligned}$$

To see a counterexample assume that

$$\begin{aligned}
m = \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_2}) &\geqslant \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = -\sqrt{\Delta_1}) = 0, \\
\Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_2}) &= \Pr(G_{ij} = 1 \mid \Delta\alpha_{ij} = \sqrt{\Delta_1}) = k,
\end{aligned}$$

which satisfies the unimodal network formation probability. Substituting these

$$\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_1) = \frac{m f_{\alpha_{ij}}(\sqrt{\Delta_1})}{f_{\alpha_{ij}}(\sqrt{\Delta_1}) + f_{\alpha_{ij}}(-\sqrt{\Delta_1})}$$
$$\Pr(G_{ij} = 1 \mid (\Delta\alpha_{ij})^2 = \Delta_2) = \frac{\epsilon f_{\alpha_{ij}}(-\sqrt{\Delta_2}) + k f_{\alpha_{ij}}(\sqrt{\Delta_2})}{f_{\alpha_{ij}}(\sqrt{\Delta_2}) + f_{\alpha_{ij}}(-\sqrt{\Delta_2})}$$

For any values of m and $f_{\alpha_{ij}}(\sqrt{\Delta_2})$, $f_{\alpha_{ij}}(\sqrt{\Delta_1})$ and k there exists large enough $f_{\Delta\alpha_{ij}}(-\sqrt{\Delta_1})$ such that

$$\frac{m f_{\alpha_{ij}}(-\sqrt{\Delta_2}) + k f_{\alpha_{ij}}(\sqrt{\Delta_2})}{f_{\alpha_{ij}}(\sqrt{\Delta_2}) + f_{\alpha_{ij}}(-\sqrt{\Delta_2})} \geqslant \frac{k f_{\alpha_{ij}}(\sqrt{\Delta_1})}{f_{\alpha_{ij}}(\sqrt{\Delta_1}) + f_{\alpha_{ij}}(-\sqrt{\Delta_1})},$$

which violates the condition given in definition of DP.

DP holds if and only if MLRP Holds

Using the Bayes Rule for continuous random variables gives

$$\begin{aligned} f_{\Delta\alpha_{ij}^2}(\Delta \mid x_i, x_j, G_{ij} = 1) &= \frac{\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j) f_{\Delta\alpha_{ij}}(\Delta \mid x_i, x_j) f(x_i, x_j)}{\Pr(G_{ij} = 1 \mid x_i, x_j) f(x_i, x_j)}, \\ &= \frac{\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j) f_{\Delta\alpha_{ij}}(\Delta \mid x_i, x_j)}{\Pr(G_{ij} = 1 \mid x_i, x_j)}. \end{aligned}$$

And similarly for $f_{\Delta\alpha_{ij}^2}(\Delta \mid x_i, x_j, G_{ij} = 0)$. By taking the ratio of the two

$$\begin{aligned} \frac{f_{\Delta\alpha_{ij}^2}(\Delta \mid x_i, x_j, G_{ij} = 0)}{f_{\Delta\alpha_{ij}^2}(\Delta \mid x_i, x_j, G_{ij} = 1)} &= \frac{\Pr(G_{ij} = 0 \mid \Delta, x_i, x_j)\Pr(G_{ij} = 1 \mid x_i, x_j)}{\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)\Pr(G_{ij} = 0 \mid x_i, x_j)}, \\ &= \frac{\big(1 - \Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)\big)\Pr(G_{ij} = 1 \mid x_i, x_j)}{\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)\big(1 - \Pr(G_{ij} = 1 \mid x_i, x_j)\big)}. \end{aligned}$$

Since the numerator is decreasing in $\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)$ and denominator is increasing in $\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)$ the right hand side is decreasing in $\Pr(G_{ij} = 1 \mid \Delta, x_i, x_j)$. Pick $\Delta_1 \geqslant \Delta_2$, MLRP implies

$$\Pr(G_{ij} = 1 \mid \Delta_2, x_i, x_j) \geqslant \Pr(G_{ij} = 1 \mid \Delta_1, x_i, x_j).$$

Using this, it is easy to see that

$$\frac{f_{\Delta\alpha_{ij}^2}(\Delta_1 \mid x_i, x_j, G_{ij} = 0)}{f_{\Delta\alpha_{ij}^2}(\Delta_1 \mid x_i, x_j, G_{ij} = 1)} \geqslant \frac{f_{\Delta\alpha_{ij}^2}(\Delta_2 \mid x_i, x_j, G_{ij} = 0)}{f_{\Delta\alpha_{ij}^2}(\Delta_2 \mid x_i, x_j, G_{ij} = 1)}.$$

Therefore, MLRP follows. Pick Δ_1 and Δ_2 such that $\Delta_1 \geqslant \Delta_2$, DP implies

$$\frac{\big(1-\Pr(G_{ij}=1 \mid \Delta_1, x_i, x_j)\big)\Pr(G_{ij}=1 \mid x_i, x_j)}{\Pr(G_{ij}=1 \mid \Delta_1, x_i, x_j)\big(1-\Pr(G_{ij}=1 \mid x_i, x_j)\big)} \geqslant \frac{\big(1-\Pr(G_{ij}=1 \mid \Delta_2, x_i, x_j)\big)\Pr(G_{ij}=1 \mid x_i, x_j)}{\Pr(G_{ij}=1 \mid \Delta_2, x_i, x_j)\big(1-\Pr(G_{ij}=1 \mid x_i, x_j)\big)},$$

which further gives

$$\frac{\big(1-\Pr(G_{ij}=1 \mid \Delta_1, x_i, x_j)\big)}{\Pr(G_{ij}=1 \mid \Delta_1, x_i, x_j)} \geqslant \frac{\big(1-\Pr(G_{ij}=1 \mid \Delta_2, x_i, x_j)\big)}{\Pr(G_{ij}=1 \mid \Delta_2, x_i, x_j)},$$

and

$$\Pr(G_{ij}=1 \mid \Delta_2, x_i, x_j) \geqslant \Pr(G_{ij}=1 \mid \Delta_1, x_i, x_j).$$

Therefore MLRP implies that the link formation probability is decreasing in Δ.

C.2 Proofs

C.2.1 Proof of Proposition 1.4.7

In this proof, I show that weak homophily is robust to an additive noise that satisfies the conditions in Assumption 3.3.1. To simplify notation, I condition on x_i and x_j throughout the proof, and drop them from the conditioning set. Expectation of $(\Delta\alpha + \Delta\epsilon)^2$ conditional on G is given by:

$$\begin{aligned}
\mathbb{E}\big[(\Delta\alpha + \Delta\epsilon)^2 \mid G=i\big] &= \mathbb{E}\big[(\Delta\alpha)^2 \mid G=i\big] + \mathbb{E}\big[\Delta\alpha\Delta\epsilon \mid G=i\big] + \mathbb{E}\big[(\Delta\epsilon)^2 \mid G=i\big], \\
&= \mathbb{E}\big[(\Delta\alpha)^2 \mid G=i\big] + \mathbb{E}\big[\Delta\alpha\mathbb{E}\big[\Delta\epsilon \mid \alpha_i, \alpha_j, G=i\big] \mid G=i\big] + \mathbb{E}\big[(\Delta\epsilon)^2 \mid G=i\big], \\
&= \mathbb{E}\big[(\Delta\alpha)^2 \mid G=i\big] + \mathbb{E}\big[\Delta\alpha\mathbb{E}\big[\Delta\epsilon \mid \alpha_i, \alpha_j] \mid G=i\big] + \mathbb{E}\big[(\Delta\epsilon)^2 \mid G=i\big], \\
&= \mathbb{E}\big[(\Delta\alpha)^2 \mid G=i\big] + \mathbb{E}\big[(\Delta\epsilon)^2 \mid G=i\big],
\end{aligned}$$

where I use the law of iterated expectations and Assumption 3.3.1(i, iii, v). To complete the proof I need to show that

$$\mathbb{E}[(\Delta\epsilon)^2 \mid G=1] = \mathbb{E}[(\Delta\epsilon)^2 \mid G=0].$$

Using Assumption 3.3.1(ii, iii), I obtain

$$\begin{aligned}
\mathbb{E}[(\Delta\epsilon)^2 \mid G=i] &= \mathbb{E}\big[\mathbb{E}[(\Delta\epsilon)^2 \mid \alpha_i, \alpha_j, G=i] \mid G=i],\\
&= \mathbb{E}\big[\mathbb{E}[(\Delta\epsilon)^2 \mid \alpha_i, \alpha_j] \mid G=i],\\
&= \mathbb{E}\big[\mathbb{E}[(\Delta\epsilon)^2] \mid G=i],\\
&= \mathbb{E}[(\Delta\epsilon)^2].
\end{aligned}$$

This, along with Assumption 3.3.1 (iv) gives

$$\mathbb{E}\big[(\Delta\alpha + \Delta\epsilon)^2 \mid x_i, x_j, G = 0\big] \geqslant \mathbb{E}\big[(\Delta\alpha + \Delta\epsilon)^2 \mid x_i, x_j, G = 1\big].$$

C.2.2 Proof of Proposition 3.3.3

First, I introduce additional notation to simplify the exposition. Throughout the proof, I condition on (x_i, x_j) and define

$$\Delta z := \Delta\alpha + \Delta\epsilon, \qquad F_{\Delta z^i}(t) := F_{\Delta z^i}(t \mid G = i).$$

For a continuously distributed random variable w, we have

$$F_{|w|}(t) = \Pr(-t \leqslant w \leqslant t) = F_w(t) - F_w(-t), \quad \text{for} \quad t \geqslant 0.$$

Therefore, Assumption 3.3.2 implies that

$$F_{\Delta\alpha^0}(t) - F_{\Delta\alpha^0}(-t) - F_{\Delta\alpha^1}(t) + F_{\Delta\alpha^1}(-t) \leqslant 0, \quad \text{for} \quad t \geqslant 0. \tag{C.2.1}$$

To prove this proposition, I need to show, using Equation (C.2.1) and Assumption 3.3.2, that

$$g(t) := F_{\Delta z^0}(t) - F_{\Delta z^0}(-t) - F_{\Delta z^1}(t) + F_{\Delta z^1}(-t) \leqslant 0, \quad \text{for} \quad t \geqslant 0.$$

By Assumption 3.3.2 (i) I can write conditional distribution of Δz using convolution as follows:

$$F_{\Delta z^i}(t) = \int\limits_{-\infty}^{\infty} F_{\Delta\alpha^i}(t-u) f_{\Delta\epsilon^i}(u) du.$$

This is because, under Assumption (3.3.2)(i), $\Delta\alpha$ and $\Delta\epsilon$ are conditionally independent from each other and the sum of two independent random variables has a distribution function given by a convolution. Substituting this, I can obtain $g(t)$ as

$$
\begin{aligned}
g(t) &= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^0}(-t-u)\Big) f_{\Delta\epsilon^0}(u)du - \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^1}(t-u) - F_{\Delta\alpha^1}(-t-u)\Big) f_{\Delta\epsilon^1}(u)du \\
&= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^0}(-t-u)\Big) f_{\Delta\epsilon}(u)du - \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^1}(t-u) - F_{\Delta\alpha^1}(-t-u)\Big) f_{\Delta\epsilon}(u)du \\
&= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^1}(t-u) f_{\Delta\epsilon}(u)du + \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^1}(-t-u) - F_{\Delta\alpha^0}(-t-u) f_{\Delta\epsilon}(u)du \\
&= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^1}(t-u)\Big) f_{\Delta\epsilon}(u)du + \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^1}(-t+\tilde{u}) - F_{\Delta\alpha^0}(-t+\tilde{u})\Big) f_{\Delta\epsilon}(-\tilde{u})d\tilde{u} \\
&= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^1}(t-u)\Big) f_{\Delta\epsilon}(u)du + \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^1}(-t+\tilde{u}) - F_{\Delta\alpha^0}(-t+\tilde{u})\Big) f_{\Delta\epsilon}(\tilde{u})d\tilde{u} \\
&= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^0}(-t+u)\Big) - \Big(F_{\Delta\alpha^1}(z-u) - F_{\Delta\alpha^1}(-t+u)\Big) f_{\Delta\epsilon}(u)du,
\end{aligned}
$$

where the second line follows from $f_{\Delta\epsilon^1} = f_{\Delta\epsilon^0} = f_{\Delta\epsilon}$ and the third line rearranges integrals. The fourth line is obtained using the change of variable, $\tilde{u} = -\tilde{u}$. The following line uses the fact that $f_{\Delta\epsilon}$ is symmetric around zero. Symmetry holds because the convolution of two independent and identically distributed random variables is symmetric about zero. Next, define $m(t)$ as

$$
m(t) := F_{\Delta\alpha^0}(t) - F_{\Delta\alpha^0}(-t) - F_{\Delta\alpha^1}(t) + F_{\Delta\alpha^1}(-t).
$$

Note that $m(t)$ is an odd function by definition, $m(t) + m(-t) = 0$, it is negative for $t > 0$ and $m(0) = 0$. Using $m(t)$ we can obtain $g(t)$ as

$$
\begin{aligned}
g(t) &= \int\limits_{-\infty}^{\infty} \Big(F_{\Delta\alpha^0}(t-u) - F_{\Delta\alpha^0}(-t+u)\Big) - \Big(F_{\Delta\alpha^1}(t-u) - F_{\Delta\alpha^1}(-t+u)\Big) f_{\Delta\epsilon}(u)du, \\
&= \int\limits_{-\infty}^{\infty} m(t-u) f_{\Delta\epsilon}(u)du.
\end{aligned}
$$

I need to show that this integral is non-negative for $t \geqslant 0$. First note that for $t = 0$

$$g(0) = \int_{-\infty}^{\infty} m(-u) f_\epsilon(u) du = 0,$$

by symmetry of $f_{\Delta\epsilon}(u)$ and the fact that m is an odd function. For $t > 0$ we can write

$$\begin{aligned}
\int_{-\infty}^{\infty} m(t-u) f_{\Delta\epsilon}(u) du = \int_{-\infty}^{\infty} m(u) f_{\Delta\epsilon}(t-u) du &= \int_{0}^{\infty} m(u) f_{\Delta\epsilon}(t-u) du + \int_{-\infty}^{0} m(u) f_{\Delta\epsilon}(t-u) du, \\
&= \int_{0}^{\infty} m(u) f_{\Delta\epsilon}(t-u) du + \int_{0}^{\infty} m(-u) f_{\Delta\epsilon}(t+u) du, \\
&= \int_{0}^{\infty} m(u) f_{\Delta\epsilon}(t-u) du - \int_{0}^{\infty} m(u) f_{\Delta\epsilon}(t+u) du, \\
&= \int_{0}^{\infty} m(u) \big(f_{\Delta\epsilon}(t-u) - f_{\Delta\epsilon}(t+u) \big) du.
\end{aligned}$$

The second line is obtained from a change of variable $\tilde{u} = -u$ and the third line uses the fact that $m(t)$ is an odd function. This derivation suggests that a sufficient condition for this integral to be negative is

$$f_{\Delta\epsilon}(t-u) - f_{\Delta\epsilon}(t+u) \geqslant 0, \quad \text{for} \quad (t,u) \in \mathbb{R}_+ \times \mathbb{R}_+.$$

Since $f_{\Delta\epsilon}(u)$ is symmetric around zero and unimodal by Assumption 3.3.2, this condition is satisfied. This completes the proof.

C.2.3 Proof of Proposition 3.3.5

In this proof, I will show that the condition given in Lemma 3.3.1 is satisfied under Assumption 3.3.3. Define $\Delta z := \Delta\alpha + \Delta\epsilon$ and let $f_{\Delta z}(t)$ be the corresponding probability density function. Using the Bayes rule for continuous random variables we have

$$f_{\Delta z^1}(t) = \frac{\mathbb{E}[G \mid \Delta z] f_{\Delta z}(t)}{E[G]}, \qquad f_{\Delta z^0}(t) = \frac{(1 - \mathbb{E}[G \mid \Delta z] f_{\Delta z}(t)}{1 - E[G]}.$$

I can write the differences of two distribution functions using these as

$$
\begin{aligned}
g(t) &:= F_{\Delta z}(t \mid G=1) - F_{\Delta z}(t \mid G=0) \\
&= \int_{-\infty}^{t} f_{\Delta z^1}(u)du - \int_{-\infty}^{t} f_{\Delta z^0}(u)du = \int_{-\infty}^{t} \frac{\mathbb{E}[G \mid \Delta\alpha] f_{\Delta\alpha}(u)}{\mathbb{E}[G]} du - \int_{-\infty}^{t} \frac{\big(1-\mathbb{E}[G \mid \Delta\alpha_{ij}]\big) f_{\Delta\alpha}(u)}{1-\mathbb{E}[G]} du, \\
&= \int_{-\infty}^{t} \left(\frac{\mathbb{E}[G \mid \Delta\alpha] f_{\Delta\alpha}(u)}{\mathbb{E}[G]} - \frac{\big(1-\mathbb{E}[G \mid \Delta\alpha]\big) f_{\Delta\alpha}(u)}{1-\mathbb{E}[G]} \right) du, \\
&= \int_{-\infty}^{t} \left(\frac{\mathbb{E}[G \mid \Delta\alpha]}{\mathbb{E}[G]} - \frac{\big(1-\mathbb{E}[G \mid \Delta\alpha]\big)}{1-\mathbb{E}[G]} \right) f_{\Delta\alpha}(u)du, \\
&= \frac{1}{\mathbb{E}[G](1-E[G])} \int_{-\infty}^{t} \left(\mathbb{E}[G \mid \Delta\alpha] - E[G]\right) f_{\Delta\alpha}(u)du.
\end{aligned}
$$

By my assumptions we have

$$
\lim_{t \to \infty} g(t) = \lim_{t \to -\infty} g(t) = 0.
$$

There exists M such that

$$
\begin{aligned}
\frac{1}{\mathbb{E}[G](1-E[G])} \int_{-\infty}^{-M} \left(\mathbb{E}[G \mid \Delta\alpha] - E[G]\right) f_{\Delta\alpha}(u)du < 0, \\
\frac{1}{\mathbb{E}[G](1-E[G])} \int_{M}^{\infty} \left(\mathbb{E}[G \mid \Delta\alpha] - E[G]\right) f_{\Delta\alpha}(u)du < 0.
\end{aligned}
$$

This implies that $g(t)$ should take positive and negative values. Therefore there exists t, t' such that $g(t) > 0$ and $g(t') > 0$. Therefore, the condition given in Lemma 3.3.1 is satisfied and I conclude that identified set is bounded.

C.2.4 Proof of Lemma 3.3.1

Without loss of generality set $x_i = 1$ and $x_j = 0$. By Assumption 3.3.3 (i) there exists $M \in \mathbb{R}$ such that $\text{supp}(\Delta y \mid x_i, x_j) \subseteq [-M, M]$. We will show that identified set satisfies

$$
B \subseteq (-2M - \beta_0, 2M - \beta_0).
$$

The identification condition states that β_0 satisfies

$$\mathbb{E}[\mathbb{1}\{|\Delta y-\beta_0|\leqslant\gamma\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{|\Delta y-\beta_0|\leqslant\gamma\}\mid x_{ij},G=0],\qquad\text{for all}\quad\gamma\geqslant 0.$$

By substituting Δy from our model this condition can be written as

$$\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\beta)|\leqslant\gamma\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\beta)|\leqslant\gamma\}\mid x_{ij},G=0]\qquad\text{for all}\quad\gamma\geqslant 0$$

Assume, as a contradiction, that $(-\infty,-2M-\beta_0)\cap B=S\neq\emptyset$. Take $\tilde{\beta}\in S$ which satisfies

$$\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\tilde{\beta})|\leqslant\gamma\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\tilde{\beta})|\leqslant\gamma\}\mid x_{ij},G=0],\qquad\text{for all}\quad\gamma\geqslant 0,$$

where $\Delta z:=\Delta\alpha+\Delta\epsilon$. This equation can be written as

$$\mathbb{E}[\mathbb{1}\{\Delta z\leqslant\gamma-(\beta_0-\beta)\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{\Delta z\leqslant\gamma-(\beta_0-\beta)\}\mid x_{ij},G=0],\qquad\text{for all}\quad\gamma\geqslant 0$$

However, this violates the condition stated in the lemma. Thus I conclude that $(2M,+\infty)$ is not in the identified set. Similarly, assume as a contradiction that $(2M-\beta_0,\infty)\cap B=S\neq\emptyset$. Take $\tilde{\beta}\in S$ which satisfies

$$\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\tilde{\beta})|\leqslant\gamma\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{|\Delta z+(\beta_0-\tilde{\beta})|\leqslant\gamma\}\mid x_{ij},G=0],$$

which can be written as

$$\mathbb{E}[\mathbb{1}\{-\Delta z-(\beta_0-\tilde{\beta})\leqslant\gamma\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{-\Delta z-(\beta_0-\tilde{\beta})\leqslant\gamma\}\mid x_{ij},G=0],].$$

This gives

$$\mathbb{E}[\mathbb{1}\{-\Delta z\leqslant\gamma+(\beta_0-\tilde{\beta})\}\mid x_{ij},G=1]\geqslant\mathbb{E}[\mathbb{1}\{-\Delta z\leqslant\gamma+(\beta_0-\tilde{\beta})\}\mid x_{ij},G=0].$$

This again violates the condition given in Lemma 3.3.1. Therefore I conclude that

$$\beta\subseteq(-2M-\beta_0,2M-\beta_0),$$

and the identified set is bounded.

Bibliography

Abel, Andrew B and Eberly, Janice C. 2002. "Investment and q with Fixed Costs: An Empirical Analysis". *Working paper. Kellog School of Management and the Wharton School of the University of Pennsylvania.*

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

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

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

Ackerberg, Daniel, Lanier Benkard, C., Berry, Steven, and Pakes, Ariel. 2007. "Chapter 63 Econometric Tools for Analyzing Market Outcomes". In *Handbook of Econometrics*, 6:4171–4276. Elsevier.

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.

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.

Andrews, Donald WK and Shi, Xiaoxia. 2013. "Inference Based on Conditional Moment Inequalities". *Econometrica* 81 (2): 609–666.

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

Aradillas-Lopez, Andres, Honoré, Bo E., and Powell, James L. 2007. "Pairwise Difference Estimation with Nonparametric Control Variables". *International Economic Review* 48 (4): 1119–1158.

Arduini, Tiziano, Patacchini, Eleonora, Rainone, Edoardo, et al. 2015. "Parametric and Semiparametric IV Estimation of Network Models with Selectivity". *Working Paper, Einaudi Institute for Economics and Finance (EIEF).*

Atalay, Enghin. 2014. "Materials Prices and Productivity". *Journal of the European Economic Association* 12 (3): 575–611.

Auerbach, Eric. 2019. "Identification and Estimation of a Partially Linear Regression Model Using Network Data". *arXiv:1903.09679*.

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*.

Baerveldt, Chris, Van Duijn, Marijtje AJ, Vermeij, Lotte, and Van Hemert, Dianne A. 2004. "Ethnic Boundaries and Personal Choice. Assessing the Influence of Individual Inclinations to Choose Intra-ethnic Relationships on Pupils' Networks". *Social Networks* 26 (1): 55–74.

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

Balat, Jorge, Brambilla, Irene, and Sasaki, Yuya. 2016. "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports". *Working Paper*.

Barrett, Garry F and Donald, Stephen G. 2003. "Consistent Tests for Stochastic Dominance". *Econometrica* 71 (1): 71–104.

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.

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.

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

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.

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.

Blume, Lawrence E., Brock, William A., Durlauf, Steven N., and Jayaraman, Rajshri. 2015. "Linear Social Interactions Models". *Journal of Political Economy* 123 (2): 444–496.

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

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

Boucher, Vincent. 2015. "Structural homophily". *International Economic Review* 56 (1): 235–264.

Boucher, Vincent and Mourifié, Ismael. 2017. "My friend far, far away: A Random Field Approach to Exponential Random Graph Models". *The Econometrics Journal* 20 (3): S14–S46.

Bramoullé, Yann, Currarini, Sergio, Jackson, Matthew O, Pin, Paolo, and Rogers, Brian W. 2012. "Homophily and Long-run Integration in Social Networks". *Journal of Economic Theory* 147 (5): 1754–1786.

Bramoullé, Yann, Djebbari, Habiba, and Fortin, Bernard. 2009. "Identification of Peer Effects through Social Networks". *Journal of Econometrics* 150 (1): 41–55.

Bruno, Michael. 1968. "Estimation of Factor Contribution to Growth Under Structural Disequilibrium". *International Economic Review* 9 (1): 49–62.

Caballero, Ricardo J and Engel, Eduardo M.R.A. 1999. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S, s) Approach". *Econometrica* 67 (4): 783–826.

Cai, Jing, De Janvry, Alain, and Sadoulet, Elisabeth. 2015. "Social Networks and the Decision to Insure". *American Economic Journal: Applied Economics* 7 (2): 81–108.

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

Chen, Xiaohong and Pouzo, Demian. 2012. "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals". *Econometrica* 80 (1): 277–321.

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mert, Duflo, Esther, Hansen, Christian, Newey, Whitney, and Robins, James. 2018a. "Double/debiased Machine Learning for Treatment and Structural Parameters". *The Econometrics Journal* 21 (1): C1–C68.

Chernozhukov, Victor, Chetverikov, Denis, and Kato, Kengo. 2018b. "Inference on Causal and Structural Parameters Using Many Moment Inequalities". *The Review of Economic Studies* 86 (5): 1867–1900.

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

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.

Collard-Wexler, Allan and De Loecker, Jan. 2016. "Production Function Estimation with Measurement Error in Inputs". *Working Paper*.

Conley, Timothy G, Hansen, Christian B, and Rossi, Peter E. 2012. "Plausibly Exogenous". *Review of Economics and Statistics* 94 (1): 260–272.

Cooper, Russell W and Haltiwanger, John C. 2006. "On the Nature of Capital Adjustment Costs". *The Review of Economic Studies* 73 (3): 611–633.

Copic, Jernej, Jackson, Matthew O, and Kirman, Alan. 2009. "Identifying Community Structures from Network Data via Maximum Likelihood Methods". *The BE Journal of Theoretical Economics* 9 (1).

Currarini, Sergio, Jackson, Matthew O, and Pin, Paolo. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation". *Econometrica* 77 (4): 1003–1045.

De Loecker, Jan, Eeckhout, Jan, and Unger, Gabriel. 2018. "The Rise of Market Power and the Macroeconomic Implications". *Working Paper*.

De Loecker, Jan, Goldberg, Pinelopi K., Khandelwal, Amit K., and Pavcnik, Nina. 2016. "Prices, Markups, and Trade Reform". *Econometrica* 84 (2): 445–510.

De Loecker, Jan and Scott, Paul T. 2016. "Estimating Market Power: Evidence from the US Brewing Industry". *Working Paper 22957. National Bureau of Economic Research.*

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

DeSouza, Sergio. 2006. "Levinsohn and Petrin's (2003) Methodology Works under Monopolistic Competition". *Economics Bulletin* 12 (6): 1–11.

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

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

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

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

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

Eberly, Janice C. 1997. "International Evidence on Investment and Fundamentals". *European Economic Review* 41 (6): 1055–1078.

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

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

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.

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

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_semina Slides_Fox.pdf.*

Gandhi, Amit, Navarro, Salvador, and Rivers, David. 2018. "On the Identification of Gross Output Production Functions". *Working Paper.*

Goldsmith-Pinkham, Paul and Imbens, Guido W. 2013. "Social Networks and the Identification of Peer Effects". *Journal of Business & Economic Statistics* 31 (3): 253–264.

Golub, Benjamin and Jackson, Matthew O. 2012. "How Homophily Affects the Speed of Learning and Best-response Dynamics". *The Quarterly Journal of Economics* 127 (3): 1287–1338.

Goolsbee, Austan and Gross, David B. 2000. "Estimating Adjustment Costs with Data on Heterogeneous Capital Goods". *Working Paper.*

Graham, Bryan S. 2016. "Homophily and Transitivity in Dynamic Network Formation". *Working Paper 22186, National Bureau of Economic Research.*

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

Grieco, Paul LE 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.

Griliches, Zvi and Mairesse, Jacques. 1995. "Production Functions: the Search for Identification". *Working Paper 5067 National Bureau of Economic Research.*

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

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

Hamermesh, Daniel S and Pfann, Gerard A. 1996. "Adjustment Costs in Factor Demand". *Journal of Economic Literature* 34 (3): 1264–1292.

Herrendorf, Berthold, Herrington, Christopher, and Valentinyi, Akos. 2015. "Sectoral Technology and Structural Transformation". *American Economic Journal: Macroeconomics* 7 (4): 104–33.

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

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.

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

Honoré, Bo E and Powell, James L. 1994. "Pairwise Difference Estimators of Censored and Truncated Regression Models". *Journal of Econometrics* 64 (1-2): 241–278.

Horváth, Lajos, Kokoszka, Piotr, and Zitikis, Ričardas. 2006. "Testing for Stochastic Dominance Using the Weighted McFadden-type Statistic". *Journal of Econometrics* 133 (1): 191–205.

Hsieh, Chih-Sheng and Lee, Lung Fei. 2016. "A Social Interactions Model with Endogenous Friendship Formation and Selectivity". *Journal of Applied Econometrics* 31 (2): 301–319.

Hsieh, Chih-Sheng and Lee, Lung-Fei. 2017. "Specification and Estimation of Network Formation and Network Interaction Models with the Exponential Probability Distribution". *Working Paper, SSRN 2967867.*

Hu, Yingyao, Huang, Guofang, and Sasaki, Yuya. 2011. "Estimating Production Functions with Robustness Against Errors in the Proxy Variables". *Working Paper.*

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

Jackson, Matthew O. 2008. "Average Distance, Diameter, and Clustering in Social Networks with Homophily". In *International Workshop on Internet and Network Economics*, 4–11. Springer.

Jackson, Matthew O and López-Pintado, Dunia. 2013. "Diffusion and Contagion in Networks with Heterogeneous Agents and Homophily". *Network Science* 1 (1): 49–67.

Jackson, Matthew O, Rogers, Brian W, and Zenou, Yves. 2017. "The Economic Consequences of Social Network Structure". *Journal of Economic Literature* 55 (1): 49–95.

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

Johnsson, Ida and Moon, Hyungsik Roger. 2017. "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach". *USC-INET Working Paper.*

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

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

Kasy, Maximilian. 2011. "Identification in triangular systems using Control Functions". *Econometric Theory* 27 (3): 663–671.

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

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

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.

Kim, Kyoo and Petrin, Amil. 2010. “Control Function Corrections for Unobserved Factors in Differentiated Product Models”.

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.

Kumar, Krishna B, Rajan, Raghuram G, and Zingales, Luigi. 1999. “What Determines Firm Size?” *Working Paper 7208 National Bureau of Economic Research.*

Levinsohn, James. 1993. “Testing the Imports-as-market-discipline Hypothesis”. *Journal of International Economics* 35 (1): 1–22.

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

Lewbel, Arthur. 2016. “The Identification Zoo-Meanings of Identification in Econometrics”. *Forthcoming on Journal of Economic Literature* 7.

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

Linton, Oliver, Maasoumi, Esfandiar, and Whang, Yoon-Jae. 2005. “Consistent Testing for Stochastic Dominance Under General Sampling Schemes”. *The Review of Economic Studies* 72 (3): 735–765.

Liu, Lili. 1993. “Entry-exit, Learning, and Productivity Change Evidence from Chile”. *Journal of Development Economics* 42 (2): 217–242.

Liu, TC and Hildebrand, George H. 1965. “Manufacturing Production Functions in the United States, 1957”. *Working Paper.*

Manski, Charles F. 1997. “Monotone Treatment Response”. *Econometrica* 65 (6): 1311–1334.

— . 1993. “Identification of Endogenous Social Effects: The Reflection Problem”. *The Review of Economic Studies* 60 (3): 531.

Manski, Charles F and Pepper, John V. 2000. “Monotone Instrumental Variables: With an Application to the Returns to Schooling”. *Econometrica* 68 (4): 997–1010.

Marschak, Jacob and Andrews, William H. 1944. “Random Simultaneous Equations and the Theory of Production”. *Econometrica* 12 (3/4): 143–205.

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

— . 2004. "Unobservable Instruments". *Working Paper.*

McPherson, Miller, Smith-Lovin, Lynn, and Cook, James M. 2001. "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology* 27 (1): 415–444.

Melitz, Marc J. 2000. "Estimating Firm-level Productivity in Differentiated Product Industries".

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America". *American Journal of Sociology* 107 (3): 679–716.

Mundlak, Yair. 1961. "Empirical Production Function Free of Management Bias". *Journal of Farm Economics* 43 (1): 44–56.

Mundlak, Yair and Hoch, Irving. 1965. "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions". *Econometrica* 33 (4): 814–828.

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

Nevo, Aviv and Rosen, Adam M. 2012. "Identification with Imperfect Instruments". *Review of Economics and Statistics* 94 (3): 659–671.

Oberfield, Ezra and Raval, Devesh. 2014. "Micro data and macro technology". *Working Paper 20452, National Bureau of Economic Research.*

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

Pavcnik, Nina. 2002. "Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants". *The Review of Economic Studies* 69 (1): 245–276.

Qu, Xi and Lee, Lung-fei. 2015. "Estimating a Spatial Autoregressive Model with An Endogenous Spatial Weight Matrix". *Journal of Econometrics* 184 (2): 209–232.

Raval, Devesh. 2019a. "Testing the Production Approach to Markup Estimation". *Working Paper.*

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

Revankar, Nagesh S. 1971. "A Class of Variable Elasticity of Substitution Production Functions". *Econometrica* 39 (1): 61–71.

Roehrig, Charles S. 1988. "Conditions for Identification in Nonparametric and Parametric Models". *Econometrica* 56 (2): 433.

Rosenbaum, Paul R and Rubin, Donald B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects". *Biometrika* 70 (1): 41–55.

Semenova, Vira. 2017. "Machine Learning for Set-Identified Linear Models". *Working Paper, arXiv:1712.10024.*

Shaked, Moshe and Shanthikumar, J George. 2007. *Stochastic Orders.* Springer Science & Business Media.

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

Shrum, Wesley, Cheek Jr, Neil H, and MacD, Saundra. 1988. “Friendship in School: Gender and Racial Homophily”. *Sociology of Education*: 227–239.

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*.

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

Van Biesebroeck, Johannes. 2003. “Productivity Dynamics with Technology Choice: An Application to Automobile Assembly”. *Review of Economic Studies* 70 (1): 167–198.

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

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

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