The Pennsylvania State University
The Graduate School

ESSAYS ON PRODUCTIVITY, UNCERTAINTY, AND FIRM ACTIVITIES

A Dissertation in
Economics
by
Shengyu Li

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2014

The dissertation of Shengyu Li was reviewed and approved* by the following:

Mark J. Roberts
Professor of Economics
Dissertation Advisor, Chair of Committee

James R. Tybout
Professor of Economics

Paul L. E. Grieco
Strumpf Early Career Professor in Economics and Assistant Professor of Economics

Spiro E. Stefanou
Professor of Agricultural Economics

Robert C. Marshall
Head, Department of Economics and Professor of Economics

*Signatures are on file in the Graduate School.

Abstract

This dissertation studies the interaction between firm heterogeneity in productivity and demand uncertainty and firm export participation. Chapter 1 develops a dynamic model to empirically study how firms' export decisions depend on productivity evolution and learning about firm-specific foreign demand. Chapter 2, a joint work with Paul L. E. Grieco and Hongsong Zhang, develops a new method to consistently estimate production functions when prices of intermediate input are heterogenous across firms but are not observed by researchers.

Chapter 1: An Empirical Structural Model of Productivity, Uncertain Demand, and Export Dynamics

This paper develops a structural model of export dynamics to empirically study how firms' market-level export decisions depend on productivity evolution and Bayesian learning about demand in foreign markets. Firms have uncertainty about foreign demand and they gradually learn about it based on the prices and quantities they observe in their individual export transactions in the Bayesian style. Firms' export decisions are dynamic and depend on the evolution of both productivity and beliefs about foreign demand. I empirically identify the role of each process in determining firm-market-level export participation and estimate the dynamic model. The identification and estimation use data on both firm shipment-level exports and firm-level production information for the Chinese ceramics industry. The empirical results indicate substantial firm heterogeneity in both productivity and demand uncertainty. Demand uncertainty is the dominant difference between potential entrants in export markets and experienced exporters. In particular, experienced exporters have higher expectations and face less uncertainty about foreign demand. Both the learning process and productivity evolution are driving forces of export participation for experienced exporters but for potential entrants the former plays a more important role. A further counterfactual exercise shows that reducing the level of uncertainty of potential entrants to that of experienced

exporters causes the number of exporters to fall by 11%.

Chapter 2: Production Function Estimation with Unobserved Input Price Dispersion

We propose a method to consistently estimate production functions in the presence of input price dispersion when intermediate input quantities are not observed. The traditional approach to dealing with unobserved input quantities—using deflated expenditure as a proxy—requires strong assumptions for consistency. Instead, we control for heterogeneous input prices by exploiting the first order conditions of the firm's profit maximization problem. We show that the traditional approach tends to underestimate the elasticity of substitution and biases estimates of the distribution parameters. Our approach applies to a general class of production functions. It can accommodate both heterogeneity in input prices and a variety of heterogeneous intermediate input types. A Monte Carlo study illustrates that the omitted price bias is significant in the traditional approach, while our method consistently recovers the production function parameters. We apply our method to a firm-level data set from Colombian manufacturing industries. The empirical results are consistent with the predictions that the use of expenditure as a proxy for quantities biases the elasticity of substitution downward. Moreover, using our preferred method, we provide evidence of significant input price dispersion and even wider productivity dispersion than is estimated using proxy methods.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest gratitude to my adviser Mark J. Roberts for his excellent guidance, great patience, and constant support, without which this dissertation would not have been possible. I especially feel grateful for every late evening that he spent on my paper, practice seminar, and even correcting my writing. His guidance and encouragement have made this a thoughtful and rewarding journey.

I am deeply grateful to my committee members Paul L. E. Grieco, Spiro E. Stefanou, and James R. Tybout. I appreciate their insightful comments, helpful remarks, and continual engagement throughout the whole process of my dissertation research. I have benefited very much from their wisdom and knowledge as I moved from an idea to a completed study.

I also appreciate very valuable comments from Neil Wallace, Stephen Yeaple, Ruilin Zhou, Theodore Papageorgiou, James Jordan, Peter Newberry.

Furthermore, I want to thank my colleagues and friends, especially Hongsong Zhang, for their generous help and insightful comments.

I owe a great debt to my parents as well as my wife, Yunyun Chen. Without your love and support, this effort would have been worth nothing. I would particularly like to thank my daughter, Clara Li, who was born before this dissertation was completed. Thank you for being a lovely and strong baby. Watching you growing up is the most enjoyable thing everyday!

All errors are my own responsibility.

Chapter 1

A Structural Model of Productivity, Uncertain Demand, and Export Dynamics

1.1 Introduction

[1]A general finding in the recent literature using firm-level trade data is that firms' export decisions are driven by unobservable firm heterogeneity in both cost and demand dimensions (e.g., Melitz (2003); Bernard, Eaton, Jensen, and Kortum (2003); Eaton, Kortum, and Kramarz (2011); Roberts, Xu, Fan, and Zhang (2012)). More productive firms enjoy lower marginal costs in production; firms with high foreign demand earn larger market shares at given prices. Both mechanisms contribute to the heterogenous export decisions observed across firms. However, some dynamic features of exporting, such as the high attrition after the first year of exporting and gradually stabilized export decisions,[2] are not reconciled by these mechanisms but are consistent with firms' behavior when they face uncertainty about foreign demand. Indeed, firms are likely to have uncertainty about demand when they enter into unfamiliar foreign markets. For example, they may be uncertain about

[1]I gratefully acknowledge the guidance, support, and encouragement from my dissertation committee: Mark Roberts, James Tybout, Paul Grieco, and Spiro Stefanou. I also appreciate valuable comments and suggestions from Neil Wallace, Stephen Yeaple, Ruilin Zhou, Theodore Papageorgiou, and Hongsong Zhang. All errors are mine. All comments are very welcome.

[2]These patterns are shown in Section 2.

how foreign customers perceive the quality of their products. But if they export, they are able to learn about it from the quantities sold in the export market, after controlling for prices. As a result, demand uncertainty induces firms to export as an experiment; but whether or not to continue exporting depends on their expected demand which changes over time based on the export outcomes they observe. Essentially, firms' export decisions do not only depend on their productivity but also rely on their demand expectations which are likely to evolve endogenously as firms export.

This paper studies the role of learning mechanisms in explaining dynamic features of exporting and contributes to the vibrant area of related research (Rauch and Watson (2003); Freund and Pierola (2010); Albornoz, Calvo Pardo, Corcos, and Ornelas (2012); Nguyen (2012); Eaton, Eslava, Jinkins, Krizan, and Tybout (2013)).[3] First, I develop a single-agent infinite-horizon dynamic model of export decisions in which the firm faces uncertainty about foreign demand and gradually learns about it from its export transactions according to Bayes' rule. In particular, the size of each individual transaction, after controlling for the price, influences the firm's expectation about foreign demand; the number of transactions in each period affects the speed of learning. Second, I take the important driving force, productivity evolution, into account when I assess the role of the learning process. That is, the firm's export decision depends on both productivity and its demand belief. Both of them evolve over time, but neither of them is observable to researchers. I empirically identify the role of each process in determining firm-market-level export participation and estimate the dynamic model using shipment-level exports and firm-level production data for the Chinese ceramics industry.

The recent empirical literature has documented firm-level demand heterogeneity as a determinant of firm performance (e.g., Foster, Haltiwanger, and Syverson (2008)). In the export context, Roberts, Xu, Fan, and Zhang (2012) find substantial firm heterogeneity in both the demand and cost dimensions with demand being more dispersed. My paper builds on this insight: firms face heterogenous demand in foreign markets. But the novelty is that they have uncertainty about foreign

[3] Some other empirical papers study the role of uncertainty and learning in a structural model. For example, see Ackerberg (2003) for a dynamic learning model to study both "informative" and "prestige" effects of advertising, and Crawford and Shum (2005) for a study of effects of uncertainty and learning in the demand for pharmaceutical drugs.

demand, thus it is the belief about demand that really influences a firm's export decision. Moreover, a forward-looking firm will expect to update its demand belief after exporting, so its demand belief endogenously evolves as the firm observes more of its own export outcomes. Specifically, I explicitly model the firm-market-specific demand faced by a firm as a demand factor. It reflects a combination of customer taste and the relative product quality difference between the firm and other suppliers, both of which introduce potential uncertainty about demand. In particular, the demand factor is used to model the intercept of the demand curve faced by the firm in the foreign market. The firm's export decision is influenced by its demand factor since a higher demand factor implies more profit at a given price. However, the firm only knows the slope but not the intercept (demand factor) of the demand curve,[4] and makes the export decision based on its belief about the demand factor. This belief can be updated based on the firm's own export experience in the foreign market. To be specific, the quantity sold in each individual transaction reflects a signal of the demand factor, after controlling for the price. The signal is noisy because it contains an idiosyncratic demand shock which is not observed by the firm. Due to the noise, demand uncertainty is not completely resolved immediately after just one transaction. Nonetheless, the firm is able to update its belief at the end of each period based on the received signals in the Bayesian style. This posterior belief then guides the firm to make its export decision in the next period. In this way, the export decision and the belief updating are endogenously correlated: the firm makes export decision based on its current demand belief knowing that its exports in this period will update the demand belief which will further affect its future export decisions.

However, the assessment of the role of the learning process will potentially be biased if another driving force of export dynamics, productivity evolution, is ignored. Productivity has been recognized as a distinct cost-side heterogeneity in studies of firm performance and survival (e.g., Baily, Hulten, Campbell, Bresnahan, and Caves (1992); Baldwin and Gorecki (1998)),[5] as well as in explaining firm-level

[4]Aw, Roberts, and Xu (2011) find that the slope of the demand curve in the foreign market, i.e., demand elasticity, is virtually identical to that of the domestic market for Taiwanese electronics industry. This result is also found for Chinese industries in this paper. Thus, compared with the intercept, firms face far less uncertainty about the slope of the demand curve in the foreign market since they have been operating in the domestic market for a long time.

[5]See Bartelsman and Doms (2000), and Syverson (2011) for reviews.

export participation (e.g., Melitz (2003); Bernard, Eaton, Jensen, and Kortum (2003); Helpman, Melitz, and Yeaple (2004)): more productive firms have lower marginal costs and tend to enter the export market while less productive firms only serve the domestic market. In particular, Eaton, Kortum, and Kramarz (2011) show that over half the variation across firms in market entry can be attributed to productivity heterogeneity. In the time series dimension, Aw, Roberts, and Xu (2011) find that firm choices of R&D and exporting have a positive effect on the firm's future productivity, which in turn drives more firms to self-select into both activities. Thus, without controlling for the change of productivity at firm-level, it is likely to attribute the effect from the productivity shock to the role of the learning process.[6] So in my model I take care of this issue. Firms' export decisions depend on two types of firm heterogeneity: productivity and the demand belief. But this places a challenge in identifying the role of the learning process from the effect of productivity evolution, since both of them are unobservable to researchers.

To identify the role of each process, I rely on two data sets. The first one is the Chinese Annual Survey of Manufacturing. It provides firm-level production information: employment, labor and material expenditures, capital stock, and domestic revenue. The second one is the Chinese Monthly Customs Transactions. It contains shipment-level exports: the export destination, quantity and price of each shipment, and shipment month. The identification strategy uses the insight that domestic revenue is only affected by productivity while export participation is influenced by both productivity and the demand belief. Thus, I utilize data from the domestic market to recover the time-varying productivity for each firm. Meanwhile, shipment-level exports contain information on how Bayesian learning occurs and allow me to recover a market-specific demand belief for each firm in each period. A model with only productivity heterogeneity predicts more productive firms export. While in my model with the two-dimensional heterogeneity, firms face greater uncertainty about demand may also export because of the larger option value of learning, even if their productivity is not high. In turn, with both productivity and demand beliefs recovered, the cross-sectional and time series patterns of export decisions identify the role of each driving force.

[6]Some other papers also illustrate the importance of productivity evolution in export decisions. For example, see Irarrazabal and Opromolla (2006).

The identification strategy suggests a structural estimation approach with two stages. In the first stage, I utilize the data from the domestic market to recover firm-level productivity, its evolution process, and firm-level marginal cost. Then I take the quantities in individual shipments after controlling for prices as demand signals received by firms and use them to update firms' demand beliefs in each period. In the second stage, I estimate an discrete choice model of export participation via the Maximum Likelihood Method with the Nested Fixed Point Algorithm as in Rust (1987). The major estimation difficulty comes from the internal calculation of the value function in each evaluation of the likelihood. Since the belief enters the value function as a two-dimensional state variable *in addition* to productivity and the aggregate demand/cost shifter, the number of state variables makes it time-consuming to use the traditional method of value function iteration to derive the underlying value function. To solve this problem, I follow the method in Nagypál (2007) to compute the value function efficiently. I first calculate the value function without uncertainty. Then I use it as an approximation of the value function with a low level of uncertainty. This in turn allows me to compute the value function with arbitrarily greater uncertainty by backward induction.

In the empirical estimation, I focus on the exports of the Chinese ceramics industry to Germany from 2000 to 2006. These firms export colorful dinnerware and ornamental articles of ceramics, such as statuettes. This industry fits the study purpose well. In this industry, most export transactions are ordinary trade, in which firms make their own decisions of production, pricing, and exporting, without facing constraints from existing contracts with foreign companies (as in processing trade). As a result, demand uncertainty, which comes from how foreign customers perceive the appeal of the product, is potentially an important issue to consider when firms make export decisions. To control for the initial condition, I divide firms into two groups, potential entrants in the foreign market and experienced exporters, according to their export status in year 2000. I allow the two groups to hold different beliefs when they first show up in the data set. The empirical results indicate substantial firm heterogeneity in both productivity and demand uncertainty. Demand uncertainty is the dominant difference between potential entrants and experienced exporters. In particular, experienced exporters have higher expectations and face less uncertainty about demand compared with potential entrants. This is reasonable, since experienced exporters may have operated in the foreign market for a long time and have learned a lot before 2000. I also find that the value difference between the decisions to export and not to export falls as demand uncertainty is resolved. This mechanism contributes to the observed high attrition after the first year of exporting.

Using the estimated model I conduct two counterfactual exercises to quantify the roles of productivity evolution and the learning process in determining firm export participation. In the first analysis, I shut down either the evolution of productivity or belief updating in order to evaluate how export participation is influenced by that process. The comparison of the two scenarios shows that, for experienced exporters, both evolutions significantly influence export participation while the learning process has a larger impact for potential entrants. In the second analysis, I experiment with the prior belief of potential entrants to study how it affects export participation. I find that reducing the level of uncertainty of potential entrants to that of experienced exporters causes the number of exporters to fall by 11%. While increasing the expectation of potential entrants to that of experienced exporters doubles the number of exporters. If potential entrants have the same prior belief as experienced exporters, then the percentage of exporters increases from 26% to 40%.

This paper is organized as follows. Section 2 motivates the paper by documenting dynamic features at firm-market level for Chinese manufacturing industries. Section 3 develops a structural model of export dynamics that incorporates heterogeneity in both productivity and Bayesian learning. Section 4 outlines an identification and estimation strategy. Section 5 describes the data sources for the empirical estimation. The estimation results are shown in Section 6. Section 7 conducts counterfactual exercises. I conclude in Section 8 with discussions for future work.

1.2 Firm-market Level Export Dynamics

This paper is empirically motivated by dynamic features of firm-market-level export participation patterns that are not easily reconciled by productivity evolution but are consistent with a model of uncertainty and Bayesian learning about foreign demand. The features presented in this section are based on the Chinese Monthly Customs Transactions data set.[7] This data set includes all export shipments of Chinese firms from 2000 to 2006. Each shipment contains export value, quantity, 8-digit HS code, shipment month, destination market, and firm identification number.

First, a large percentage of firms drop out after the first year of exporting to a market. I follow Eaton, Eslava, Jinkins, Krizan, and Tybout (2013) and consider the export cohorts that began exporting to the U.S. in a particular year after at least one year of no exporting. In Table 1.1, each column reports the percentages of firms within each cohort that chose to export after the year of entry. The pattern shows that there is a significant percentage (around 30%) of firms that drop out of the U.S. market after the first year of exporting. However, conditional on survival in the second year, the percentage of exporters stays roughly the same. Aggregate demand shocks cannot explain the large attrition rate after the first entry, since the high attrition appears in all cohorts that entered in different years and in other destinations. Did it result from a negative firm-level productivity shock? Using firm-level production data, I find that, for firms that dropped out after the first year of exporting, 45% of them were experiencing decreases in relative productivity (measured as output per worker) while 55% of them were experiencing increases in relative productivity. Thus, although it is possible that productivity evolution plays a role, it is unlikely that it alone can capture the entire significant attrition.

Second, the export decisions of new exporters *gradually* become stable over time. To ensure that the firms I am looking at are more likely to be new exporters in the U.S. market, I consider the firms that started to export to the U.S. in the last quarter of 2001 but did not export to the U.S. prior to that. I calculate the percentage of these firms that switched their export status (from export to not to export, or vice versa) in each quarter after 2002. Figure 1.1 shows that overall this percentage becomes smaller over time. This implies that new exporters' export decisions become more stable over time. More importantly, the percentage *gradually* becomes stable after several continuing significant drops. This pattern is not limited to the U.S. market or the firm cohort that began to export in

[7] The patterns are for medium and large manufacturing firms that also show up in the Chinese Annual Survey of Manufacturing.

a particular year, but is a universal feature that also appears in other export cohorts and in other destinations. Productivity evolution can account for the stabilized export participation, but it alone is not capable to capture the continuing significant drops in the beginning. Nonetheless, this feature is consistent with an environment of uncertainty and Bayesian learning about demand, in which firms update their beliefs about foreign demand based on their export outcomes.[8]

These dynamic features show that it is possible that both productivity evolution and Bayesian learning about demand contribute to the observed patterns, but neither of them alone can explain all patterns. This motivates me to incorporate both processes as driving forces to reconcile these patterns and study how they influence firm-level export dynamics separately.

1.3 The Model

In this section, I develop a structural model of firm-market-level export dynamics with heterogenous firms learning about demand curves they face in foreign markets. A novelty of the model is that it considers how the two processes, productivity evolution and the evolution of the demand belief, affect a firm's export participation. In particular, the firm learns about its foreign demand from the prices and quantities it observes in its export transactions in the Bayesian style, and in turn the firm's belief about foreign demand endogenously evolves as it exports. I first provide an overview of the model then specify the details.

1.3.1 An Overview

Consider an industry with I single-product firms and an infinite-period horizon. There are a total of J foreign markets in addition to the domestic market. Firms are heterogenous in both the cost and demand dimensions. In the cost dimension, firms are different in productivity, a measure of efficiency in converting input to output. Higher productivity implies lower marginal cost in production. In the demand

[8]Potentially, a simpler model without Bayesian learning is to allow the one-shot learning and serially correlated idiosyncratic demand, which is observed by firms but is unobservable to researchers. However, such a model will predict the switching percentage to drop very shapely after the first entry and be stable afterward, which is not consistent with the observation in Figure 1.1.

dimension, firms are heterogenous in the product quantity demanded at a given price, which results from a combination of customer tastes and the relative product quality difference between the firm and international competitors. I refer to this heterogeneity, which is not captured by the aggregate demand shifter, as the firm's demand factor. The demand factor is allowed to be different across foreign markets for each firm, in order to capture different consumer tastes and international sellers in different markets. For each firm, productivity evolves exogenously from one period to another,[9] but the demand factor in each market is constant.

In each period t, each firm makes decisions of production, pricing, and export participation in each foreign market and the domestic market.[10] Two types of information are crucial for it to make these decisions. The first one is its productivity level, and the second is its demand factor in each foreign market. At the beginning of each period t, firms observe the realization of their productivity ω_{it}. However, in contrast to the current literature (e.g. Roberts, Xu, Fan, and Zhang (2012)), this paper assumes the foreign demand factor is firm-market specific and unknown to firms. As a result, each firm makes the export decision based on its current belief about its demand factor in each foreign market at the beginning of period t.

For the sake of simplicity, I assume that the firm's export decisions for different markets are independent.[11] If the firm decides to export to market j, then it pays either a fixed cost (c_{it}^{fj}) or a sunk entry cost (c_{it}^{sj}), depending on its export status in market j in the last period $t-1$. Once in market j, firm i receives (and fulfills) n_{it}^{j} orders from customers in that market. By observing the quantity demanded at a given price in *each* transaction, the firm can learn about its demand factor

[9]If firms endogenously choose productivity, by investing in R&D for example, then more productive firms are more likely to conduct R&D (Doraszelski and Jaumandreu (2013)). The assumption of exogenous productivity evolution will attribute the impact from R&D to the high persistence of productivity. I abstract from the endogenous productivity evolution because of lack of R&D data in this paper. I discuss the future research along this direction in the conclusion section.

[10]I abstract from the decision on enter or exit production, and assume that each firm in production will serve the domestic market.

[11]Independence of export decisions does not necessarily imply the export participation outcomes are independent. Actually, the outcomes are correlated because of productivity of manufacturing goods for different markets is the same within the firm. Similar independence assumption is employed in the recent literature (e.g., Eaton, Kortum, and Kramarz (2011)).In the industry I will consider, the correlation of export participation in two destination is only around 0.1.

in market j. In particular, the belief about the demand factor is updated at the end of period t in the Bayesian style, based on these observations. However, if the firm decides not to export in market j, then it does not pay the fixed/sunk cost, and there is no export profit generated from that market. More importantly, the belief regarding the demand factor in that market will not be updated at the end of period t because no transaction has happened (or been observed).

Several points need to be clarified. First, the learning processes of different firms in different markets are assumed to be independent. This assumption is motivated by data. This paper study the kind of industries in which firms mainly conduct ordinary trade and demand uncertainty is potentially an important issue when firms make export decisions. For these industries, including ceramics industry, fiber optics industry, and many others, it is common that firms have large sales in one destination but no sale in another. For example, the correlation coefficients between export revenue across destinations are on average 0.15 in the ceramics industry. This implies the firm export revenue is not highly correlated in different destinations. This is consistent with the assumption that the underlying demand factor is firm-market specific, thus the independent learning assumption is reasonable, especially for markets with great cultural and geographic distinctions.[12] Second, it is feasible to consider that productivity evolution is affected by export participation; however, export participation in a *single* market usually has an insignificant effect on productivity evolution.[13] Thus, I do not consider such productivity gains from exporting in this paper. However, the belief about the demand factor in each market is endogenously related to the export participation in that market. Third, the (period by period) export growth at firm level is captured by the increase in the number of orders that the firm receives in a period (i.e., n_{it}^j will increase if the firm keeps exporting). This captures the possibility that the stock of customers who are familiar with the firm or the firm's distribution

[12]Some papers relax this assumption, but they either consider a finite-horizon model (e.g., Albornoz, Calvo Pardo, Corcos, and Ornelas (2012)), or one-shot learning (e.g., Nguyen (2012)). Also, they do not estimate a structural model, which is important to quantify the endogenous dynamic relationship between export participation and learning. A direction for future work is to relax this assumption, but assuming the demand factor has the global scope as in Albornoz, Calvo Pardo, Corcos, and Ornelas (2012).

[13]I considered an AR(1) process of productivity evolution, in which there is productivity gain from exporting: $\omega_{it} = g_0 + g_1\omega_{it-1} + g_2 e_{it-1}^j + \epsilon_{it}$, where e_{it-1}^j is an indicator of export decision of last period in a specific market j. It turns out that g_2 is usually insignificant.

network is increasing over time if the firm keeps exporting.

The timing is summarized as follows:

t — Choose e_{it}^{j} — Update to b_{it}^{j}

Observe ω_{it}, b_{it-1}^{j}, c_{it}^{fj}, c_{it}^{sj} — Receive, fulfill n_{it}^{j} orders — $t+1$

where ω_{it} is firm i's productivity in period t, b_{it-1}^{j} represents the belief about the demand factor in market j at the beginning of period t (or at the end of period $t-1$), and $e_{it}^{j} = 0$ or 1 represents the export participation in market j in period t.

In what follows, I describe the elements of the model in detail by considering the firm's static decisions (on production and pricing), dynamic decisions (on exporting), as well as the evolution of the firm's belief about foreign demand.

1.3.2 Static Decisions

The firm's static decision is to set prices for the domestic market and foreign markets (where it decides to export) so that it maximizes its current profit in each period, after observing its productivity, capital stock, and aggregate demand in each market.

1.3.2.1 Cost of Production

Each firm i faces a constant short-run marginal cost to produce its product for domestic and foreign markets. The logarithm of the marginal cost for period t is written as:[14]

$$\ln C_{it} = \gamma_0 + \gamma_w \ln W_{it} + \gamma_k \ln K_{it} - \omega_{it}, \tag{1.1}$$

where W_{it}, K_{it}, and ω_{it} are the is the wage rate, capital stock, and productivity of firm i in period t respectively. Within each firm and each period, the productivity is the same to produce goods for domestic market as well as all foreign markets.[15]

[14] A similar specification is used in Aw, Roberts, and Xu (2011).

[15] I assume the technology employed in manufacturing products for different markets is the same within the firm in each period. Nonetheless, it is possible that firms ship products with different quality (and different marginal costs) to different destinations as documented in Bastos and Silva (2010). In this case, I can use a set of market dummies in the marginal cost function to control for the destination effect. These market dummies can be identified using the difference in average export prices across markets. I exclude the dummies in the specification since I will consider ceramic industry, and there is not much difference in producing product for different

That is, the firm faces the same short-run marginal cost in all markets for the entire period t. However, the marginal cost does vary across firms and over time. The variation is captured by two sources of heterogeneity. The first one is the observable heterogeneity of K_{it} and W_{it}, which is exogenously given in each period.[16] The second source is captured by ω_{it}, the firm-time specific productivity. It is also observed by the firm, but is not observable by researchers. ω_{it} is assumed to evolve according to an exogenous first-order Markov process:

$$\omega_{it} = g(\omega_{it-1}) + \epsilon_{it}, \tag{1.2}$$

where ϵ_{it} is an innovation term that is independently drawn from $N(0, \sigma_\epsilon)$.

One important feature of this specification is that the cost-side heterogeneity both across firms and over time is taken into account when I examine the export dynamics with a learning process regarding demand in a foreign market.

1.3.2.2 Demand and Pricing

The domestic market is assumed to be monopolistically competitive. In particular, firm i faces the Dixit-Stiglitz type demand curve in each period t:

$$Q_{it}^D = Q_t^D \Big(\frac{P_{it}^D}{P_t^D}\Big)^{\eta^D} \equiv \Phi_t^D (P_{it}^D)^{\eta^D}, \tag{1.3}$$

where Q_t^D and P_t^D are industrial-level output and price index respectively and η^D is the demand elasticity. Note that this demand function describes an aggregate relationship between price and quantity: the total amount of firm i's product sold in the domestic market in the entire period t depends on the industrial aggregate Φ_t^D, firm i's price P_{it}^D, and the constant demand elasticity η^D.

After observing Φ_t^D and its marginal cost C_{it}, the firm sets a price to maximize

markets.

[16] In reality capital stock may evolve endogenously, and more productive firms may choose to invest in the capital and reduce their marginal costs. This leads to firms with high productivity and capital more likely to export. But in the empirical application, I consider a relatively short period and the investment in the capital is lumpy, so I follow the literature and treat the capital as exogenous.

its profit in the domestic market in period t:

$$\max_{P_{it}^D} \Phi_t^D (P_{it}^D)^{\eta^D} (P_{it}^D - C_{it}). \quad (1.4)$$

The first order condition implies

$$P_{it}^D = \frac{\eta^D}{1+\eta^D} C_{it}. \quad (1.5)$$

Thus the domestic revenue for period t is (in logarithm):

$$\ln R_{it}^D = (\eta^D+1)\ln(\frac{\eta^D}{\eta^D+1}) + \ln \Phi_t^D + (\eta^D+1)(\gamma_0 + \gamma_w \ln W_{it} + \gamma_k \ln K_{it} - \omega_{it}). \quad (1.6)$$

That is, the domestic revenue can be written as a function of demand elasticity, industrial aggregate, capital stock, and productivity. Notice that although ω_{it} is referred as productivity, it is essentially the combination of physical productivity which influences production cost directly, as well as the effects from product characteristics (e.g., quality) that would affect the quantity of product demanded in all markets.

For each foreign market j, the firm's demand is heterogenous in two aspects. First, firms are different in the expected number of orders they will receive in period t. This reflects differences in the stocks of customers and the sizes of firms' distribution networks. Specifically, if firm i decides to export to market j, then the number of orders it will receive, n_{it}^j, is modeled as an exogenously draw from a truncated Poisson distribution with parameter λ_{it}^j.[17] Note that λ_{it}^j is the expected number of orders in period t. In reality, exporting does not only generate profit but also builds up customer stock over time. To capture this feature, I assume that firms exporting in the last period are likely to receive more orders in the current period. That is, $n_{it}^j \sim \text{Poisson}(\lambda_{it}^j)$, and λ_{it}^j evolves over time according to[18]

$$\ln \lambda_{it}^j = \psi_0 + \psi_1 \ln(\tilde{n}_{it-1}^j) + \psi_2 e_{it-1}^j, \quad (1.7)$$

[17] I use the truncated Poisson to model the number of orders because I assume that, if the firm exports to the foreign market, it will receive at least one order.

[18] Arkolakis (2010) provides a theory of market penetration costs in which paying higher costs allows firms to reach an increasing number of consumers in a country. But in this paper I assume this reduce form evolution to keep the analysis tractable.

where $\tilde{n}^j_{it-1}$ is the number of transactions in the most recent period up to period $t-1$ and e^j_{t-1} is the indicator of export status in period $t-1$.[19] Suppose $\psi_2 > 0$, then this implies that firms continuing to export are likely to have contact with more customers and build up their customer stocks, which will then lead to export growth via increasing in the number of future transactions.[20] Note that λ^j_{it} is different across firms and over time but is known by the firm when it makes its export decision.

The second aspect of demand heterogeneity is the quantity demanded at a given price in a single order. With a slight abuse of the notation, I use n as the index of orders (or transactions) of firm i in market j and period t. For each firm i, the quantity sold in each order n in market j is endogenously determined by price P^j_{in}, an aggregate demand shifter ϕ^j_t, and an idiosyncratic demand shock which can be decomposed as $\zeta^j_{in} = \xi^j_i + u^j_{in}$. Here ξ^j_i measures the average demand specific to firm i that is not captured by the aggregate demand shifter ϕ^j_t. It captures all sources of firm i's demand heterogeneity that are unique to market j. For example, it may reflect a combination of customer tastes and the relative product quality difference between the firm and local/other international suppliers in market j. I refer to ξ^j_i as the demand factor, which is a constant over time within firm i and market j but is different across firms and markets. u^j_{in} is an unexpected idiosyncratic demand shock associated with order n and is (i.i.d.) drawn from $N(0, \sigma^j_u)$. Specifically, the quantity demanded in transaction n is given by the demand curve:[21]

$$Q^j_{in} = \phi^j_t (P^j_{in})^{\eta^j} e^{\zeta^j_{in}}, \tag{1.8}$$

where η^j is the demand elasticity in market j.

The firm's profit maximization problem for each specific transaction n is to set a

[19]For example, if the firm exported in period $t-1$ with n_{it-1} transactions, then $\tilde{n}_{it-1} = n_{it-1}$ and $e_{it-1} = 1$. However, if the firm did not export in period $t-1$, but exported in period $t-2$ with n_{it-2} transactions, then $\tilde{n}_{it-1} = n_{it-2}$ and $e_{it-1} = 0$. For firms that never exported, then I use the average number of transactions of exporters as a proxy of $\tilde{n}_{it-1}$ in the empirical estimation.

[20]Note that this evolution process is stationary conditional on the firm's export decision.

[21]Note that this implies an analogy of *aggregate* expected demand curve as in the domestic market: $E(Q^j_{it}) = \lambda^j_{it}\phi^j_t(P^j_{it})^{\eta^j} E(e^{\xi^j_i})e^{\sigma^2_u/2} \equiv \Phi^j_{it}(P^j_{it})^{\eta^j}$, where the firm's heterogeneity in total foreign demand from market j in period t, i.e. Φ^j_{it}, consists of λ^j_{it} and $E(e^{\xi^j_i})$. A similar specification is employed in Eaton, Eslava, Jinkins, Krizan, and Tybout (2013).

price that maximizes the profit in this transaction. At the beginning of period t, the firm has not observed the demand shock ζ_{in}^j yet, but it has observed its marginal cost C_{it} and the aggregate demand shifter ϕ_t^j. Thus, the profit maximization problem for transaction n is

$$\max_{P_{in}^j} \phi_t^j (P_{in}^j)^{\eta^j} (P_{in}^j - C_{it}) \mathrm{E}(e^{\zeta_{in}^j}). \tag{1.9}$$

Note that $\mathrm{E}(e^{\zeta_{in}^j})$ does not affect the optimal pricing rule, but it does influence the expected quantity demanded as well as the expected profit in this transaction. Specifically, the first order condition implies the following pricing rule:

$$P_{in}^j = \frac{\eta^j}{1+\eta^j} C_{it}. \tag{1.10}$$

Thus, the expected profit for a single transaction n is

$$E(\pi_{in}^j) = \phi_t^j \frac{-1}{1+\eta^j} \left[\frac{\eta^j}{1+\eta^j}\right]^{\eta^j} C_{it}^{1+\eta^j} E(e^{\zeta_{in}^j}). \tag{1.11}$$

Consequently, the expected total export profit for the entire period t is the sum of profit generated by all transactions in period t:

$$\begin{aligned} E(\Pi_{it}^j) &= E\Big(\sum_{n=1}^{n_{it}} \pi_{in}^j\Big) = E\Big(\sum_{n=1}^{n_{it}^j} \phi_t^j \frac{-1}{1+\eta^j} \left[\frac{\eta^j}{1+\eta^j}\right]^{\eta^j} C_{it}^{1+\eta^j} e^{\zeta_{in}^j}\Big) \\ &= \lambda_{it}^j \phi_t^j \frac{-1}{1+\eta^j} \left[\frac{\eta^j}{1+\eta^j}\right]^{\eta^j} C_{it}^{1+\eta^j} e^{\sigma_u^2/2} E(e^{\xi_i^j}). \end{aligned} \tag{1.12}$$

The last equation holds because $E(e^{\zeta_{in}^j}) = e^{\sigma_u^2/2} E(e^{\xi_i^j})$ and the expected number of orders is λ_{it}^j. Therefore, the expected total export profit for period t depends on the aggregate demand shifter ϕ_t^j, the marginal cost C_{it}^j, the expected number of transactions λ_{it}^j, and the expectation of the demand factor $E(e^{\xi_i^j})$. However, the firm *may* have no previous sales in this market before and thus faces uncertainty about the demand factor ξ_i^j. Thus, it is the belief about ξ_i^j that determines $E(e^{\xi_i^j})$, which in turn influences the expectation of export profit in the entire period t.

Nonetheless, the firm observes the quantity demanded in each order n *after* exporting, which reflects the realization of ζ_{in}^j:

$$\zeta_{in}^j = \ln Q_{in}^j - \ln \phi_t^j - \eta^j \ln P_{in}^j. \tag{1.13}$$

That is, (Q_{in}^j, P_{in}^j) contains information about ζ_{in}^j. Since ζ_{in}^j can be viewed as a signal of ξ_i^j, the firm is able to learn about its underlying ξ_i^j from these signals and subsequently updates its belief about ξ_i^j. As a result, the belief regarding ξ_i^j evolves as the firm exports, and this comes to affect the expected value of exporting in the future. The next subsection explains this learning process in greater detail.

1.3.3 Demand Uncertainty and Bayesian Learning by Exporting

In this section, I characterize uncertainty and Bayesian learning about foreign demand by exporting in a single market j. Since I assume the learning processes of different firms in different markets are independent, I omit the superscript for market j in order to simplify the notation.

I start to model the learning process by specifying firms' knowns and unknowns. The objective of the learning process is the demand factor ξ_i, which is unknown by firm i.[22] However, the firm knows the distribution of ξ for the entire industry in that market, and this serves as a prior belief about its specific ξ_i. More generally, I allow for observable heterogeneity in the prior belief across firms and model it as a function of firm-market characteristics. For example, firms may believe the demand is higher in markets with larger populations; also, some firms may have operated in the foreign market for a long time and have learnt a lot (thus with a low uncertainty) before I observe them in the data set. Formally, each firm holds a (known) prior belief $\xi_i \sim N(m_{i0}, \sigma_{i0})$ at the beginning of the initial period.[23] $m_{i0} = h_m(x_{i0}, z)$ is the prior expectation of the demand factor ξ_i while $\sigma_{i0} = h_\sigma(x_{i0}, z)$ captures the initial uncertainty, where x_{i0} is the firm's characteristics, such as age

[22] Another way to introduce the demand uncertainty is to assume λ_{it} is unknown by the firm. However, the variance of the number of orders is large even within a firm, and this suggests that the learning speed is too slow to be consistent with data. In contrast, modeling the demand factor with uncertainty fits the data better.

[23] Note that "the initial period" means the first period when a firm appears in the data set. It is not necessary to be the period of its first export.

and ownership, and z is the foreign market's characteristics, such as population and GDP.[24] In particular, if $\sigma_{i0} = 0$, then the firm has no uncertainty about the demand factor.

The firm can learn the true value of the demand factor ξ_i by observing ζ_{in} in each of its transactions. In particular, assuming it is exporting, each firm i observes ζ_{in} as a signal of the demand factor ξ_i. Note that $\zeta_{in} = \xi_i + u_{in}$, but ξ_i and the unexpected idiosyncratic demand shock u_{in} are not separately observed. Thus, the value of ξ_i is not immediately revealed because of the noise u_{in}. However, the firm knows the distribution of the noise: $u_{in} \sim N(0, \sigma_u)$. This knowledge enables the firm to update its belief according to Bayes' rule after observing a series of signals.[25] The standard deviation σ_u determines the speed of learning. In the extreme case where $\sigma_u = 0$, ξ_i can be accurately revealed after just one transaction. However, if σ_u is large, then the firm needs more signals to achieve a given level of accuracy.

Specifically, given the belief at the beginning of period t as $\xi_i \sim N(m_{it-1}, \sigma_{it-1})$, if firm i has decided not to export in period t, then the firm will not observe any new signals of its demand factor and its belief will be the same as it was at the beginning of period $t+1$. However, if the firm has decided to export and conducts n_{it} transactions in period t, then the belief will be updated after receiving n_{it} pieces of information $\{\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{in_{it}}\}$ about the true demand factor in that period. Consequently, the posterior belief *at the end of period* t is given by $\xi_i \sim N(m_{it}, \sigma_{it})$, where[26]

$$m_{it} = \begin{cases} \dfrac{\sigma_u^2 m_{it-1} + \sigma_{it-1}^2 \tilde{\zeta}_{it}}{\sigma_u^2 + \sigma_{it-1}^2 n_{it}}, & \text{if exported in period } t \\ m_{it-1}, & \text{otherwise} \end{cases} \tag{1.14}$$

[24] Adding more characteristics, say age, may help to control for the heterogeneity in the prior belief. For example, a firm established 30 years ago may have learnt its demand factor; while a recently founded firm may have never exported to the foreign market before and faces greater uncertainty. In the empirical estimation, to reduce the estimation burden, I will use the first year export status as an initial condition to control for heterogeneity in the prior beliefs, but it is readily be extended to incorporate more firm characteristics.

[25] I can also allow the firm to update the belief after each transaction, but since the firm only makes export decision in the beginning of each period, the two specifications are equivalent.

[26] See DeGroot (2005), Chapter 9.

and

$$\sigma_{it}^2 = \begin{cases} \dfrac{\sigma_{it-1}^2 \sigma_u^2}{\sigma_u^2 + \sigma_{it-1}^2 n_{it}}, & \text{if exported in period } t \\ \sigma_{it-1}^2, & \text{otherwise} \end{cases} \tag{1.15}$$

and

$$\tilde{\zeta}_{it} = \sum_{n=1}^{n_{it}} \zeta_{in}. \tag{1.16}$$

Alternatively, the above equations can be written in terms of the initial belief and the entire history of signals received until period t: $\{\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{iN_{it}}\}$ (where $N_{it} = \sum_{\tau=1}^{t} n_{i\tau}$):

$$m_{it} = \frac{\sigma_u^2 m_{i0} + \sigma_{i0}^2 \overline{\zeta}_{it}}{\sigma_u^2 + N_{it}\sigma_{i0}^2}, \tag{1.17}$$

and

$$\sigma_{it}^2 = \frac{\sigma_{i0}^2 \sigma_u^2}{\sigma_u^2 + N_{it}\sigma_{i0}^2}, \tag{1.18}$$

and

$$\overline{\zeta}_{it} = \sum_{n=1}^{N_{it}} \zeta_{in}. \tag{1.19}$$

Note that, in each period t, the belief about ξ_i is characterized by two variables: the mean and the standard deviation of the belief. The mean represents the expectation of the demand factor ξ_i and may fluctuate over time depending on the entire history of signals that the firm received. However, the standard deviation, which measures the magnitude of uncertainty, is strictly decreasing as the firm receives more signals. As a result, for some firms the expectation of the belief will increase if the firms keep observing large quantities purchased at given prices, but for others the expectation may decrease if their transaction quantities are at lower levels than expected. For all firms, however, standard deviations will keep falling if the firms keep exporting, which implies that uncertainty is decreasing over time as the firms receive more signals.

1.3.4 Dynamic Decision – Export with Learning about demand

In this section, I characterize a forward-looking firm's export participation with learning about the demand factor in a single market j. Again, I omit the superscript for market j to simplify the notation. The term "export" or "not to export" means "export" or "not to export" to a *specific* market j. *Within* market j, the export decision and the learning process are endogenously related. The firm's export decision depends on its current belief regarding the demand factor; moreover if the firm decides to export, then it will expect the belief in the next period to be updated according to the signals received from exporting. Hence, the export decision is dynamic not only because of the sunk entry cost that the firm has to pay if it did not export in the last period but also because of this endogenous learning process. However, the assessment of uncertainty and the learning process will be biased if the persistence introduced by the sunk entry cost is ignored. Thus, following the literature (e.g., Roberts and Tybout (1997); Das, Roberts, and Tybout (2007); Aw, Roberts, and Xu (2011)), I take the effect of the sunk entry cost into account by assuming: if the firm decides to export but it did not export in the last period, it must pay a sunk cost c^s_{it}; otherwise, it pays a fixed cost c^f_{it}. I assume that c^f_{it} and c^s_{it} are independently drawn from distributions $G_f(\cdot)$ and $G_s(\cdot)$, respectively.

At the beginning of period t, given the current belief $N(m_{it-1}, \sigma_{it-1})$ regarding the demand factor, the expected total export profit in period t (before considering the fixed/sunk cost) is the sum of profit from all transactions in period t:

$$\begin{aligned} E\Big[\Pi(s_{it}, e_{it-1})\Big] &= E\Big(\sum_{n=1}^{n_{it}} \pi(s_{it})\Big) = E\Big(\sum_{n=1}^{n_{it}} \phi_t \frac{-1}{1+\eta}\Big[\frac{\eta}{1+\eta}\Big]^{\eta} C_{it}^{1+\eta} e^{\zeta_{in}}\Big) \\ &= \lambda_{it}\phi_t \frac{-1}{1+\eta}\Big[\frac{\eta}{1+\eta}\Big]^{\eta} C_{it}^{1+\eta} \exp(m_{it-1} + \sigma^2_{it-1}/2 + \sigma^2_u/2), \\ &= \lambda_{it}\tilde{\phi}_{it} \exp(m_{it-1} + \sigma^2_{it-1}/2 + \sigma^2_u/2 - (\eta+1)\omega_{it}), \end{aligned} \tag{1.20}$$

where $\lambda_{it} = \exp(\psi_0 + \psi_1 \log(\tilde{n}_{it-1}) + \psi_2 e_{it-1})$, $\tilde{\phi}_{it}$ is a combination of cost and demand shifters: $\tilde{\phi}_{it} = \phi_t \frac{-1}{1+\eta}\left[\frac{\eta}{1+\eta}\right]^{\eta} e^{\gamma_0(1+\eta)} W_{it}^{\gamma_w(1+\eta)} K_{it}^{\gamma_k(1+\eta)}$, and e_{it-1} is the dummy variable indicating the export status in period $t-1$. The set of state variables is summarized in $s_{it} = (\tilde{\phi}_{it}, \tilde{n}_{it-1}, \omega_{it}, m_{it-1}, \sigma_{it-1})$.[27]

Thus, the expected total export profit in period t depends on the expected number of orders, productivity, the current belief about the demand factor, and an aggregate demand/cost shifter $\tilde{\phi}_{it}$. In particular, if the firm exported in the last period, then it will expect an increase in the number of orders that it will receive in this period. This implies an increase in both the total export volume and profit for firm i in period t.

Also, the current belief characterized by $(m_{it-1}, \sigma_{it-1})$ is a part of the state variables, since it affects the expected profit. More specifically, the expected profit is increasing in both the mean m_{it-1} and the standard deviation σ_{it-1} of the current belief, holding other variables fixed. This implies that, if the firm has a higher expectation of the demand factor then it expects more profit in this period; also, if the firm faces greater uncertainty about the demand factor then it is more likely to export because of the option value of learning. This feature is a result of the assumption that profit is an increasing and convex function of the demand shock ζ_{it}, which is commonly assumed in the literature and also employed in this model.

The timing of the entire model is summarized as follows:

1. At the beginning of period t, the firm observes (s_{it}, e_{it-1}), where e_{it-1} is a dummy variable indicating whether the firm exported or not in period $t-1$;

2. After observing its fixed cost draw c_{it}^{f} or sunk cost draw c_{it}^{s}, the firm decides whether to export or not (i.e., choose $e_{it} = 0$ or 1), based on its current state (s_{it}, e_{it-1}) which includes its current productivity and belief about the demand factor characterized by $(m_{it-1}, \sigma_{it-1})$;

3. If the firm decides to export, it pays the fixed cost c_{it}^{f} if it exported in the last period or pays the sunk cost c_{it}^{s} otherwise. During period t, the firm receives and fulfills n_{it} orders from customers in the foreign market. The quantity and the price of each order is determined by equations 1.8 and 1.10 respectively. The pricing decision is static. The firm is able to observe a signal ζ_{in} about ξ_i in each order after exporting. At the end of this period, a series of signals $\{\zeta_{i1}, \zeta_{i2}, \ldots, \zeta_{in_{it}}\}$ is observed, which is then used to update its belief about ξ_i. The posterior belief is (m_{it}, σ_{it}), according to equations 1.14 and 1.15;

[27] Note that e_{it-1} is also a part of the state variables, but I pull it out from s_{it} to simplify the later notations.

4. If the firm decides not to export, then there is no export profit from this market in this period; there is also no update to the belief at the end of this period (i.e., $m_{it} = m_{it-1}$ and $\sigma_{it} = \sigma_{it-1}$);

5. Period $t+1$ begins and all other state variables are updated. In particular, its productivity evolves to ω_{it+1}.

Given the timing, I model the firm's dynamic export participation using a Bellman equation. I denote the expected value function at the beginning of each period t before observing the fixed cost draw or the sunk cost draw as $V_t(s_{it}, e_{it-1})$. The firm will choose to export to this market if the expected total (current plus future) payoff is greater than the cost (i.e., fixed cost or sunk cost) it must pay. Thus, the Bellman equation is given by

$$\begin{aligned} & V_t(s_{it}, e_{it-1}) \\ & = E_{c^f, c^s} \max \begin{cases} \delta E[V_{t+1}(s_{it+1}, 0) | s_{it}, e_{it-1}], & \text{if} \quad e_{it} = 0 \\ \\ E[\Pi(s_{it}, e_{it-1})] - e_{it-1} c^f_{it} - (1 - e_{it-1}) c^s_{it} + \\ \delta E[V_{t+1}(s_{it+1}, 1) | s_{it}, e_{it-1}], & \text{if} \quad e_{it} = 1 \end{cases} \end{aligned} \tag{1.21}$$

where δ is the discount rate, and

$$\begin{aligned} & E[V_{t+1}(s_{it+1}, e_{it}) | s_{it}, e_{it-1}] \\ & = \int V_{t+1}(s_{it+1}, e_{it}) dF(s_{it+1} | s_{it}, e_{it-1}, e_{it}) \\ & = \int V_{t+1}(s_{it+1}, e_{it}) dF_\omega(\omega_{it+1} | \omega_{it}) dF_b(m_{it}, \sigma_{it}, \tilde{n}_t | s_{it}, e_{it-1}, , e_{it}), \end{aligned} \tag{1.22}$$

and $F_\omega(\omega_{it+1}|\omega_{it})$ and $F_b(m_{it}, \sigma_{it}, \tilde{n}_t | s_{it}, e_{it-1}, e_{it})$ are the transition probabilities of the four key state variables (productivity, the mean and the standard deviation of the belief, and the number of orders in the most recent period). It is important to note that the transition of productivity is independent from the export decision. However, the transition of the belief about the demand factor is affected by the export decision as well as the number of orders actually received in period t. Also, there is no direct correlation between the two transition probabilities, $F_\omega(\cdot|\cdot)$ and

$F_b(\cdot|\cdot)$. I now turn to specify the transition probabilities of all four key state variables.[28]

First, $F_\omega(\omega_{it+1}|\omega_{it})$ is the distribution of productivity in period $t+1$, given the productivity in period t. Specifically, given productivity evolution in equation 1.2, ω_{it+1} is drawn from $N(g(\omega_{it}), \sigma_\epsilon)$. That is,

$$F_\omega(\omega_{it+1}|\omega_{it}) = N(g(\omega_{it}), \sigma_\epsilon). \tag{1.23}$$

Second, $F_b(m_{it}, \sigma_{it}, \tilde{n}_t|s_{it}, e_{it-1}, e_{it})$ is the joint distribution of the belief and $\tilde{n}_t$ at the beginning of period $t+1$, given the current state s_{it} and the export decision (e_{it-1}, e_{it}). Note that the distribution of the belief and $\tilde{n}_t$ are not independent, since the updated belief is related to the number of received signals which are contained in orders. Specifically, this joint probability is determined according to equations 1.14 and 1.15, and I explain it in detail as follows.

Given the belief at the beginning of period t as

$$\xi_i \sim N(m_{it-1}, \sigma_{it-1}), \tag{1.24}$$

then the posterior belief at the end of period t (or the beginning of period $t+1$) remains the same if the firm has decided not to export in period t. That is, the transition probability is degenerate:

$$F_b(m_{it} = m_{it-1}, \sigma_{it} = \sigma_{it-1}, \tilde{n}_{it} = \tilde{n}_{it-1}|s_{it}, e_{it-1}, e_{it} = 0) = 1.$$

On the other hand, if the firm has decided to export, and suppose for the entire period t it receives a total of n_{it} signals $\{\zeta_{i1}, \zeta_{i2}, \dots, \zeta_{in_{it}}\}$, then the updated belief is

$$\xi_i \sim N(m_{it}, \sigma_{it}), \tag{1.25}$$

where

$$m_{it} = m_{it-1} + n_{it}\frac{\sigma_{it}^2}{\sigma_u^2}\Big(\frac{1}{n_{it}}\sum_{n=1}^{n_{it}}(\xi_i + u_{in}) - m_{it-1}\Big) \tag{1.26}$$

[28]Since the other state variable $\tilde{\phi}_t$ is an aggregate index of capital stock and the year dummy, I following Aw, Roberts, and Xu (2011) to assume the firm forms a rational perception of the sequence of $\tilde{\phi}_t$.

and

$$\sigma_{it}^2 = \frac{\sigma_{it-1}^2 \sigma_u^2}{n_{it}\sigma_{it-1}^2 + \sigma_u^2}. \tag{1.27}$$

Since m_{it-1} and σ_{it-1} are known at period t, the transition depends on random variables (n_{it}, ξ_i, u_{in}). Note that the distributions of n_{it} and u_{in} are known, and ξ_i is believed to be distributed as the current belief of firm i (i.e. $N(m_{it-1}, \sigma_{it-1})$). Thus, *conditional on* n_{it}, the distribution of m_{it} is

$$m_{it} \sim N\Big(m_{it-1}, n_{it}\frac{\sigma_{it}^2}{\sigma_u^2}\sqrt{\sigma_u^2/n_{it} + \sigma_{it-1}^2}\Big) \equiv F_m(m_{it}|n_{it}, m_{it-1}, \sigma_{it-1}, e_{it} = 1), \tag{1.28}$$

and the distribution of σ_{it} is degenerate:

$$F_\sigma(\sigma_{it} = \frac{\sigma_{it-1}^2 \sigma_u^2}{n_{it}\sigma_{it-1}^2 + \sigma_u^2}|n_{it}, m_{it-1}, \sigma_{it-1}, e_{it} = 1) = 1. \tag{1.29}$$

Since n_{it} is drawn from the truncated Poisson distribution with known parameter $\lambda_{it} = \exp(\psi_0 + \psi_1 \log(\tilde{n}_{it-1}) + \psi_2 e_{it-1})$,[29] the probability to receive n_{it} orders in period t is

$$F_n(n_{it}|s_{it}, e_{it-1}) = \frac{\lambda_{it}^{n_{it}} e^{-\lambda_{it}}}{n_{it}!(1 - e^{-\lambda_{it}})}. \tag{1.30}$$

Therefore, the joint transition probability of $(m_{it}, \sigma_{it}, \tilde{n}_{it})$ is given by

$$\begin{aligned} &F_b(m_{it}, \sigma_{it}, \tilde{n}_{it}|s_{it}, e_{it-1}, e_{it}) \\ &= F_m(m_{it}|n_{it}, s_{it}, e_{it} = 1)F_\sigma(\sigma_{it}|n_{it}, s_{it}, e_{it} = 1)F_n(n_{it}|s_{it}, e_{it-1}) \end{aligned} \tag{1.31}$$

Thus, if the firm decides not to export in period t, then there is no transition in the belief. But if the firm decides to export, then the joint transition probability of the belief consists of the distributions of m_{it} and σ_{it}, conditional on the realization of n_{it} and the current state of the belief $(m_{it-1}, \sigma_{it-1})$.

In this way, I incorporate heterogeneity in both productivity and the demand belief into a model of exporting. The firm's export decision depends on both productivity evolution and the evolution of the demand belief. In particular, productivity evolves exogenously and influences the export decision by affecting the marginal cost of production. The demand belief evolves endogenously, depend-

[29] Note that λ_{it} is known given the state (s_{it}, e_{it-1}).

ing on the export decision and export outcomes. The next section demonstrates the strategy of identifying the effects of the two processes in determining export dynamics and estimating the structural model.

1.4 Identification and Estimation Strategy

There are two major sources driving the export dynamics in this model. The first one is the evolution of productivity, and the second one is the evolution of the belief regarding the demand factor. Both sources are heterogeneous across firms and over time, and neither of them is observable to researchers. I rely on two sets of data to identify the role of each source. The first data set provides firm-level production information, which includes employment, labor and material expenditures, capital stock, and domestic revenue for each firm in each period. The second data set contains firm shipment-level exports, including the export destination, quantity and price of each shipment, and shipment month. The strategy for identification is to utilize the fact that productivity affects both domestic revenue and export participation while the demand belief only influences export participation.

To be specific, domestic revenue depends on current productivity; it does not depend on the evolution of the belief in any of the foreign markets. In particular, high productivity implies a low marginal cost, which allows the firm to set a low price and increases domestic revenue. Thus, the relationship between domestic revenue and productivity enables me to recover the productivity for each firm in each period. In turn, with the recovered firm-level productivity, I am able to estimate productivity evolution before considering the dynamic export decision.[30]

The observed export participation in each foreign market, together with the shipment-level prices and quantities, allows me to recover the demand belief of each firm in each period (up to a set of parameters to be estimated) and to identify the learning process in each foreign market. Specifically, the price and quantity sold in each shipment imply a demand signal received by the firm, which is then

[30] It is possible to allow productivity gain from exporting to a *single* foreign market, similar to Aw, Roberts, and Xu (2011). This will not break the identification of productivity from domestic sales, but will make the analysis of export participation complicated. Since no significant productivity gain found in the industry I am going to study, I assume there is no such "learning by doing" in exporting for simplification.

used to recover the expectation of the belief. The number of shipments in each period allows me to recover the magnitude of uncertainty of the belief. Export participation depends on both the expectation and uncertainty of the belief, but in different ways. The probability of exporting is increasing in the expectation, which fluctuates over time depending on the value of the demand signals received. However, since the option value of learning is decreasing when uncertainty is resolved over time, the export probability is decreasing in the number of received signals in a deterministic style, holding other factors fixed. In particular, a model with only productivity heterogeneity predicts more productive firms export. While in my model with the two-dimensional heterogeneity, firms face more uncertainty about demand may also export because of the large option value of learning even if their productivity is not high enough. Thus, with both productivity and demand beliefs being recovered, the cross-sectional and time series patterns of export decisions identify the role of each driving force.

The estimation approach is inspired by the identification strategy. I divide the full set of parameters into a set of static parameters and a set of dynamic parameters. I first estimate the static parameters, time-varying productivity, and demand signals received by each firm. Then I estimate the dynamic parameters.

1.4.1 Estimation of Static Parameters, Productivity, and Demand Signals

As the first step, I estimate the set of static parameters: demand elasticity in the domestic market (η^D) and each foreign market (η^j), and the marginal cost parameters. In addition, for each firm, I recover its time-varying productivity and demand signals received in each market and each period. I use firm-level production information to estimate the marginal cost function and to recover productivity. Then I aggregate shipment-level exports into firm-market-level exports, and estimate the demand elasticities from the relationship between total variable costs and firm-market-level exports. Finally, I recover the market-level demand signals for each firm using shipment-level exports. These demand signals will be used to update firms demand beliefs in the dynamic estimation stage. The implementation of this strategy is specified as follows.

First, I estimate the marginal cost parameters and productivity using the firm-level production and domestic sale information. Specifically, the domestic revenue function (1.6) implies

$$\ln R_{it}^D = (\eta^D+1)\ln(\frac{\eta^D}{\eta^D+1}) + \ln \Phi_t^D + (\eta^D+1)(\gamma_0^D + \gamma_w \ln W_{it} + \gamma_k \ln K_{it} - \omega_{it}) + v_{it}, \tag{1.32}$$

where v_{it} is the measurement error. Note that the firm's productivity can be correlated with its capital stock. Thus, to control for the unobservable productivity ω_{it}, I follow Olley and Pakes (1996) and Levinsohn and Petrin (2003) to rewrite the unobserved productivity in terms of related observable variables. In particular, firms' choice of variable material and labor inputs, M_{it} and L_{it}, depends on the level of productivity and the demand beliefs about export markets. Since I assume that marginal cost is constant in output, the relative input will not be a function of total output and thus not depend on demand beliefs about export markets. Also, if technology differences are not Hick's neutral, then the differences in the mix of the two inputs across firms and over time will reflect differences in productivity level.[31] Thus, I can write the unobserved productivity as a function of the relative input, conditional on the capital stock level: $\omega_{it} = \omega(K_{it}, M_{it}/L_{it})$. Then I combine the demand elasticity terms into an intercept $\tilde{\gamma}_0^D$ and use a set of time dummies, $\tilde{\Phi}_t$, to capture the domestic industrial aggregate Φ_t^D. Thus, the above equation can be written as:

$$\ln R_{it}^D = \tilde{\gamma}_0^D + \sum_{t=1}^{T} \gamma_t \tilde{\Phi}_t + \gamma_w \ln W_{it} + f(K_{it}, M_{it}/L_{it}) + v_{it}, \tag{1.33}$$

where $f(K_{it}, M_{it}/L_{it}) = (1+\eta^D)(\gamma_K \ln K_{it} - \omega(K_{it}, M_{it}/L_{it}))$ is a function of capital stock, material input and labor input. I parameterize function f as a cubic polynomial of its arguments. Now the error term v_{it} is uncorrelated with the right-hand-side variables. Thus, I use ordinary least square regression to obtain the estimates. An important output from the regression is the fitted value of function f, which is denoted as $\hat{f}_{it}$. This is an estimate of $(1+\eta^D)(\gamma_K \ln K_{it} - \omega_{it})$. That

[31] Non-Hicks neutral productivity has been found in a large empirical literature. See Stevenson (1980) for a model using plant-level data. Aw, Roberts, and Xu (2011) also utilize the same idea to recover productivity.

is,

$$\hat{f}_{it} = (1 + \eta^D)(\gamma_K \ln K_{it} - \omega_{it}). \tag{1.34}$$

Then, I follow Olley and Pakes (1996) to construct a series of productivity measures for each firm, by utilizing the productivity evolution process. In particular, the first-order Markov process of productivity evolution is specified as

$$\omega_{it} = g_0 + g_1\omega_{it-1} + \epsilon_{it}. \tag{1.35}$$

Substitute $\omega_{it} = -\frac{1}{\eta^D+1}\hat{f}_{it} + \gamma_K \ln K_{it}$ into the above evolution process and get

$$\hat{f}_{it} = -(\eta^D+1)g_0 + g_1\hat{f}_{it-1} + (\eta^D+1)\gamma_K \ln K_{it} - g_1(\eta^D+1)\gamma_K \ln K_{it-1} - (\eta^D+1)\epsilon_{it}. \tag{1.36}$$

Again, the error term is uncorrelated with all right-hand-side variables. Thus, this equation can be estimated by nonlinear least squares, since the function is nonlinear in g_1 and $(\eta^D + 1)\gamma_K$. The key parameters estimated in this equation are $g_0^* = (\eta^D + 1)g_0$, g_1, and $\gamma_K^* = (\eta^D + 1)\gamma_K$. Note that η^D, g_0, and γ_K are not separately identified. However, if η^D is known, then $g_0 = \frac{g_0^*}{\eta^D+1}$ and $\gamma_K = \frac{\gamma_K^*}{\eta^D+1}$ are immediately recovered. More importantly, from equation 1.34, I can recover productivity as $\omega_{it} = -\frac{1}{\eta^D+1}\hat{f}_{it} + \gamma_K \ln K_{it}$ with knowledge of η^D.

To estimate η^D, I follow Aw, Roberts, and Xu (2011) and utilize the relationship between the total variable cost (TVC) and domestic revenue (R^D) as well as the total export revenue in each market (X^j). Since the marginal cost of production is the same for domestic sales and exports to all markets, the first order conditions for profit maximization of the domestic and foreign markets imply that the total variable cost is an elasticity-weighted combination of total revenue in each market. Specifically, for each firm i and each period t:

$$\begin{aligned} TVC_{it} &= Q_{it}^D C_{it} + \sum_{j=1}^{J} Q_{it}^j C_{it} \\ &= (1 + \frac{1}{\eta^D})Q_{it}^D P_{it}^D + \sum_{j=1}^{J}(1 + \frac{1}{\eta^j})Q_{it}^j P_{it}^j \qquad (1.37) \\ &= (1 + \frac{1}{\eta^D})R_{it}^D + \sum_{j=1}^{J}(1 + \frac{1}{\eta^j})X_{it}^j, \end{aligned}$$

where Q_{it}^j is the total quantity exported to market j by firm i in period t, X_{it}^j is the corresponding total revenue, and R_{it}^D is the total revenue in the domestic market. Note that although the firm may export to market j with multiple transactions, the demand function (i.e., equation 1.8) implies that the optimal price is proportional to the marginal cost, as shown in equation 1.10. Thus the second equality in the above equation holds. Therefore, the following empirical equation can be used to estimate η^D as well as all η^js:

$$TVC_{it} = (1 + \frac{1}{\eta^D})R_{it}^D + \sum_{j=1}^{J}(1 + \frac{1}{\eta^j})X_{it}^j + \upsilon_{it}, \tag{1.38}$$

where υ_{it} is the measurement error.

Hence, up to now I have obtained the key estimates $\hat{g}_0$, $\hat{g}_1$, $\hat{\gamma}_K$, $\hat{\eta}^D$, and all $\hat{\eta}^j$s. Also, it is straightforward to recover the productivity for each firm in each period as $\omega_{it} = -\frac{1}{\hat{\eta}^D+1}\hat{f}_{it} + \hat{\gamma}_K \ln K_{it}$.

In addition, with the estimates of all $\hat{\eta}^j$s, I can recover the aggregate demand shifter ϕ_t^j in the foreign demand function (i.e., equation 1.8) using shipment-level exports via ordinary least squares. Specifically, the logarithm of the demand function in foreign market j is

$$\ln Q_{in}^j = \ln \phi_t^j + \hat{\eta}^j \ln P_{in}^j + \zeta_{in}^j, \tag{1.39}$$

where $\zeta_{in}^j = \xi_i^j + u_{in}^j$ is the error term with demand factor ξ_i^j. Since the demand factor is uncorrelated with the aggregate demand shifter, the regression produces unbiased estimates of ϕ_t^j.[32] The regression also provides an estimate of the standard deviation of the noise by using the demeaned version of equation 1.39 to eliminate the unobservable firm-fixed effect ξ_i^j:[33]

$$\hat{\sigma}_u^j = \sqrt{\frac{1}{\sum_i N_i^j}\sum_{i,n}(\Delta \ln Q_{in}^j - \Delta \ln \hat{\phi}_t^j - \hat{\eta}^j \Delta \ln P_{in}^j)^2}, \tag{1.40}$$

[32]To avoid potential endogeneity problem, I use a balanced data in this estimation. Also, the time dummies and the firm demand factor is not separately identified. I normalize the first year dummy as 10, but this does not affect the dynamic estimation stage.

[33]The demeaned version of equation 1.39 implies $u_{it}^j = \Delta \ln Q_{in}^j - \Delta \ln \hat{\phi}_t^j - \hat{\eta}^j \Delta \ln P_{in}^j + \overline{u}_i^j$, where $\overline{u}_i^j = \frac{1}{N_i^j}\sum_n u_{in}^j$.

where N_i^j is the total number of transactions from firm i to market j, $\Delta \ln Q_{in}^j = \ln Q_{in}^j - \frac{1}{N_i^j}\sum_n \ln Q_{in}^j$, $\Delta \ln P_{in}^j = \ln P_{in}^j - \frac{1}{N_i^j}\sum_n \ln P_{in}^j$, and $\Delta \ln \hat{\phi}_t^j = \ln \hat{\phi}_t^j - \frac{1}{N_i^j}\sum_n \ln \hat{\phi}_t^j$.

After obtaining these estimates, I can recover the demand signals received by firms as

$$\zeta_{in}^j = \ln Q_{in}^j - \ln \hat{\phi}_t^j - \hat{\eta}^j \ln P_{in}^j. \tag{1.41}$$

Then I use demand signals to update firms' market-specific demand beliefs according to Bayes' rule as specified in equations 1.17 and 1.18.[34]

To sum up, after the static estimation, three objectives have been achieved. First, I have estimated the parameters in the marginal cost function. Second, I have obtained a productivity measure for each firm in each period, and the productivity evolution process has been recovered. Third, I have recovered the market-specific demand signals received by each firm in each period. These signals enable me to write firms' demand beliefs as a function of initial beliefs up to a set of parameters to be estimated in the dynamic stage.

1.4.2 Estimation of Dynamic Parameters

The set of dynamic parameters includes the parameterized initial belief function $h_m(\cdot|\beta_m)$, $h_\sigma(\cdot|\beta_\sigma)$, distributions of fixed cost $G_f(\cdot|\beta_f)$ and sunk cost $G_s(\cdot|\beta_s)$, and the parameters for Poisson parameter evolution ψ_0, ψ_1, and ψ_2 in equation 1.7, where $\beta_m, \beta_\sigma, \beta_f, \beta_s$ are vectors of parameters associated with the corresponding functions.

1.4.2.1 Estimation Details

I estimate the dynamic parameters via the Maximum Likelihood Method. The likelihood is constructed from data on the discrete choice of export participation together with the number of transactions for each firm in each period and each market. To simplify notation, I only consider the exports to a single market j and omit the market superscript, since the export decisions and learning processes in different markets are independent for each firm. However, it is straightforward to

[34]Notice that the year dummies and demand factors are not separately identified. Without loss of generality, I normalize the first year dummy as zero.

extend the estimation to all foreign markets. In particular, for each firm i in each period t, I observe the export participation dummy e_{it}. If the firm exported in period t, then $e_{it} = 1$, and I can observe the total number of transactions n_{it}, as well as the price (P_{in}) and quantity (Q_{in}) in each transaction n.[35] If the firm did not export to that market in period t, then $e_{it} = 0$ and no transaction happened.

The full likelihood consists of two partial likelihoods. The first one is about the number of transactions n_{it} of each firm i in each period t. The parameters involved include the parameters for the evolution process of λ_{it}, which is summarized in the vector $\theta_1 = (\psi_0, \psi_1, \psi_2)$. The first partial likelihood is given by the truncated Poisson probability since the number of transactions is assumed to be greater than zero if the firm exports:

$$\ell^1(n_{it}; \theta_1) = \frac{\lambda_{it}^{n_{it}} e^{-\lambda_{it}}}{n_{it}!(1 - e^{-\lambda_{it}})}, \tag{1.42}$$

where λ_{it} is specified as equation 1.7.

Given θ_1, the second partial likelihood is about the discrete choice of export participation e_{it}, conditional on (s_{it}, e_{it-1}). The parameters involved are parameters for initial beliefs and distributions of fixed and sunk entry costs, which are summarized in the vector $\theta_2 = (\beta_m, \beta_\sigma, \beta_f, \beta_s)$. The second partial likelihood is given by:

$$\begin{aligned} &\ell^2(e_{it}; \theta_1, \theta_2) \\ &= e_{it} \Pr(e_{it} = 1 | s_{it}, e_{it-1}; \theta_1, \theta_2) + (1 - e_{it}) \Pr(e_{it} = 0 | s_{it}, e_{it-1}; \theta_1, \theta_2) \end{aligned} \tag{1.43}$$

where $\Pr(e_{it} = 1 | s_{it}, e_{it-1}; \theta_1, \theta_2)$ is the conditional probability of exporting. It depends on the parameters θ_1 and θ_2. Given the parameterized distributions of fixed cost and sunk cost, the conditional probability of exporting is

$$\begin{aligned} &\Pr(e_{it} = 1 | s_{it}, e_{it-1}, \theta_1, \theta_2) \\ &= \Pr\Big(E[\Pi(s_{it}, e_{it-1})] + \delta E[V(s_{it+1}, 1) | s_{it}, e_{it-1}] - \delta E[V(s_{it+1}, 0) | s_{it}, e_{it-1}] \\ &> e_{it-1} c^f + (1 - e_{it-1}) c^s]\Big), \end{aligned} \tag{1.44}$$

[35]Note the superscript of market j is suppressed, and I use n as an index of transactions.

where δ is the discount factor, $E[\Pi(s_{it}, e_{it-1})]$ is the period export profit implied by equation 1.20, and $E[V(s_{it+1}, e_{it})|s_{it}, e_{it-1}]$ is the expected value of exporting at the beginning of period $t+1$, as described in the Bellman equation. Note that the belief affects this probability through both the expected period export revenue $E[\Pi(s_{it}, e_{it-1})]$ and the expected future value of exporting $E[V(s_{it+1}, e_{it})|s_{it}, e_{it-1}]$.

Thus, the full likelihood is the product of the two partial likelihoods:

$$\ell^f(e_{it}, n_{it}; \theta_1, \theta_2) = \ell^1(n_{it}|\theta_1)\ell^2(e_{it}|\theta_1, \theta_2). \tag{1.45}$$

I denote the full set of dynamic parameters as $\theta = (\theta_1, \theta_2)$. Then, θ can be estimated via maximizing the likelihood:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i,t} \log(\ell^f(e_{it}, n_{it}; \theta_1, \theta_2)). \tag{1.46}$$

It is computationally difficult to obtain an estimate of θ in this way. The evaluation of the full likelihood requires solving for the value function. Given the high dimension of the state variables, this is a major computational burden in the estimation. Also, θ_1 increases the dimension of parameters to be estimated. This in turn makes the estimation even more time-consuming. To solve this issue, I reduce the computational burden in two directions. First, I adopt the strategy proposed by Rust (1987) to estimate the parameters by three stages, which is described in the remainder of this subsection. Second, I follow Nagypál (2007) to compute the value function, which is specified in the next subsection in detail.

Like Rust (1987), I estimate the parameters in three stages. The first stage estimates θ_1 via the first partial likelihood. In particular, I estimate θ_1 using the data on the number of transactions and export participation of each firm in each period. To be specific, θ_1 is obtained as

$$\hat{\theta}_1 = \arg\max_{\theta_1} \sum_{i,t} \log(\ell^1(n_{it}; \theta_1)). \tag{1.47}$$

Note that in this estimation, there is no need to calculate the value function. In this way, the first stage provides an estimate of $\theta_1 = (\psi_0, \psi_1, \psi_2)$.

Then, with the estimated $\hat{\theta}_1$, the second stage is to estimate θ_2 via the second partial likelihood:

$$\hat{\theta}_2 = \arg\max_{\theta_2} \sum_{i,t} \log(\ell^2(e_{it}; \hat{\theta}_1, \theta_2)). \tag{1.48}$$

This stage requires the internal computation of the value function at each evaluation of ℓ^2, according to the Nested Fixed Point Algorithm. The detailed algorithm of value function computation will be described in the next subsection. However, the obvious advantage of this estimation is that there are fewer parameters to be estimated, which reduces the estimation burden.

The third stage is to use the estimated $(\hat{\theta}_1, \hat{\theta}_2)$ as an initial starting value to produce an efficient estimate of θ via the full likelihood:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i,t} \log(\ell^f(e_{it}, n_{it}; \theta_1, \theta_2)). \tag{1.49}$$

This stage also involves the internal calculation of the value function for each evaluation of ℓ^f. Drawing on the work of Rust (1987), this estimation yields a consistent estimator of asymptotic covariance matrix for θ. Nonetheless, the estimate of θ from this stage is usually identical to the estimates from the first two stages.

Since the above estimations require the internal calculation of the value function in each evaluation of the likelihood, it is important to compute it efficiently. I now turn to the algorithm of value function computation in greater detail.

1.4.2.2 Value Function Computation

Now I describe the algorithm used to solve for the value function, given a set of parameters θ.

First, I construct the demand belief for each firm in each period using the demand signals recovered in equation 1.41. I consider the belief as a two-dimensional state variable in the value function. Specifically, the full set of state variables includes: an aggregate state $\tilde{\phi}_{it}$, the number of transactions in the most recent period $\tilde{n}_{it-1}$, productivity ω_{it}, belief state m_{it-1} and σ_{it-1}. As a result, the curse of dimensionality makes it difficult to follow the traditional value function iteration method to derive the underlying value function. To solve this problem, I follow the method implemented in Nagypál (2007) to compute the value function. This method utilizes the fact that the magnitude of uncertainty (measured by the standard deviation σ_{it} of the belief) is strictly decreasing in the number of transactions (signals). Thus, at the limit state, the firm will have no uncertainty about the demand factor and $\sigma_{it} = 0$. The value function at this limit state has one fewer state variable. More importantly, the transition of the belief state is degenerate: $\Pr(m_{it} = m_{it-1} = \xi_i) = 1$ and $\Pr(\sigma_{it} = 0) = 1$. These features significantly simplify the computation of the value function at the limit state. Once this is computed, I approximate the firm's value function after observing a large number of signals (so that σ_{it} is small enough, e.g., $\sigma^* = 0.01$) as the value function at the limit state. Using this approximated value function with uncertainty σ^*, I calculate the value function with arbitrarily greater uncertainty (any larger σ_{it}) via backward induction. The details of this procedure are explained in A.

1.5 Data

I will use the model developed above to study how the two unobservable driving processes separately explain the firm's export participation. To do this, I utilize two major pieces of information to identify productivity evolution and Bayesian learning about foreign demand from each other. The first one is the firm-level input and output production information, and the second one is the shipment-level exports. Thus, I draw data from two sources.

The first source is the Chinese Monthly Customs Transactions data set. This data set includes all export shipments of Chinese firms from 2000 to 2006. It is common that firms have multiple shipments to a single destination market within a year. Across all industries, the average number of shipments in a year is 8.4 and the standard deviation is 31. Each shipment contains shipment value, quantity, 8-digit HS code, type of trade, destination market, shipment month, and firm identification number. Such detailed information makes it possible to analyze how firms' export participation is endogenously correlated to Bayesian learning about demand. In particular, this distinguishes the export decisions and the export transactions in different destination markets and allows me to recover the demand signals received by firms from the shipment-level quantities and unit prices. Comparing commonly used annual export data sets, this enables me to track the change of export patterns in each specific market over time, and in turn to attribute such dynamic patterns separately to heterogeneity in productivity and Bayesian learning.

One important feature of exports from Chinese manufacturing firms is that a significant portion of transactions is processing trade, in which domestic firms' intermediate material and even related technology are directly supplied by foreign firms. The domestic firms are more like long-term contractors rather than active exporters, and their export decisions are less likely to be affected by uncertainty in foreign markets once the firms obtain contracts. Thus, in this paper I focus on transactions of ordinary trade, in which firms make their own decisions on production, pricing, and exporting, without being constricted by the existing contracts with foreign upstream suppliers. In particular, the model will be estimated using the data for the ceramics industry. This industry produces sanitation ceramics, special ceramics, and daily-used ceramics. This industry fits the study purpose well because these firms export a very concentrated product line. The major products for exporting are colorful dinnerware and ornamental articles of ceramics such as statuettes. This means that they can be viewed as single-product firms, each of which produces a differentiated product. More importantly, most export transactions in this industry are ordinary trade. In particular, in 2006, the ordinary trade in this industry generated revenue of more than 390 million US dollars, which accounts for 90% of all trade types. This suggests that uncertainty and learning about foreign demand are potentially important when firms make export decisions.

The second data source is the Chinese Annual Survey of Manufacturing from the Chinese National Bureau of Statistics. This data set provides detailed annual firm-level production information of all medium and large manufacturing firms that had total annual sales of more than $600,000 from 2000 to 2006. The primary variables include firm-level domestic revenue, labor wage, employment, material input, and capital stock. This information is used to construct a firm-time specific productivity measure, as well as a marginal production cost function with observable cost shifters such as capital stock and wage rate.

The sets of firm identification numbers are not the same in the two data sets since they are collected by different agencies. However, I am able to match the two data sets according to the recorded firm name, phone number, zip code, and some other identifying variables. About 114,000 out of 278,000 firms in the custom data set are matched (around 41%).[36] In the estimation of static parameters, I use the input and output information for all firms in the ceramics industry, but the estimation of dynamic parameters is based on the data for firms that can be identified from both data sets.

1.6 Empirical Results

I estimate the structural model developed above using shipment-level exports from the Chinese ceramics industry to Germany, supplemented with firm-level production data as described the last section. However, this estimation can be easily extended to other industries and foreign markets. I have chosen this particular industry because it fits the study purpose well for the aforementioned reasons. In the estimation, I use the 2000-2006 data on the 394 firms that can be identified from both the shipment-level export data set and the annual production data set.

1.6.1 Estimates of Static Parameters and Productivity

The static parameters include the parameters in the marginal cost function (1.1), productivity evolution (1.2), and the demand function in foreign markets (1.8).

Table 1.2 shows the estimates of the marginal cost and productivity parameters. Note that coefficient γ_k measures the elasticity of capital in the marginal cost. The negative estimate (-0.063) means that when capital stock is larger, the marginal cost is lower. The estimate of γ_w is positive (0.056), which implies that firms with higher wage rates incur larger marginal cost in production. More importantly, the estimated parameters of the AR(1) productivity evolution are both positive and significant. The high estimate of g_1, 0.832, implies that productivity evolves persistently. Thus, firms with high productivity in this year will expect to have high productivity in the next year. However, the standard deviation of the innovation term is $\sigma_\epsilon = 0.095$. This implies that there are still significant unexpected productivity shocks that shift firm productivity. Thus, it is necessary to

[36]Note that the custom data records all transactions for all firms, while the annual survey data only records medium and large manufacturing firms, thus the percentage of matched firms conditional on medium and large manufacturing firms should be larger than 41%.

take the evolution of productivity into account in the investigation of the learning process.

Table 1.3 shows the estimates of the domestic demand elasticity as well as demand elasticities for the ten foreign markets that account for 61% of total exports. For simplicity, I assume the other markets share the same demand elasticity, which is represented by η^{oth}. This essentially simplifies the estimation equation to

$$TVC_{it} = (1 + \frac{1}{\eta^D})R_{it}^D + \sum_{j=1}^{10}(1 + \frac{1}{\eta^j})X_{it}^j + (1 + \frac{1}{\eta^{oth}})X_{it}^{oth} + \upsilon_{it}, \tag{1.50}$$

where υ_{it} is the measurement error. As shown in Table 1.3, this simplification does not pose any problem since the elasticities are very similar across the domestic market and all ten foreign markets.[37] These estimates indicate that the demand elasticities range from 3.96 to 4.18, implying markups of price over the marginal cost of 31.4 to 33.7 percent. Given that these markets spread all over the world, the result suggests that there is not much difference across different markets in terms of demand elasticity. An important implication is that firms are unlikely to face uncertainty about the demand elasticity in a foreign market. The reason is that the firm's operation in the domestic market or any foreign market is enough to allow them to know the demand elasticity across all other foreign markets. Thus, the assumption made throughout this paper – firms know the slope but have uncertainty about the intercept of the demand curve – is reasonable.

The estimated standard deviation of signal noise u_{it} for Germany is $\hat{\sigma}_u = 1.59$. It measures the the informativeness of each individual signal. If $\hat{\sigma}_u$ is high, then each transaction contains less effective information about the true demand factor ξ_i; accordingly, it takes the firm more transactions to reach a certain accuracy of the belief. In particular, the estimated $\hat{\sigma}_u = 1.59$ implies that the learning speed is slow and it takes 49 transactions to reduce the variance (σ_{it}^2) from 1 to 0.05.

[37]The result from Aw, Roberts, and Xu (2011) also shows nearly identical elasticity of the domestic market and the aggregated foreign market using Taiwanese electronics industry.

1.6.2 Estimates of Dynamic Parameters

The dynamic parameters include the mean and standard deviation of the initial belief, the parameterized distributions of fixed and sunk costs, and parameters for the order process. The model is estimated with shipment-level export data from the Chinese ceramics industry to Germany from 2000 to 2006.

Since the data set only includes exports from year 2000, I do not observe firms' export status before 2000. Thus, I use year 2000 as an initial condition and allow firms to hold different beliefs when they first show up in the data, since some firms may have operated in the foreign market for a long time which I do not observe. In particular, I divide firms into two groups: potential entrants in the foreign market and experienced exporters, according to whether or not they exported to Germany in 2000.[38] The potential entrants are assumed to have an initial belief characterized by $N(m_{00}, \sigma_{00})$, while the experienced exporters are assumed to have an initial belief characterized by $N(m_{10}, \sigma_{10})$. These are initial beliefs that are held by firms when they first appear in the data set. The fixed and the sunk entry costs are assume to be drawn from Exponential distributions, with mean $\overline{c}^f$ for the fixed cost and mean $\overline{c}^s$ for the sunk cost, respectively. These distribution parameters are assumed to be the same for both potential entrants and experienced exporters.

The estimate of (ψ_0, ψ_1, ψ_2) from the Maximum Likelihood estimation (1.47) is given in Table 1.4. All estimates are positive and significant. $\hat{\psi}_1 = 0.626$ implies that the persistence of the number of transactions is high. Firms with more transactions in the past period will also have more transactions in this period, if they choose to export. Moreover, the positive and significant estimate of ψ_2 suggests that firms that continue to export will expect around 65% more transactions in the current period than firms that did not export in the last period. This suggests that there is a significant "building-up" effect of customer stock that leads to export growth at firm level.

The estimates of initial beliefs and distribution parameters for the fixed and sunk costs are obtained from the Maximum Likelihood estimation (1.48) and are

[38]Here I assume firms within each group have the same initial belief. However, it is possible to allow initial belief to be heterogeneous within each group. For example, firm age, capital stock and ownership are observable in the data set and can be used to control for the heterogeneity in the initial belief by writing the initial belief as a function of these observable characteristics.

reported in Table 1.5. In particular, the estimates of initial beliefs for the two groups deliver two interesting implications. First, as expected, experienced exporters tend to have higher initial expectations, and this implies that their export participation is associated with high expectations about foreign demand. Second, experienced exporters face less uncertainty. This is reasonable since unlike potential entrants these firms may have operated in the foreign market for a long time and have learned a lot before year 2000. Notice that the estimate of standard deviation of signal noise $\hat{\sigma}_u = 1.59$ is significantly larger than the standard deviations of initial beliefs. This implies that the learning speed is slow. In particular, as shown in Figure 1.2, it takes 39 more transactions (signals) for a new exporter to reach the same accuracy (0.236) of the belief as an experienced exporter.

Export participation can be influenced by demand uncertainty through two different channels. The first channel is the variation of the expectation based on the observed signals (transaction outcomes). A high expectation implies a high expected value of exporting which encourages the firm to export. The second channel is the option value of learning. As shown in Figure 1.3, the value difference between the two choices, to export and not to export, is decreasing and concave as demand uncertainty disappears, holding other factors constant. As the firm exports for a longer time and conducts more transactions with foreign buyers, demand uncertainty resolves over time and the option value of learning from exporting decreases. Since the export decision relies on the comparison between this value difference and the fixed or sunk cost that the firm incurs, the decreasing option value of learning contributes to the stylized fact that many firms quit exporting within a short period after entry. However, this does not necessarily mean that all firms are less likely to export over time, as their updated expectations vary according to the signals received.

Table 1.6 compares the differences between exporters and non-exporters in terms of productivity and the demand expectation. Note that the expectation about the demand factor is $E(\exp(\xi)) = \exp(m + \sigma^2/2)$. I find that firms export if their productivity is high or if they have better expectation about foreign demand. In particular, the average productivity of exporters is 0.089, which means exporters are 16% (i.e. $\exp(0.145)$) more efficient in production than non-exporters. Also, on average, $m + \sigma^2/2$ is higher for exporters (0.516) than non-exporters (-0.019).

This implies that exporters expect 71% (i.e., exp(0.535)) more demand than non-exporters. It suggests that expected demand is indeed a determinant of the export decision.

However, the heterogeneity in productivity and the demand belief looks differently in potential entrants and experienced exporters. Figure 1.4 shows the kernel densities of productivity for potential entrants and experienced exporters respectively.[39] Although there is significant productivity heterogeneity within both groups, the mean difference between the two groups is small: 0.131. This implies that on average experienced exporters are 14% (i.e., exp(0.131)) more efficient in production than potential entrants. In turn, this suggests that experienced exporters expect 49% more profit resulted from high productivity, compared with potential entrants. Figure 1.5 shows the kernel densities of the demand belief (summarized as $m + \sigma^2/2$) for the two groups respectively. The result indicates substantial firm heterogeneity in the demand belief. Consistent with their initial beliefs, experienced exporters hold more optimistic and less dispersed beliefs about their demand, compared with potential entrants. More importantly, the demand belief difference between the two groups is large. In particular, on average experienced exporters expect 88% more profit resulted from high demand, compared with potential entrants. These comparisons suggest that although experienced exporters are superior to potential entrants in both the expected demand and productivity, the former is the dominant difference.

These findings suggest productivity and the demand belief play different roles in export participation for potential entrants and experienced exporters. In particular, a firm decides to export may because of high productivity or high demand expectation, both of which lead to high expected profit of exporting. The estimated model allows me to decompose the expected profit difference between exporters and non-exporters into two components: a productivity component and a demand component. These two components can be further compared between potential entrants and experienced exporters. Table 1.7 shows the detailed comparison. To put productivity and the demand belief in the same scale of expected profit and thus render them comparable, I multiply productivity by $(\eta + 1)$ since productivity enters the expected profit as $(\eta + 1)\omega$. It turns out that the aver-

[39] Productivity estimates across firms and over time are pooled together to obtain the densities.

age productivity difference (i.e., $\Delta(\eta+1)\omega$) between exporters and non-exporters in the potential-entrant group is 0.259, and its counterpart in the experienced-exporter group is slightly larger (0.268). This suggests for both group, exporters expect about 30% more export profit than non-exporters because of higher productivity. On the other hand, the average expected demand difference between exporters and non-exporters in the potential-entrant group is 0.656 (corresponding to 93% more profit), but its counterpart for the experienced-exporter group is significantly smaller (0.226) (corresponding to 25% more profit). This implies that the major cross-sectional difference between exporters and non-exporters in the experienced-exporter group is productivity, but for the potential-entrant group, heterogeneity in expected demand is the dominant factor. This result is reasonable and expected, since experienced exporters face less uncertainty about their foreign demand thus productivity becomes the main determinant of export participation. However, potential entrants face much more demand uncertainty, which encourages them to enter into the market although their productivity is not as high as that of experienced exporters.

Above results indicate substantial heterogeneity in both productivity and the demand belief, and show that both processes are driving forces of export dynamics. However, the two processes play different roles for potential entrants and experienced exporters. In the next section, I employ the estimated model to investigate how the two forces affect export participation separately for both potential entrants and experienced exporters.

1.7 Counterfactual Analysis

I conduct two sets of counterfactual analysis about productivity evolution and the belief updating process to study how these processes influence the export participation of potential entrants and experienced exporters. In the first exercise, I shut down either evolution to evaluate how export participation is influenced by that process. In the second exercise, I experiment with the initial belief of potential entrants to study how it affects export participation.

I consider two scenarios in the first set of analysis. First, I shut down the belief evolution by assuming that each firm knows its demand factor and faces no uncertainty. The demand factors of potential entrants and experienced exporters are drawn from the distribution of each group, $N(\hat{m}_{00}, \hat{\sigma}_{00})$ and $N(\hat{m}_{10}, \hat{\sigma}_{10})$, respectively. Note that there is heterogeneity of the demand factor across firms, but the demand factor is fixed over time within a firm and there is no uncertainty. Moreover, I allow the productivity of each firm to evolve as assumed by the model. The predicted percentages of exporters in the two groups as well as all firms are shown in the third row of Table 1.8. The difference between the predicted percentage and actual percentage in data tells us how export participation is affected by heterogeneity in the demand belief. In particular, the percentages of exporters in potential entrants and experienced exporters decrease by 16% and 7%, respectively, and the percentage of exporters in all firms decreases by 12%. This means that the evolution of the demand belief is indeed a driving force of exporting.

The second scenario is to eliminate productivity evolution by replacing the productivity of each firm with its lowest productivity. Note that this only eliminates the time-series productivity heterogeneity within a firm, but it still allows for productivity heterogeneity across firms. I also leave the demand belief to evolve as assumed by the model. The predicted percentages of exporters are shown in the last row of Table 1.8. The difference between the predicted and actual percentages shows how export participation is affected by time-series productivity heterogeneity. The percentages of exporters in potential entrants and experienced exporters decrease by 3% and 6%, respectively, and the percentage for all firms decreases by 4%. This means that the evolution of productivity has a more significant effect on experienced exporters than potential entrants. However, the degrees by which the percentages decrease are less than those in the first scenario, where the demand belief evolution is controlled for. The comparison of the two scenarios implies that, for experienced exporters both forces significantly influence export participation while for potential entrants the belief evolution plays a more important role.

Of course, as shown in the second row of Table 1.8, the percentages of exporters predicted by the original model are very close to the actual percentages in data. This means the results in the above analysis are reliable.

In the second set of experiments, I adjust the initial belief of potential entrants to study how it affects export participation. Productivity still evolves as described in the model. In particular, I make the following three adjustments, and the results are summarized in Table 1.9. First, I assume that potential entrants have the same initial expectation as experienced exporters (which is higher), and leave their uncertainty unchanged. As a result, the percentage of exporters of potential entrants increases from 26% to 51%, almost doubled. This means the expectation of the belief does have a significant effect on export participation. Second, I adjust the standard deviation of the belief, which measures the demand uncertainty, to be the one of the experienced exporters (which is smaller), and let their expectations remain the same. As expected, the percentage of exporters decreases to 23% (which corresponds to a decrease of 11% in the number of exporters), since the option value of learning is smaller. Lastly, I make potential entrants have the same initial belief as experienced exporters. The model predicts that the percentage of exporters will increase to 40%. Note that although the increase in the expectation and the decrease in uncertainty have opposite effects, the combination of the two, in this case, has a net positive effect. That is, if potential entrants hold the same belief as experienced exporters, then they are more likely to export.

1.8 Conclusion and Discussion

In this paper, I develop a structural model of export dynamics with productivity evolution and Bayesian learning about demand in foreign markets to study how the two evolution processes affect firms' export participation. In particular, I allow firms to have inaccurate beliefs about foreign demand and to learn about it through their own individual export transactions. The model contributes to the current literature on export dynamics by allowing firms' export decisions to depend on two heterogenous factors, productivity and the demand belief: while productivity exogenously evolves over time, the demand belief is endogenously related to the export decision, and is updated periodically based on the demand signals received after exporting.

I then apply the model to the shipment-level export data set of the Chinese ceramics industry to investigate the magnitude of uncertainty and examine how the resolution of uncertainty and productivity evolution affect export participation. A general finding is that exporters and non-exporters are different in both productivity and the demand belief. The demand belief heterogeneity is the dominant difference between potential entrants in export markets and experienced exporters. Moreover, productivity and the demand belief play different roles for potential entrants and experienced exporters. The major cross-sectional difference between exporters and non-exporters for experienced firms is productivity, but for potential entrants, demand uncertainty is the dominant factor.

In the counterfactual analysis, I investigate how the two evolution processes influence export dynamics. The result confirms that, the learning process is more important for potential entrants while for experienced exporters both productivity and the learning process are driving forces of export participation. In particular, it predicts that reducing the level of uncertainty of potential entrants to that of experienced exporters causes the number of exporters to fall by 11%. It also show that the learning mechanism indeed contributes to the large attrition rate after the first year of exporting and gradually stabilized export decisions observed in data.

This paper demonstrates that both productivity evolution and the learning process contribute to the observed export dynamics, and shows how to utilize shipment-level exports and firm-level production information to estimate the role of each process. These empirical results are obtained from studying the exports of the Chinese ceramics industry to Germany. However, much about what I claim for this industry can be generalize to other destinations and industries. With the results for more destinations and industries, it is possible to connect export dynamics and the learning process to destination-specific and industrial-specific characteristics.

This paper also implies promising future research. One direction is to examine how the productivity improvement is endogenously related to Bayesian learning about foreign demand. That is, instead of assuming that productivity exogenously evolves over time, the extension recognizes that firms may improve their productivity levels by conducting R&D investment. On the one hand, firms export decisions are driven by both their demand beliefs and productivity; on the other hand, firms with high demand expectations may choose to conduct R&D which in turn increases the propensity of exporting and thus has a further impact on their demand beliefs. In this circumstance, there could be a positive linkage between the demand belief and R&D investment. As the first step, I will examine this linkage using reduced form regressions with data on firm-level production, R&D, and shipment-level exports. Then I will take the linkage into a structural dynamic model to further examine how a forward-looking firm's choice of R&D is related to its belief about foreign demand.

Another interesting extension would be to study additional dynamic patterns of exporting firms such as the sequential exporting documented in Albornoz, Calvo Pardo, Corcos, and Ornelas (2012): firms often start selling to a single country; if it is successful, they tend to expand to other markets. My model can capture this feature by allowing that learning processes across different markets are correlated. As a result, the export outcomes observed in one market are informative regarding the demand in another market. The potential contribution of this extension to the current literature is two-fold. First, it is able to rationalize the gradually stabilized export dynamics through Bayesian learning. Second, it takes productivity evolution into account in the analysis. This is important since without controlling for productivity evolution, it is likely to attribute the effect from the productivity shock to the role of Bayesian learning. However, the major difficulty of this extension is computational: there are more belief state variables to keep track of in the estimation. One possible simplification is to assume the firm's demand is at global scope rather than market-specific. This is left as future work.

Appendices

A Solve for Value Function

I follow the method implemented in Nagypál (2007) to compute the value function. The first step in implementing the method is to solve for the value function at the limit state ($\sigma_{it} = 0$), i.e., the value function when the firm knows its exact ξ_i. This can be achieved by value function iteration. Specifically, I omit the notation for firm i and denote the new state variables without σ_{it} as $\tilde{s}_t = (\tilde{\phi}, \tilde{n}_{t-1}, \omega_t, m)$. Note that at the limit state, the firm knows its own demand factor, so there is no transition of state m. Now the value function $V(\tilde{s}_t, e_{t-1})$ with no uncertainty can be computed by iterating the following Bellman equations implied by equation 1.21:

$$\begin{aligned}
V(\tilde{s}_t, 0) = \Pr & \Big[E\Pi(\tilde{s}_t, 0) - c^s + \delta EV(\tilde{s}_{t+1}, 1) < \delta EV(\tilde{s}_{t+1}, 0)\Big] \times \delta EV(\tilde{s}_{t+1}, 0) \\
& + \Pr\Big[E\Pi(\tilde{s}_t, 0) - c^s + \delta EV(\tilde{s}_{t+1}, 1) > \delta EV(\tilde{s}_{t+1}, 0)\Big] \\
& \times \Big[E\Pi(\tilde{s}_t, 0) - E(c^s|e_t = 1) + \delta EV(\tilde{s}_{t+1}, 1)\Big]
\end{aligned} \tag{51}$$

and

$$\begin{aligned}
V(\tilde{s}_t, 1) = \Pr & \Big[E\Pi(\tilde{s}_t, 1) - c^f + \delta EV(\tilde{s}_{t+1}, 1) < \delta EV(\tilde{s}_{t+1}, 0)\Big] \times \delta EV(\tilde{s}_{t+1}, 0) \\
& + \Pr\Big[E\Pi(\tilde{s}_t, 1) - c^f + \delta EV(\tilde{s}_{t+1}, 1) > \delta EV(\tilde{s}_{t+1}, 0)\Big] \\
& \times \Big[E\Pi(\tilde{s}_t, 1) - E(c^f|e_t = 1) + \delta EV(\tilde{s}_{t+1}, 1)\Big].
\end{aligned} \tag{52}$$

where $E\Pi(\tilde{s}_t, 1)$ is the period export profit implied by equation 1.20, and

$$\begin{aligned}
& EV(\tilde{s}_{t+1}, e_t) \\
& \equiv E[V(\tilde{s}_{t+1}, e_t)|\tilde{s}_t, e_{t-1}] \\
& = \sum_n \left[\Pr(\tilde{n}_t = n|s_{it}, e_{it-1}) \int V(\tilde{s}_{t+1}, e_t) dF_\omega(\omega_{t+1}|\omega_t) \right],
\end{aligned} \tag{53}$$

where $F_\omega(\cdot)$ and $\Pr(\cdot|s_{it}, e_{it-1})$ are the transition probabilities of ω_t and $\tilde{n}_t$, respectively.

Then $V(\tilde{s}_t, e_{t-1})$ is used as an approximation of the value function with low enough uncertainty (i.e., a small enough σ_{it}). Note that the standard deviation σ_{it} is a decreasing function of the number of signals, and there is a one-to-one mapping between the σ_{it} and the number of signals, given the initial state of σ_{it} (i.e., σ_{i0} in the initial belief). Thus, we use the number of signals as an index of uncertainty. I denote $V(\tilde{s}_t, e_{t-1}, N)$ as the value function after receiving a total of N signals. Denote N^* such that $\sigma_{iN^*} < \epsilon$ as the cutoff number of signals after which the value function with uncertainty is approximated by the value function with no uncertainty. That is, for $N \geq N^*$,

$$V(\tilde{s}_t, e_{t-1}, N) \approx V(\tilde{s}_t, e_{t-1}). \tag{54}$$

Now I have the value function at one level of uncertainty: $V(\tilde{s}_t, e_{t-1}, N^*) = V(\tilde{s}_t, e_{t-1})$. Then I can recover the value function with an arbitrary higher uncertainty ($V(\tilde{s}_t, e_{t-1}, N)$ with $N \leq N^*$) using backward induction, which is implied from the Bellman equation:

$$\begin{aligned} V(\tilde{s}_t, 0, N_t) = \Pr &\Big[E\Pi(\tilde{s}_t, 0, N_t) - c^s + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t) < \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t)\Big] \\ &\times \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t) \\ &+ \Pr\Big[E\Pi(\tilde{s}_t, 0, N_t) - c^s + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t) > \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t)\Big] \\ &\times \Big[E\Pi(\tilde{s}_t, 0, N_t) - E(c^s|e_t = 1) + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t)\Big], \end{aligned} \tag{55}$$

and

$$\begin{aligned} V(\tilde{s}_t, 1, N_t) = \Pr &\Big[E\Pi(\tilde{s}_t, 1, N_t) - c^f + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t) < \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t)\Big] \\ &\times \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t) \\ &+ \Pr\Big[E\Pi(\tilde{s}_t, 1, N_t) - c^f + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t) > \delta EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t)\Big] \\ &\times \Big[E\Pi(\tilde{s}_t, 1, N_t) - E(c^f|e_t = 1) + \delta EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t)\Big]. \end{aligned} \tag{56}$$

where N_t and N_{t+1} are the total number of signals received up to the beginning of period t and $t+1$, and $EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t)$ and $EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t)$ are shorthand notations for the expected value functions, which are explained as follows.

If the firm decides not to export in period t, then no additional signal will be received. That is, $N_{t+1} = N_t$, and there is no transition of m_{t+1} or $\tilde{n}_t$: $m_{t+1} = m_t$ and $\tilde{n}_t = \tilde{n}_{t-1}$. Thus,

$$\begin{aligned} &EV(\tilde{s}_{t+1}, 0, N_{t+1}|N_t) \\ &= \int V(\tilde{s}_{t+1}, 0, N_t) dF_\omega(\omega_{t+1}|\omega_t). \end{aligned} \tag{57}$$

But if the firm decides to export in period t, then it may receive a total of n_t additional signals in period t. The total number of signals will increase from N_t to $N_{t+1} = N_t + n_t$. Note that n_t becomes the most recent number of transactions:

$\tilde{n}_t = n_t$. The transition of the belief state specified in equation 1.31 implies:

$$
\begin{aligned}
&EV(\tilde{s}_{t+1}, 1, N_{t+1}|N_t) \\
&= E[V(\tilde{s}_{t+1}, 1, N_{t+1})|\tilde{s}_t, e_{t-1}, N_t] \\
&= \sum_{n_t=1}^{\infty} \Pr(n_t|\tilde{s}_t, e_{t-1}) E[V(\tilde{s}_{t+1}, 1, N_t + n_t)|\tilde{s}_t, e_{t-1}, N_t] \\
&= \sum_{n_t=1}^{\infty} \left[\Pr(n_t|\tilde{s}_t, e_{t-1}) \int V(\tilde{s}_{t+1}, 1, N_t + n_t) dF_\omega(\omega_{t+1}|\omega_t) dF_m(m_t|n_t, m_{t-1}, N_t, 1) \right],
\end{aligned}
\tag{58}
$$

where $\Pr(n_t|\tilde{s}_t, e_{t-1})$, $F_\omega(\omega_{t+1}|\omega_t)$ and $F_m(m_t|n_t, m_{t-1}, N_t, e_t = 1)$ are the transition probabilities of the number of transactions, the productivity and the expectation in the belief state respectively, each of which is given by equation 1.30, 1.23, and 1.28.

In this way, I recover all the relevant value function with uncertainty, which is indexed by N.

Figure 1.1. Percentage of export status switches

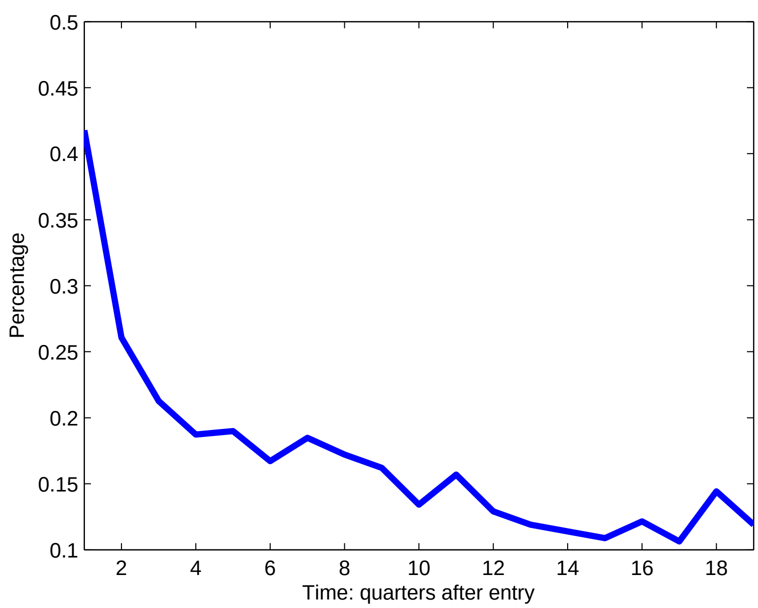


Figure 1.2. The resolution of uncertainty of demand

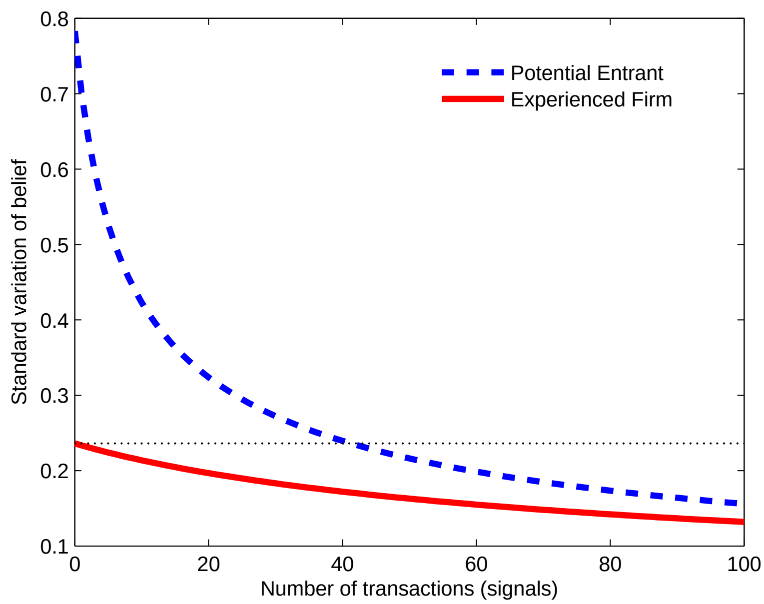


Figure 1.3. Value difference between export and not to export

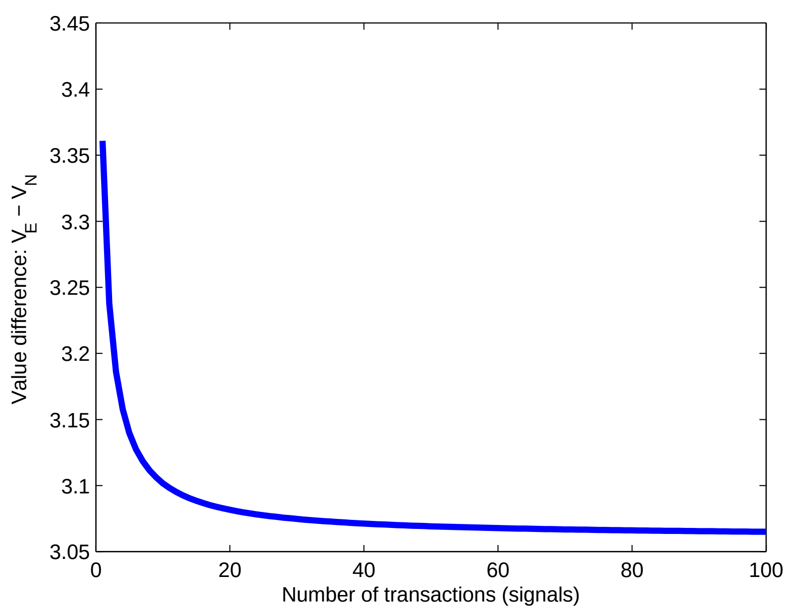


Figure 1.4. Heterogeneity in productivity, by group

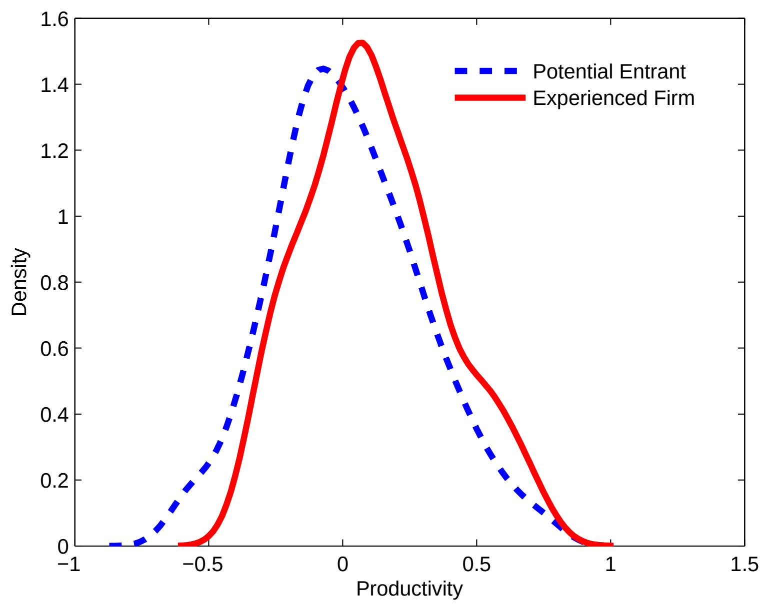


Figure 1.5. Heterogeneity in expected foreign demand, by group

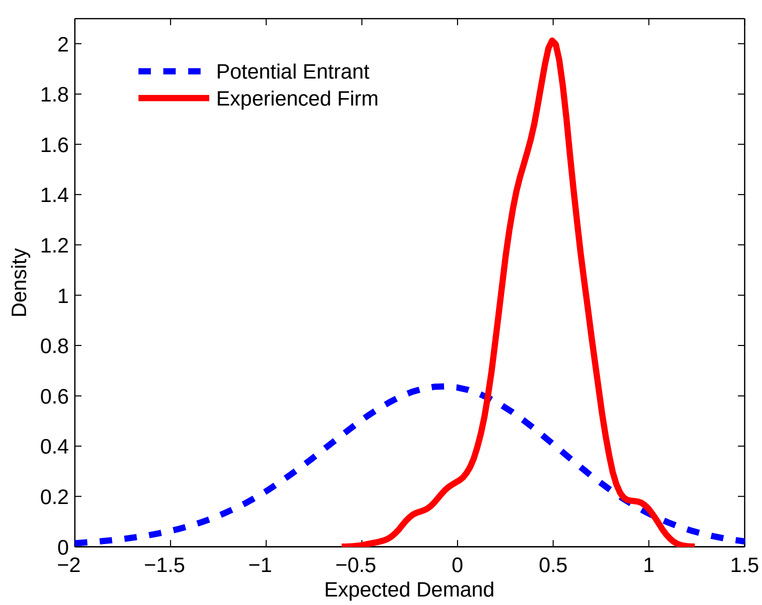


Table 1.1. Percentage of exporters to U.S., by entry cohort

	Year of entry			
Year	2001	2002	2003	2004
2001	100%	-	-	-
2002	76%	100%	-	-
2003	74%	70%	100%	-
2004	75%	68%	65%	100%
2005	76%	72%	65%	68%
2006	75%	68%	60%	69%

[1] Percentages are based on the total number of firms in each cohort. For example, 75% in the second column and sixth row means that among firms entered in year 2001, 75% of them exported in year 2004.

[2] All transaction types are included.

Table 1.2. Estimates of marginal cost and productivity parameters

Parameter	Estimate	Std. Err.
γ_w	0.056^{***}	(0.003)
γ_k	-0.063^{***}	(0.003)
g_0	0.055^{***}	(0.010)
g_1	0.832^{***}	(0.005)
R^2	0.87	
#Obs	6613	

[1] The parameters are estimated for Ceramic Industry from Chinese Annual Survey of Manufacturing. Each observation is a year.

Table 1.3. Demand elasticity estiamtes

Parameter	Estimate	Std. Err.
$1+1/\eta^{D}$	0.758^{***}	(0.005)
$1+1/\eta^{USA}$	0.753^{***}	(0.005)
$1+1/\eta^{ITA}$	0.753^{***}	(0.005)
$1+1/\eta^{JPN}$	0.749^{***}	(0.005)
$1+1/\eta^{DEU}$	0.753^{***}	(0.006)
$1+1/\eta^{ARE}$	0.756^{***}	(0.005)
$1+1/\eta^{GBR}$	0.752^{***}	(0.005)
$1+1/\eta^{ESP}$	0.748^{***}	(0.005)
$1+1/\eta^{AUS}$	0.760^{***}	(0.006)
$1+1/\eta^{CAN}$	0.761^{***}	(0.006)
$1+1/\eta^{HKG}$	0.756^{***}	(0.005)
$1+1/\eta^{oth}$	0.755^{***}	(0.005)
R^2	0.97	
#Obs	2947	

[1] The parameters are estimated for Ceramic Industry for matched firms both in Chinese Annual Survey of Manufacturing and Chinese Customs Transactions. Each observation is a year.

Table 1.4. Estimates of dynamic parameters (1)

Parameter	$\hat{\psi}_0$	$\hat{\psi}_1$	$\hat{\psi}_2$
Estimate	0.341^{***}	0.626^{***}	0.502^{***}
Std. Err.	(0.009)	(0.028)	(0.100)

Table 1.5. Estimates of dynamic parameters (2)

Parameter	$\hat{m}_{00}$	$\hat{m}_{10}$	$\hat{\sigma}_{00}$	$\hat{\sigma}_{10}$	$\hat{c}^f$	$\hat{c}^s$
Estimate	-0.344	0.516	0.783	0.236	0.654	3.155
Std. Err.	(0.083)	(0.140)	(0.107)	(0.051)	(0.072)	(0.308)

Table 1.6. Export decision: productivity v.s. expected demand

	ω	$m + \sigma^2/2$
Export	0.089	0.516
Not to Export	-0.056	-0.019
Difference	0.145	0.535

[1] Numbers in the table are average values. Each observation is a firm-year combination.

Table 1.7. Within group comparison: profitability decomposition

Difference	$\Delta(\eta + 1)\omega$	$\Delta(m + \sigma^2/2)$
Potential Entrants	0.259	0.656
Experienced Exporters	0.268	0.226

[1] Numbers in the table are average differences between exporters and non-exporters. Each observation is a firm-year combination.

Table 1.8. Counterfactuals: learning v.s. productivity

Restriction	Pr(**export**|**potential**)	Pr(**export**|**experienced**)	Pr(**export**)
Data	26%	75%	39%
Model	26%	74%	39%
No Uncertainty	22% (↓ 16%)	69% (↓ 7%)	34% (↓ 12%)
No Prod. Evolution	25% (↓ 3%)	70% (↓ 6%)	38% (↓ 4%)

Table 1.9. Counterfactuals: role of initial belief

	Data	Model	Adjusted Expectation	Adjusted Uncertainty	Both
Pr(export)	26%	26%	51% (↑ 96%)	23% (↓ 11%)	40% (↑ 54%)

Chapter 2

Production Function Estimation with Unobserved Input Price Dispersion

2.1 Introduction

[1]In applications of production function estimation, many datasets do not contain a specific accounting of intermediate input prices and quantities, but instead provide only information on the total expenditure on material inputs. This presents a challenge for consistent estimation when input prices are not homogeneous across firms or when different firms have access to different types of inputs (for example, parts of varying quality). To address this issue, many previous studies assume a homogenous intermediate input is purchased from a single, perfectly competitive market. This assumption facilitates the use of input expenditures as a proxy for quantities (e.g., Levinsohn and Petrin, 2003). However, if this assumption does not hold—for example, if transport costs create price heterogeneity across geography—then the traditional proxy-based estimator is inconsistent. The logic of the inconsistency is straightforward: input price heterogeneity will be observed by firms who respond to price differences both by substituting across inputs and adjusting their total output, causing an endogeneity problem that cannot be controlled for using a Hicks-neutral structural error term. Even in a narrowly defined industry, perfect competition in input markets is not likely to hold, so the proxy approach is clearly not ideal. Fortunately, observed variation in labor input quantities, together with

[1]Joint work of Paul L. E. Grieco, Shengyu Li, and Hongsong Zhang. The authors would like to thank participants at the 22nd Annual Meeting of the Midwest Econometrics Group, the 11th Annual International Industrial Organization Conference, the 2013 North American Meetings of the Econometric Society, and the 2013 Conference of the European Association for Research in Industrial Economics for very helpful comments. In addition, we benefited from thoughtful comments provided by Robert Porter, Mark Roberts, David Rivers, James Tybout, and two anonymous referees. We are also grateful to Mark Roberts and James Tybout for providing the data used in the empirical application. All errors are the authors' own responsibility.

labor and materials expenditures, contains useful information on the intermediate input price variation across firms. By utilizing this variation within a structural model of firms maximizing profits, we introduce a method to consistently estimate firms' production function in the presence of unobserved intermediate input price heterogeneity.

The omitted price problem for production function estimation was first recognized by Marschak and Andrews (1944). They proposed the use of expenditures and revenues as proxies for input and output quantities under the assumption that prices were homogeneous across firms. In practice, the literature has documented significant dispersions in both input and output prices across firms and over time (Dunne and Roberts, 1992; Roberts and Supina, 1996, 2000; Beaulieu and Mattey, 1999; Bils and Klenow, 2004; Ornaghi, 2006; Foster, Haltiwanger, and Syverson, 2008; Kugler and Verhoogen, 2012). Klette and Griliches (1996) show the consequence of ignoring the output price dispersion is a downward bias in the scale estimate of production function.[2] The effect of input price dispersion is slightly more complicated. Using a unique data set containing both inputs price and quantity data, Ornaghi (2006) documents input price bias under the Cobb-Douglas production function. In Section 2.2, we discuss how input price dispersion also biases both the output elasticity and substitution parameters in more general specifications.[3]

A typical data set for production function estimation contains firm-level revenue, intermediate (i.e., material) expenditure, total wage expenditure, capital stock, investment, and additional wage rate/labor quantity. However, quantities and prices for intermediate input are typically not available. The basic idea of our approach is to exploit the first order conditions of firms' profit maximization to impute the unobserved physical quantities of inputs from their expenditures.[4] We then use this recovered physical quantity of intermediate inputs to replace the physical quantity of inputs in the estimation.

Due to these data restrictions, almost all approaches treat materials as a single,

[2]Klette and Griliches (1996) provide a structural approach for controlling for output price variation, we incorporate their approach into our model which additionally controls for input price variation. Of course, because we assume profit maximization, it is important that our model include a demand function so that we can derive the firm's first order conditions.

[3]Ornaghi (2006) considers the Cobb-Douglas specification of the production function, where the elasticity of substitution is assumed to be fixed at one and leaves how price dispersion may bias the elasticity of substitution unexplored. Section 2.2 shows how input price dispersion also biases substitution parameters in more flexible production function specifications, which we confirm in our Monte Carlo study.

[4]To be precise, we recover a quality-adjusted index for the physical quantity of materials used by the firm. The associated materials price also represents a quality-adjusted price. In Section 2.4 we extend the model to consider the case where the firm chooses from several unobserved input types. Our procedure follows the common practice of assuming that observed inputs (labor and capital) are homogeneous to production. See Fox and Smeets (2011) for a study on the role of input heterogeneity in production function estimation.

homogenous input. However, firms often purchase a wide variety of intermediate inputs at a variety of prices. For example, a clothing manufacturer may purchase high or low quality cloth, in addition to a variety of threads, buttons, etc. In the case where firms are selecting a vector of intermediate inputs, datasets typically contain only total input expenditure. In Section 2.4, We show how our method can be extended to account for a vector of unobserved heterogeneous inputs by imputing a quantity index as well as a quality-adjusted input price index for each firm. We confirm in Monte Carlo experiments that our method works well even when firms face a complicated input choice across several potential inputs.

In addition, our method recovers the underlying input price distribution, an important source of heterogeneity across firms. Accounting for input price heterogeneity can give rise to richer explanations of firm policies. For example, if firms' exit decisions are modeled as a cutoff in firm productivity levels, input price heterogeneity would imply that less productive firms may remain in the market when they have access to lower input prices.

The idea of exploiting the first order conditions of profit maximization is also employed in many other studies. Taking materials inputs as given, Gandhi, Navarro, and Rivers (2013) use the transformed first-order conditions of the firm's profit maximization problem to estimate the elasticity of substitution and separate the non-structural errors as their first step in their production function estimation procedure. Doraszelski and Jaumandreu (2013), also assuming labor and materials quantities are observed, use the first-order conditions of labor and material choices to impute the unobserved productivity. Together with a Markov assumption on productivity evolution, this identifies the production function parameters. Katayama, Lu, and Tybout (2009) use the first-order conditions for profit maximization to construct a welfare-based firm performance measure based on Bertrand-Nash equilibrium in a differentiated product industry when input and output prices are unobserved, instead of the traditional productivity measure. Epple, Gordon, and Sieg (2010) instead develop a procedure using the first order condition of the indirect profit function to estimate the housing supply function. Zhang (2012) uses first order conditions as constraints to directly control for structural errors to estimate a production function with biased technology change in Chinese manufacturing industries. De Loecker (2011), De Loecker and Warzynski (2012), and De Loecker, Goldberg, Khandelwal, and Pavcnik (2012) also use the first order condition of labor choice and/or material choice of profit maximization to estimate firm-level markup. The recovered markup is then used to analyze firm performance in international trade. Santos (2012) instead uses the first order condition of labor and material choices to recover demand shocks by adding a timing restriction on the sequence of input choices. Our work is also related to production estimation literature based on factor share regression (Klein, 1953; Solow, 1957; Walters, 1963), which also uses expenditure data to estimate production function using first order conditions.[5]

Our method is closest to Doraszelski and Jaumandreu (2013) and Gandhi, Navarro, and Rivers (2013). These papers also assume that both material and labor choices are static and use the first order conditions of profit maximization as constraints to identify production parameters. Our method differs from these papers in both the data requirement and how we back out the unobserved productivity. Doraszelski and Jaumandreu (2013) use both wage and material prices to directly back out the unobserved productivity using a timing restriction. Our method, without requiring the observation of material price (or quantity), uses the relationship between labor and materials expenditures and quantities to help back out productivity and the unobserved material quantity (and material price). Gandhi, Navarro, and Rivers (2013) show that, when materials quantities are directly observed, it is possible to use first order conditions to non-parametrically identify the production function. We rely on a parametric approach, but avoid the need to observe material quantities directly.

Our method applies to a very general set of production function parameterizations. After employing our assumptions to control for productivity and unobserved price heterogeneity, we are able to estimate the model parameters via a generalized methods of moments (GMM) estimator encompassing the restrictions from the revenue function and (potentially) additional moments. We provide two procedures to fully recover the structural parameters for the most common production function specifications: CES and translog. In addition to the restrictions derived from the revenue function directly, the CES approach relies on cross-sectional restrictions implied by the CES form, while the translog relies on panel assumptions, making use of the common assumption that productivity evolves according to a first-order Markov process. This later approach is available for extremely flexible production specifications.

We demonstrate our approach using the CES production function specification, and evaluate it by carrying out a Monte Carlo study that compares its performance to the traditional estimator and an "oracle" estimator that observes input prices and quantities directly. The results show that our approach recovers the true parameters very well. In contrast, the traditional approach causes systematic biases in the parameter estimates. In particular, the elasticity of substitution is underestimated in the traditional approach as predicted, and the distribution parameters are also biased. This bias could mislead researchers attempting to

[5]Share regression can consistently (but may be biased as Walters (1963) point out) recover the production parameters when firms are price-takers in output market and technology shows constant return to scale. For many applications, Cobb-Douglas is a good approximation of production function. However, as is well known, it implies constant expenditure share for static inputs, even when firms face different input prices. This is not the case in the micro-level data, which usually suggests a large dispersion of expenditure shares among firms. So, we think it will be more realistic to recognize the dispersion of expenditure share, especially when the purpose is to consider firm behavior.

make policy recommendations. For example, in a trade policy setting, this bias could result in erroneous counterfactual estimates of demand and supply changes of *all* inputs and outputs due to a proposed change to tariff rates on imported intermediate inputs.

We apply our approach to a plant-level data set from Colombian manufacturing industries and compare our results with those derived using the traditional estimator. The results are consistent with both our predictions and the results of the Monte Carlo experiments. That is, compared with our method, the elasticity of substitution from the traditional approach is consistently lower. Moreover, the distribution parameter estimates of the traditional method differ significantly from those of our method.

Our results indicate significant input price dispersion in all industries, providing further indication of the importance of controlling for unobserved price heterogeneity. The recovered distribution and evolution of intermediate prices are similar to that for studies in which input prices are directly observed (e.g., Atalay, 2012). We also find a positive correlation between intermediate input prices, wages and productivity, also corroborating earlier studes (Kugler and Verhoogen, 2012). Finally, the distribution of productivity estimated using our approach is even wider than using traditional approaches, suggesting that there is more productivity dispersion in Colombian manufacturing than previously thought.

The following section reviews the omitted input and output price biases in detail. Section 2.3 introduces a model with unobserved price heterogeneity and outlines our procedure to consistently estimate the model. Section 2.4 then extends the model to multiple heterogeneous inputs. Section 2.5 presents Monte Carlo experiments that evaluate the performance of our estimator and confirm the biases in traditional methods when unobserved price heterogeneity is present. We apply our method to a data set on Colombian manufacturing in Section 2.6, and conclude in Section 2.7.

2.2 Omitted Price Biases

In ideal cases where physical quantities of input and output are available, they can be used directly in the estimation of production functions (Eslava, Haltiwanger, Kugler, and Kugler, 2004; Ornaghi, 2006; Grieco and McDevitt, 2012). However, many datasets contain information on the total expenditure on intermediate inputs but not a specific accounting of their prices and quantities. In this case, the traditional approach is to use the deflated value (by industry-level price indices) of inputs and output (De Loecker, 2011) as proxy of physical quantities. This procedure implicitly requires that firms operate in perfectly competitive input markets so that all firms in the industry face the same prices. However, markets are more likely to be imperfectly competitive and are characterized by heterogenous features. For example, transportation costs may create input price differences between firms based on geography. Firm-level input and output prices vary across firms and over time, which impact the firm-level input choice. Consequently, the traditional approach will induce bias in the estimation (Klette and Griliches, 1996; Van Beveren, 2010).

Consider a production function in logarithm form $q_{jt} = f(x_{jt}, \theta_0)$, where q_{jt} is the log firm-level physical quantity of output produced by a vector of log physical input x_{jt}, and θ_0 is the parameter of interest. For commonly available firm-level production data sets, q_{jt} and x_{jt} are not available. Instead, we observe the (deflated) revenue $r_{jt} = q_{jt} + p^q_{jt}$ and (deflated) input expenditures $v_{jt} = x_{jt} + p^x_{jt}$, where p^q_{jt} and p^x_{jt} are the log firm-level output and input prices which are deflated by industry-level price indices. In the traditional estimation, q_{jt} and x_{jt} are often substituted by r_{jt} and v_{jt}. Thus, while the true model is,

$$r_{jt} = f(v_{jt} - p^x_{jt}, \theta_0) + p^q_{jt} + \epsilon_{jt},$$

the estimated model is,

$$r_{jt} = f(v_{jt}, \hat{\theta}) + u_{jt},$$

where ϵ_{jt} are i.i.d measurement errors, and u_{jt} contains both the measurement error ϵ_{jt} and the omitted terms p^q_{jt} and p^x_{jt}. Note that if input and output prices vary across firms, and input x_{jt} is chosen after these prices are observed, then the input expenditure v_{jt} is correlated with both input and output prices. Ignoring this will result in an inconsistent estimator of θ_0. We now provide two specific examples to illustrate the effect of unobserved price dispersion on the estimation of production functions.

Example 2.1:[6] Consider the Cobb-Douglas production function,

$$q_{jt} = \beta_0 + \beta_l \ell_{jt} + \beta_m m_{jt} + \epsilon_{jt},$$

where lower case letters represent the logarithm value and for simplicity ϵ_{jt} is an i.i.d. measurement error.[7] In commonly available data sets, where q_{jt} and m_{jt} are not available, revenue $r_{jt} = q_{jt} + p^q_{jt}$ and material expenditure $v^m_{jt} = m_{jt} + p^m_{jt}$ are used as proxy. Thus, the model is estimated via,

$$r_{jt} = \beta_0 + \beta_\ell \ell_{jt} + \beta_m v^m_{jt} + \underbrace{p^q_{jt} - \beta_m p^m_{jt} + \epsilon_{jt}}_{u_{jt}},$$

where $u_{jt} = p^q_{jt} - \beta_m p^m_{jt} + \epsilon_{jt}$ is the error term. The endogeneity problem arises since $E(v^m_{jt} u_{jt}) \neq 0$. For simplicity, suppose β_ℓ is known, then the estimated coefficient

[6] Ornaghi (2006) shows a similar example.

[7] Here we disregard the presence of unobserved productivity difference across firms for clarity. Our model in section 3 will account for the unobserved productivity.

for material is given by,[8]

$$\text{plim } \hat{\beta}_m = \beta_m + \frac{cov(v^m, p^q - \beta_m p^m)}{var(v^m)} = \beta_m + \frac{cov(v^m, p^q)}{var(v^m)} - \beta_m \frac{cov(v^m, p^m)}{var(v^m)}.$$

The estimate is inconsistent if the last two additional terms are non-zero. Note that input expenditures and output price are usually negatively correlated,[9] we expect $\frac{cov(v^m,p^q)}{var(v^m)} < 0$. Similarly, if the correlation between materials price and materials expenditure is positive, then $\frac{cov(v^m,p^m)}{var(v^m)} > 0$. In this case, we have $\hat{\beta}_m < \beta_m$. So the material coefficients will be underestimated if quantities of material and output are substituted by deflated values. However, in the case that material expenditure and material price are negatively related, $\frac{cov(v^m,p^m)}{var(v^m)} < 0$, the direction of the bias is ambiguous. In both cases, input price bias is present as long as material choice depends on the omitted material price. Finally, it is easy to see that the same mechanism will induce inconsistency in the more flexible translog functional form, which also leads to a log-linear specification of the production function. ■

Example 2.2: For the CES production function, ignoring input price heterogeneity can bias the elasticity of substitution between inputs. To see this, consider the CES production function (without capital for simplicity),

$$Q_{jt} = F(M_{jt}, L_{jt}) = [\alpha M_{jt}^{\gamma} + (1-\alpha) L_{jt}^{\gamma}]^{\frac{1}{\gamma}},$$

where $\gamma = \frac{\sigma - 1}{\sigma}$, and σ is the constant elasticity of substitution which is defined as

$$\sigma = -\frac{\partial \ln(M/L)}{\partial \ln(F_M/F_L)},$$

where F_M and F_L are the marginal output of material and labor. If (deflated) expenditures (denoted as $E_M = P_M M$ and $E_L = P_L L$) are used instead of firm-level quantities, then the estimated model is

$$Q_{jt} = \hat{F}(E_{M_{jt}}, E_{L_{jt}}) = [\hat{\alpha} E_{M_{jt}}^{\hat{\gamma}} + (1-\hat{\alpha}) E_{L_{jt}}^{\hat{\gamma}}]^{\frac{1}{\hat{\gamma}}},$$

where $\hat{\gamma} = \frac{\hat{\sigma}-1}{\hat{\sigma}}$. In particular, the elasticity of substitution $\hat{\sigma}$ is calculated using labor expenditure and material expenditure which is different from the original

[8] If β_ℓ is not known and β_ℓ and β_m are estimated jointly, then $\hat{\beta}_\ell$ will also be biased.

[9] Klette and Griliches (1996, 346) lists cases where this is true.

definition:

$$
\begin{aligned}
\hat{\sigma} &= -\frac{\partial \ln(E_M/E_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})} = -\frac{\partial \ln(E_M/E_L)}{\partial \ln(F_M/F_L)} \times \frac{\partial \ln(F_M/F_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})} \\
&= \left(-\frac{\partial \ln(M/L)}{\partial \ln(F_M/F_L)} - \frac{\partial \ln(P_M/P_L)}{\partial \ln(F_M/F_L)}\right) \times \frac{\partial \ln(F_M/F_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})} \\
&= (\sigma - 1) \times \frac{\partial \ln(F_M/F_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})}
\end{aligned}
$$

The last equation holds because $\sigma = -\frac{\partial \ln(M/L)}{\partial \ln(F_M/F_L)}$ by definition and $\frac{\partial \ln(P_M/P_L)}{\partial \ln(F_M/F_L)} = 1$ if the firm chooses M and L to minimize its variable cost. This equation shows that the bias comes from two sources. The first source is the use of (E_M, E_L) rather than (M, L) for the elasticity of substitution. That is,

$$
\frac{\partial \ln(E_M/E_L)}{\partial \ln(F_M/F_L)} \neq \frac{\partial \ln(M/L)}{\partial \ln(F_M/F_L)}.
$$

The second source of bias is due to inconsistency between the estimated production function and the true production function due to the use of (E_M, E_L) instead of (M, L) in estimation, i.e.,

$$
\frac{\partial \ln(F_M/F_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})} \neq 1.
$$

If $\frac{\partial \ln(F_M/F_L)}{\partial \ln(\hat{F}_{E_M}/\hat{F}_{E_L})}$ is less than 1, then the elasticity of substitution is underestimated ($\hat{\sigma} < \sigma$). The intuition is that because of cost minimization, the physical input ratio will change in a direction against the change in input price ratio. As a result, the change in the input expenditure ratio E_M/E_L is offset partially by the change in P_M/P_L. For example, suppose there is an increase in material price, so P_M/P_L rises. With cost minimization, the physical input M will be partially substituted by labor, thus M/L drops. However, the percentage drop in the expenditure ratio E_M/E_L is less than that in M/L, because $E_M/E_L = (P_M/P_L) \cdot (M/L)$ and P_M/P_L rose. As a result, the estimated elasticity of substitution measured using expenditure proxies will be biased. Note that this also implies the distribution parameter (α) is also biased. ■

These examples show that when firms face heterogeneous input prices, traditional production function estimates that rely on materials expenditure to proxy for the quantity of materials are inconsistent. The bias affect estimates of both output elasticity and the elasticity of substitution between inputs. It is easy to see how this inconsistency could lead to misleading counterfactual analysis to policy questions of interest. For example, consider a proposed tariff increase on some intermediate input: researchers who estimate the production function while failing to account for input price heterogeneity would mis-predict both the change in firm output and the degree of substitution from the imported input to other inputs resulting from the tariff. In the rest of the paper, we propose and demonstrate a structural approach which uses information on the relative expenditure on inputs to control for unobserved input price and consistently estimate the production function.

2.3 Estimation with Unobserved Price Dispersion

In this section, we introduce a model of firms' decision-making in a standard monopolistically competitive output market. The goal is to find an approach to estimating the production function when we have the commonly available data on output value, all inputs values, wage rate and labor quantity. Instead of substituting quantities with deflated values, our approach exploits the first order conditions implied by profit maximization to impute unavailable physical quantities of intermediate inputs from expenditures.

2.3.1 The General Model

To fix ideas, we first present our approach for general forms of production and demand functions and outline our approach for consistent estimation. In the next subsections, we will show how our approach can be used to estimate two commonly used specifications of the production function: CES and translog.

Suppose at each period t, each firm j produces a single product using labor (L_{jt}), intermediate material (M_{jt}), and capital (K_{jt}) using the production function,

$$Q_{jt} = \exp(\omega_{jt})F(L_{jt}, M_{jt}, K_{jt}; \theta),$$

where Q_{jt} is the output, ω_{jt} is a Hicks-neutral productivity shock observed by the firm (but not by researchers), and θ is the set of parameters in the production function. For notational clarity, we assume that materials is a scalar input, however we show that this can be generalized in Section 2.4.

The inverse demand function is,

$$P_{jt} = P_t(Q_{jt}; \eta),$$

where P_{jt} is the output price and η is the set of parameters in the demand function. The inverse demand function $P_t(\cdot, \cdot)$ is continuous and decreasing in its first coordinate. We allow the demand to be different over time. We make the following assumptions:

Assumption 1 (Smooth Production Function)**.** *Production function $F(\cdot)$ is known up to a finite dimensional parameter θ, strictly increasing in inputs, and continuously differentiable up to second order. For $i \in \{M, L\}, \lim_{i\to\infty} \frac{\partial F}{\partial i} = 0$ and $\lim_{i\to 0} \frac{\partial F}{\partial i} = \infty$.*

Assumption 2 (Exogenous Input Prices)**.** *Firms are price takers in input markets. Suppliers use linear pricing, but input prices are allowed to be different across firms and over time.*

Assumption 3 (Profit Maximization)**.** *After observing their productivity draw, ω_{jt}, firms optimally choose labor and material inputs to maximize the profit in each period. The firm's capital stock for period t is chosen prior to the revelation of ω_{jt}.*

Several points are worth highlighting. Some previous literature (e.g., Arellano and Bond (1991), Ackerberg, Caves, and Frazer (2006)) allows adjustment costs in labor, but an implicit assumption on homogenous input price is required for consistency when only input expenditure is available to researchers. In this paper, we assume that both labor and material inputs are flexibly chosen at the beginning of each period, as in Levinsohn and Petrin (2003) and Doraszelski and Jaumandreu (2013). In addition, as in Olley and Pakes (1996), we assume capital is pre-fixed in the short run. However, in contrast to the previous literature, labor and material input choices depend on idiosyncratic input prices. This is an additional source of firm heterogeneity along with the well-known Hicks-neutral technology shifter, ω_{jt}. The assumption that firms are price takers does not preclude firms being offered different prices on the basis of their size (i.e., capital stock), productivity, or negotiating ability, but does assume that firms do not receive "quantity discounts," which would endogenously affect purchasing decisions. As shown in the previous section, ignoring the variation in input prices and using the deflated input expenditures as proxies of physical inputs will introduce a bias into the estimation.

To control for price and productivity differences across firms, we use an explicit model of profit maximization to construct estimation equations involving only commonly available expenditure data. Using first order conditions implies stronger assumptions in terms of static profit maximization/cost minimization by choosing material and labor. In particular, we exploit the structure of the first order conditions to impute materials prices and quantities from expenditure data.

While, relative to Olley and Pakes (1996), we strengthen some assumptions by requiring profit maximization, we are able to relax others. Because we use the first order conditions to recover the unobserved productivity, ω_{jt}, we will not need to use a "proxy" (such as investment) to recover it. Indeed, investment will not be used in our procedure at all, so there is no need for an invertability condition on the investment function. Instead, materials quantities and productivity will be jointly recovered from the two first order conditions.

2.3.2 Recovering the Unobserved Input Prices

We assume the econometrician observes revenue $R_{jt} = P_{jt}Q_{jt}$, inputs expenditure $E_{Mjt} = P_{M_{jt}}M_{jt}$, wage rate $P_{L_{jt}}$, number of workers or number of working hours L_{jt}, and capital stock K_{jt}. But she does not observe the material inputs prices or quantities ($P_{M_{jt}}$ and M_{jt}) or output prices and quantities (P_{jt} and Q_{jt}).

At the beginning of each period, after observing capital stock, K_{jt}, productivity shock, ω_{jt}, and idiosyncratic input prices $P_{L_{jt}}$ and $P_{M_{jt}}$, firm j chooses its own labor and materials inputs to maximize its period profit. The firm's static decision problem is:

$$\begin{aligned} &\max_{L_{jt},M_{jt}} P_t(Q_{jt};\eta)Q_{jt} - P_{Ljt}L_{jt} - P_{Mjt}M_{jt} \\ &\text{s.t. } Q_{jt} = \exp(\omega_{jt})F(L_{jt}, M_{jt}, K_{jt}; \theta). \end{aligned}$$

The corresponding first order conditions are,

$$\begin{aligned} \exp(\omega_{jt})F_{L_{jt}}\left[P_t(Q_{jt};\eta) + Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right] &= P_{L_{jt}}, \\ \exp(\omega_{jt})F_{M_{jt}}\left[P_t(Q_{jt};\eta) + Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right] &= P_{M_{jt}}. \end{aligned} \tag{2.1}$$

Given the Inada conditions of Assumption 1, we know an interior solution to these first order conditions exists. Dividing the two first order conditions, multiplying both sides by $\frac{L_{jt}}{M_{jt}}$, and rearranging yield,

$$\frac{F_{L_{jt}}L_{jt}}{F_{M_{jt}}M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} = 0, \tag{2.2}$$

where $F_{L_{jt}}$ and $F_{M_{jt}}$ are the partial derivatives of F with respective to labor and material, and $E_{L_{jt}} = P_{L_{jt}}L_{jt}$ and $E_{M_{jt}} = P_{M_{jt}}M_{jt}$ are expenditures on labor and material, which are observed in the data.

(2.2) is the key to our approach. It relates labor and the intermediate input, given that they are optimally chosen to maximize profits. (2.2) is always satisfied at the firm choice of $(M_{jt}, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt})$, given that firms choose their inputs optimally using profit maximization. So the key question is whether (2.2) places enough restriction on the unobserved material quantity M_{jt} so that we can recover it from the observed firm choices, $(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt})$, up to production function parameter θ. The following proposition gives conditions under which we are able to impute M_{jt} from the observed data.

Proposition 1. *Define,*

$$z(M_{jt}; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) \equiv \frac{F_{L_{jt}}L_{jt}}{F_{M_{jt}}M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}},$$

For a given observation of $(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt})$ *and parameter vector* θ*, suppose either* $\frac{\partial z}{\partial M} > 0$ *or* $\frac{\partial z}{\partial M} < 0$ *for all* $M \in (0, \infty)$ *such that* $z(M; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0$*. Then there exists a unique* M^* *that satisfies,*

$$z(M^*; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0.$$

Once we recover $M^*_{jt} = M^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta)$, we can replace the unobserved intermediate inputs M_{jt} in (2.1) to back out the productivity shock as $\omega^*_{jt} = \omega^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta)$. The proof of Proposition 1 is straightforward and provided in A. It can be used to show conditions under which a unique materials quantity is recoverable for both of the two special cases we consider, CES and Translog. For more general models, it is possible the condition will not hold and multiple materials quantity-price combinations may satisfy the first order conditions. In this case, the model is partially identified. For the remainder of this paper, we will assume that M^* can be uniquely recovered.

The following three examples illustrate Proposition 1. In the first example, we show that the condition in proposition 1 is *not* satisfied for the Cobb-Douglas specification. This is because, as is well-known, the Cobb-Douglas production function assumes that expenditure shares are constant within the data, eliminating the source of variation we need to separate input prices and quantities. In Example 3.2, we show that for the more general CES production function, we can impute materials quantities as long as the elasticity of substitution is not 1 (the Cobb-Douglass case). We are able to test for this restriction by observing whether the expenditure shares are constant across firms in the data. Finally, we show in example 3.3 that under the translog specification, M^* can be recovered under a simple invertibility condition.

Example 3.1 (Cobb-Douglas Production Function)
For Cobb-Douglas production function

$$Q_{jt} = \exp(\omega_{jt}) F(L_{jt}, M_{jt}, K_{jt}; \theta) = \exp(\omega_{jt}) L_{jt}^{\alpha_L} M_{jt}^{\alpha_M} K_{jt}^{\alpha_K},$$

where $\theta = (\alpha_L, \alpha_M, \alpha_K)$, it is straightforward to show that,

$$\begin{aligned} z(M_{jt}, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta) &= \frac{F_{L_{jt}} L_{jt}}{F_{M_{jt}} M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\ &= \frac{\alpha_L K_{jt}^{\alpha_K} L_{jt}^{\alpha_L - 1} M_{jt}^{\alpha_M} L_{jt}}{\alpha_M K_{jt}^{\alpha_K} L_{jt}^{\alpha_L} M_{jt}^{\alpha_M - 1} M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\ &= \frac{\alpha_L}{\alpha_M} - \frac{E_{L_{jt}}}{E_{M_{jt}}}. \end{aligned}$$

In this case, $z(\cdot)$ does not vary with M_{jt} (e.g., $\frac{\partial z}{\partial M_{jt}} = 0$), so unobserved materials cannot be recovered from (2.2). The intuition is that, because the elasticity of substitution is fixed at one, when the relative inputs price $(\frac{P_L}{P_M})$ changes firms always choose labor and material such that the percentage increase (or decrease) of the labor-material ratio $(\frac{L}{M})$ equals the percentage decrease (or increase) of relative price $(\frac{P_L}{P_M})$. As a result, the expenditure ratio $\frac{E_{L_{jt}}}{E_{M_{jt}}}$ remains constant $(\frac{\alpha_L}{\alpha_M})$. In this case, we cannot separate the price and quantity of materials from the information on expenditure ratio $\frac{E_{L_{jt}}}{E_{M_{jt}}}$. ■

Of course, because we observe the expenditure ratio between materials and labor in the data, it is easy to verify that it is not constant across all firms. As long as there is variation, it is reasonable to specify a production function that allows the ratio to vary and use this variation to impute materials prices and quantities.

Example 3.2 (CES Production Function)

Consider CES production function,[10]

$$Q_{jt} = \exp(\omega_{jt}) F(L_{jt}, M_{jt}, K_{jt}; \theta) = \exp(\omega_{jt})[\alpha_L L_{jt}^{\gamma} + \alpha_M M_{jt}^{\gamma} + \alpha_K K_{jt}^{\gamma}]^{\frac{1}{\gamma}}, \quad (2.3)$$

where $\gamma = \frac{\sigma-1}{\sigma}$ (σ is the elasticity of substitution), and $\theta = (\alpha_L, \alpha_M, \alpha_K, \sigma)$. We can show that,

$$\begin{aligned}
& z(M_{jt}, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) \\
= \; & \frac{F_{L_{jt}} L_{jt}}{F_{M_{jt}} M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\
= \; & \frac{\exp(\omega_{jt})[\alpha_L L_{jt}^{\gamma} + \alpha_M M_{jt}^{\gamma} + \alpha_K K_{jt}^{\gamma}]^{\frac{1}{\gamma}-1} \alpha_L L_{jt}^{\gamma-1} L_{jt}}{\exp(\omega_{jt})[\alpha_L L_{jt}^{\gamma} + \alpha_M M_{jt}^{\gamma} + \alpha_K K_{jt}^{\gamma}]^{\frac{1}{\gamma}-1} \alpha_M M_{jt}^{\gamma-1} M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\
= \; & \frac{\alpha_L L_{jt}^{\gamma}}{\alpha_M M_{jt}^{\gamma}} - \frac{E_{L_{jt}}}{E_{M_{jt}}}
\end{aligned}$$

Taking the derivative of $z(\cdot)$ with respect to M_{jt} we yields,

$$\frac{\partial z}{\partial M_{jt}} = -\gamma \frac{\alpha_L L_{jt}^{\gamma}}{\alpha_M M_{jt}^{\gamma+1}},$$

It is clear that the sign of $\frac{\partial z}{\partial M_{jt}}$ is determined by $-\gamma$ only.[11] Therefore, as long as

[10] Below we will work with a normalized form of the CES production function, but we use an unnormalized form here for expositional simplicity.

[11] When $\gamma = 0$, the CES function is equivalent to the Cobb-Douglas case (Example 3.1) and

$\gamma \neq 0$ (i.e., $\sigma \neq 1$), we can recover the unobserved material from (2.2).

$$M^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta) = \left(\frac{\alpha_L}{\alpha_M}\frac{E_{M_{jt}}}{E_{L_{jt}}}\right)^{\frac{1}{\gamma}} L_{jt}.$$

Intuitively, it is possible to infer information about material quantity M_{jt} from the inputs expenditure ratio $\frac{E_{M_{jt}}}{E_{L_{jt}}}$. This feature, together with the definition $E_{Mjt} = P_{Mjt}M_{jt}$ help us separate the quantity M_{jt} and material price P_{Mjt} from each other. ■

Example 3.3 (Translog Production Function)

Finally, we consider the translog production function specification,

$$\begin{aligned} Q_{jt} &= \exp(\omega_{jt}) F(L_{jt}, M_{jt}, K_{jt}; \theta) \qquad (2.4)\\ &= \exp(\omega_{jt}) \exp\Big\{\alpha_k \ln K_{jt} + \alpha_l \ln L_{jt} + \alpha_m \ln M_{jt} \\ &\quad + \frac{1}{2}\alpha_{kk}(\ln K_{jt})^2 + \frac{1}{2}\alpha_{ll}(\ln L_{jt})^2 + \frac{1}{2}\alpha_{mm}(\ln M_{jt})^2 \\ &\quad + \alpha_{kl}(\ln K_{jt})(\ln L_{jt}) + \alpha_{km}(\ln K_{jt})(\ln M_{jt}) + \alpha_{lm}(\ln L_{jt})(\ln M_{jt})\Big\} \end{aligned}$$

Where $\theta = (\alpha_k, \alpha_l, \alpha_m, \alpha_{kk}, \alpha_{ll}, \alpha_{mm}, \alpha_{kl}, \alpha_{km}, \alpha_{lm})$ are the structural parameters. The translog is a more flexible generalization of the Cobb-Douglas production function which allows for the elasticity of substitution to be a function of the inputs. Under this specification,

$$\begin{aligned} z(M_{jt}, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) &= \frac{F_{L_{jt}}L_{jt}}{F_{M_{jt}}M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\ &= \frac{\alpha_l + \alpha_{ll}\ln L_{jt} + \alpha_{kl}\ln K_{jt} + \alpha_{ml}\ln M_{jt}}{\alpha_m + \alpha_{mm}\ln M_{jt} + \alpha_{km}\ln K_{jt} + \alpha_{ml}\ln L_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \\ &= \frac{S_{Ljt}}{S_{Mjt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}} \end{aligned}$$

Where S_{Ljt} and S_{Mjt} are the numerator and denominator of the first term, respectively. The derivative with respect to M_{jt} is,

$$\frac{\partial z}{\partial M_{jt}} = \frac{1}{M_{jt}S_{Mjt}}\left(\alpha_{ml} - \alpha_{mm}\frac{S_{Ljt}}{S_{Mjt}}\right).$$

we cannot recover the unobserved materials from (2.2). This case must be excluded from the parameter set Θ.

So the sign is determined by $\alpha_{ml} - \alpha_{mm}\frac{S_{Ljt}}{S_{Mjt}}$.[12] At any solution where $z(\cdot) = 0$, we know $\frac{S_{Ljt}}{S_{Mjt}} = \frac{E_{L_{jt}}}{E_{M_{jt}}}$, so for any M such that $z(M, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0$,

$$\text{sign}\left(\frac{\partial z}{\partial M_{jt}}\right) = \text{sign}\left(\alpha_{ml} - \alpha_{mm}\frac{E_{Ljt}}{E_{Mjt}}\right),$$

which does not vary with M_{jt} given $(L_{jt}, K_{jt}, E_{Ljt}, E_{Mjt}; \theta)$. Applying Proposition 1, we can recover M_{jt} as long as $\alpha_{ml} - \alpha_{mm}\frac{E_{Ljt}}{E_{Mjt}} \neq 0$.[13] In this case the closed form for M^*_{jt} is,

$$M^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta) = \tag{2.5}$$
$$\exp\left(\frac{\frac{E_{Ljt}}{E_{Mjt}}(\alpha_m + \alpha_{km}\ln K_{jt} + \alpha_{ml}\ln L_{jt}) - (\alpha_l + \alpha_{ll}\ln L_{jt} + \alpha_{kl}\ln K_{jt})}{(\alpha_{ml} - \alpha_{mm}\frac{E_{Ljt}}{E_{Mjt}})}\right).$$

Finally, we note that for higher order translog specifications, a similar procedure may be available, but it will necessitate finding the roots of a polynomial (rather than linear) equation in M, introducing the possibility that $z(\cdot) = 0$ may have multiple solutions.[14] ■

2.3.3 Estimation

We now turn to estimation of the parameters of the production function, θ, and the inverse demand function, η, using commonly available data on output value, all input values, and the wage rate. If our assumptions are satisfied, the intermediate input quantity can be uniquely imputed as,

$$M^*_{jt} = M^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta). \tag{2.6}$$

Once the quantity of material M_{jt} is recovered, we can plug it back into either of the first order conditions and recover the unobserved productivity,

$$\omega^*_{jt} = \omega^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, M^*_{jt}; \theta). \tag{2.7}$$

[12] We know $S_{M_{jt}}$ is positive because it is proportional to the product of the marginal product of materials and M_{jt}.

[13] A sufficient condition for $\alpha_{ml} - \frac{E_{L_{jt}}}{E_{M_{jt}}}\alpha_{mm} \neq 0$ to always hold is if $\alpha_{mm}\alpha_{ml} < 0$.

[14] It would be ideal to be able to impute M^* for a nonparametric production function. Unfortunately, our method requires a parametric approach to recovering M^*, as is illustrated by these examples. Nonetheless, we can in principle accommodate an arbitrarily flexible parametric specification with appropriate restrictions to guarantee uniqueness.

Different from Olley and Pakes (1996), here we recover the unobserved productivity parametrically from the firm's first order conditions.[15] There are several advantages to this method. First, the estimation does not require the investment data. Second, there is no need to rely on invertibility of the investment policy function, which may be problematic when adjustment costs generate lumpiness in the optimal investment policy. Moreover, our method of controlling for endogeneity does not require the Markov assumption on the productivity evolution process. Of course, assumptions on the productivity evolution process may still be needed in identifying the production function, as we will discuss below. Finally, as we fully exploit the structural assumptions of the parametric production function and corresponding first order conditions to recover unobserved productivity and material quantity, as in Doraszelski and Jaumandreu (2013), we do not have to rely on nonparametric methods to estimate these functions. As a result, even when both expenditures may be functions of the same set of variables in equations (2.8) (implicit in ω^*_{jt}) and (2.13) below, we still have identification as long as labor and materials expenditures are not perfectly correlated.[16] Ackerberg, Caves, and Frazer (2006) also discussed this possibility (page 16, version of December 28, 2006).

Since output quantities are not directly observed, we follow Klette and Griliches (1996) and use the revenue function as the estimating equation. The revenue function is,

$$R_{jt} = \exp(u_{jt}) P_t \left(Q_{jt}; \eta\right) Q_{jt}.$$

Where R_{jt} is the observed revenue of the firm, $Q_{jt} = e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt}; \theta)$ is the predicted quantity of physical output based on observed inputs and the model parameters (θ, η), and u_{jt} is a mean-zero revenue error term which incorporates measurement error as well as demand and productivity shocks that are unanticipated by the firm. Taking the logarithm of the revenue function yields,

$$\ln R_{jt} = \ln P_t \left(e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt}; \theta); \eta\right) + \ln \left[e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt}; \theta)\right] + u_{jt} \quad (2.8)$$

In this equation, the unobserved productivity and material quantity, ω^*_{jt} and M^*_{jt}, are recovered as functions of observed variables as in equations (2.6) and (2.7). The only remaining unobservable, u_{jt}, is unknown to the firm and is uncorrelated with the observed inputs.

To simplify notations, denote $\mathbf{w}_{jt} \equiv (L_{jt}, E_{M_{jt}}, E_{L_{jt}}, K_{jt})$, and $r_{jt} \equiv \ln R_{jt}$, and

[15] Zhang (2012) uses a similar approach to allow for biased technical change in the production function.

[16] The data in our application shows that the labor-material expenditure ratio has large variation across firms (as required by our empirical model), supporting the idea that these two expenditures are not perfectly correlated.

$\beta \equiv (\theta, \eta) \in \mathbb{R}^D$. Define

$$f(\mathbf{w}_{jt};\beta) = \ln P_t\left(e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt};\theta);\eta\right) + \ln\left[e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt};\theta)\right]$$

Therefore, the true parameter β^0 solves the following nonlinear least squares problem,

$$\min_{\beta} E\Big[(r_{jt} - f(\mathbf{w}_{jt};\beta))^2\Big]. \tag{2.9}$$

Of course, we have not yet shown that β^0 is identified. Indeed, in both of our primary examples we need additional restrictions to identify β^0. In order to accommodate these additional restrictions, we cast the non-linear least square problem in terms of the generalized method of moments (GMM) via its first order conditions. To be specific, the first order conditions of the non-linear least squares (2.9) are,

$$E\Bigg[\nabla_\beta f(\mathbf{w}_{jt};\beta)\Big(r_{jt} - f(\mathbf{w}_{jt};\beta)\Big)\Bigg] = 0,$$

where $\nabla_\beta f(\mathbf{w}_{jt};\beta)$ is the $D \times 1$ vector of partial derivatives with respective to β.

The GMM framework allows us to easily add additional restrictions in a manner similar to Wooldridge (2009). For the CES, these restrictions are related to aggregate measures and do not involve any additional assumptions. For the translog, we rely on the additional assumption that productivity moves according to a Markov process to provide moment restrictions with which to identify the remaining parameters. This second approach is quite general and can provide identifying restrictions for many functional forms (including the CES, if it were necessary). In both cases, these restrictions can be imposed in terms of moment conditions:

$$E[\mathbf{h}(\mathbf{x}_{jt};\beta)] = 0,$$

where $\mathbf{h}(\mathbf{x}_{jt};\beta)$ is a $S \times 1$ dimension function regarding observable exogenous variables $\mathbf{x}_{jt}$ (which may include $\mathbf{w}_{jt}$) and the parameter vector β. Define $\Phi(\beta)$ as a $D \times D$ Hessian matrix,

$$\Phi(\beta) = E\Bigg[\Big(\nabla_\beta f(\mathbf{w}_{jt};\beta)\Big)\Big(\nabla_\beta f(\mathbf{w}_{jt};\beta)\Big)'\Bigg],$$

and define $\Psi(\beta)$ as a $S \times D$ matrix,

$$\Psi(\beta) = E\Big[\nabla_\beta \mathbf{h}(\mathbf{x}_{jt};\beta)\Big].$$

Finally, define the $(D+S) \times D$ matrix $V(\beta) = [\Phi(\beta);\Psi(\beta)]$, we can now provide conditions for identification of β^0.

Proposition 2. *Suppose there exists an open neighborhood of $\beta_0 \in \Gamma$ in which both $\Phi(\beta)$ and $\Psi(\beta)$ have a constant rank. Then β_0 is locally identifiable if and only if $V(\beta_0)$ has rank D.*[17]

The proof of this proposition, which is established in B, is a direct application of Theorem 2 in Rothenberg (1971). Komunjer (2012) provides conditions for global identification in the context of non-linear moment equalities models, of which our model is a special case. Identification clearly relies on the structural information provided through firms first order conditions. In recent work, Gandhi, Navarro, and Rivers (2013) have established nonparametric identification of production functions when input prices are assumed to be homogeneous. Under heterogenous input prices, it is difficult to recover the unobserved M_{jt} from (2.2) without a parametric form of the production function. Moreover, the issue of multiple possible materials quantities satisfying (2.2) becomes more severe, leading to the possibility of partial identification.

With identification conditions established, we can estimate all parameters via GMM:

$$\hat{\beta} = \text{argmin}_\beta \left[\frac{1}{n} \sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta) \right]' W \left[\frac{1}{n} \sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta) \right], \tag{2.10}$$

where $\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta) = [\nabla_\beta f(\mathbf{w}_{jt}; \beta)(r_{jt} - f(\mathbf{w}_{jt}; \beta)); \mathbf{h}(\mathbf{x}_{jt}; \beta)]$ and W is a positive semi-definite weight matrix. When the problem is over-identified, we use two-step GMM to obtain the optimal weight matrix. B discusses consistency and the asymptotic distribution of this estimator.

The following two subsections illustrate how to implement the estimation with additional restrictions for two common production function specifications: CES and translog production functions.

2.3.4 Implementation with CES specification

As presented in the examples above, we consider a CES production function with an elasticity of substitution σ.[18] It has been commonly recognized that the CES production function needs to be normalized to give meaningful identification of its parameters. A branch of the literature has analyzed the importance and the method of normalization (de La Grandville, 1989; Klump and de La Grandville, 2000; Klump and Preissler, 2000; de La Grandville and Solow, 2006; Leon-Ledesma,

[17] Local identification is defined in Rothenberg (1971).

[18] We follow the literature in assuming constant returns to scale in this specification. This assumption can be relaxed by adding a scale parameter. This does not affect the estimation procedure but make the scale parameter and demand elasticity not separately identified. However, if Markov process of productivity is assumed, then one can easily (using (45) in Appendix D) separately identify these two parameters using the variation in industrial-level output and prices.

McAdam, and Willman, 2010). We follow this literature and normalize the CES production function according to the geometric mean.[19] Specifically, let the baseline point for our normalization be the geometric mean of $(Q_{jt}, L_{jt}, M_{jt}, K_{jt})$, denoted as $\overline{Z} = (\overline{Q}, \overline{L}, \overline{M}, \overline{K})$ where $\overline{X} = \sqrt[n]{X_1 X_2 \cdots X_n}$.[20] Then the normalized CES production function can be written as,

$$Q_{jt} = e^{\omega_{jt}} \overline{Q} \left[\alpha_L \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} + \alpha_M \left(\frac{M_{jt}}{\overline{M}} \right)^{\gamma} + \alpha_K \left(\frac{K_{jt}}{\overline{K}} \right)^{\gamma} \right]^{\frac{1}{\gamma}}, \tag{2.11}$$

where $\gamma = \frac{\sigma-1}{\sigma}$ and $\alpha_L, \alpha_M, \alpha_K$ are the distribution parameters, which sum to 1, $\alpha_L + \alpha_M + \alpha_K = 1$. The normalization has three advantages for our purposes. First, it scales the level of inputs according to an industry average, eliminating the effect of units on the parameters. Second, the geometric mean of capital and labor $(\bar{K}, \bar{L})$ are computable using the observed data, and will be convenient to use in constructing an additional restriction to identify the distribution parameters.[21] Third, this scaling gives the distribution parameters a precise interpretation. Specifically, they are the marginal return to inputs (in normalized units) for a firm with the geometric mean level of inputs, productivity, and input prices.[22]

Since analysts typically observe revenues rather than output quantities, we follow Klette and Griliches (1996) and also specify a classic Dixit-Stiglitz demand function,

$$\frac{Q_{jt}}{Q_t} = \left(\frac{P_{jt}}{P_t} \right)^{\eta}, \tag{2.12}$$

where Q_t and P_t are industry-level output quantity and price in period t, and η is the demand elasticity. As discussed earlier, the CES production function satisfies the condition of Proposition 1 if $\sigma \neq 1$ (i.e., $\gamma \neq 0$).

Given our specification for the production function (2.11) and demand function (2.12), we follow our proposed procedure from the previous section to derive the estimating revenue equation,[23]

$$\ln R_{jt} = \ln \frac{\eta}{1+\eta} + \ln \left[E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^{\gamma} \right) \right] + u_{jt}. \tag{2.13}$$

It is easy to see that while (2.13) provides identification of the elasticity of sub-

[19]For the details of this normalization and how we implement it in this paper, see C.

[20]In principle, any point $Z_0 = (Q_0, L_0, M_0, K_0)$ (which satisfies normalization conditions, i.e., (39)-(41) in C) can be chosen as the baseline point, for example a default choice could be $(1, \ldots, 1)$. The entire CES production function is identified up to the knowledge of the baseline point.

[21]Of course, $\bar{M}$ is not computable using the observed data, since we do not observe M_{jt} for any firm. The implication of this is that we will recover materials usage relative to the geometric mean $(M_{jt}/\bar{M})$ instead of materials directly (M_{jt}).

[22]Note that the normalized input of the baseline point is simply $(1, 1, 1)$.

[23]See D for the complete derivation.

stitution and the slope of the demand curve, it does not identify the distribution parameters. This is due to the substitution of our structural equation for the unobserved materials inputs. Fortunately, two additional restrictions allow us to identify the distribution parameters. The first is simply the adding up constraint of the distribution parameters, the second is implied by the first order conditions of profit maximization. To see this, note that the first order conditions for each firm j at period t are

$$\lambda\overline{\alpha}_L e^{\omega_{jt}}\overline{Q}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}}\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma-1}\frac{1}{\overline{L}}=P_{L_{jt}},$$

and

$$\lambda\overline{\alpha}_M e^{\omega_{jt}}\overline{Q}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}}\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma-1}\frac{1}{\overline{M}}=P_{M_{jt}},$$

where λ is the lagrangian multiplier. The ratio of the two equations yields,

$$\frac{\alpha_L(L_{jt}/\overline{L})^{\gamma}}{\alpha_M(M_{jt}/\overline{M})^{\gamma}}=\frac{P_{L_{jt}}L_{jt}}{P_{M_{jt}}M_{jt}}\equiv\frac{E_{L_{jt}}}{E_{M_{jt}}}. \tag{2.14}$$

This equation holds for each firm j at each period t. Taking the geometric mean of (2.14) across all observations implies,[24]

$$\frac{\alpha_M}{\alpha_L}=\frac{\overline{\overline{E}}_M}{\overline{\overline{E}}_L}, \tag{2.15}$$

where $\overline{\overline{E}}_M$ and $\overline{\overline{E}}_L$ are the geometric mean of $E_{M_{jt}}$ and $E_{L_{jt}}$ respectively. Because expenditures on materials and labor are observed in the data for all observations, the right hand side of this restriction can be directly computed.

Therefore, the model can be estimated via the following nonlinear least square estimation with restrictions:

$$\begin{aligned}
\hat{\beta} &= \operatorname{argmin}_{\beta}\sum_{jt}\left[\ln R_{jt}-\ln\frac{\eta}{1+\eta}-\ln\left\{E_{M_{jt}}+E_{L_{jt}}\left(1+\frac{\alpha_K}{\alpha_L}\left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}}\right)^{\gamma}\right)\right\}\right]^2\\
&\quad \text{subject to} \quad \frac{\alpha_M}{\alpha_L}=\frac{\overline{\overline{E}}_M}{\overline{\overline{E}}_L}, && (2.16)\\
&\quad \alpha_L+\alpha_M+\alpha_K=1, && (2.17)
\end{aligned}$$

where $\beta=(\eta,\alpha_L,\alpha_M,\alpha_K,\gamma)$.

Alternatively, the problem can be cast in a GMM framework as in (2.10). Write

[24] Recall that the geometric mean of a ratio is the ratio of geometric means.

the nonlinear equation (2.13) as $r_{jt} = f(\mathbf{w}_{jt};\beta) + u_{jt}$, where $f(\mathbf{w}_{jt};\beta)$ is the right hand side of (2.13) without u_{jt}. The restrictions (2.16) and (2.17) can be viewed as degenerate moment restrictions:

$$E\left[\mathbf{h}(\mathbf{x}_{jt};\beta)\right] \equiv E\begin{bmatrix}\overline{E}_M\alpha_L - \overline{E}_L\alpha_M \\ \alpha_L + \alpha_M + \alpha_K - 1\end{bmatrix} = 0.$$

We show in B that $V(\beta)$ has full column rank. Thus, we can estimate the parameters via GMM:

$$\hat{\beta} = \text{argmin}_\beta \left[\frac{1}{n}\sum_{jt}\mathbf{m}(\mathbf{w}_{jt},\mathbf{x}_{jt};\beta)\right]' W \left[\frac{1}{n}\sum_{jt}\mathbf{m}(\mathbf{w}_{jt},\mathbf{x}_{jt};\beta)\right], \tag{2.18}$$

where $\mathbf{m}(\mathbf{w}_{jt},\mathbf{x}_{jt};\beta) = [\nabla_\beta f(\mathbf{w}_{jt};\beta)(r_{jt} - f(\mathbf{w}_{jt};\beta)); \mathbf{h}(\mathbf{x}_{jt};\beta)]$ and W is a positive semi-definite weight matrix. Since the model is just-identified in the CES specification, the GMM implementation is equivalent to the nonlinear least square estimation with constraints.[25] With all of the production function parameters estimated we can solve for each firm's productivity level from (2.7), which is reduced to (45) in Appendix D.

2.3.5 Implementation with Translog specification

We now turn to the translog specification, which was introduced in (2.4). As with the CES implementation, we assume demand follows a Dixit-Stiglitz specification (2.12). Again, not all parameters of the production function are identified by the revenue equation. However we can use time series assumptions as additional restrictions to identify the remaining parameters. In particular, we follow Olley and Pakes (1996) to assume that firm productivity follows first order Markov process in order to generate the needed moment restrictions.

Example 3.3 shows how to recover materials. We can insert (2.5) into one of the first order conditions to uniquely recover the unobserved productivity ω^*_{jt},

$$\omega^*_{jt} = \frac{1}{1+1/\eta}[-\ln(1+1/\eta) + \ln P_{Ljt} - \ln F_{Ljt} - 1/\eta \ln F_{jt}]. \tag{2.19}$$

That is, productivity can be written as a known function of $(L_{jt}, K_{jt}, E_{Ljt}, E_{Mjt})$ up to parameters (θ, η), since $\ln F_{Ljt}$ and $\ln F_{jt}$ are functions of these variables.

Substituting both (2.5) and (2.19) into the log revenue function yields our

[25] Of course, if more restrictions were available and necessary, then the model is over identified and we can still cast the estimation in terms of GMM as proposed in (2.10).

estimating equation,

$$
\begin{aligned}
\ln R_{jt} &= -\ln(1+1/\eta) + \ln P_{Ljt} - \ln F_{Ljt} + \ln F_{jt} + u_{jt} \\
&= -\ln\left(1+\frac{1}{\eta}\right) + \ln\left(E_{M_{jt}} - \frac{\alpha_{mm}}{\alpha_{ml}} E_{L_{jt}}\right) \\
&\quad - \ln\left[\left(\alpha_m - \alpha_l \frac{\alpha_{mm}}{\alpha_{ml}}\right) + \left(\alpha_{km} - \alpha_{kl}\frac{\alpha_{mm}}{\alpha_{ml}}\right)\ln K_{jt}\right. \\
&\quad \left. + \left(\alpha_{lm} - \alpha_{ll}\frac{\alpha_{mm}}{\alpha_{ml}}\right)\ln L_{jt}\right] + u_{jt},
\end{aligned}
$$

which can be rewritten, in the notation of the general model (2.10), as

$$
r_{jt} = f(\mathbf{w}_{jt}; \beta) + u_{jt}, \tag{2.20}
$$

where β is the vector of parameters, including η and all α's.

It is clear that only nonlinear combinations of production and demand parameters are identified from the revenue equation. Moreover, α_k, α_{mk}, and α_{kk}, are canceled when computing $\ln F_{jt} - \ln F_{Ljt}$. Therefore, we need additional restrictions to help identify all production and demand parameters separately, either from the cross section restriction or the time series restriction. With the translog specification, cross sections restriction are not easy to find. Instead we follow Olley and Pakes (1996) and Doraszelski and Jaumandreu (2013) in using time series restrictions on productivity to help identify remaining parameters.[26] In particular, we assume that productivity follows a first order Markov process,

$$
\omega_{jt} = g(\omega_{jt-1}) + \epsilon_{jt}, \tag{2.21}
$$

where ϵ_{jt}, the productivity innovation, is independent of capital as well as variable labor and material input at time $t-1$. Given β, we can calculate ω^*_{jt} from (2.19) and estimate $\hat{g}$ from (2.21).[27] Then, we can define,

$$
\epsilon_{jt}(\mathbf{w}_{jt}; \beta) = \omega^*_{jt} - \hat{g}(\omega^*_{jt-1}).
$$

By the assumption on the productivity innovation process, $\epsilon_{jt}(\mathbf{w}_{jt}; \beta)$ must be uncorrelated with the firms information set at time $t-1$. This provides additional restrictions that, together with the revenue equation (2.20), allow us to identify all the parameters. Let $\mathbf{x}_{jt}$ be a vector of instruments which are independent of ϵ_{jt}, and define $\mathbf{h}(\mathbf{x}_{jt}; \beta) = \epsilon_{jt}(\mathbf{w}_{jt}; \beta)\mathbf{x}_{jt}$. Thus we can construct a set of moment

[26] From (2.19), we know that α_k, α_{mk}, and α_{kk} are not cancelled in ω^*_{jt}.

[27] In principle this can be a parametric or non-parametric regression, depending on the assumptions on g. If it is parametric, then we could easily incorporate this estimation into our GMM approach and estimate β and the parameters of g in a single step.

conditions, $E[\mathbf{h}(\mathbf{x}_{jt};\beta)] = 0$. The dimension of $\mathbf{x}_{jt}$ should be large enough so that $V(\beta)$ has full column rank and all parameters are identified.[28]

Following the general model (2.10), define the set of moment conditions as, $E[\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt};\beta)] = 0$, where

$$\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt};\beta) = [\nabla_\beta f(\mathbf{w}_{jt};\beta)(r_{jt} - f(\mathbf{w}_{jt};\beta)); \epsilon_{jt}(\mathbf{w}_{jt};\beta)\mathbf{x}_{jt}].$$

We estimate all parameters via the following GMM sample analogue,

$$\hat{\beta} = \text{argmin}_\beta \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt};\beta)\right]' W \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt};\beta)\right].$$

While we use the translog as a motivating example, this procedure can be directly applied to more general (parametric) production functions as long as the productivity shock is additively separable from F, and can be assumed to follow a Markov process.

2.4 Multiple Materials Inputs

So far, we have followed the literature in assuming that firms purchase a single homogeneous intermediate input. Indeed, the ability to treat the imputed firm-specific price and quantity choices as quality-adjusted scalars representing a single homogenous input is critical since our demand specification assumes that outputs are horizontally differentiated.[29] In reality, intermediate input expenditures are an aggregate of a wide variety of different input goods. Ideally, an analyst would be able to account for each of these goods separately in the production function. Unfortunately, datasets typically contain only total input expenditure, not information on the various types used, much less prices and quantities for each. With such limited data, it is clearly not possible to learn the impact of individual inputs. However, if the effect of inputs on production can be summarized through a homogeneous materials index function, we show that the remaining production function parameters can consistently be recovered using only total expenditure information.

To be specific, suppose the firm may use up to D_M different types of materials. Denote the vector of materials quantities used in production as $M_{jt} = (M_{1jt}, M_{2jt}, \ldots, M_{D_M jt})$. These input types may be entirely different input goods

[28] A valid set of instruments could be

$$\mathbf{x}_{jt+1} = \left(\ln\left(\frac{X_{jt}}{\overline{X}}\right), \left(\ln\left(\frac{X_{jt}}{\overline{X}}\right)\right)^2\right)$$

where $X = L_{jt}, K_{jt}, E_{Mjt}$. See B for additional details.

[29] We thank an anonymous referee for making this point.

(thread versus fabric) or the same input good of different quality (cotton versus polyester fabric). However, only the total expenditure on all components $E_{M_{jt}} = \sum_{d=1}^{D_M} P_{M_{djt}} M_{djt}$, rather than each specific component M_{djt}, is known to the econometrician. Assume inputs enter into the production function as,

$$Q_{jt} = e^{\omega_{jt}} F(L_{jt}, \mu(M_{jt}), K_{jt}; \theta), \tag{2.22}$$

where $\mu : \mathbb{R}^{D_M} \to \mathbb{R}$ is a homogeneous index function which summarizes the contribution of all materials inputs to production.[30] As part of the production function, we assume that μ is known to the firm. While this structure allows materials to substitute for each other in an unknown manner, it does restrict the substitution patterns between materials and other production inputs, namely labor and capital. As a result, we can allow for vertically or horizontally differentiated materials and treat them as different elements of the materials vector in M_{jt}. The corresponding idiosyncratic material prices for each component is summarized in price vector $P_{Mjt} = (P_{M1jt}, P_{M2jt}, \ldots, P_{MD_Mjt})$, which is observed by firms but not by researchers.

The firm's static optimization problem is now to choose L_{jt} and the vector M_{jt} to maximize the profit given productivity and capital stock:

$$\begin{aligned} &\max_{L_{jt}, M_{jt}} P_t(Q_{jt}; \eta) Q_{jt} - P_{Ljt} L_{jt} - P_{M_{jt}} M_{jt} \\ &\text{s.t. } Q_{jt} = \exp(\omega_{jt}) F(L_{jt}, \mu(M_{jt}), K_{jt}; \theta). \end{aligned}$$

The first order conditions for L_{jt} and all components of vector M_{jt} are:

$$\exp(\omega_{jt}) F_{L_{jt}} \left[P_t(Q_{jt}; \eta) + Q_{jt} \frac{\partial P_t(Q_{jt}; \eta)}{\partial Q_{jt}} \right] = P_{L_{jt}}, \tag{2.23}$$

$$\exp(\omega_{jt}) F_{\mu_{jt}} \left[P_t(Q_{jt}; \eta) + Q_{jt} \frac{\partial P_t(Q_{jt}; \eta)}{\partial Q_{jt}} \right] \mu_d(M_{jt}) = P_{Md_{jt}}, \tag{2.24}$$

for $\forall d = 1, 2, \ldots, D_M$, where $\mu_d(M_{jt}) = \frac{\partial \mu(M_{jt})}{\partial M_{djt}}$.

Denote the optimal choice of the firm as L^*_{jt} and vector M^*_{jt}. Thus the total expenditure on materials, which is observed by the researcher, is $E^*_{M_{jt}} = \sum_{d=1}^{D_M} P_{Mdjt} M^*_{djt}$. Define the material price index as $P_{\mu jt} = \frac{E^*_{M_{jt}}}{\psi(M^*_{jt})}$, where $\psi(M^*_{jt}) = \sum_{d=1}^{D_M} M^*_{djt} \mu_d(M^*_{jt})$. Using this price index, the information in (2.24) can be summarized into a single equation by multiplying (2.24) by M^*_{djt}, summing them up

[30] In order for Assumption 1 to continue to hold, we assume that μ is differentiable in M_{jt}. However we are able to allow for some non-differentiability in μ, as we will see in the Monte Carlo experiment in Section 2.5.6.

across d, and dividing it by $\psi(M^*_{jt})$,

$$\exp(\omega_{jt})F_{\mu_{jt}}\left[P_t(Q_{jt};\eta)+Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right]=P_{\mu_{jt}}. \tag{2.25}$$

This equation together with (2.23) can be viewed as the first order conditions of the firm's optimization problem if it faced labor price $P_{L_{jt}}$ and a materials price $P_{\mu_{jt}}$ for *single* material μ. The following proposition shows how our method can be adapted to an unknown vector of inputs.

Proposition 3. *Suppose the index function $\mu_{jt}=\mu(M_{jt}):\mathbb{R}^{D_M}\to\mathbb{R}$ is homogeneous of degree $\kappa>0$. Then given parameter θ, the firm's optimal choices of input quantities and expenditure satisfy the following equation:*

$$z(\mu_{jt},L_{jt},K_{jt},E_{Mjt},E_{Ljt};\theta,\kappa)\equiv\frac{F_{L_{jt}}L_{jt}}{F_{\mu_{jt}}\mu_{jt}}-\frac{E_{L_{jt}}}{\frac{1}{\kappa}E_{M_{jt}}}=0, \tag{2.26}$$

In addition, this equation admits a unique solution $\mu^*(L_{jt},K_{jt},E_{Mjt},E_{Ljt};\theta,\kappa)$ when the conditions of Proposition 1 hold for $z(\cdot)$ as defined in (2.26)*.*

The proof of this proposition is provided in A and directly follows the Euler's Theorem for homogeneous functions and Proposition 1. We can now substitute μ^* into the revenue function as with M^* in Section 2.3, albeit with κ as an additional scale parameter. In some specifications (e.g., translog), κ may not be separately identified from the production function parameters. In this case κ can be scale normalized to 1 without loss of generality, since it is absorbed in the primary parameters of the production function. In other cases (e.g., CES), κ can be identified through the revenue function, where it represents returns to scale of the materials aggregator index μ.[31] In this case, estimation still follows our method from Section 2.3, except that now we substitute the material expenditure with $\frac{1}{\kappa}E_{M_{jt}}$ and estimate κ as an additional parameter of the production function. For example, in the CES specification of Section 2.3.4, we can employ the following revenue equation:[32]

$$\ln R_{jt}=\ln\frac{\eta}{1+\eta}+\ln\left[\frac{1}{\kappa}E_{M_{jt}}+E_{L_{jt}}\left(1+\frac{\alpha_K}{\alpha_L}\left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}}\right)^{\gamma}\right)\right]+u_{jt}.$$

[31]Whether or not κ is identified depends on whether or not the production function already accounts for scale effects on materials independent of other inputs. That is, suppose the production function is $F(L,M,K;\theta)$ and consider the alternative $\tilde{F}(L,M,K;\kappa,\theta)=F(L,M^{\kappa},K;\theta)$. In $\tilde{F}(\cdot)$, κ may or may not be identified depending on the form of F. If κ is not identified, normalizing $\kappa=1$ simply returns the researcher to the original specification. For theoretical reasons, some researchers may still want to impose that $\kappa=1$ even when it is formally identified. This would be essentially equivalent to assuming constant returns to scale in the materials aggregator μ.

[32]We conducted a Monte Carlo experiment to verify this result with $\mu(M_{jt})=M^{\kappa}_{jt}$, and our method works very well in recovering the primary parameter θ as well as κ. Results are available upon request.

In sum, our method still works if the effect of material inputs on production can be summarized by a homogeneous materials index function.[33] As we would expect, the functional form of $\mu(\cdot)$ is not identified without more information, but its functional form (indeed, even its dimension) is not needed to recover the other production parameters, θ.

Although we assume that $\mu(\cdot)$ is homogeneous to make our method work, this still allows a vast set of flexible functional forms that may incorporate both vertically and horizontally differentiated materials inputs. We verify the validity of this approach through a Monte Carlo experiment in Section 2.5.6.

2.5 Monte Carlo Experiment

This section presents Monte Carlo experiments that evaluate the performance of our method, and show how it corrects for input price heterogeneity. For brevity, we concentrate on the CES version of the estimator. We first describe the data generation process, then estimate the model in three different ways based on assumed data availability. At the end of this section, we examine the performance of our method when the firm chooses from a vector of inputs.

2.5.1 Data Generation

Using the CES specification of the production function (2.11) and a Dixit-Siglitz demand system (2.12), we generate N replications of simulated data sets, given a set of true parameters of interest (η, σ, α_L, α_M and α_K). In each replication, there are J firms in production for T periods. We simulate a sequence of productivity for each firm (ω_{jt}) from an AR(1) process. We allow the firm's capital stock (K_{jt}) to evolve based on an investment rule (investment is denoted as I_{jt}) that depends on its productivity and capital stock,

$$\log(I_{jt}) = \xi\omega_{jt} + (1-\xi)\log(K_{jt}).$$

Finally, we allow input prices ($P_{L_{jt}}$ and $P_{M_{jt}}$) to vary across firms and over time. Table 1 lists the underlying parameters used to generate the data set. A full description of the data generating process is provided in E.

Given these variables and industrial-level outputs and prices (Q_t and P_t), we derive a sequence of optimal choices of labor and material inputs (L_{jt} and M_{jt} with corresponding input expenditures $E_{L_{jt}}, E_{M_{jt}}$), the optimal output quantity (Q_{jt}), price (P_{jt}) and revenue (R_{jt}) for firm j in each period t. In this way, we generate

[33] If $\mu(\cdot)$ is not homogenous, then the "total material expenditure" implied by (2.25) together with (2.23) will be $\mu(M^*_{jt})P_{\mu_{jt}} = \frac{\mu(M^*_{jt})}{\psi(M^*_{jt})}E_{M^*_{jt}}$ which is not observable (since generally $\frac{\mu(M^*_{jt})}{\psi(M^*_{jt})}$ is not a constant). In this case, information on expenditure alone is insufficient to control for variation in materials inputs even if prices are homogenous across firms.

a data set of $\{\omega_{jt}, K_{jt}, I_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, R_{jt}, Q_t, P_t\}$ for each firm j and period t. All these variables are observable to firms, however, usually only a subset of them are available to researchers.

2.5.2 Our Method

We first estimate the model with our method. In this case, we assume a researcher observes $\{K_{jt}, L_{jt}, E_{L_{jt}}, E_{M_{jt}}, R_{jt}, P_t, Q_t\}$ for each firm and each period. The researcher is not required to observe firm's investment, material input quantity, physical outputs quantity or, of course, productivity. As described in the previous section, we exploit the first order conditions to impute firm-level material quantities from labor quantities and expenditures. This approach allows us to estimate the production function while controlling for unobserved price dispersion. We will evaluate our method by comparing our estimates with the true parameters, as well as with those derived from two alternative estimation methods that require additional data.

2.5.3 Traditional Method with Direct Proxy

For our first point of comparison, we estimate the model using a direct proxy method that is substitutes $E_{M_{jt}}$ for M_{jt}. The method follows Olley and Pakes (1996) in using a control function approach to utilize investment data to control for unobserved productivity. Traditionally, researchers have used deflated expenditure on materials inputs to proxy for intermediate input quantities when applying this and similar methods (e.g., Levinsohn and Petrin, 2003), and we follow that practice here. We will refer to this method as the "OP" procedure, although we should emphasize that it is the direct proxy, rather than the OP procedure, that is introducing the bias. In contrast with our own method, the OP procedure takes output quantities as observable. Hence there will be no output price bias and any resultant bias is caused by the substitution of physical material input by its deflated cost.[34]

Specifically, we assume that researchers using this method can observe a set of data $\{K_{jt}, L_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, I_{jt}, P_t, Q_t\}$ and estimate parameters via the (logarithm) production function:

$$\ln\left(\frac{Q_{jt}}{\overline{Q}}\right) = \left\{\omega_{jt} + \frac{1}{\gamma}\ln\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma} + \alpha_M\left(\frac{E_{M_{jt}}}{\overline{E}_M}\right)^{\gamma} + \alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]\right\} + u_{jt}, \tag{2.27}$$

where the error term u_{jt} accounts for measurement error of revenue and productivity shocks that are unanticipated by the firm.

[34] We could easily incorporate a revenue function into this procedure. We do not in order to emphasize that the direct proxy is the cause of the resulting bias.

To control for endogeneity bias, Olley and Pakes (1996) assume that productivity follows a first order Markov process. Following our data generating process, we are more specific and assume that productivity follows an AR(1) specification,

$$\omega_{jt+1} = g_0 + g_1\omega_{jt} + \epsilon_{jt+1}.$$

Since the true data generating process is in fact AR(1), this rules out specification error associated with the productivity evolution process, so that the Monte Carlo focuses on the bias caused by dispersion in input prices. Within our data generating process, the investment decision is a function of current capital stock and the unobservable heterogenous productivity and therefore, the OP method can approximate the productivity by a control function of investment and capital stock: $\omega_{jt} = \omega_t(I_{jt}, K_{jt})$. Substituting it into (2.27) yields,[35]

$$\ln\left(\frac{Q_{jt}}{\overline{Q}}\right) = \phi(L_{jt}, E_{M_{jt}}, K_{jt}, I_{jt}, \Phi_t) + u_{jt},$$

where Φ_t represents time dummies to capture aggregate investment shifters. This equation can be estimated non-parametrically. This estimation is consistent since the right-hand-side variables are all uncorrelated with u_{jt}. We estimate ϕ using the method of sieves.[36] Denote $\hat{\phi}_{jt}$ as the fitted value of $\phi(L_{jt}, E_{M_{jt}}, K_{jt}, I_{jt}, \Phi_t)$. Then productivity can be expressed as,

$$\omega_{jt} = \hat{\phi}_{jt} - \frac{1}{\gamma}\ln\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma} + \alpha_M\left(\frac{E_{M_{jt}}}{\overline{E}_M}\right)^{\gamma} + \alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right].$$

Substituting ω_{jt+1} and ω_{jt} into the evolution process of productivity, we obtain,

$$\epsilon_{jt+1} = \omega_{jt+1} - (g_0 + g_1\omega_{jt}).$$

Note that ϵ_{jt+1} is uncorrelated with variables up to period t, so we can construct the set of moment conditions with which we estimate the model's parameters,

$$E(\epsilon_{jt+1}\mathbf{x}_{jt+1}) = 0, \tag{2.28}$$

where $\epsilon_{jt+1}(\eta, \tau, \gamma, g_0, g_1) = \omega_{jt+1} - (g_0 + g_1\omega_{jt})$, and $\mathbf{x}_{jt+1}$ is a combination of variables that are uncorrelated with the innovation term in period $t+1$, e.g.,

[35] In contrast to the original OP paper, we follow Ackerberg, Caves, and Frazer (2006) in recovering the labor and materials parameter out of the second stage of the OP estimation to avoid collinearity issues in the first stage.

[36] In practice, we model $\phi(\cdot)$ with a cubic function with interactions.

$L_{jt}, E_{L_{jt}}, E_{M_{jt}}, K_{jt}, K_{jt+1}$.[37]

2.5.4 Oracle-OP Procedure

Finally, we compare our method to a first-best case when input quantities are actually observed. We refer to this as the "Oracle-OP" case as it uses the Olley and Pakes (1996) inversion to recover productivity but uses the actual materials input quantities instead of a proxy. That is, we observe a full data set which consists of $\{K_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, I_{jt}, P_t, Q_t\}$ for each firm and each period. This enables us to estimate the production function in (2.27) without using expenditure as a proxy. The only difference between the oracle case and previous OP's procedure is that material quantity is not substituted by its proxy, since the true quantity is observable in this case. In comparison to our own method, this method requires that the researcher observes investment, output quantity, and materials input quantities.

2.5.5 Results

The results of the Monte Carlo experiments for three separate elasticities of substitution are presented in Table 2.2. For each method, the listed parameter represents the median estimate of the 1000 Monte Carlo replications with standard errors in parenthesis. The square brackets contain the root mean squared error of the estimates. Across all parameterizations, our method recovers the parameters well. In contrast, the elasticity of substitution (σ) and α_K are severely underestimated by OP. This corresponds to the biases due to input price dispersion as documented in Section 2.2. The results for the oracle-OP method confirm that when input price heterogeneity is observed, the bias is eliminated. Interestingly, it appears there is little loss in efficiency between the oracle-OP method and the method we propose, despite the fact that we do not use investment, output quantity, or input quantities. Of course, our method makes use of the additional structure implied by the firm's first order conditions, which is not used within the OP framework.

To further investigate the performance of the estimators, Figure 1 plots the density of $\hat{\sigma}$ for the three cases. The dashed line represents the true value of σ. Clearly, our method generates estimates that are concentrated around the true elasticity of substitution. However, the proxy-OP method produces biased estimates of σ. This bias is economically significant, implying an elasticity of substitution up to 20% lower than the true value. As expected, when we allow the researcher to observe

[37] In the Monte Carlo experiment, we choose

$$\mathbf{x}_{jt+1} = \left(\ln\left(\frac{X_{jt}}{\bar{X}}\right), \ln\left(\frac{K_{jt+1}}{\bar{K}}\right), \left(\ln\left(\frac{X_{jt}}{\bar{X}}\right)\right)^2, \left(\ln\left(\frac{K_{jt+1}}{\bar{K}}\right)\right)^2\right)$$

where $X = L_{jt}, K_{jt}, E_{Ljt}, E_{Mjt}$, to serve as instruments.

input quantities directly, the oracle-OP method performs well. The validation of the expenditure proxy requires no heterogeneity in input prices across firms. When heterogeneity is present, the unobserved input price dispersion will bias the estimation. The advantage of our method is that it provides a consistent way of imputing unobservable input quantities from observable expenditures of both materials and labor, which is necessary to control for the input price heterogeneity.

In addition to controlling for input price dispersion, our method allows the researcher to recover estimates of the unobserved input prices. In short, material quantities and prices can be imputed from (44) (in Appendix C). Figure 2.2 presents the kernel density estimation of the imputed material prices from our method and compares it to the true density of material prices in the Monte Carlo.[38] It shows that the imputed material price density matches the true density quite well.

2.5.6 Multiple Materials Inputs

The earlier Monte Carlo experiments assumed a scalar materials input. Here, we allow firms to endogenously select a vector of inputs, as in Section 2.4, to provide an illustration of how our method works under these conditions.

For consistency, we again use the basic CES formulation, and only emphasize the introduction of multiple materials here. Specifically, in addition to labor and capital stock, firms may choose to use any combination of three components (M_1, M_2, M_3) of materials in production. They enter into the production function though the index function μ,[39]

$$\mu(M_{jt}) = \max\left(\left[(\delta M_{1jt})^{\gamma_1} + M_{3jt}^{\gamma_1}\right]^{1/\gamma_1}, \left[M_{2jt}^{\gamma_2} + M_{3jt}^{\gamma_2}\right]^{1/\gamma_2}\right), \tag{2.29}$$

where $\gamma_1 = \frac{\sigma_1 - 1}{\sigma_1}$ and $\gamma_2 = \frac{\sigma_2 - 1}{\sigma_2}$. The functional form of μ is observable to the firms but not to the econometrician. The experiment's basic structure is inspired by the following scenario:[40] M_1 and M_2 are vertically differentiated versions of the same type of input (e.g., two versions of the same part). They differ in their quality such that the efficiency for M_1 in the production process is $\delta < 1$, while the efficiency for M_2 is normalized to be one. They also differ in regard to their substitutability with the third component (σ_1 and σ_2 may not be equal). They are produced within an imperfectly competitive industry and their prices are P_{M_1} and P_{M_2} offered to all firms.[41] The functional firm of $\mu(M_{jt})$ implies that the firm

[38] We present the case for true $\sigma = 2.5$, but the results from other cases are very similar.

[39] Note that this function is homogenous of degree 1, as seems reasonable given our scenario.

[40] We thank a referee for suggesting a version of this Monte Carlo design.

[41] We have also experimented with allowing these prices to be heterogeneous and have also been able to successfully recover the production function parameters.

will optimally use either M_1 or M_2 since using both can provide no benefit over only purchasing one. The third component is a homogenous material M_3, with idiosyncratic price $P_{M_{3jt}}$. The production function is,

$$Q_{jt} = e^{\omega_{jt}}\overline{Q}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma} + \alpha_M\left(\frac{\mu(M_{jt})}{\overline{M}}\right)^{\gamma} + \alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}}. \tag{2.30}$$

The firm observes the price vector, $(P_{M_1}, P_{M_2}, P_{M_{3jt}})$, and optimally chooses its vector of inputs, $(M_{1jt}, M_{2jt}, M_{3jt})$. However, only total materials expenditure $E_M = \sum_{d=1}^{3} P_{M_{djt}} M_{djt}$ is observed by the researcher, who is attempting to recover the production function parameters $(\alpha_L, \alpha_M, \alpha_K, \gamma)$, as well as the distribution of $\mu(M_{jt})$, its price index $P_{\mu jt}$, and productivity ω_{jt}.

Importantly, due to the difference in their elasticities of substitution with M_3, the price of M_3 will affect the optimal decision to employ M_1 or M_2 in production (see Figure 2.3). Each firm has a cutoff point $\tilde{P}_{M_{3jt}}$, and the choice of quality levels depends on whether it faces a price for M_3 above or below this cutoff. To generate our data, we solve the optimization problem for each firm j and period t, and obtain the input demand $(M^*_{1jt}, M^*_{2jt}, M^*_{3jt}, L^*_{jt})$[42], which is then substituted into the demand function and the production function to calculate the other endogenous variables. Therefore, we have generated the entire data set of firm-level variables for each firm j and period t: $\{\omega_{jt}, K_{jt}, I_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, R_{jt}, Q_t, P_t\}$, where $M_{jt} = (M^*_{1jt}, M^*_{2jt}, M^*_{3jt})$. Then we estimate the model with our method only using data set on $\{K_{jt}, L_{jt}, E_{L_{jt}}, E_{M_{jt}}, R_{jt}, Q_t, P_t\}$.

Table 2.3 presents the result for N replications. As discussed earlier, the form of $\mu(M)$ is not identified, but we find that our method recovers the primary parameters (i.e., all α's, σ and η) very well. Also, the material quantity index and price index can be recovered. In Figure 2.4, we compare the recovered material quantity index $\hat{\mu}(M)$ and price index $\hat{P}_\mu$ with the true indexes $\mu(M)$ and P_μ. Both show that our method recovers the underlining material quantity index and price index accurately in this example.

2.6 Application: Colombian Data

To evaluate the performance of our estimator using real data, we apply our method to a dataset of Colombian manufacturing firms from 1981 to 1989, which was collected by the Departamento Administrativo Nacional de Estadistica (DANE).[43] This application serves two purposes. The first is to compare our results with those found using the traditional proxy method to account for unobserved materials inputs. The second is to illustrate additional information which can be recovered

[42] Either M^*_{1jt} or M^*_{2jt} should be zero because of the substitution between the two.
[43] For a detailed introduction to the data set, see Roberts and Tybout (1997).

using our method, including the distribution of input prices and their relationship to productivity.

This dataset contains detailed information about firm-level revenue (R), labor and material input expenditure (E_L and E_M), capital stock (K), employment (L), the wage bill (E_L), and investment (I). However, firm-level price information about input and output is not available. Moreover, only total expenditure on "raw materials, materials and packaging" is available (E_M) rather than total quantities (M). This includes expenditure on raw materials such as cloth and gasoline, but does not include consumption of electrical energy, "general expenses" such as professional services and advertising, or "industrial expenses" such as spare or replacement parts, all of which are reported separately. It is extremely common in the literature to treat materials as a homogenous input (e.g. Levinsohn and Petrin, 2003) and our approach can be interpreted as following this tradition. However, as shown in Section 2.4 it can also be employed if firms are optimally choosing from a vector of heterogenous inputs. In this case, the imputed input price represents the shadow price of increasing the use of inputs in production. This is important since material expenditure represents the sum of several different input types that may vary across firms even within an industry.[44]

First, we estimate the model using our method using the CES specification of the production function normalized at the geometric mean as illustrated in Section 2.3.4. As a primary basis of comparison, we also estimate the production function using materials expenditure as a proxy for materials inputs as in Olley and Pakes (1996). To focus on the impact of input price heterogeneity, we control for output price bias by incorporating a demand function in this approach, as suggested in Klette and Griliches (1996). Thus, the only difference between the two estimators is in their treatment of input quantities. We refer to the second method as OP-KG in the text and tables. Of course, there are many other approaches that may be used to estimate production functions. In Appendix F, we compare our method to several alternative approaches, including employing first order conditions to recover productivity while using the proxy approach for materials and following well-known panel data methods (Arellano and Bond, 1991).

Estimates for four large industries are displayed in Table 2.4: clothing, bakery products, printing and publishing, and metal furniture.[45] In all these industries, the estimate of the elasticity of substitution is significantly lower using the OP-KG method compared with the results from our method. This is consistent with both our intuition about the bias generated by unobserved price dispersion and the patten shown in the Monte Carlo experiments. Moreover, the elasticity of substitution estimates are significantly greater than one in all industries when using

[44] For simplicity, we normalize k (the degree of homogeneity of μ) to be 1 in the application.

[45] We have estimated the model for a wide variety of industries and found these results to be representative with respect to the performance of the estimators. Additional results are available by request.

our method. This implies that production function is not likely Cobb-Douglas in these industries. The results support the conclusion that ignoring input price dispersion would lead to inconsistent estimates of elasticities of substitution, and that our method is capable of controlling for unobserved price dispersion.

Biased estimates of elasticity of substitution (σ) using the OP-KG method will contaminate estimates of the distribution parameters. However, the direction of the bias is unclear. We find that our method produces estimates of α_K that are at least 30 percent larger, and sometimes more than twice as large, as the estimates of α_K using the OP-KG method. These results mirror the findings from the Monte Carlo study, where α_K is also underestimated by the traditional method. It appears that ignoring price dispersion is likely to lead researchers to underestimate the degree of capital intensity in production.

A key output from production function estimation is the implied productivity distribution of firms within an industry. We find that there are substantial differences in the estimates of this distribution between the two methods. Figure 2.5 shows the productivity distributions estimated using our method and the OP-KG method for each of the four industries.[46] For all industries, the productivity distribution in OP-KG is more concentrated than using our method to control for price dispersion. The result is most stark for the bakery products industry, where our implied distribution has an inter-quartile range that is 3.4 times as wide as that using the OP-KG method. But even in the clothing industry, where the two productivity distributions are most similar, our distribution has an inter-quartile range more than 60 percent larger than is found using OP-KG. This suggests that omitting the unobserved input price dispersion tends to underestimate the firm heterogeneity in productivity. One possible reason might be a positive correlation between input prices and productivity, which we report below. Intuitively, positive correlation between the productivity and input prices could bias productivity estimates since a firm with low productivity tends to use low-price materials. In the OP-KG method, where all firms are assumed to have the same material price, the total material quantity used by low-productivity firms is underestimated, resulting in overestimates of their productivity. Similarly, OP-KG would underestimate the productivity for high-productivity firms facing high prices. As a result, OP-KG, by not controlling for the unobserved input prices, would underestimate the degree of productivity dispersion within the industry. A large literature, recently reviewed by Syverson (2011), is devoted to understanding and explaining heterogeneity of productivity among firms.[47] Our finding indicates that the “true” productivity

[46]Figure 2.5 follows Olley and Pakes (1996) in defining productivity as the sum of ω_{it}, which is known to the firm when it chooses labor and materials, and u_{it} which is unanticipated productivity and measurement error. In F, we compare the distributions of $\omega_{it} + u_{it}$ with only anticipated productivity, ω_{it}. We find that for both methods the distributions are fairly similar, implying that the bulk of productivity dispersion is anticipated by firms.

[47]An earlier review of this literature is provided by Bartelsman and Doms (2000).

heterogeneity may be even larger than is indicated by estimations that fail to control for unobserved input price dispersion.

In addition to results on the production function and the distribution of productivity, our method also provides estimates of the unobserved input prices and quantities across firms. Because these prices and quantities are recovered from the first order condition, they reflect quality-adjusted quantity indices and the imputed prices are purged of the effect of quality differences. In Figure 2.6, we present the kernel density estimations of imputed material prices (in logarithm) from our method for each of the four industries pooled across all years. In all industries, the distributions of input prices are quite spread out, indicating that price dispersion is substantial. Our findings are partially corroborated by studies, such as Ornaghi (2006) and Atalay (2012), which observe input prices directly and also find significant dispersion. Since our input prices are quality adjusted they suggest that even after considering quality differences across inputs, which could account for significant dispersion when inputs are directly observed, there is still large dispersion in the input prices.

We are also able to use our method to analyze the dynamics of input price dispersion. While it is not assumed in our estimation, we would expect a significant amount of persistence in firms' input prices over time. To check this, we fit the input prices to a simple AR(1) process to analyze their persistence. Table 2.5 shows the estimated persistence with standard error. In all four industries, there is quite high persistence with mean around 0.75, which is close to the persistence reported in Atalay (2012) in which firm-level input prices and quantities are available. Thus, firms that are able to secure low prices today are likely to be able to secure them again in the future. This gives us some confidence that our imputed prices do not simply reflect estimation error, but are a persistent feature of firms.

Finally, we examine the joint relationship between input prices and productivity in our sample of firms. As shown in Table 5, the imputed input price is positively correlated with the recovered productivity. That is, higher productivity firms tend to pay higher input prices. As mentioned above, this correlation is one reason why our method indicates a higher degree of productivity dispersion than we see in traditional methods that assume input prices are homogeneous. Table 5 also reports the correlation between input prices and observed wages, and again finds a positive correlation—high productivity firms pay more for both labor and materials. These results are consistent with Kugler and Verhoogen (2012), who directly examine data on input prices and compares them with productivity estimates. In explaining their result, Kugler and Verhoogen (2012) emphasize the quality complementarity hypothesis—input quality and plant productivity are complementary in generating output. However, because we recover the input prices using the marginal contribution of inputs in production, our recovered input price is quality-adjusted, ruling out the quality-complementarity explanation. Even so, we find a positive correlation between input prices and productivity. This indicates that alternative factors, such as plant-specific demand shocks or market power in input sectors, as discussed in Kugler and Verhoogen (2012), may also contribute to the dispersion of input prices within industries.

2.7 Conclusion

We analyze the problem of unobserved input prices and quantities in the estimation of production functions. Simply using expenditures as a proxy for quantities is likely to bias production function estimates in the presence of input price heterogeneity. To account for unobserved price dispersion, we introduce a method which exploits the first order conditions of profit maximization and imputes unobservable firm-level quantities of outputs and inputs from observable data on revenue and expenditures. We show how to apply our method using the commonly used CES production function specification.

To validate our method, we conduct Monte Carlo experiments to evaluate the performance of our estimation method. The results confirm that ignoring unobserved price dispersion biases the estimation when deflated values are used as proxies of quantities. In contrast, our method recovers the true parameters very well.

We further show that these differences matter in real data by applying the methods to a dataset on the Colombian manufacturing sector. The results are in line with theory and the Monte Carlo study. Specifically, the elasticity of substitution is significantly lower compared with our method when using the expenditure proxy. In addition, our results confirm the presence of unobserved price dispersion, and indicate that input prices and firm productivity are positively correlated. As a result, we find significantly more productivity dispersion in the industries we study than would be uncovered using a traditional estimator. The results suggest that our method of imputing unobservable firm-level input quantities from observable expenditures is important to effectively control for input price dispersion and consistently estimate the production function and the degree of dispersion in productivity.

Appendices

A Recovering M_{jt}

In this appendix, we provide proofs for Proposition 1 and Proposition 3 which show the conditions which allow us to recover materials price and quantities in the single material input and multiple materials input cases respectively.

Proposition 1 Define,

$$z(M_{jt}; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) \equiv \frac{F_{L_{jt}} L_{jt}}{F_{M_{jt}} M_{jt}} - \frac{E_{L_{jt}}}{E_{M_{jt}}},$$

For a given observation of $(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt})$ and parameter vector θ, suppose either $\frac{\partial z}{\partial M} > 0$ or $\frac{\partial z}{\partial M} < 0$ for all $M \in (0, \infty)$ such that $z(M; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0$. Then there exists a unique M^* that satisfies,

$$z(M^*; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0.$$

Proof. First, note that the existence of a solution to $z(\cdot)$ is implied Assumption 1, which guarantees that the first order conditions hold for some $(L, M) \in \mathbb{R}^2_+$. Uniqueness is shown by contradiction. Suppose that there are multiple M in the set $\mathcal{M} = \{M : z(M, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0\}$, and for all of them $\frac{\partial z}{\partial M} > 0$. Take any two consecutive solutions $M', M'' \in \mathcal{M}$ such that M'' is the smallest member of $\mathcal{M}$ that is larger than M'. Take a sequence converging to M' from above, we know there exist some m' such that $z(m', L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) > 0$. Likewise, take a sequence converging to M'' from below, we know there exist some m'' such that $z(m', L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) < 0$. Because $z(\cdot)$ is continuous in M, which is guaranteed by Assumption 1, there must be some $m^* \in (m', m'')$ such that $z(m^*, L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta) = 0$. However, this contradicts M' and M'' being consecutive members of $\mathcal{M}$. □

Proposition 3 (Multiple Inputs) Suppose the index function $\mu_{jt} = \mu(M_{jt}) : \mathbb{R}^{D_M} \to \mathbb{R}$ is homogeneous of degree $\kappa > 0$. Then given parameter θ, the firm's optimal choices of input quantities and expenditure satisfy the following equation:

$$z(\mu_{jt}; L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}, \theta, \kappa) \equiv \frac{F_{L_{jt}} L_{jt}}{F_{\mu_{jt}} \mu_{jt}} - \frac{E_{L_{jt}}}{\frac{1}{\kappa} E_{M_{jt}}} = 0. \tag{31}$$

In addition, this equation admits a unique solution $\mu^*(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt}; \theta, \kappa)$ when the conditions of Proposition 1 hold for $z(\cdot)$ as defined in (31).

Proof. Note that the firm's optimal choices of input quantities and expenditure satisfy the first order conditions (2.23) and (2.25) developed in the main body of the paper. Taking the ratio of the two equations produces

$$\frac{F_{L_{jt}}}{F_{\mu_{jt}}} = \frac{P_{L_{jt}}}{P_{\mu_{jt}}},$$

which implies

$$\frac{F_{L_{jt}} L_{jt}}{F_{\mu_{jt}} \mu_{jt}} = \frac{E_{L_{jt}}}{P_{\mu_{jt}} \mu_{jt}}, \tag{32}$$

Recall $P_{\mu_{jt}}$ is the material price index which is defined as

$$P_{\mu jt} = \frac{E_{M_{jt}}}{\sum_{d=1}^{D} M_{djt} \mu_d(M_{jt})}.$$

The Euler's Theorem for homogeneous functions implies that

$$\sum_{d=1}^{D} M_{djt} \mu_d(M_{jt}) = \kappa \mu(M_{jt}),$$

given $\mu(\cdot)$ is homogenous of degree κ. Therefore, $P_{\mu jt} = \frac{E_{M_{jt}}}{\kappa \mu(M_{jt})}$. That is, $P_{\mu_{jt}} \mu_{jt} = \frac{1}{\kappa} E_{M_{jt}}$. Substitute it into (32) gives (31).

Recovering the value of μ uniquely is a direct application of Proposition 1. Note that in the single input case, μ is just the identity function, which is homogeneous of degree 1. Thus in this special case we know $\kappa = 1$. □

B Identification and Asymptotics

In this appendix, we provide the proof of Proposition 2 the general identification condition for the model and discuss identification of the CES and Translog production function specifications. Finally, we present the asymptotic distribution of our GMM-based estimator.

Proposition 2 *Suppose there exists an open neighborhood of $\beta_0 \in \Gamma$ in which both $\Phi(\beta)$ and $\Psi(\beta)$ have a constant rank. Then β_0 is locally identifiable if and only if $V(\beta_0)$ has rank D.*

Proof. Let the true model is specified as, $r_{jt} = f(\mathbf{w}_{jt}; \beta_0) + u_{jt}$, where,

$$f(\mathbf{w}_{jt}; \beta_0) = \left\{ \ln P_t \left(e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt}; \theta_0); \eta_0 \right) + \ln \left[e^{\omega^*_{jt}} F(L_{jt}, M^*_{jt}, K_{jt}; \theta_0) \right] \right\}$$

as in (2.8).

Without loss of generality, assume u_{jt} has normal distribution with mean zero and unit variance.[48] The logarithmic density function of the sample $\{(r_{jt}, \mathbf{w}_{jt})\}_{jt}$ is $-\frac{1}{2}\sum_t [r_{jt} - f(\mathbf{x}_{jt}, \beta_0)]^2$. Thus, for a given β, $\Phi(\beta)$ defined in Proposition 2 is the information matrix. The additional restrictions that are utilized for identification are $E[\mathbf{h}(\mathbf{x}_{jt}, \beta)] = 0$, with Jacobean matrix $\Psi(\beta)$. Thus, our model fits the nonlinear regression framework of Rothenberg (1971) and we can apply Theorem 2 in Rothenberg (1971) to show local identification. □

Next, we consider identification in the context of the CES and the Translog specifications. To see that the parameters of the CES production function are not identified by the revenue equation alone, note that for $\beta = (\eta, \alpha_L, \alpha_M, \alpha_K, \gamma)$. The information matrix is,

$$\Phi(\beta) = \begin{bmatrix} E[f_\eta f_\eta] & E[f_\eta f_{\alpha_L}] & 0 & E[f_\eta f_{\alpha_K}] & E[f_\eta f_\gamma] \\ E[f_{\alpha_L} f_\eta] & E[f_{\alpha_L} f_{\alpha_L}] & 0 & E[f_{\alpha_L} f_{\alpha_K}] & E[f_{\alpha_L} f_\gamma] \\ 0 & 0 & 0 & 0 & 0 \\ E[f_{\alpha_K} f_\eta] & E[f_{\alpha_K} f_{\alpha_L}] & 0 & E[f_{\alpha_K} f_{\alpha_K}] & E[f_{\alpha_K} f_\gamma] \\ E[f_\gamma f_\eta] & E[f_\gamma f_{\alpha_L}] & 0 & E[f_\gamma f_{\alpha_K}] & E[f_\gamma f_\gamma] \end{bmatrix}.$$

Note that the rank of $\Phi(\beta)$ is 3 since $\frac{f_{\alpha_K}}{f_{\alpha_L}} = -\frac{\alpha_L}{\alpha_K}$ is a constant. In particular, the rank of the sub matrix containing columns 2, 3, and 4 is one.

To see how the additional restrictions aid in identification, recall that $\Psi(\beta)$ is specified as,

$$\Psi(\beta) = \begin{bmatrix} 0 & -\frac{\alpha_M}{\alpha_L^2} & \frac{1}{\alpha_L} & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

It is clear that the rank of the sub matrix containing column 2, 3, and 4 of $\Phi(\beta)$ is two. Thus, the matrix $V(\beta_0) = [\Phi(\beta_0); \Psi(\beta_0)]$ has full column rank five. So all parameters are locally identified.

A similar exercise can be carried out for the translog case. Denote

$$\beta = (\eta, \alpha_k, \alpha_l, \alpha_m, \alpha_{kk}, \alpha_{ll}, \alpha_{mm}, \alpha_{kl}, \alpha_{km}, \alpha_{lm}),$$

so there are ten parameters to identify. The rank of $\Phi(\beta)$ is four. With additional six moment restrictions as specified in the paper, $\Psi(\beta_0)$ has column rank six (assuming the instruments are not perfectly collinear). Thus, $V(\beta_0)$ has rank ten, and all parameters are locally identified.

[48] This assumption is used only to fit the notation and terminology of Rothenberg (1971), if u_t is not normally distributed then $-\frac{1}{2}\sum_{jt}[r_{jt} - f(\mathbf{w}_{jt}, \beta_0)]^2$ is not a logarithmic density. However $\Phi(\beta)$ is still the information matrix of a nonlinear least squares problem and the remainder of the proof is unchanged.

Now we establish the estimators' consistency and state the asymptotic distribution. Since our estimator is an extremum estimator, we need (a) identification; (b) uniform convergence of the objective function for consistency. Since (a) has been established, we focus on (b).

Define

$$\Omega(\beta) = E\Big[\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\Big]' W E\Big[\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\Big],$$

and

$$\Omega_n(\beta) = \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\right]' W \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\right],$$

where n is the number of observations.

Assume the true parameter $\beta_0 \in \Gamma$, and $\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)$ is continuous in $\beta \in \Gamma$ with probability one. Also, assume $E\Big[\sup_\beta |\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)|\Big] < \infty$. Then, the Uniform Law of Large Number implies:

$$\sup_\beta \left|\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta) - E\Big[\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\Big]\right| = o_p(1).$$

This in turn implies the uniform convergence of the objective function:

$$\sup_\beta \Big|\Omega_n(\beta) - \Omega(\beta)\Big| = o_p(1).$$

Therefore, the estimator defined for the general model (2.10) is consistent. For asymptotic distribution of the estimator, define

$$A = E\Big[\frac{\partial \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)}{\partial \beta'}\Big|_{\beta=\beta_0}\Big],$$

and

$$B = E\Big[\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta_0)\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta_0)'\Big].$$

These matrices can be estimated by their empirical analogues. If the weight matrix W is a consistent estimator of B^{-1} the asymptotic distribution is,

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, [A'B^{-1}A]^{-1}).$$

C The Standard CES Normalization

The purpose of this appendix is to show how the CES function is normalized in the literature, and compare this normalization with the one we adopt.

Motivation of Normalization

It has been commonly recognized that the CES production function need to be normalized to give meaningful identification of its parameters. There is a branch of literature analyzing the importance and the method of normalization, which includes de La Grandville (1989), Klump and de La Grandville (2000), Klump and Preissler (2000), de La Grandville and Solow (2006), and Leon-Ledesma, McAdam and Willman (2010).

The current literature has illustrated the key motivation of the normalization in details for two-factor-input production function (see Brown and de Cani (1963), Klump and Preissler (2000) and Leon-Ledesma, McAdam and Willman (2010)). However, we will work with three-factor-input production function, $Q = F(L, M, K)$. It is defined as a linear homogeneous function in which the elasticity of substitution between any two factors is a constant. The idea and motivation of the standard normalization procedure can be easily extended to our case. To see this, let us follow the literature by stating the definition of elasticity of substitution σ:

$$\begin{cases} \dfrac{\partial \ln(M/L)}{\partial \ln(F_L/F_M)} = \sigma \\ \dfrac{\partial \ln(K/L)}{\partial \ln(F_L/F_K)} = \sigma \end{cases}$$

This definition provides us with a second-order partial differential equation system. Given the assumption of the linear homogenous function, the general solution of the equation system is given by,

$$Q = F(L, M, K) = \lambda_1[L^\gamma + \lambda_2 M^\gamma + \lambda_3 K^\gamma]^{\frac{1}{\gamma}},$$

where $\gamma = \frac{\sigma-1}{\sigma}$, and λs are three arbitrary constants of integration emerging in the process of solving the differential equation system. One particular functional form used in the literature is obtained by taking $\tilde{\alpha}_L = \frac{1}{1+\lambda_2+\lambda_3}$, $\tilde{\alpha}_M = \frac{\lambda_2}{1+\lambda_2+\lambda_3}$, $\tilde{\alpha}_K = 1 - \tilde{\alpha}_L - \tilde{\alpha}_M$ and $C = \lambda_1(1 + \lambda_2 + \lambda_3)^{\frac{1}{\gamma}}$, thus

$$Q = F(L, M, K) = C[\tilde{\alpha}_L L^\gamma + \tilde{\alpha}_M M^\gamma + \tilde{\alpha}_K K^\gamma]^{\frac{1}{\gamma}}.$$

Here $\tilde{\alpha}_L$, $\tilde{\alpha}_M$ and $\tilde{\alpha}_K$ are referred as distribution parameters.[49] However, one can obtain different function forms by taking different specifications for λs. Each of these forms is called a family of CES functions. Examples of different families include ones used in Pitchford (1960), Arrow et al. (1961), and David and van de Klundert (1965). Therefore, as shown in the literature, a common baseline point is needed to compare different families of CES functions whose members are distinguished only by different elasticities of substitution. To this end, one needs to fix baseline point for the level of production (Q_0), factor inputs (L_0, M_0, K_0), and

[49] We use $\tilde{\alpha}$'s to denote the un-normalized (or "original") distribution parameters, while α's are reserved for the normalized distribution parameters, unless otherwise noticed.

the marginal rates of substitution (μ_{ML0}, μ_{KL0}), which are equal to the price ratios (P_{M_0}/P_{L_0}, P_{K_0}/P_{L_0}) because of the cost minimization.[50] For detailed motivation of normalization, refer to La Grandville (1989) and Leon-Ledesma, McAdam, and Willman (2010).

Standard Normalization Procedure

We follow de La Grandville (1989) and Leon-Ledesma, McAdam, and Willman (2010) to illustrate the normalization of the three-factor-input CES function. Given the elasticity of substitution σ, for any baseline point

$$Z_0 = (L_0, M_0, K_0, Q_0, \mu_{ML0}, \mu_{KL0}),$$

there are four equations about four parameters that characterize one particular family of CES functions:

$$\tilde{\alpha}_L + \tilde{\alpha}_M + \tilde{\alpha}_K = 1, \tag{33}$$

$$\left(\frac{F_M}{F_L}\right)_0 = \frac{\tilde{\alpha}_M}{\tilde{\alpha}_L}\left(\frac{L_0}{M_0}\right)^{1-\gamma} = \mu_{ML0} \equiv \frac{P_{M_0}}{P_{L_0}}, \tag{34}$$

$$\left(\frac{F_K}{F_L}\right)_0 = \frac{\tilde{\alpha}_K}{\tilde{\alpha}_L}\left(\frac{L_0}{K_0}\right)^{1-\gamma} = \mu_{KL0} \equiv \frac{P_{K_0}}{P_{L_0}}, \tag{35}$$

$$Q_0 = C[\tilde{\alpha}_L L_0^{\gamma} + \tilde{\alpha}_M M_0^{\gamma} + \tilde{\alpha}_K K_0^{\gamma}]^{\frac{1}{\gamma}}. \tag{36}$$

The equations (34) and (35) are implied by cost minimization. Note that the validation of (35) implicitly assumes the optimal choice of capital stock in the short run. The last equation holds since Q_0 is the physical output produced by its corresponding factor inputs. De La Grandville (1989) provides a graphical representation of the normalization. He shows that, after normalization all CES functions in the same family share the common baseline point of tangency, although their elasticities of substitution are different. Therefore, the purpose of normalization is to compare different CES functions in a meaningful way: on the one hand, different families of CES functions can be characterized by different baseline points, on the other hand, the members of each family sharing common baseline point are distinguished only by different elasticities of substitution.

These four equations imply a solution of four parameters:

$$\tilde{\alpha}_L(\sigma, Z_0) = \frac{P_{L0} L_0^{\frac{1}{\sigma}}}{P_{M0} M_0^{\frac{1}{\sigma}} + P_{L0} L_0^{\frac{1}{\sigma}} + P_{K0} K_0^{\frac{1}{\sigma}}},$$

$$\tilde{\alpha}_M(\sigma, Z_0) = \frac{P_{M0} M_0^{\frac{1}{\sigma}}}{P_{M0} M_0^{\frac{1}{\sigma}} + P_{L0} L_0^{\frac{1}{\sigma}} + P_{K0} K_0^{\frac{1}{\sigma}}},$$

[50] Note that P_{K_0} is the user price of capital, which usually is not accurately measured. To this end, we will extend the normalization to cases where P_{K_0} is not available.

$$\tilde{\alpha}_K(\sigma, Z_0) = \frac{P_{K0}K_0^{\frac{1}{\sigma}}}{P_{M0}M_0^{\frac{1}{\sigma}} + P_{L0}L_0^{\frac{1}{\sigma}} + P_{K0}K_0^{\frac{1}{\sigma}}},$$

$$C(\sigma, Z_0) = Q_0 \left[\frac{P_{L0}L_0^{\frac{1}{\sigma}} + P_{M0}M_0^{\frac{1}{\sigma}} + P_{K0}K_0^{\frac{1}{\sigma}}}{P_{L0}L_0 + P_{M0}M_0 + P_{K0}K_0} \right]^{\frac{\sigma}{\sigma-1}}.$$

Note that given the elasticity of substitution, the value of parameters depend on the choice of baseline point Z_0. Hence, comparing any two CES functions is not informative unless they are specified with the same baseline point.

Substituting the value of these parameters into the original function, we obtain:

$$Q = C(\sigma, Z_0)[\tilde{\alpha}_L(\sigma, Z_0)L^\gamma + \tilde{\alpha}_M(\sigma, Z_0)M^\gamma + \tilde{\alpha}_K(\sigma, Z_0)K^\gamma]^{\frac{1}{\gamma}}.$$

After re-parameterizations, one particular family of CES production function with corresponding normalized parameters is given by

$$Q = Q_0 \left[\alpha_{L0} \left(\frac{L}{L_0} \right)^\gamma + \alpha_{M0} \left(\frac{M}{M_0} \right)^\gamma + \alpha_{K0} \left(\frac{K}{K_0} \right)^\gamma \right]^{\frac{1}{\gamma}},$$

where:

$$\begin{cases} \alpha_{L0} = \dfrac{E_{L_0}}{E_{L_0} + E_{M_0} + E_{K_0}} \\ \alpha_{M0} = \dfrac{E_{M_0}}{E_{L_0} + E_{M_0} + E_{K_0}} \\ \alpha_{K0} = 1 - \alpha_{L0} - \alpha_{M0} \end{cases}$$

and $E_{L_0} = P_{L_0}L_0$, $E_{M_0} = P_{M_0}M_0$ and $E_{K_0} = P_{K_0}K_0$ are expenditures of labor, material and capital respectively[51]. Hence a normalized CES function is characterized by the baseline point Z_0 and elasticity of substitution σ: while each baseline point specifies a family of CES production functions, the members of each family sharing a common baseline values are distinguished only by different elasticities of substitution. The normalized distribution parameters now solely depend on the baseline point. Thus they can be prefixed before the estimation if normalization equations (34)-(35) hold (thus the normalization is valid).

Our CES Normalization

In the standard normalization literature, capital is assumed to be a static input which is chosen optimally in each period. However, in practice, capital may be chosen dynamically. For this reason, we extend the standard normalization approach to allow that capital is not running at the cost-minimizing level in the short run.

[51]Note that the expenditure on capital E_{K_0} is different from the capital stock K_0. But they are related by $E_{K_0} = P_{K_0}K_0$, where P_{K_0} is the user price of capital stock.

Specifically, although capital could be optimally chosen in the long run, the user price of capital (P_K, if available) may not reflect the marginal cost of capital in the short run. To this end, we assume the choice of capital can deviate from the short-run optimal value by certain magnitude of τ which is treated as a parameter to be estimated. This extension also allows for additional flexibility to deal with situations when the user cost of capital service ($E_K = P_K K$) is not available.

We start from the original production function

$$Q = \exp(\tilde{\omega})F(L, M, K) = \exp(\tilde{\omega})[\tilde{\alpha}_L L^\gamma + \tilde{\alpha}_M M^\gamma + \tilde{\alpha}_K K^\gamma]^{\frac{1}{\gamma}}, \tag{37}$$

where $\tilde{\omega}$ is the firm-level productivity.

As suggested by Leon-Ledesma, McAdam, and Willman (2010), the baseline point is chosen as the geometric sample mean:

$$\overline{Z} = (\overline{L}, \overline{M}, \overline{K}, \overline{Q}, \overline{\mu}_{ML}),$$

where $\overline{\mu}_{ML}$ is the average marginal rate of substitution between material and labor (i.e., $\overline{P}_M/\overline{P}_L$).

Note that the choice of the baseline value specifies a family of CES functions. Given the baseline value, the equations that characterize this family are:

$$\tilde{\alpha}_L + \tilde{\alpha}_M + \tilde{\alpha}_K = 1, \tag{38}$$

$$\left(\frac{F_M}{F_L}\right)_{\overline{Z}} = \frac{\tilde{\alpha}_M}{\tilde{\alpha}_L}\left(\frac{\overline{L}}{\overline{M}}\right)^{1-\gamma} = \overline{\mu}_{ML}, \tag{39}$$

$$\left(\frac{F_K}{F_L}\right)_{\overline{Z}} = \frac{\tilde{\alpha}_K}{\tilde{\alpha}_L}\left(\frac{\overline{L}}{\overline{K}}\right)^{1-\gamma} = \left(\tau\frac{\overline{E}_L}{\overline{E}_K}\right)\overline{\mu}_{KL} = \tau\frac{\overline{L}}{\overline{K}}, \tag{40}$$

$$\overline{Q} = e^{\overline{\omega}}[\tilde{\alpha}_L \overline{L}^\gamma + \tilde{\alpha}_M \overline{M}^\gamma + \tilde{\alpha}_K \overline{K}^\gamma]^{\frac{1}{\gamma}}, \tag{41}$$

where $\overline{\omega}$ is the "average" productivity associated with producing $\overline{Q}$ by $(\overline{L}, \overline{M}, \overline{K})$.

Here τ in (40) is introduced as an inefficiency parameter to measure the mean deviation of capital stock from its optimal level. This extension is important for multiple reasons compared with the standard normalization procedure. First, by introducing such an additional flexible parameter, we allow for the case when the capital stock is not optimally chosen in the short run (although it could be optimal in the long run). Specifically, when $\tau = \frac{\overline{E}_K}{\overline{E}_L}$, the marginal rate of substitution of labor and capital at the baseline point is equal to the price ratio, which implies the capital stock is indeed optimally chosen; when $\tau \neq \frac{\overline{E}_K}{\overline{E}_L}$, the actual capital deviates from the optimal amount. We will not specify the value of τ but leave it to be revealed by data as a parameter to estimate. Second, in our empirical application, such a flexible parameter enables us to deal with situations where the average "price" (or the user cost) of capital stock $\overline{P}_K$ (or $\overline{E}_K$) is not available or accurately measured. In other words, instead of assuming that $\overline{P}_K$ or $\overline{E}_K$ is known, we let it be absorbed in the parameter τ which can be estimated from data.

Given γ and τ, the distribution parameters implied by the equations (38), (39) and (40) are given by:

$$
\begin{aligned}
\tilde{\alpha}_L(\gamma,\tau) &= \frac{\frac{\overline{E}_L}{\overline{L}^\gamma}}{\frac{\overline{E}_L}{\overline{L}^\gamma}+\frac{\overline{E}_M}{\overline{M}^\gamma}+\tau\frac{\overline{E}_L}{\overline{K}^\gamma}} \\
\tilde{\alpha}_M(\gamma,\tau) &= \frac{\frac{\overline{E}_M}{\overline{M}^\gamma}}{\frac{\overline{E}_L}{\overline{L}^\gamma}+\frac{\overline{E}_M}{\overline{M}^\gamma}+\tau\frac{\overline{E}_L}{\overline{K}^\gamma}} \\
\tilde{\alpha}_K(\gamma,\tau) &= 1-\tilde{\alpha}_L(\gamma,\tau)-\tilde{\alpha}_M(\gamma,\tau)
\end{aligned}
\tag{42}
$$

As in the standard normalization procedure, we plug the distribution parameters into the original CES function to obtain the normalized CES function after re-parametrization:

$$
Q = e^{\omega}\overline{Q}\left[\alpha_L\left(\frac{L}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}},
$$

where

$$
\begin{aligned}
\alpha_L &= \frac{\overline{E}_L}{\overline{E}_L+\overline{E}_M+\tau\overline{E}_L} \\
\alpha_M &= \frac{\overline{E}_M}{\overline{E}_L+\overline{E}_M+\tau\overline{E}_L}, \\
\alpha_K &= 1-\alpha_L-\alpha_M
\end{aligned}
\tag{43}
$$

and

$$
\omega = \tilde{\omega}-\overline{\omega}.
$$

Note that, these equations imply $\frac{\alpha_K}{\alpha_L}=\tau$, which is why we define the ratio of α_K and α_L as τ in (2.13). In addition, this normalization places restrictions on the value of α's via (43) which is used to help identify all α's as shown in the paper.

D Details of Implementation for CES Specification

In this appendix, we describe details of implementation and precisely define our estimator for the normalized CES production function.

Each firm j chooses labor and material quantities to maximize the profit in each period t, given its capital stock and productivity. The firm's static problem is:

$$
\max_{L_{jt},M_{jt}} P_t(Q_{jt})Q_{jt} - P_{L_{jt}}L_{jt} - P_{M_{jt}}M_{jt},
$$

where

$$Q_{jt} = e^{\omega_{jt}} \overline{Q} \left[\alpha_L \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} + \alpha_M \left(\frac{M_{jt}}{\overline{M}} \right)^{\gamma} + \alpha_K \left(\frac{K_{jt}}{\overline{K}} \right)^{\gamma} \right]^{\frac{1}{\gamma}},$$

and

$$\frac{Q_{jt}}{Q_t} = \left(\frac{P_{jt}}{P_t} \right)^{\eta}.$$

Note that L_{jt}, M_{jt} and Q_{jt} are physical quantities of labor and material input and output respectively. The first order conditions with respective to labor and material are

$$\frac{1+\eta}{\eta} \frac{\partial Q_{jt}}{\partial L_{jt}} \frac{Q_{jt}^{1/\eta} P_t}{Q_t^{1/\eta}} = P_{L_{jt}},$$

$$\frac{1+\eta}{\eta} \frac{\partial Q_{jt}}{\partial M_{jt}} \frac{Q_{jt}^{1/\eta} P_t}{Q_t^{1/\eta}} = P_{M_{jt}}.$$

Note that $E_{L_{jt}} = P_{L_{jt}} L_{jt}$ and $E_{M_{jt}} = P_{M_{jt}} M_{jt}$, and plug the demand function into above equations we obtain:

$$\frac{1+\eta}{\eta} \frac{\partial Q_{jt}}{\partial L_{jt}} \frac{L_{jt}}{Q_{jt}} = \frac{E_{L_{jt}}}{R_{jt}},$$

$$\frac{1+\eta}{\eta} \frac{\partial Q_{jt}}{\partial M_{jt}} \frac{M_{jt}}{Q_{jt}} = \frac{E_{M_{jt}}}{R_{jt}},$$

where $R_{jt} = P_{jt} Q_{jt}$ is the revenue for firm j at period t.

Take the ratio with respective to both sides of the equations, and we can solve for material quantity:

$$\frac{M_{jt}}{\overline{M}} = \left[\frac{\alpha_L E_{M_{jt}}}{\alpha_M E_{L_{jt}}} \right]^{\frac{1}{\gamma}} \frac{L_{jt}}{\overline{L}}. \tag{44}$$

This implies that material quantity can be imputed from observables ($E_{L_{jt}}$, $E_{M_{jt}}$, and L_{jt}) up to unknown parameters. Substitute this M_{jt} in the first order condition for labor and solve for ω_{jt}, we have

$$e^{-\frac{1+\eta}{\eta}\omega_{jt}} = \alpha_L \frac{1+\eta}{\eta} \frac{P_t}{Q_t^{1/\eta}} \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} \frac{\overline{Q}^{\frac{1+\eta}{\eta}}}{E_{L_{jt}}} \left[\alpha_L \left(\frac{E_{L_{jt}} + E_{M_{jt}}}{E_{L_{jt}}} \right) \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} + \alpha_K \left(\frac{K_{jt}}{\overline{K}} \right)^{\gamma} \right]^{\frac{1}{\gamma\eta} + \frac{1}{\gamma} - 1} \tag{45}$$

Note that the imputed ω_{jt} is also a function of observables.

Plug the imputed M_{jt} and ω_{jt} into the revenue equation:

$$R_{jt} = \exp(u_{jt}) P_t(Q_{jt}) Q_{jt},$$

where u_{jt} is the measurement error.

After some algebra we have,

$$\ln R_{jt} = \ln \frac{\eta}{1+\eta} + \ln \left[E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^{\gamma} \right) \right] + u_{jt}.$$

Therefore, the model can be estimated via the following nonlinear least square estimation with restrictions:

$$\begin{aligned} \hat{\beta} \;\; &= \;\; \operatorname{argmin} \sum_{jt} \left[\ln R_{jt} - \ln \frac{\eta}{1+\eta} - \ln \left\{ E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^{\gamma} \right) \right\} \right]^2 \\ & \text{subject to} \qquad \frac{\alpha_M}{\alpha_L} = \frac{\overline{E}_M}{\overline{E}_L}, \qquad (46) \\ & \qquad\qquad\qquad \alpha_L + \alpha_M + \alpha_K = 1, \qquad (47) \end{aligned}$$

where $\hat{\beta} = (\hat{\eta}, \hat{\alpha}, \hat{\gamma})$.

As discussed in the paper, this nonlinear least square estimation with constraint is equivalent to the GMM estimator defined in (2.18). However, the nonlinear least square estimation is easier to implement with re-parametrization. In particular, we define $\tau = \alpha_K/\alpha_L$ so that the two constricts (46) and (47) implies (43). The nonlinear least square estimation give us estimate $(\hat{\eta}, \hat{\tau}, \hat{\gamma})$ with standard errors. In turn leads to the estimate of the entire set of parameters of interest, $(\hat{\eta}, \hat{\alpha}_L, \hat{\alpha}_M, \hat{\alpha}_K, \hat{\sigma})$, where $\hat{\alpha}$'s follow from (43) and $\hat{\sigma} = 1/(1-\hat{\gamma})$. The associated standard errors are derived according to the delta method.[52]

E Monte Carlo Description

In this appendix, we outline the data generating process for the Monte Carlo experiments. Specifically, the Monte Carlo experiments consist of N replications of simulated data sets, given a set of true parameters of interest (η, σ, α_L, α_M and α_K).[53] In each replication, we simulate a sequence of productivity (ω_{jt}), idiosyncratic input prices ($P_{L_{jt}}$ and $P_{M_{jt}}$), and capital stock (K_{jt}) for each firm j over time. Given these variables and random shocks, we derive a sequence of optimal choices of labor and material inputs (L_{jt} and M_{jt}), the optimal output quantity (Q_{jt}) and price (P_{jt}) for firm j in each period t.

In each replication, there are J firm in production for T periods. The evolution

[52] We have also implemented the estimator using a direct GMM approach and the results are almost identical.

[53] To simplify the notation, we drop the tilde over the distribution parameters in this section as well as the report in the associated tables. But please note that the data is generated from the "original" production function (37).

process of productivity for each firm is assumed to be a first order Markov process:

$$\omega_{jt+1} = g_0 + g_1\omega_{jt} + \varepsilon^{\omega}_{jt+1},$$

where $\varepsilon^{\omega}_{jt+1}$ is the innovation shock realized in period $t+1$, which is assumed to be a normally distributed i.i.d. error term with zero mean and standard deviation $sd(\varepsilon^{\omega})$. The initial productivity of each firm (ω_{j0}) is drawn from a normal distribution of mean ω_0 and standard deviation $se(\omega_0)$.

The investment rule and the capital evolution process are set as,

$$\log(I_{jt}) = \xi\omega_{jt} + (1-\xi)\log(K_{jt}),$$

$$K_{jt+1} = K_{jt} + I_{jt},$$

where $\xi \in (0,1)$ is an arbitrary weight. The initial capital stock of each firm (K_{j0}) is drawn from a normal distribution of mean K_0 and standard variance $se(K_0)$.

The idiosyncratic labor and material input prices ($P_{L_{jt}}$ and $P_{M_{jt}}$) are generated as follows:

$$P_{L_{jt}} = \overline{P}_{L_t} e^{\varepsilon^{PL}_{jt}},$$

$$P_{M_{jt}} = \overline{P}_{M_t} e^{\varepsilon^{PM}_{jt}},$$

where $\overline{P}_{L_t}$ and $\overline{P}_{M_t}$ are the industrial-level labor and material prices in period t, which can be drawn from $N(\overline{P}_L, sd(\overline{P}_L))$ and $N(\overline{P}_M, sd(\overline{P}_M))$ independently, or set to 1 for simplicity as in our implementation. ε^{PL}_{jt} and ε^{PM}_{jt} are deviations from the industrial-level input prices, which are independently drawn from $N(0, sd(\varepsilon^{PL}))$ and $N(0, sd(\varepsilon^{PM}))$ respectively.

For the demand side, we assume the industrial-level output quantities are generated by,

$$Q_t = r^t Q_0 e^{\varepsilon^{Q}_t},$$

where r is the growth rate of the industrial-level output quantity, Q_0 is the initial industrial-level output quantity, and $\varepsilon^{Q}_t \sim N(0, se(\varepsilon^{Q}))$ is an independent random shock. The industrial-level output prices are generated by,

$$P_t = Q_t^{1/\eta},$$

where η is the demand elasticity.[54]

Now we have simulated $\{\omega_{jt}, K_{jt}, I_{jt}, P_{L_{jt}}, P_{M_{jt}}, Q_t, P_t\}$ for each firm j and period t. Given these variables, we can derive the optimal labor and material input choices (L_{jt} and M_{jt}) and the corresponding output quantity (Q_{jt}) for each firm j and period t according to the first order conditions associated with the firm's static profit maximization problem. Specifically, the optimal labor input is derived

[54] We set P_t and Q_t as 1 for simplicity in the implementation.

as,

$$L_{jt} = \left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{1}{\gamma-1}} M_{jt}, \tag{48}$$

where the material input M_{jt} is given by,

$$M_{jt} = \left[\frac{(e^{-\omega_{jt}} Q_{jt})^{\gamma} - \alpha_K K^{\gamma}}{\alpha_M + \alpha_L \left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{\gamma}{1-\gamma}}}\right]^{\frac{1}{\gamma}}, \tag{49}$$

and Q_{jt} is the solution of the following equation:

$$\frac{\eta+1}{\eta}\left(\frac{P_t}{Q_t^{\frac{1}{\eta}}}\right) Q_{jt}^{\frac{1}{\eta}} = e^{-\omega_{jt}} \frac{\left[P_{M_{jt}} + P_{L_{jt}}\left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{1}{\gamma-1}}\right]}{\left[\alpha_M + \alpha_L\left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{\gamma}{1-\gamma}}\right]^{\frac{1}{\gamma}}} \left(1 - \alpha_K K^{\gamma} (e^{-\omega_{jt}} Q_{jt})^{-\gamma}\right)^{\frac{1}{\gamma}-1}. \tag{50}$$

Given the derived variables and underlying true parameters, (50) is only about Q_{jt}. It is easy to verify that (50) implies a unique solution for Q_{jt} since given $\eta < -1$, the left hand side is decreasing in Q_{jt} while the right hand side is increasing in Q_{jt}. Denote the solution of the equation as Q^*_{jt}. Once we obtain Q^*_{jt}, we can derive the corresponding L_{jt} and M_{jt} from (48) and (49). Hence, the expenditures of input are given by $E_{L_{jt}} = P_{L_{jt}} L_{jt}$ and $E_{M_{jt}} = P_{M_{jt}} M_{jt}$. The observed output with a measurement error is given by

$$Q_{jt} = Q^*_{jt} e^{\varepsilon^q_{jt}},$$

where $\varepsilon^q_{jt} \sim N(0, sd(\varepsilon^q))$ is the measurement error. At last, firm level output price P_{jt} is derived by equation

$$\frac{Q_{jt}}{Q_t} = \left(\frac{P_{jt}}{P_t}\right)^{\eta},$$

and the firm-level revenue is obtained by

$$R_{jt} = P_{jt} Q_{jt}.$$

Hence, we have generated a data set of

$$\{\omega_{jt}, K_{jt}, I_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, R_{jt}, Q_t, P_t\}$$

for each firm j and period t.

F Additional Application Results

Comparison to other estimation methods

While we use the OP-KG estimation method as our primary basis of comparison in the main body of the paper, there are many other approaches to estimating production functions. In this appendix, we compare our method to three additional approaches. First, we implement a simple nonlinear least squares estimator for the production function which proxies for materials with expenditure and also ignores the presence of heterogeneity. Second we use an approach that uses the first order conditions to control for productivity, but continues to use a materials expenditure to proxy for materials quantities. Finally, we compare our estimator to a panel data estimator a la Arellano and Bond (1991), where the productivity term includes a fixed effect and an AR(1) process.

First, we estimate the model with naive nonlinear least square estimation, in which the material expenditure is used as a proxy of quantity and the productivity is lumped into the additive error term. Specifically, the following model is estimated:[55]

$$\ln\left(\frac{R_{jt}}{\overline{R}}\right) = \zeta_t + \frac{1}{\gamma}\frac{1+\eta}{\eta}\ln\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma} + \alpha_M\left(\frac{E_{M_{jt}}}{\overline{E}_M}\right)^{\gamma} + \alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right] + u_{jt}, \tag{51}$$

where $\zeta_t = \ln\left(\frac{P_t}{\overline{P}_t}\right) - \frac{1}{\eta}\ln\left(\frac{Q_t}{\overline{Q}_t}\right)$, and $\gamma = \frac{\sigma-1}{\sigma}$ and σ is the elasticity of substitution. P_t and Q_t are industry-level output price and quantity. Note that u_{jt} contains both the productivity and measurement error.

The result is shown in the second column of Table A.1 under the title 'NLLS'. For comparison purposes, the first column reproduces our estimates from Table 2.4, while the final column reproduces the OP-KG estimates. We can immediately see that controlling for productivity is essential to producing reasonable estimates of the demand parameter η, which has the wrong sign and an extremely high magnitude under the NLLS specification.

Secondly, we estimate the model with nonlinear least square estimation, with the proxy of material quantity (i.e., material expenditure) and the productivity imputed from the first order condition of labor input. To be specific, with the productivity imputed from the first order condition of labor input, the revenue equation can be derived similarly to D:

$$\ln R_{jt} = \ln\frac{\eta}{1+\eta} + \ln\left[E_{L_{jt}}\frac{\alpha_M}{\alpha_L}\left(\frac{M_{jt}/\overline{M}}{L_{jt}/\overline{L}}\right)^{\gamma} + E_{L_{jt}}\left(1 + \frac{\alpha_K}{\alpha_L}\left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}}\right)^{\gamma}\right)\right].$$

Since M_{jt} is not observed, we use its proxy $E_{M_{jt}}$. Thus, the following empirical

[55]To make results comparable, we estimate this normalized revenue equation instead of production function.

equation is estimated:

$$\ln R_{jt} = \ln \frac{\eta}{1+\eta} + \ln \left[E_{L_{jt}} \frac{\alpha_M}{\alpha_L} \left(\frac{E_{M_{jt}}/\overline{E}_M}{L_{jt}/\overline{L}} \right)^{\gamma} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^{\gamma} \right) \right] + e_{jt}$$

with normalization restriction (43), where e_{jt} is the measurement error.

The result from this table is presented in column 3 of Table A.1 under the title 'Prod.'. Controlling for productivity generates a reasonable demand parameter, as opposed to the earlier approach. However, we see that now, the elasticity of substitution parameter is substantially larger than in all other methods. This is intuitive. Note that the first order conditions for labor and material implies that $E_{L_{jt}} \frac{\alpha_M}{\alpha_L} \left(\frac{M_{jt}/\overline{M}}{L_{jt}/\overline{L}} \right)^{\gamma} = E_{M_{jt}}$. The difference between this estimation and our method is that we utilize this relationship rather than using a proxy of material quantity. As shown in the table, the elasticity of substitution is significantly larger than our estimates, because the variance of $E_{L_{jt}} \frac{\alpha_M}{\alpha_L} \left(\frac{E_{M_{jt}}/\overline{E}_M}{L_{jt}/\overline{L}} \right)^{\gamma}$ is larger ($1.2 \sim 3$ times) than the variance of $E_{M_{jt}}$.

Third, we estimate a CES version of persistent panel data method (Arellano and Bond, 1991; Blundell and Bond, 2000), with an AR(1) term and a fixed effect. In particular, consider the empirical equation (51), but now the error term is decomposed as

$$u_{jt} = \beta_j + \nu_{jt} + \epsilon_{jt},$$

and

$$\nu_{jt} = \rho \nu_{jt-1} + v_{jt},$$

where ϵ_{jt} is the i.i.d. measurement error and v_{jt} is the i.i.d. innovation term.

Following the persistent panel data method, we take the quasi-difference

$$D_t(\eta, \sigma, \tau, \rho) = u_{jt} - \rho u_{jt-1},$$

$$D_{t-1}(\eta, \sigma, \tau, \rho) = u_{jt-1} - \rho u_{jt-2},$$

where u_{jt} is given by[56]

$$u_{jt} = \ln \left(\frac{R_{jt}}{\overline{R}} \right) - \left\{ \zeta_t + \frac{1}{\gamma} \frac{1+\eta}{\eta} \ln \left[\alpha_L \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} + \alpha_M \left(\frac{E_{M_{jt}}}{\overline{E}_M} \right)^{\gamma} + \alpha_K \left(\frac{K_{jt}}{\overline{K}} \right)^{\gamma} \right] \right\},$$

where $\zeta_t = \ln \left(\frac{P_t}{\overline{P}_t} \right) - \frac{1}{\eta} \ln \left(\frac{Q_t}{\overline{Q}_t} \right)$.

Then we construct the following moment condition for GMM estimation:

$$E\Big[D_t(\eta, \sigma, \tau, \rho) - D_{t-1}(\eta, \sigma, \tau, \rho) | E_{L_{jt-2}}, E_{M_{jt-2}}, K_{jt-2}, K_{jt-2}^2 \Big] = 0.$$

[56] Note that all α's are parameterized as in (43), as a function of τ.

The result is reported in the fourth column of A.1 under the title 'AB'. Like the OP-KG estimator, we would expect this the elasticity of substitution to be biased downward using this approach, due to the way the expenditure proxy for materials is employed. In fact, we do see that this estimator, like OP-KG estimates a smaller $\hat{\sigma}$ relative to our method.

Comparing Productivity Measures

As with other structural approaches to production function estimation, there are two potential approaches to defining "productivity" in our model. In the body of the paper, we follow the most common approach, and report the distribution of $\widehat{\omega_{it} + u_{it}}$ which is the residual from the production function itself. This represents the sum of productivity anticipated by the firm as well as unanticipated productivity and potential measurement error in revenues. Alternatively, we could employ (2.7) to recover $\hat{\omega}_{it}$ alone from the system of first order conditions. This approach includes only an estimate of productivity anticipated by the firm when it makes its labor and materials decision. A similar approach could be recover $\hat{\omega}_{it}$ alone using the OP-KG procedure. However in this case, the anticipated productivity relaters to the firm's expectation of productivity when making the *investment* decision which occurs later than the hiring decision according to the timing assumptions.

It is interesting to see whether these different definitions of the productivity distribution yield substantially different results. We investigate this in Figure A.1, which plots the two distributions for the two different methods for the Clothing industry. (Results for other industries are similar and are available by request). We see that while there is a substantial difference across methods (as is also visible in Figure 2.5), the difference across definitions for a given method is relatively small. It is particularly small when using our method. This implies that the bulk of the variance in the distribution of productivity is due to anticipated productivity differences, which further supports the importance of controlling for productivity differences when estimating production functions.

Finally, Figure A.2 presents the two distributions for our method only across all four industries. It shows that the result that the two distributions are quite similar is robust across the four industries we consider in the main body of the paper.

Table 2.1. Monte Carlo Parameter Values

Single material	Description	Value
η	Demand elasticity	-4
σ	Elasticity of substitution	0.8, 1.5, 2.5
α_L	Distribution parameter of labor	0.4
α_M	Distribution parameter of material	0.4
α_K	Distribution parameter of capital	0.2
g_0	Parameter in productivity evolution	0.2
g_1	Parameter in productivity evolution	0.95
$sd(\varepsilon^{\omega})$	Standard deviation of productivity innovation	0.01
$sd(\varepsilon^{PL})$	Standard deviation of labor price shock	0.2
$sd(\varepsilon^{PM})$	Standard deviation of material price shock	0.2
$sd(\varepsilon^{q})$	Standard deviation of output measurement error	0.01
T	Number of periods	10
J	Number of firms	100
N	Number of Monte Carlo replications	1000
Multiple materials[1]		
σ	Elasticity of substitution across primary inputs	1.5
P_{M1}	Price of $M1$ (constant)	0.1
P_{M2}	Price of $M2$ (constant)	0.18
$sd(\varepsilon^{PM3})$	Standard deviation of price shock of $M3$	0.2
δ	Effective factor of $M1$	0.65
σ_1	Elasticity of substitution between $M1$ and $M3$	2.1
σ_2	Elasticity of substitution between $M2$ and $M3$	2.0

[1] Only parameters different from the single material input setting are listed.

Table 2.2. Monte Carlo: estimated results from three different methods[1]

	True	$\sigma = 0.8$			$\sigma = 1.5$			$\sigma = 2.5$		
		Us	OP	Oracle	Us	OP	Oracle	Us	OP	Oracle
$\hat{\eta}$	-4.00	-3.994	-	-	-3.999	-	-	-3.999	-	-
		(0.156)	-	-	(0.047)	-	-	(0.014)	-	-
		[0.160]	-	-	[0.048]	-	-	[0.014]	-	-
$\hat{\sigma}$		0.800	0.669	0.802	1.501	1.208	1.502	2.499	1.988	2.495
		(0.007)	(0.174)	(0.030)	(0.018)	(0.214)	(0.048)	(0.037)	(0.372)	(0.068)
		[0.007]	[0.146]	[0.013]	[0.018]	[0.299]	[0.025]	[0.037]	[0.516]	[0.037]
$\hat{\alpha}_L$	0.40	0.400	0.483	0.400	0.400	0.478	0.400	0.400	0.459	0.401
		(0.005)	(0.007)	(0.007)	(0.002)	(0.005)	(0.005)	(0.001)	(0.010)	(0.003)
		[0.005]	[0.083]	[0.003]	[0.002]	[0.078]	[0.002]	[0.001]	[0.059]	[0.001]
$\hat{\alpha}_M$	0.40	0.400	0.484	0.400	0.400	0.479	0.400	0.400	0.460	0.400
		(0.005)	(0.007)	(0.007)	(0.002)	(0.005)	(0.005)	(0.001)	(0.011)	(0.003)
		[0.005]	[0.084]	[0.003]	[0.002]	[0.079]	[0.002]	[0.001]	[0.060]	[0.001]
$\hat{\alpha}_K$	0.20	0.200	0.033	0.199	0.200	0.043	0.201	0.200	0.082	0.199
		(0.006)	(0.015)	(0.013)	(0.008)	(0.009)	(0.009)	(0.005)	(0.021)	(0.005)
		[0.010]	[0.166]	[0.005]	[0.003]	[0.157]	[0.005]	[0.005]	[0.118]	[0.003]

[1] The table reports the medians of N replications of each case. The standard errors are included in the parentheses and root mean squared errors are in the square brackets.

[2] $\hat{\tau}$ is endogenously implied by the choice of capital stock. True value of $\hat{\tau}$ for the three cases are 0.326, 0.209 and 0.173 respectively.

Figure 2.1. Monte Carlo experiments: true $\sigma = 0.8, 1.5, 2.5$

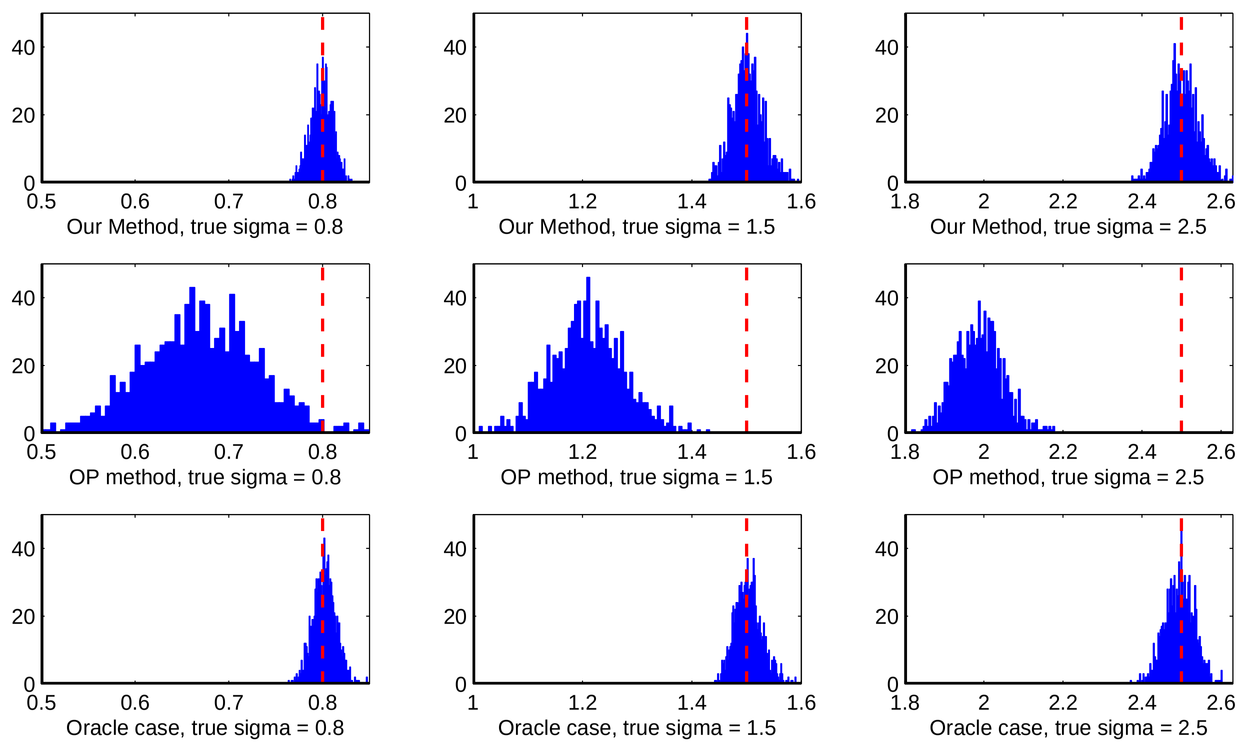


Figure 2.2. Kernel density estimation: imputed material prices v.s. true material prices

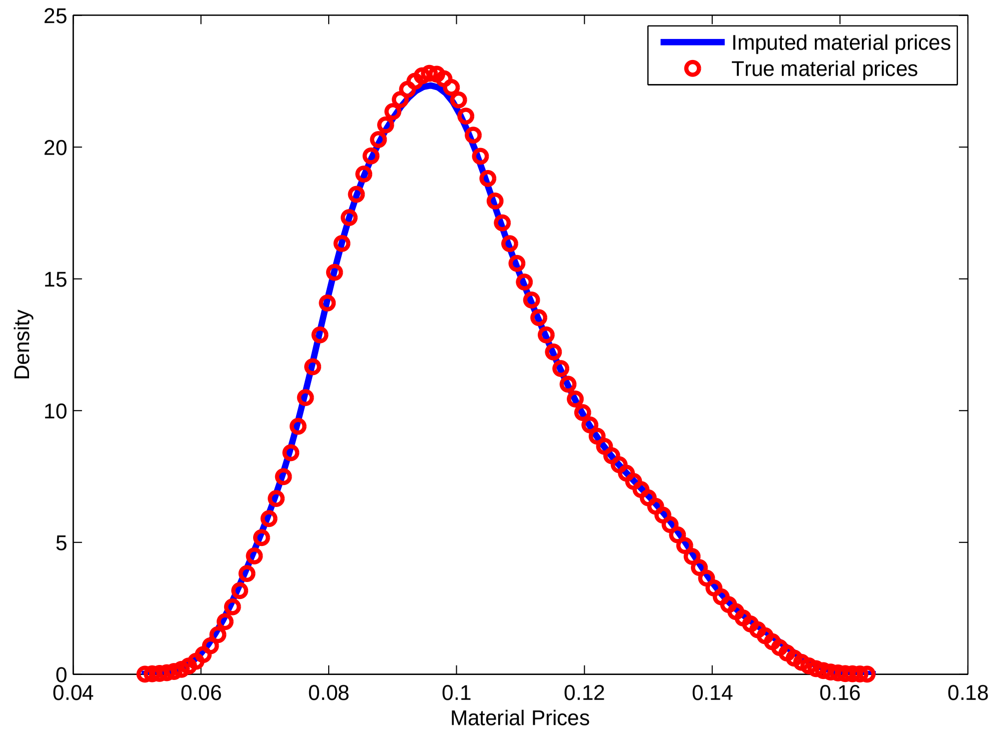


Figure 2.3. Profit difference in choosing different quality levels

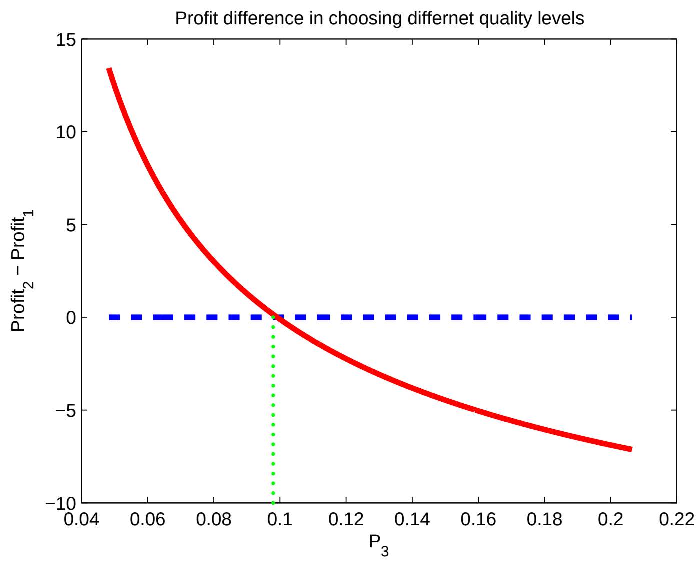


Table 2.3. Multiple materials Monte Carlo: parameter estimates[1]

Parameter	$\hat{\eta}$	$\hat{\sigma}$	$\hat{\alpha}_L$	$\hat{\alpha}_M$	$\hat{\alpha}_K$
True	-4.000	1.500	0.400	0.400	0.200
Estimation	-3.997	1.497	0.400	0.400	0.200
SE	(0.001)	(0.026)	(0.004)	(0.005)	(0.008)
RMSE	[0.052]	[0.027]	[0.003]	[0.001]	[0.003]

[1] The table reports the medians of $N = 1000$ replications of each case. The standard errors are included in the parentheses and root mean squared errors are in the square brackets.

Figure 2.4. Multiple materials Monte Carlo: true and recovered material and material price index

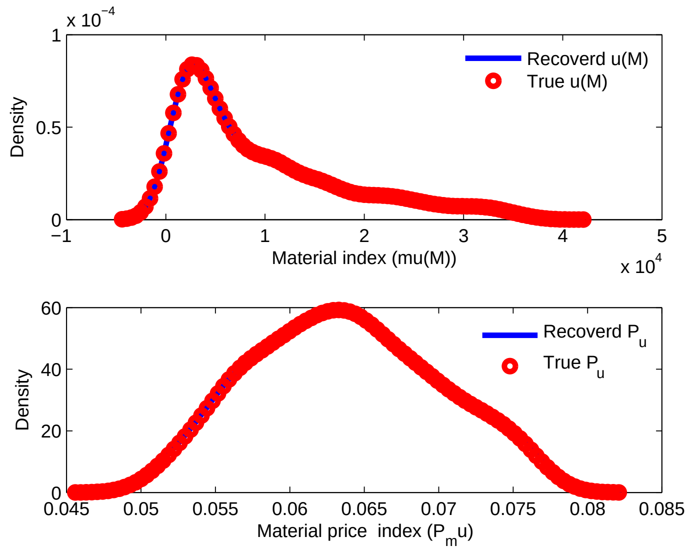


Table 2.4. Estimated results for Colombian industries

	Clothing		Bakery Products		Printing & Pub.		Metal Furniture	
	Us	OP-KG	Us	OP-KG	Us	OP-KG	Us	OP-KG
$\hat{\eta}$	-5.768	-8.465	-5.231	-5.253	-4.659	-12.161	-5.518	-6.947
	(0.121)	(1.544)	(0.188)	(0.417)	(0.236)	(5.434)	(0.433)	(2.686)
$\hat{\sigma}$	1.948	0.361	1.443	0.401	2.555	0.593	1.772	0.393
	(0.234)	(0.018)	(0.117)	(0.011)	(0.405)	(0.054)	(0.379)	(0.045)
$\hat{\alpha}_L$	0.361	0.371	0.244	0.251	0.372	0.381	0.300	0.304
	(0.002)	(0.001)	(0.002)	(0.000)	(0.005)	(0.004)	(0.005)	(0.005)
$\hat{\alpha}_M$	0.601	0.618	0.705	0.725	0.537	0.549	0.637	0.647
	(0.003)	(0.001)	(0.006)	(0.001)	(0.007)	(0.005)	(0.010)	(0.011)
$\hat{\alpha}_K$	0.038	0.011	0.050	0.025	0.091	0.070	0.064	0.049
	(0.004)	(0.002)	(0.007)	(0.002)	(0.013)	(0.009)	(0.015)	(0.016)
$\hat{g}_0$	0.008	0.101	0.039	0.148	-0.025	0.211	-0.033	0.219
	(0.010)	(0.011)	(0.015)	(0.011)	(0.015)	(0.039)	(0.024)	(0.072)
$\hat{g}_1$	0.695	0.972	0.822	0.955	0.906	0.950	0.824	0.877
	(0.014)	(0.010)	(0.012)	(0.005)	(0.019)	(0.014)	(0.026)	(0.026)
#Obs	5763		2269		2377		903	

Figure 2.5. Kernel density estimation of productivity $(\widehat{\omega_{jt} + u_{jt}})$

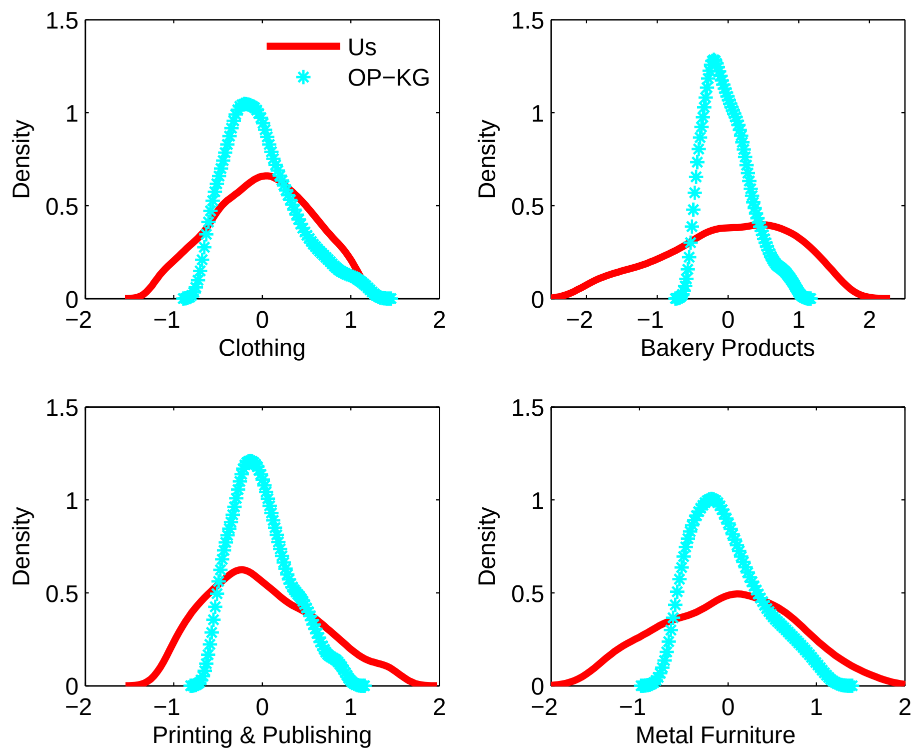


Table 2.5. Persistence of imputed material prices

	Persistence	Standard Error
Clothing	0.77	0.20
Bakery Products	0.77	0.21
Printing & Publishing	0.82	0.18
Metal Furniture	0.68	0.19

Table 2.6. Correlations between imputed productivity and input prices in logarithm

	$\mathrm{corr}(\hat{\omega}, \log(\hat{P}_M))$	$\mathrm{corr}(\hat{\omega}, \log(P_L))$	$\mathrm{corr}(\log(\hat{P}_M), \log(P_L))$
Clothing	0.76	0.60	0.26
Bakery Products	0.93	0.58	0.40
Printing & Publishing	0.68	0.88	0.68
Metal Furniture	0.85	0.65	0.48

Figure 2.6. Kernel density estimation of imputed material prices in logarithm

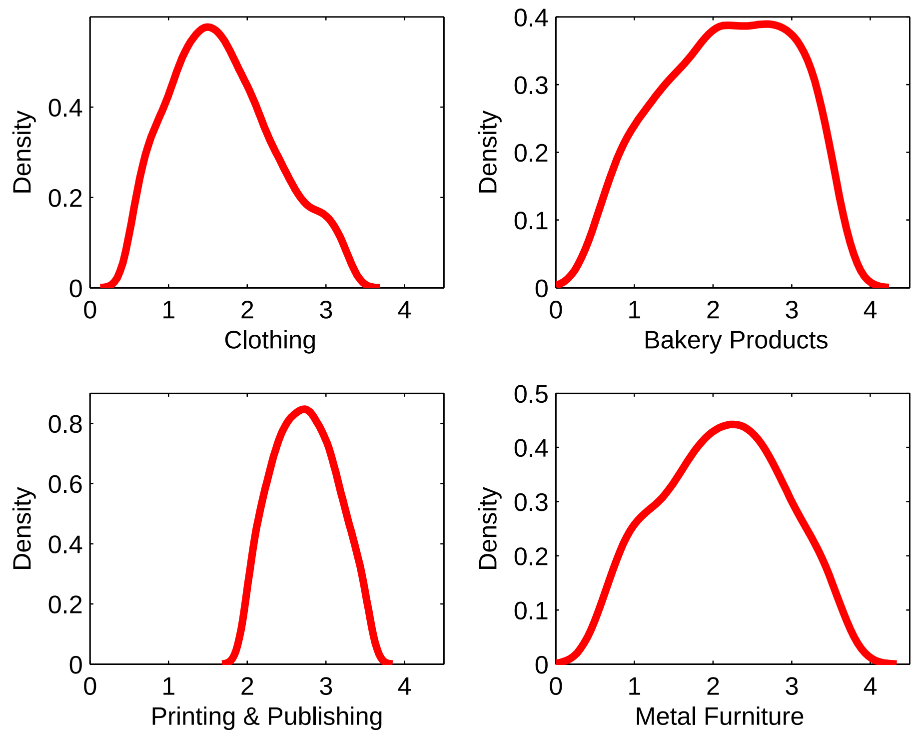


Table A.1. Estimated results by various methods for Colombian industries

	Clothing					Bakery Products				
	Us	NLLS	Prod.	AB	OP-KG	Us	NLLS	Prod.	AB	OP-KG
$\hat{\eta}$	-5.768	109.423	-5.020	-2.416	-8.465	-5.231	24.564	-4.650	-2.546	-5.253
	(0.121)	(39.317)	(0.105)	(0.373)	(1.544)	(0.188)	(1.927)	(0.156)	(0.484)	(0.417)
$\hat{\sigma}$	1.948	1.880	6.826	0.569	0.361	1.443	1.411	4.369	0.363	0.401
	(0.234)	(0.077)	(0.292)	(0.229)	(0.018)	(0.117)	(0.104)	(0.233)	(0.298)	(0.011)
$\hat{\alpha}_L$	0.361	0.372	0.370	0.358	0.371	0.244	0.253	0.254	0.244	0.251
	(0.002)	(0.001)	(0.001)	(0.037)	(0.001)	(0.002)	(0.001)	(0.001)	(0.025)	(0.000)
$\hat{\alpha}_M$	0.601	0.620	0.617	0.597	0.618	0.705	0.731	0.734	0.704	0.725
	(0.003)	(0.002)	(0.002)	(0.062)	(0.001)	(0.006)	(0.002)	(0.004)	(0.072)	(0.001)
$\hat{\alpha}_K$	0.038	0.008	0.013	0.045	0.011	0.050	0.016	0.012	0.053	0.025
	(0.004)	(0.004)	(0.003)	(0.100)	(0.002)	(0.007)	(0.003)	(0.005)	(0.097)	(0.002)
$\hat{g}_0$	0.008		0.048		0.101	0.039		0.029		0.148
	(0.010)		(0.010)		(0.011)	(0.015)		(0.015)		(0.011)
$\hat{g}_1$	0.695		0.771		0.972	0.822		0.873		0.955
	(0.014)		(0.018)		(0.010)	(0.012)		(0.022)		(0.005)
$\hat{\rho}$				0.547					0.264	
				(0.071)					(0.191)	
#Obs	5763					2269				
	Printing & Publishing					Metal Furniture				
	Us	NLLS	Prod.	AB	OP-KG	Us	NLLS	Prod.	AB	OP-KG
$\hat{\eta}$	-4.659	14.142	-3.857	-2.189	-12.161	-5.518	10.514	-4.186	-6.280	-6.947
	(0.236)	(0.986)	(0.135)	(1.739)	(5.434)	(0.433)	(0.918)	(0.210)	(6.713)	(2.686)
$\hat{\sigma}$	2.555	2.218	3.326	0.313	0.593	1.772	1.395	5.600	0.429	0.393
	(0.405)	(0.186)	(0.145)	(0.405)	(0.054)	(0.379)	(0.090)	(0.598)	(0.212)	(0.045)
$\hat{\alpha}_L$	0.372	0.379	0.389	0.298	0.381	0.300	0.295	0.319	0.233	0.304
	(0.005)	(0.003)	(0.004)	(0.070)	(0.004)	(0.005)	(0.002)	(0.003)	(0.057	(0.005
$\hat{\alpha}_M$	0.537	0.546	0.560	0.429	0.549	0.637	0.627	0.677	0.494	0.647
	(0.007)	(0.004)	(0.005)	(0.101)	(0.005)	(0.010)	(0.005)	(0.005)	(0.121)	(0.011)
$\hat{\alpha}_K$	0.091	0.075	0.051	0.273	0.070	0.064	0.078	0.004	0.273	0.049
	(0.013)	(0.006)	(0.009)	(0.171)	(0.009)	(0.015)	(0.008)	(0.008)	(0.178)	(0.016)
$\hat{g}_0$	-0.025		-0.018		0.211	-0.033		-0.003		0.219
	(0.015)		(0.015)		(0.039)	(0.024)		(0.024)		(0.072)
$\hat{g}_1$	0.906		0.933		0.950	0.824		0.863		0.877
	(0.019)		(0.017)		(0.014)	(0.026)		(0.033)		(0.026)
$\hat{\rho}$				0.358					-0.303	
				(0.380)					(0.462)	
#Obs	2377					903				

Figure A.1. Comparison: densities of $\hat{\omega}$ and $\hat{\omega} + \hat{u}$ – large demonstration

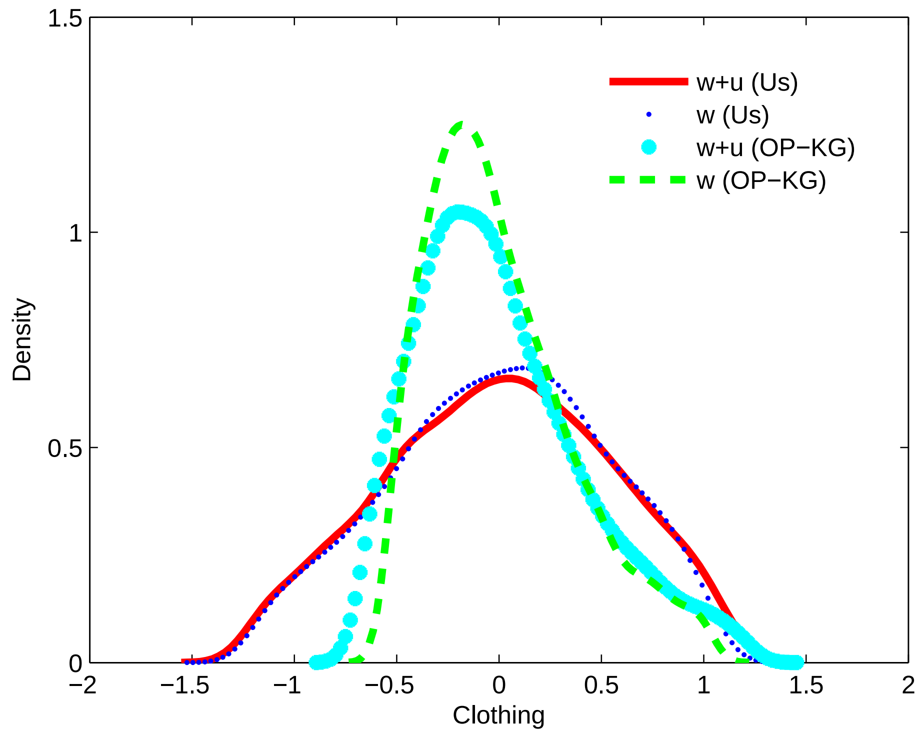
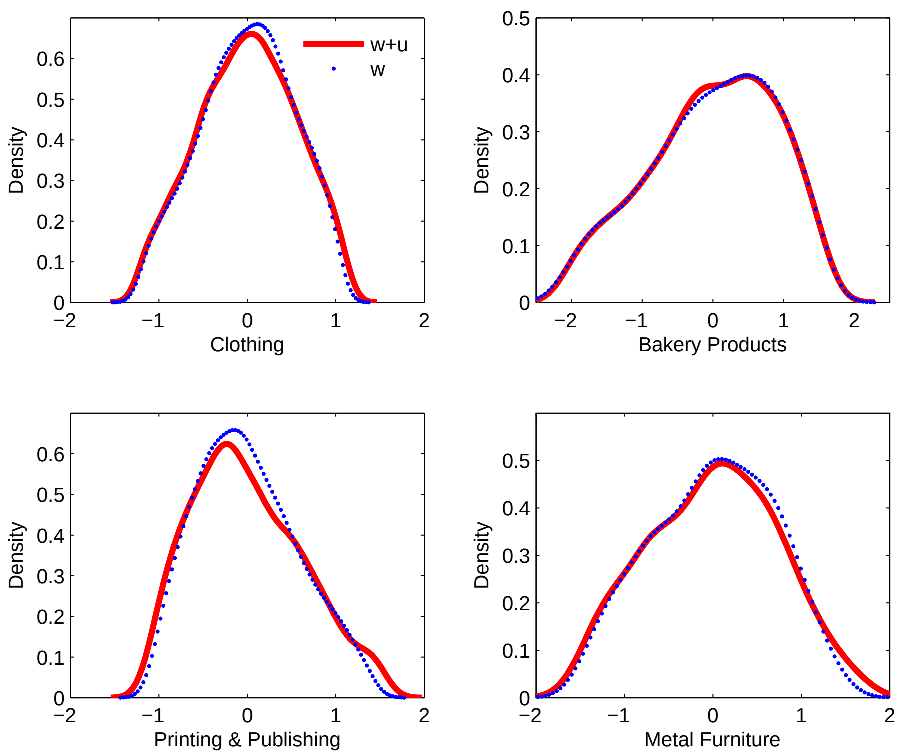


Figure A.2. Comparison: densities of $\hat{\omega}$ and $\hat{\omega} + \hat{u}$ from our method – small figures



Bibliography

Ackerberg, D. A. (2003): "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review*, 44(3), 1007–1040.

Ackerberg, D. A., K. Caves, and G. Frazer (2006): "Structural Identification of Production Function," UCLA, Deloitte and Touche, Rotman School of Management.

Albornoz, F., H. F. Calvo Pardo, G. Corcos, and E. Ornelas (2012): "Sequential exporting," *Journal of International Economics*, 88(1), 17–31.

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 58(2), 277–297.

Arkolakis, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of political economy*, 118(6), 1151–1199.

Atalay, E. (2012): "Materials Prices and Productivity," University of Chicago.

Aw, B. Y., M. Roberts, and D. Y. Xu (2011): "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 101, 1312–1344.

Baily, M. N., C. Hulten, D. Campbell, T. Bresnahan, and R. E. Caves (1992): "Productivity dynamics in manufacturing plants," *Brookings papers on economic activity. Microeconomics*, 1992, 187–267.

Baldwin, J. R., and P. Gorecki (1998): *The dynamics of industrial competition: A North American perspective*. Cambridge University Press.

Bartelsman, E., and M. Doms (2000): "Understanding Producivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

BASTOS, P., AND J. SILVA (2010): "The quality of a firm's exports: Where you export to matters," *Journal of International Economics*, 82(2), 99–111.

BEAULIEU, J., AND J. MATTEY (1999): "The effects of General Inflation and Idiosynchratic Cost Shocks on Within-Commodity Price Dispersion: Evidence from Microdata," *Review of Economics and Statistics*, 81(2), 205–216.

BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, pp. 1268–1290.

BILS, M., AND P. KLENOW (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.

BLUNDELL, R., AND S. BOND (2000): "GMM estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 19, 321–340.

CRAWFORD, G. S., AND M. SHUM (2005): "Uncertainty and learning in pharmaceutical demand," *Econometrica*, 73(4), 1137–1173.

DAS, S., M. J. ROBERTS, AND J. R. TYBOUT (2007): "Market Entry Costs, Producer Heterogeneity and Export Dynamics," *Econometrica*, 75(3), 837–873.

DE LA GRANDVILLE, O. (1989): "In Quest of the Slutsky Diamond," *American Economic Review*, 79(3), 468–481.

DE LA GRANDVILLE, O., AND R. M. SOLOW (2006): "A Conjecture on General Means," *Journal of Inequalities in Pure and Applied Mathematics*, 7(1).

DE LOECKER, J. (2011): "Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity," *Econometrica*, Vol. 79, No. 5, pp. 1407–1451.

DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2012): "Prices, Markups and Trade Reform," *NBER Working Paper 17925.*

DE LOECKER, J., AND F. WARZYNSKI (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102(6), 2437–71.

DEGROOT, M. H. (2005): *Optimal statistical decisions*, vol. 82. Wiley-Interscience.

DORASZELSKI, U., AND J. JAUMANDREU (2013): "R&D and productivity: Estimating endogenous productivity," *The Review of Economic Studies.*

Dunne, T., and M. J. Roberts (1992): "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," in *Empirical Studies in Industrial Organization: Essays in Honor of Leonard W. Weiss*, ed. by D. Audretsch, and J. Sigfried. Kluwer Academic Publishers.

Eaton, J., M. Eslava, D. Jinkins, C. J. Krizan, and J. Tybout (2013): "A Search and Learning Model of Export Dynamics," Working Paper, Department of Economics, The Pennsylvania State University.

Eaton, J., S. Kortum, and F. Kramarz (2011): "An anatomy of international trade: Evidence from French firms," *Econometrica*, 79(5), 1453–1498.

Epple, D., B. Gordon, and H. Sieg (2010): "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 100(3), 905–924.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler (2004): "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75, 333–371.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Fox, J., and V. Smeets (2011): "Does Input QUality Drive Measured Productivity Differences in Firm Productivity," *International Economic Review*, 52(4), 961–989.

Freund, C., and M. Pierola (2010): "Export entrepreneurs: evidence from Peru," *World Bank Policy Research Working Paper Series, Vol.*

Gandhi, A., S. Navarro, and D. Rivers (2013): "On the Identification of Production Functions: How Heterogeneous is Productivity?," University of Wisconsin-Madison and University of Western Ontario.

Grieco, P., and R. McDevitt (2012): "Productivity and Quality in Health Care: Evidence from the Dialysis Industry," Penn State Univ.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): "Export versus FDI with Heterogeneous Firms," *The American Economic Review*, 94(1), 300–316.

Irarrazabal, A. A., and L. D. Opromolla (2006): "Hysteresis in export markets," *New York University, Working Paper.*

KATAYAMA, H., S. LU, AND J. TYBOUT (2009): "Firm-Level Productivity Studies: Illusions and a Solution," *International Journal of Industrial Organization*, 27(3), 403–413.

KLEIN, L. R. (1953): "A Textbook of Econometrics," *Evanston, Illinois, and White Plains, New York: Row, Peterson and Company*, pp. pp. ix, 355.

KLETTE, T. J., AND Z. GRILICHES (1996): "The inconsistency of common scale estimators when output prices are unobserved and endogenous," *Journal of Applied Econometrics*, 11(4), 343–361.

KLUMP, R., AND O. DE LA GRANDVILLE (2000): "Economic Growth and the Elasticity of Subsitituion: Two Theorems and Some Suggestions," *American Economic Review*, 90(1), 282–291.

KLUMP, R., AND H. PREISSLER (2000): "CES Production Functions and Economic Growth," *Scandanavian Journal of Economics*, 102(1), 41–56.

KOMUNJER, I. (2012): "Global Identification in Nonlinear Models with Moment Restrictions," *Econometric Theory*, 28, 719–729.

KUGLER, M., AND E. VERHOOGEN (2012): "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 79(1), 307–339.

LEON-LEDESMA, M. A., P. MCADAM, AND A. WILLMAN (2010): "Identifying the Elasticity of Substitition with Biased Technical Change," *American Economic Review*, 100(4), 1330–1357.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70(2), pp. 317–341.

MARSCHAK, J., AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12(3/4), 143–205.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

NAGYPÁL, É. (2007): "Learning by doing vs. learning about match quality: Can we tell them apart?," *The Review of Economic Studies*, 74(2), 537–566.

NGUYEN, D. X. (2012): "Demand uncertainty: Exporting delays and exporting failures," *Journal of International Economics*, 86(2), 336–344.

OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), pp. 1263–1297.

ORNAGHI, C. (2006): "Assessing the Effects of Measurement Errors on the Estimation of Production Functions," *Journal of Applied Econometrics*, 21(6), 879–891.

RAUCH, J. E., AND J. WATSON (2003): "Starting small in an unfamiliar environment," *International Journal of Industrial Organization*, 21(7), 1021–1042.

ROBERTS, M., AND J. TYBOUT (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87, 545–564.

ROBERTS, M., D. XU, X. FAN, AND S. ZHANG (2012): "A Structural Model of Demand, Cost, and Export Market Selection for Chinese Footwear Producers," NBER Working Paper 17725.

ROBERTS, M. J., AND D. SUPINA (1996): "Ouput price, markups, and producer size," *European Economic Review*, 40, 909–921.

——— (2000): "Ouput Price and Markup Dispersion in Micro Data: The Roles of Producer Heterogeneity and Noise," in *Advances in Applied Microeconomics*, ed. by M. Baye, vol. 9. JAI Press.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

SANTOS, C. D. (2012): "Identifying demand shocks from production data," Tillburg University.

SOLOW, R. M. (1957): "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, Vol. 39, pp. 312–320.

STEVENSON, R. (1980): "Measuring technological bias," *The American Economic Review*, 70(1), 162–173.

SYVERSON, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–65.

VAN BEVEREN, I. (2010): "Total Facter Productivity Estimation: A Practical Review," *Journal of Economic Surveys*, 26, 98–128.

WALTERS, A. A. (1963): "Production and Cost Functions: An Econometric Survey," *Econometrica*, 31(1/2), pp. 1–66.

Wooldridge, J. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104, 112–114.

Zhang, H. (2012): "Biased Technology and Contribution of Technological Change to Economic Growth: Firm-Level Evidence from China," working paper, Pennsylvania State University.

Shengyu Li (Citizenship: China, F1 Visa)

Contact Information	Department of Economics The Pennsylvania State University 401 Kern Building University Park, PA, 16802	Cell: (814) 441-9564 E-mail: sul223@psu.edu Homepage: http://shengyuli.weebly.com/

Research Interests
Primary: Industrial Organization
Secondary: Applied Microeconomics

Education
Ph.D., Economics, the Pennsylvania State University, May 2014.
M.A., Economics, Shanghai University of Finance and Economics, June 2009.
B.A., Economics and Mathematics (with distinction), Wuhan University, June 2007.

Ph.D. Thesis
"Essays on Productivity, Uncertainty, and Firm Activities"
Thesis Advisor: Mark Roberts.

Working Papers
"A Structural Model of Productivity, Uncertain Demand, and Export Dynamics", 2013.
"Production Function Estimation with Unobserved Input Price Dispersion", with Paul Grieco and Hongsong Zhang, 2013.
"Technology Distance and FDI Spillovers with Factor-Biased Technology", with Hongsong Zhang. Work in progress.

Research Experience
RA, Prof. Mark Roberts, Penn State, Summer 2013 - Fall 2013.
RA, Prof. Guoqian Tian, Shanghai University of Finance and Economics (China), Fall 2007 - Spring 2009.

Teaching Experience
Instructor, Introduction to Econometrics, Penn State, Summer 2011.
TA, Empirical Methods in Economics (Graduate Level), Penn State, Spring 2013.
TA, Econometrics (Graduate Level), Penn State, Fall 2010 - Fall 2012.
TA, Intermediate Macroeconomics, Penn State, Fall 2009 - Spring 2010.

Conference Presentations
The Econometric Society North American Summer Meetings, CA, U.S., June 13 - 16, 2013.
The 11th Annual International Industrial Organization Conference, MA, U.S., May 17 - 19, 2013.
The 22nd Annual Meeting Of the Midwest Econometrics Group, KY, U.S., September 18 - 19, 2012.

Honors and Awards
Bao-steel Scholarship (merit-based), Bao-steel Education Fund, Shanghai, China, 2009.
Second Prize of National Graduates Mathematical Contest in Modeling, China, 2008.
Outstanding Graduates, Wuhan, China, 2007.
First Class Award, Wuhan University, 2003-2007.

Skills
- English (fluent), Chinese (native).
- Matlab, Stata.