Production Function Estimation with Unobserved Input Price Dispersion*

Paul L. E. Grieco[a], Shengyu Li[b], and Hongsong Zhang[c]

[a]The Pennsylvania State University
[b]Durham University
[c]The University of Hong Kong

December 30, 2014

Abstract

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. Our approach applies to a wide class of production functions and can be extended to accommodate 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 prediction 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.

Keywords: *production functions, unobserved price bias, productivity dispersion*

*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 Andrés Aradillas-López, Robert Porter, Joris Pinkse, 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. Correspondence: Grieco: `paul.grieco@psu.edu`, Li: `shengyu.li@durham.ac.uk`, Zhang: `hszhang@hku.hk`.

1 Introduction

In applications of production function estimation, many datasets do not contain a specific accounting of intermediate input prices and quantities, but instead only provide information on the total expenditure on intermediate inputs (i.e., materials). 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: firms will 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 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' profit-maximization decisions, 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.[1] The effect of input

[1]Klette and Griliches (1996) provide a structural approach for controlling for output price variation, we incorporate

price dispersion is similar. Using a unique data set containing both inputs price and quantity data, Ornaghi (2006) documents input price bias under the Cobb-Douglas production function.

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 inputs are often not available. The basic idea of our approach is to exploit the first order conditions of firms' profit maximization to recover the unobserved physical quantities of inputs from their expenditures.[2] We then use this recovered physical quantity of intermediate inputs to consistently estimate the model parameters. We illustrate our approach using the constant elasticity of substitution (CES) production function as our leading example. We then briefly discuss how the technique can be applied to more general production function specifications and incorporate the possibility that materials expenditure represents the aggregation of a vector of different intermediate inputs. These extensions are fully developed in the supplemental material.

Our model allows firms to be heterogenous in two unobserved dimensions: they have different total factor productivity and face different intermediate inputs prices. We are able to recover the joint distribution of unobserved heterogeneity and find that both productivity and input prices are important sources of heterogeneity across firms. Accounting for input price heterogeneity can give rise to richer explanations of firm policies. For example, if input prices are persistent, firms' exit decisions should be modeled as a cutoff in productivity levels and input prices, implying that relatively 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 has been employed in many other studies. Assuming homogeneous input prices, Gandhi, Navarro, and Rivers (2013) use the transformed first-order conditions of the firm's profit maximization problem to estimate the

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.

[2]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 2.2.2 we extend the model to consider the case where the firm chooses from several unobserved intermediate 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.

elasticity of substitution and separate the non-structural errors as the 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 recover 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—an alternative to traditional productivity measures—based on Bertrand-Nash equilibrium. Epple, Gordon, and Sieg (2010) develop a procedure using the first order condition of the indirect profit function to estimate the housing supply function. Zhang (2014) uses first order conditions as constraints to directly control for structural errors to estimate a production function with non-neutral technology shocks 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) 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 the earlier production function 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.[3]

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

[3] A share regression can consistently recover the production parameters when firms are price-takers in output market and technology shows constant return to scale (but may be biased as Walters (1963) points out). For many applications, Cobb-Douglas is a good approximation of production function. However, 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. Therefore, we think it is more realistic to recognize the dispersion of expenditure share, especially when the purpose is to consider firm behavior.

the observation of material price (or quantity), uses the relationship between expenditures and quantities of labor and materials 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 or input prices are homogenous, 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 or assume homogeneity.

We demonstrate our approach 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 well. In contrast, the traditional proxy approach causes systematic biases in the parameter estimates. In particular, the elasticity of substitution is underestimated in the proxy approach. This is an intuitive implication of unobserved input price bias as expenditure variation reflects the combined impact of price differences and quantity differences. Moreover, the distribution parameters are also biased. This bias could mislead researchers attempting to 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 studies (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 introduces a model with unobserved price heterogeneity and outlines our procedure to consistently estimate the model parameters. Section 3 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 4, and conclude in Section 5.

2 A Model with Unobserved Price Dispersion

In this section, we introduce a model of firms' decision-making in a standard monopolistically competitive output market. We use this model to show how the production function parameters can be identified and estimated without resorting to a proxy for unobserved materials quantities. Instead of substituting quantities with deflated expenditure, our approach exploits the first order conditions implied by profit maximization to recover unavailable physical quantities of intermediate inputs from expenditures and labor input quantity. For ease of exposition, the following section presents the model for the CES production function specification with a scalar intermediate input. Section 2.2 discuses extensions to general parametric forms and a vector of intermediate inputs. These extensions, as well as an illustration applying our method to the translog production function specification, are fully developed in the Supplemental Material.

2.1 The CES Production Function with Scalar Input

In this section, we present our approach for the constant elasticity production function and Dixit-Stiglitz demand.[4] It has been commonly recognized that the CES production function needs to be normalized to give meaningful interpretation of its parameters. A branch of the literature

[4]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 the scale parameter and demand elasticity are not separately identified without additional assumptions. For example, if Markov process of productivity is assumed, the scale parameter and the demand elasticity can be separately identified. We demonstrate the use of a Markov timing assumption to aid identification in the Supplemental Material (Online Appendix 2).

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, McAdam, and Willman, 2010). We follow this literature and normalize the CES production function according to the geometric mean.[5] Specifically, in each period t, a firm j produces a quantity Q_{jt} of a single homogeneous output using labor (L_{jt}), intermediate material (M_{jt}), and capital (K_{jt}) via the production function. 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 $\overline{X} = \sqrt[n]{X_1 X_2 \cdots X_n}$.[6] Then the normalized CES production function can be written as,

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

where ω_{jt} is a Hicks-neutral productivity shock observed by the firm (but not by researchers). The parameters to estimate are $\theta = (\gamma, \alpha_L, \alpha_M, \alpha_K)$. The elasticity of substitution (σ) is determined by γ, where $\gamma = \frac{\sigma - 1}{\sigma}$. The distribution parameters $\alpha_L, \alpha_M, \alpha_K$ are restricted to sum to 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 labor and capital ($\overline{L}, \overline{K}$) are computable using the observed data, and will be convenient to use in constructing an additional restriction to identify the distribution parameters.[7] 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.

Firms are monopolistically competitive and face an inverse demand function which we assume is Dixit-Stiglitz,

$$P_{jt} = P_t(Q_{jt}; \eta) = P_t \left(\frac{Q_{jt}}{Q_t} \right)^{\frac{1}{\eta}}, \tag{2}$$

[5] For the detail of this normalization and how we implement it in this paper, see the supplemental material (Online Appendix 4).

[6] In principle, any point $Z_0 = (Q_0, L_0, M_0, K_0)$ (which satisfies normalization conditions in Online Appendix 4) can be chosen as the baseline point, for example a default choice could be $(1, 1, 1, 1)$.

[7] Of course, neither $\overline{M}$ nor $\overline{Q}$ is computable using the observed data, since we do not observe M_{jt} or Q_{jt} for any firm. This has two implications. First, we will recover materials usage relative to the geometric mean ($M_{jt}/\overline{M}$) instead of materials directly (M_{jt}). Second, the absolute level of ω_{jt} absorbs $\overline{Q}$, however, since $\overline{Q}$ is a constant, the change and dispersion of ω_{jt} over time and across firms are still meaningful.

where Q_t and P_t are industry-level output quantity and price in period t, and $\eta < -1$ is the demand elasticity. We make the following assumptions:

Assumption 1 (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. Prices have strictly positive support.*

Assumption 2 (Profit Maximization)**.** *After observing their productivity draw, ω_{jt}, and firm-specific input prices, 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}.*

Assumption 3 (Data)**.** *The researcher observes revenue R_{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}. However, she does not observe firms' productivity ω_{jt}, or the prices and quantities of either outputs (i.e., P_{jt} and Q_{jt}) or materials inputs (i.e., $P_{M_{jt}}$ and M_{jt}). All these variables are observed (or chosen) by the firm.*

Assumption 1 is our primary departure from the earlier literature, it weakens the typical assumption that input prices are homogeneous when they are not observed. The assumption that firms are price takers does not preclude them 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.

Assumption 2 is common in the literature, it is needed since our approach relies on profit maximization conditions. One restriction of Assumption 2 is that it assumes labor and materials are both fully flexible. Some in the literature (e.g., Arellano and Bond, 1991; Ackerberg, Caves, and Frazer, 2006) allow adjustment costs in labor, but their methods require an implicit assumption on homogenous input price 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 quasi-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 in addition to the well-known Hicks-neutral technology shifter, ω_{jt}.

Finally, Assumption 3 merely formalizes our assumption that only materials expenditure, rather than materials prices and quantities are observed. Here we assume that materials is a scalar, homogeneous input. We will provide a structural interpretation of our estimates in the case where materials expenditure is an aggregation of expenditure on several different unobserved materials types in Section 2.2.2.

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.

Given our assumptions, the firm chooses its own labor and material input quantities to maximize its period profit after observing capital stock, K_{jt}, productivity shock, ω_{jt}, and input prices $P_{L_{jt}}$ and $P_{M_{jt}}$. The firm's decision problem is:

$$\begin{aligned} &\max_{L_{jt}, M_{jt}} P_t(Q_{jt};\eta)Q_{jt} - P_{Ljt}L_{jt} - P_{Mjt}M_{jt} \\ &\text{s.t. } Q_{jt} = \exp(\omega_{jt})F(L_{jt}, M_{jt}, K_{jt};\theta). \end{aligned} \tag{3}$$

The corresponding first order conditions are,

$$\begin{aligned} \exp(\omega_{jt})F_{L_{jt}}\left[P_t(Q_{jt};\eta) + Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right] &= P_{L_{jt}}, \\ \exp(\omega_{jt})F_{M_{jt}}\left[P_t(Q_{jt};\eta) + Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right] &= P_{M_{jt}}, \end{aligned} \tag{4}$$

where $F_{L_{jt}}$ and $F_{M_{jt}}$ are the partial derivatives of $F(\cdot)$—the CES production function given in (1)—with respect to labor and material. Given our assumptions on the production and demand functions, a finite solution to the profit maximization problem (3) exists.[8] Dividing the two first

[8]We provide assumptions for more general forms of the production and demand functions in the supplemental

order conditions, multiplying both sides by $\frac{L_{jt}}{M_{jt}}$, and rearranging yields,

$$\frac{F_{L_{jt}} L_{jt}}{F_{M_{jt}} M_{jt}} = \frac{\alpha_L}{\alpha_M} \left(\frac{(L_{jt}/\overline{L})}{(M_{jt}/\overline{M})} \right)^{\gamma} = \frac{E_{L_{jt}}}{E_{M_{jt}}} \tag{5}$$

where $E_{L_{jt}} = P_{L_{jt}} L_{jt}$ and $E_{M_{jt}} = P_{M_{jt}} M_{jt}$ are expenditures on labor and material and the first equality makes use of our CES specification.

Equation (5) is the key to our approach. It relates the ratio of input quantities to the ratio of input expenditures, which is observable to the researcher. Given that firms choose their inputs optimally at an interior solution of profit maximization, (5) is always satisfied at the firm choice of (L_{jt}, M_{jt}). The key question is whether (5) places enough restrictions on the unobserved material quantity M_{jt} so that we can uniquely recover it from the observed data, $(L_{jt}, K_{jt}, E_{Mjt}, E_{Ljt})$, up to production function parameters. For the CES specification, it is straightforward to see that as long as $\gamma \neq 0$, the unique (normalized) level of materials that solves (5) is,

$$\frac{M_{jt}}{\overline{M}} = \left(\frac{\alpha_L}{\alpha_M} \frac{E_{M_{jt}}}{E_{L_{jt}}} \right)^{\frac{1}{\gamma}} \frac{L_{jt}}{\overline{L}}. \tag{6}$$

Intuitively, variation in the expenditure ratio, coupled with the first order conditions for materials and labor use, provides information that can be used to separate materials prices and quantities. The key exception is when $\gamma = 0$—the special Cobb-Douglas case of unit elasticity of substituion. The failure of the method under Cobb-Douglas is instructive: 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 the relative price ($\frac{P_L}{P_M}$). As a result, the expenditure ratio $\frac{E_{L_{jt}}}{E_{M_{jt}}}$ remains constant ($\frac{\alpha_L}{\alpha_M}$), materials quantity drops out of (5), and we cannot separate the price and quantity of materials from the information on the expenditure ratio. However, for all other elasticities of substitution, variation in the expenditure ratio reflects the optimal response to changes in the price ratio, and can be used together with observed labor inputs to uncover materials quantities and prices. Fortunately, it is easy to test for the Cobb-Douglas case by checking whether or not

material (Online Appendix 1).

the expenditure ratio does in fact vary in the data. As long as this variation exists, our approach illustrates how to make use of it to consistently estimate the production function parameters. Next, to recover the unobserved productivity term, we can substitute (6) into the first order condition for labor in (4),[9]

$$\omega_{jt} = \frac{\eta}{1+\eta} \log \left\{ \frac{1}{\alpha_L} \frac{\eta}{1+\eta} \frac{Q_t^{1/\eta}}{P_t} \left(\frac{L_{jt}}{\overline{L}} \right)^{-\gamma} \frac{E_{L_{jt}}}{\overline{Q}^{\frac{1+\eta}{\eta}}} \left[\alpha_L \left(\frac{E_{L_{jt}} + E_{M_{jt}}}{E_{L_{jt}}} \right) \left(\frac{L_{jt}}{\overline{L}} \right)^{\gamma} + \alpha_K \left(\frac{K_{jt}}{\overline{K}} \right)^{\gamma} \right]^{-\frac{1}{\gamma}\left(1+\frac{1}{\eta}\right)} \right\}. \tag{7}$$

We now derive the primary estimating equation. Since output quantities are not directly observed, we follow Klette and Griliches (1996) and use the revenue function to estimate both demand and production parameters. The revenue equation is,

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

Where R_{jt} is the observed revenue of the firm, Q_{jt} 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,

$$\ln R_{jt} = \ln P_t \left(e^{\omega_{jt}} F(L_{jt}, M_{jt}, K_{jt}; \theta); \eta \right) + \ln \left[e^{\omega_{jt}} F(L_{jt}, M_{jt}, K_{jt}; \theta) \right] + u_{jt}.$$

Given our specification for the production function (1) and demand function (2), we use (6) to substitute out materials and (7) to substitute out ω_{jt} to derive,[10]

$$\ln R_{jt} = \ln \frac{\eta}{1+\eta} + \ln \left[E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^{\gamma} \right) \right] + u_{jt}. \tag{8}$$

It is easy to see that while (8) provides identification of the elasticity of substitution 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

[9] See Appendix A for the derivation of ω_{jt} in the CES case.
[10] See Appendix A for the complete derivation.

allow us to identify the distribution parameters. The first is simply the adding up constraint of the distribution parameters; the second is implied by profit maximization. To see this, recall that the following equality holds for every observation,

$$\frac{\alpha_L (L_{jt}/\overline{L})^\gamma}{\alpha_M (M_{jt}/\overline{M})^\gamma} = \frac{P_{L_{jt}} L_{jt}}{P_{M_{jt}} M_{jt}} \equiv \frac{E_{L_{jt}}}{E_{M_{jt}}}. \tag{9}$$

Taking the geometric mean of (9) across all observations implies,[11]

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

where $\overline{E}_M$ and $\overline{E}_L$ 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:

$$\begin{aligned}
\hat{\beta} \quad &= \quad \text{argmin}_\beta \sum_{jt} \left[\ln R_{jt} - \ln \frac{\eta}{1+\eta} - \ln \left\{ E_{M_{jt}} + E_{L_{jt}} \left(1 + \frac{\alpha_K}{\alpha_L} \left(\frac{K_{jt}/\overline{K}}{L_{jt}/\overline{L}} \right)^\gamma \right) \right\} \right]^2 \\
& \quad \text{subject to} \qquad \frac{\alpha_M}{\alpha_L} = \frac{\overline{E}_M}{\overline{E}_L}, \qquad (10) \\
& \qquad\qquad\qquad\quad \alpha_L + \alpha_M + \alpha_K = 1, \qquad (11)
\end{aligned}$$

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

To make identification more transparent, the problem can alternatively be cast in a GMM framework. Write the nonlinear equation (8) as $r_{jt} = f(\mathbf{w}_{jt}; \beta) + u_{jt}$, where $f(\mathbf{w}_{jt}; \beta)$ is the right hand side of (8) without u_{jt}. The restrictions (10) and (11) can be viewed as degenerate moment restrictions:

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

Stacking these together with the moments from the revenue function, we have a vector of moments,

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

$\mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta) = [\nabla_\beta f(\mathbf{w}_{jt}; \beta)(r_{jt} - f(\mathbf{w}_{jt}; \beta)); \mathbf{h}(\mathbf{x}_{jt}; \beta)]$ which give us the following GMM problem,

$$\hat{\beta} = \text{argmin}_\beta \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\right]' W \left[\frac{1}{n}\sum_{jt} \mathbf{m}(\mathbf{w}_{jt}, \mathbf{x}_{jt}; \beta)\right]. \tag{12}$$

To see that (12) identifies all of the parameters, define the matrix,

$$\Phi(\beta) = E\left[\Big(\nabla_\beta f(\mathbf{w}_{jt}; \beta)\Big)\Big(\nabla_\beta f(\mathbf{w}_{jt}; \beta)\Big)'\right] = \begin{bmatrix} E[f_\eta f_\eta] & E[f_\eta f_{\alpha_L}] & 0 & E[f_\eta f_{\alpha_K}] & E[f_\eta f_\gamma] \\ E[f_{\alpha_L} f_\eta] & E[f_{\alpha_L} f_{\alpha_L}] & 0 & E[f_{\alpha_L} f_{\alpha_K}] & E[f_{\alpha_L} f_\gamma] \\ 0 & 0 & 0 & 0 & 0 \\ E[f_{\alpha_K} f_\eta] & E[f_{\alpha_K} f_{\alpha_L}] & 0 & E[f_{\alpha_K} f_{\alpha_K}] & E[f_{\alpha_K} f_\gamma] \\ E[f_\gamma f_\eta] & E[f_\gamma f_{\alpha_L}] & 0 & E[f_\gamma f_{\alpha_K}] & E[f_\gamma f_\gamma] \end{bmatrix}.$$

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, define $\Psi(\beta)$ as,

$$\Psi(\beta) = E\Big[\nabla_\beta \mathbf{h}(\mathbf{x}_{jt}; \beta)\Big] = \begin{bmatrix} 0 & -\frac{\alpha_M}{\alpha_L^2} & \frac{1}{\alpha_L} & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

It is clear that the rank of the sub matrix containing column 2, 3, and 4 of $\Phi(\beta)$ is two. Thus, the information matrix for the GMM problem at the true parameter β_0, $V(\beta_0) = [\Phi(\beta_0); \Psi(\beta_0)]$, has full column rank. Following Rothenberg (1971), we can conclude that all parameters are locally identified. Since the specification is just-identified, the GMM implementation with any positive semi-definite weight matrix is equivalent to the nonlinear least square estimation with constraints. We show consistency and present the asymptotic distribution of this estimator for the general parametric case in the supplemental material (Online Appendix 1).

2.2 Extensions

2.2.1 General Parametric Forms

While we have focused on the CES production function for concreteness, our approach broadly applies to a large class of parametric production functions. In this section, we briefly discuss how the technique generalizes to other parametric specifications. Full details, as well as a separate discussion of implementation for the translog production function are provided in the supplemental material.

There are two places where the parametric form plays a key role. The first is (5), which we use to back out the unobserved materials quantity. For the CES case (and the translog case in Online Appendix 2) we are able to derive a closed form for materials. In general, we need to ensure that (5) exhibits a unique solution. The supplemental material (Online Appendix 1) offers conditions which imply the existence of a unique solution for materials.

The second place where we appeal to the CES specification is in the derivation of (10), which exploits the relationship between the geometric mean and the CES functional form to derive an additional moment condition based on the geometric mean expenditure ratio in the population. While this moment restriction is particular to the CES specification, similar population moments may be available for other functional forms. More generally, timing assumptions may also be used to provide additional moments. For example, in Online Appendix 2, we show how the additional assumption that productivity moves according to a Markov process—which is commonly employed in the production function literature—can provide moment restrictions with which to identify all the parameters for the translog specification. This second approach can be used with any functional form to provide identifying moment restrictions (including the CES, if it were necessary).

2.2.2 Multiple Materials Inputs

We have followed the literature in assuming that firms purchase a single homogeneous intermediate input. Indeed, the ability to treat the recovered 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.[12] In reality, intermediate input expenditures are an aggregate of a wide variety of different input goods. Unfortunately, datasets typically contain only the total material 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 material expenditure information.

To be specific, suppose the firm may use up to D_M different types of materials. Denote the vector of material 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 (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 researchers. Assume inputs enter into the production function as,

$$Q_{jt} = e^{\omega_{jt}} F(L_{jt}, \mu(M_{jt}), K_{jt}; \theta), \tag{13}$$

where $\mu : \mathbb{R}_+^{D_M} \to \mathbb{R}_+$ is a homogeneous index function which summarizes the contribution of all materials inputs to production.[13] As part of the production function, we assume that μ is known to the firm. The corresponding idiosyncratic material prices for each component is summarized in price vector $P_{Mjt} = (P_{M1jt}, P_{M2jt}, \ldots, P_{MD_M jt})$, 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, input prices, and capital stock. The first order conditions for L_{jt} and

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

[13]We assume that firms optimally purchase a positive amount of all goods so (15) holds. To accommodate the possibility that some firms do not use some inputs, we can allow for a discrete choice between homogeneous production technologies, e.g., $\mu(M_{jt}) = \max(\mu^1(M_{jt}^1), \mu^2(M_{jt}^2))$ where $M_{jt} = (M_{jt}^1, M_{jt}^2)$ and $\mu^1(\cdot)$ and $\mu^2(\cdot)$ are homogeneous functions of the same degree. Then, only the first order conditions with respect to the profit maximizing technology are relevant. We will use this more general setup in the Monte Carlo experiment in Online Appendix 3.

all components of the vector M_{jt} are:

$$\exp(\omega_{jt})F_{L_{jt}}\left[P_t(Q_{jt};\eta)+Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right] = P_{L_{jt}}, \tag{14}$$

$$\exp(\omega_{jt})F_{\mu_{jt}}\left[P_t(Q_{jt};\eta)+Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right]\mu_d(M_{jt}) = P_{Md_{jt}}, \qquad \forall d=1,2,\ldots,D_M \tag{15}$$

where $\mu_d(M_{jt})=\frac{\partial\mu(M_{jt})}{\partial M_{djt}}$.

Denote the optimal choice of the firm as L^*_{jt} and the 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 (15) can be summarized into a single equation by multiplying (15) by M^*_{djt}, summing across d, and dividing it by $\psi(M^*_{jt})$,

$$\exp(\omega_{jt})F_{\mu_{jt}}\left[P_t(Q_{jt};\eta)+Q_{jt}\frac{\partial P_t(Q_{jt};\eta)}{\partial Q_{jt}}\right]=P_{\mu_{jt}}. \tag{16}$$

This equation together with (14) can be viewed as the first order conditions of the firm's optimization problem if it faced labor price $P_{L_{jt}}$ and a material price $P_{\mu_{jt}}$ for *single* material input μ. Our method can now be applied as in the scalar materials input case. We show this formally in Proposition 3 of the supplemental materials (Online Appendix 3). 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, 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 for a complicated functional form of $\mu(\cdot)$ through a Monte Carlo experiment in the supplemental materials (Online Appendix 3). Moreover, the results of our empirical application in Section 4 can be interpreted either through the traditional lens of a homogeneous materials input or the more general assumptions of a vector of unknown inputs with a homogeneous aggregator assumption.

3 Monte Carlo Experiments

This section presents Monte Carlo experiments that evaluate the performance of our method, and show how it corrects for input price heterogeneity. We first describe the data generation process, then estimate the model in three different ways based on assumed data availability.

3.1 Data Generation

Here we briefly describe the data generation process used in the Monte Carlo; a full description is provided in Appendix B. Using the CES specification of the production function (1) and a Dixit-Siglitz demand system (2), 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. For each firm, we simulate a sequence of productivity ω_{jt} and input prices ($P_{L_{jt}}$ and $P_{M_{jt}}$) over time. 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. 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,

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

Which is compatible with the assumptions of Olley and Pakes (1996) and this paper (although our approach does not make use of the investment decision).

In this way, we generate a data set of $\{\omega_{jt}, K_{jt}, I_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, R_{jt}\}$ 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. Table 1 lists the underlying parameters used to generate the data set.

3.2 Our Method

We first estimate the model with our method. In this case, we assume a researcher observes $\{K_{jt}, L_{jt}, E_{L_{jt}}, E_{M_{jt}}, R_{jt}\}$ 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 recover firm-level material quantities from labor quantities and expenditures. This approach allows us to estimate the production function while controlling for unobserved price and productivity 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.

3.3 Traditional Method with Direct Proxy

For our first point of comparison, we estimate the model using a direct proxy method that 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 "proxy-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 method, the proxy-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.[14]

Specifically, researchers using this method observe $\{K_{jt}, L_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, I_{jt}\}$ and estimate parameters via the (logarithm) production function:

$$\ln\left(\frac{Q_{jt}}{\overline{Q}}\right) = \frac{1}{\gamma}\ln\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma} + \alpha_M\left(\frac{E_{M_{jt}}}{\overline{E}_M}\right)^{\gamma} + \alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right] + \omega_{jt} + u_{jt}, \tag{17}$$

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

It is well-known that the unobservable firm-level heterogeneity ω_{jt} causes transmission bias. To control for endogeneity bias, Olley and Pakes (1996) assume that productivity follows a first

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

order Markov process. Following our data generating process, we are more specific and assume that productivity follows an AR(1) process,

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

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 this into (17) yields,[15]

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

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.[16] 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 K_{jt+1} and variables up to period t, so we can construct the

[15] 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.

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

set of moment conditions with which we estimate the model's parameters,

$$E(\epsilon_{jt+1}\mathbf{x}_{jt+1}) = 0, \tag{18}$$

where $\epsilon_{jt+1}(\alpha_L, \alpha_M, \alpha_K, \gamma, g_0, g_1) = \omega_{jt+1} - (g_0 + g_1\omega_{jt})$, and $\mathbf{x}_{jt+1}$ is a vector of variables that are uncorrelated with the innovation term in period $t+1$, e.g., $L_{jt}, E_{L_{jt}}, E_{M_{jt}}, K_{jt}, K_{jt+1}$.[17]

3.4 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 $\{K_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, I_{jt}\}$ for each firm and each period. This enables us to estimate the production function in (17) without using expenditure as a proxy. The only difference between the oracle case and the previous proxy-OP procedure is that material quantity is not substituted by its proxy, since the true quantity is observable in this case. In comparison to our method, this method requires that the researcher observes investment, output quantity, and materials input quantities.

3.5 Results

The results of the Monte Carlo experiments for three different true elasticities of substitution are presented in Table 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 proxy-OP. 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

[17] In the Monte Carlo experiment, we choose

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

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

the oracle-OP method and the method we propose, despite the fact that we do not use investment, output quantity, or material 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. The intuition for a downward bias in the elasticity of substitution is straightforward. 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 prices P_M/P_L may induce an opposing change in the input quantity ratio M/L, but the effects are partially offset when only the expenditure ratio $E_M/E_L = (P_M/P_L) \times (M/L)$ is observed.[18] As expected, when we allow the researcher to observe input quantities directly, the oracle-OP method performs well.

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 recovered from (6). Figure 2 presents the kernel density estimation of the recovered material prices from our method and compares it to the true density of material prices in the Monte Carlo.[19] It shows that the recovered material price density matches the true density quite well.

4 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).[20] This application serves two purposes. The first is to compare our results with those found using the traditional proxy method to account for

[18]We formally derive the bias in the elasticity of substitution for the CES specification in a working paper version of this paper.

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

[20]For a detailed introduction to the data set, see Roberts and Tybout (1997).

unobserved material input quantities. The second is to illustrate additional information which can be recovered 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), and investment (I). However, firm-level price information about material input and output is not available. Moreover, only total expenditure on "raw materials, materials and packaging" (E_M) rather than total quantities (M) is available. 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 2.2.2, our method can also be employed if firms are optimally choosing a vector of heterogenous material inputs. In this case, the recovered material price index represents the shadow price of increasing the use of material 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.[21]

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. 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). We refer to this 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 the supplemental materials (Online Appendix 5), 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).

[21] For the sake of simplicity as well as the comparability to the traditional proxy-based methods, we normalize κ (the degree of homogeneity of μ) to be 1 in the application.

Estimates for four large industries are displayed in Table 3: clothing, bakery products, printing and publishing, and metal furniture.[22] 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 pattern shown in the Monte Carlo experiments. Moreover, the elasticity of substitution estimates are significantly greater than one in all industries when using 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 will lead to inconsistent estimates of elasticities of substitution, and that our method is capable of controlling for unobserved price dispersion.

Biased estimates of the 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 proxy-OP method. They suggest 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 3 shows the productivity distributions estimated using our method and the OP-KG method for each of the four industries.[23] 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.

[22]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.

[23]Figure 3 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 the supplemental materials (Online Appendix 5), 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.

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 in Table 5. 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.[24] Our finding indicates that the "true" productivity 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 recovered prices are purged of the effect of quality differences. In Figure 4, we present the kernel density estimations of recovered 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 and identified through variation in firms expenditure ratios, they suggest that quality differences alone may not fully account for input price dispersion.

We are also able to use our method to analyze the dynamics of input price dispersion. While

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

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. The results are reported in Table 4. In all four industries, there is quite high persistence with mean around 0.75, which is close to the persistence reported in Atalay (2012) where 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 recovered 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 recovered 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 positive correlation between input prices and productivity within industries.

5 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 to jointly recover firm-level materials quantities and prices together with productivity from observable data on revenues, labor quantities, and expenditures.

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 larger productivity dispersion in the industries we study than would be uncovered using a traditional proxy-based estimator.

References

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

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.

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

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

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.

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

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.

Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity: Estimating Endogenous Productivity," *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.

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.

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.

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.

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.

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.

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.

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. 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.

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.

Syverson, C. (2011): "What Determines Productivity?," *Journal of Economic Literature*, 49(2), 326–65.

Walters, A. A. (1963): "Production and Cost Functions: An Econometric Survey," *Econometrica*, 31(1/2), pp. 1–66.

Zhang, H. (2014): "Non-neutral Technology, Firm Heterogeneity, and Labor Demand," Hong Kong University.

Appendices

Appendix A Details of Implementation for CES Specification

In this appendix, we explicitly derive 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:

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

where the production function is,

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

and the inverse demand function is,

$$P_{jt} = P_t\left(\frac{Q_{jt}}{Q_t}\right)^{1/\eta}.$$

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 respect 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:

$$\frac{M_{jt}}{\overline{M}} = \left[\frac{\alpha_L E_{M_{jt}}}{\alpha_M E_{L_{jt}}}\right]^{\frac{1}{\gamma}}\frac{L_{jt}}{\overline{L}}. \tag{19}$$

This implies that material quantity can be recovered 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 we have

$$e^{\frac{1+\eta}{\eta}\omega_{jt}}\alpha_L\frac{1+\eta}{\eta}\frac{P_t}{Q_t^{1/\eta}}\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma-1}\frac{\overline{Q}^{\frac{1+\eta}{\eta}}}{\overline{L}}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma\eta}+\frac{1}{\gamma}-1}=P_{L_{jt}}. \tag{20}$$

Multiply both sides by L_{jt} and use $E_{L_{jt}} = L_{jt}P_{L_{jt}}$,

$$e^{\frac{1+\eta}{\eta}\omega_{jt}}\alpha_L\frac{1+\eta}{\eta}\frac{P_t}{Q_t^{1/\eta}}\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}\overline{Q}^{\frac{1+\eta}{\eta}}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma\eta}+\frac{1}{\gamma}-1}=E_{L_{jt}}. \tag{21}$$

Now substitute (6) into above equation and re-arrange,

$$e^{\frac{1+\eta}{\eta}\omega_{jt}}=\frac{1}{\alpha_L}\frac{\eta}{1+\eta}\frac{Q_t^{1/\eta}}{P_t}\left(\frac{L_{jt}}{\overline{L}}\right)^{-\gamma}\frac{E_{L_{jt}}}{\overline{Q}^{\frac{1+\eta}{\eta}}}\left[\alpha_L\left(\frac{E_{L_{jt}}+E_{M_{jt}}}{E_{L_{jt}}}\right)\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{-\frac{1}{\gamma}(1+\frac{1}{\eta})} \tag{22}$$

Which can be solved for ω_{jt} to yield (7) in the main text. However, it will be more straightforward to substitute (22) directly when deriving the estimating equation. Which we turn to now.

First, we plug the production function and the inverse demand function into the revenue equation,

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

where u_{jt} is the measurement error, and get

$$\begin{aligned}R_{jt} &= \exp(u_{jt})\frac{P_t}{Q_t^{\frac{1}{\eta}}}Q_{jt}^{1+\frac{1}{\eta}}\\ &= \exp(u_{jt})\frac{P_t}{Q_t^{\frac{1}{\eta}}}e^{\frac{1+\eta}{\eta}\omega_{jt}}\left[\alpha_L\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_M\left(\frac{M_{jt}}{\overline{M}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}(1+\frac{1}{\eta})}.\end{aligned}$$

Plug in (6) into above equation and we have:

$$R_{jt} = \exp(u_{jt})\frac{P_t}{Q_t^{\frac{1}{\eta}}}e^{\frac{1+\eta}{\eta}\omega_{jt}}\left[\alpha_L\left(\frac{E_{L_{jt}}+E_{M_{jt}}}{E_{L_{jt}}}\right)\left(\frac{L_{jt}}{\overline{L}}\right)^{\gamma}+\alpha_K\left(\frac{K_{jt}}{\overline{K}}\right)^{\gamma}\right]^{\frac{1}{\gamma}(1+\frac{1}{\eta})}. \tag{23}$$

Using (22) to substitute out $e^{\frac{1+\eta}{\eta}\omega_{jt}}$,

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

Take logs to arrive at the estimating equation,

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

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

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

As discussed in the paper, this nonlinear least square estimation with constraint is equivalent to the GMM estimator defined in (12).

Appendix B 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).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.

There are J firm in production for T periods. The evolution 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 $sd(\omega_0)$.

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 deviation $sd(K_0)$.

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

$$P_{L_{jt}} = P_{L_t} e^{\varepsilon_{jt}^{PL}},$$

$$P_{M_{jt}} = P_{M_t} e^{\varepsilon_{jt}^{PM}},$$

where shocks ε_{jt}^{PL} and ε_{jt}^{PM} are deviations from the industrial-level input prices P_{L_t} and P_{M_t} (which are set to be constant 0.1 for simplicity), and these shocks are independently drawn from $N(0, sd(\varepsilon^{PL}))$ and $N(0, sd(\varepsilon^{PM}))$ respectively.

After simulating $\{\omega_{jt}, K_{jt}, I_{jt}, P_{L_{jt}}, P_{M_{jt}}\}$ for each firm j and period t,[25] we 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 as,

$$L_{jt} = \left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{1}{\gamma-1}} M_{jt}, \tag{24}$$

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

$$M_{jt} = \left[\frac{(e^{-\omega_{jt}}Q_{jt})^{\gamma} - \alpha_K K^{\gamma}}{\alpha_M + \alpha_L\left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{\gamma}{1-\gamma}}}\right]^{\frac{1}{\gamma}}, \tag{25}$$

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

$$\frac{\eta+1}{\eta}\left(\frac{P_t}{Q_t^{\frac{1}{\eta}}}\right) Q_{jt}^{\frac{1}{\eta}} = e^{-\omega_{jt}} \frac{\left[P_{M_{jt}} + P_{L_{jt}}\left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{1}{\gamma-1}}\right]}{\left[\alpha_M + \alpha_L\left(\frac{\alpha_M P_{L_{jt}}}{\alpha_L P_{M_{jt}}}\right)^{\frac{\gamma}{1-\gamma}}\right]^{\frac{1}{\gamma}}} \left(1 - \alpha_K K^{\gamma}(e^{-\omega_{jt}}Q_{jt})^{-\gamma}\right)^{\frac{1}{\gamma}-1}. \tag{26}$$

Given the derived variables and underlying true parameters, (26) is only about Q_{jt}. It is easy to verify that (26) 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 (24) and (25). Hence, the expenditures of input are given by $E_{L_{jt}} = P_{L_{jt}} L_{jt}$ and $E_{M_{jt}} = P_{M_{jt}} M_{jt}$. At last, firm level output

[25] In addition, for simplicity, we normalize the industrial level index P_t and Q_t as 1 since we do not focus on aggregate shocks.

price P_{jt} is calculated by inverting Dixit-Stiglitz demand,

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

and the firm-level revenue is

$$R_{jt} = P_{jt} Q^*_{jt} \exp(u_{jt}).$$

where $u_{jt} \sim N(0, sd(u))$ is the measurement error, or unanticipated productivity or demand shock.

Hence, we have generated a data set of $\{\omega_{jt}, K_{jt}, I_{jt}, L_{jt}, M_{jt}, E_{L_{jt}}, E_{M_{jt}}, Q_{jt}, R_{jt}\}$ for each firm j and period t.[26]

[26] In the OP and Oracle-OP cases, researchers observe output quantity a measurement error. So we generate the observed quantity as $Q_{jt} = Q^*_{jt} e^{\varepsilon^q_{jt}}$ where $\varepsilon^q_{jt} \sim N(0, sd(\varepsilon^q))$ is the measurement error. To make the error terms consistent, we set $sd(\varepsilon^q) = sd(u)$.

Table 1: Monte Carlo Parameter Values[1]

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
ξ	Parameter in the investment rule	0.2
$sd(K_0)$	Standard deviation of initial capital stock (in logarithm)	0.05
$sd(\omega_0)$	Standard deviation of initial productivity	0.05
$sd(\varepsilon^\omega)$	Standard deviation of productivity innovation (ε^ω)	0.01
$\overline{P}_L$	Mean price of labor	0.1
$\overline{P}_M$	Mean price of materials	0.1
$sd(P_L)$	Standard deviation of labor price	0.02
$sd(P_M)$	Standard deviation of material price	0.02
$sd(u)$	Standard deviation of revenue measurement error (u)	0.01
T	Number of periods	10
J	Number of firms	100
N	Number of Monte Carlo replications	1000
Multiple materials[2]	**Description**	**Value**
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
$\overline{P}_{M3}$	Mean price of $M3$	0.1
$sd(P_{M3})$	Standard deviation of $M3$ price	0.02
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] Labor and materials prices are log-normally distributed.

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

Table 2: Monte Carlo: Estimated results from three different methods[1]

	True	$\sigma = 0.8$			$\sigma = 1.5$			$\sigma = 2.5$		
		Us	Proxy-OP	Oracle-OP	Us	Proxy-OP	Oracle-OP	Us	Proxy-OP	Oracle-OP
$\hat{\eta}$	-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.

Figure 1: Monte Carlo experiments: true $\sigma = 0.8, 1.5, 2.5$

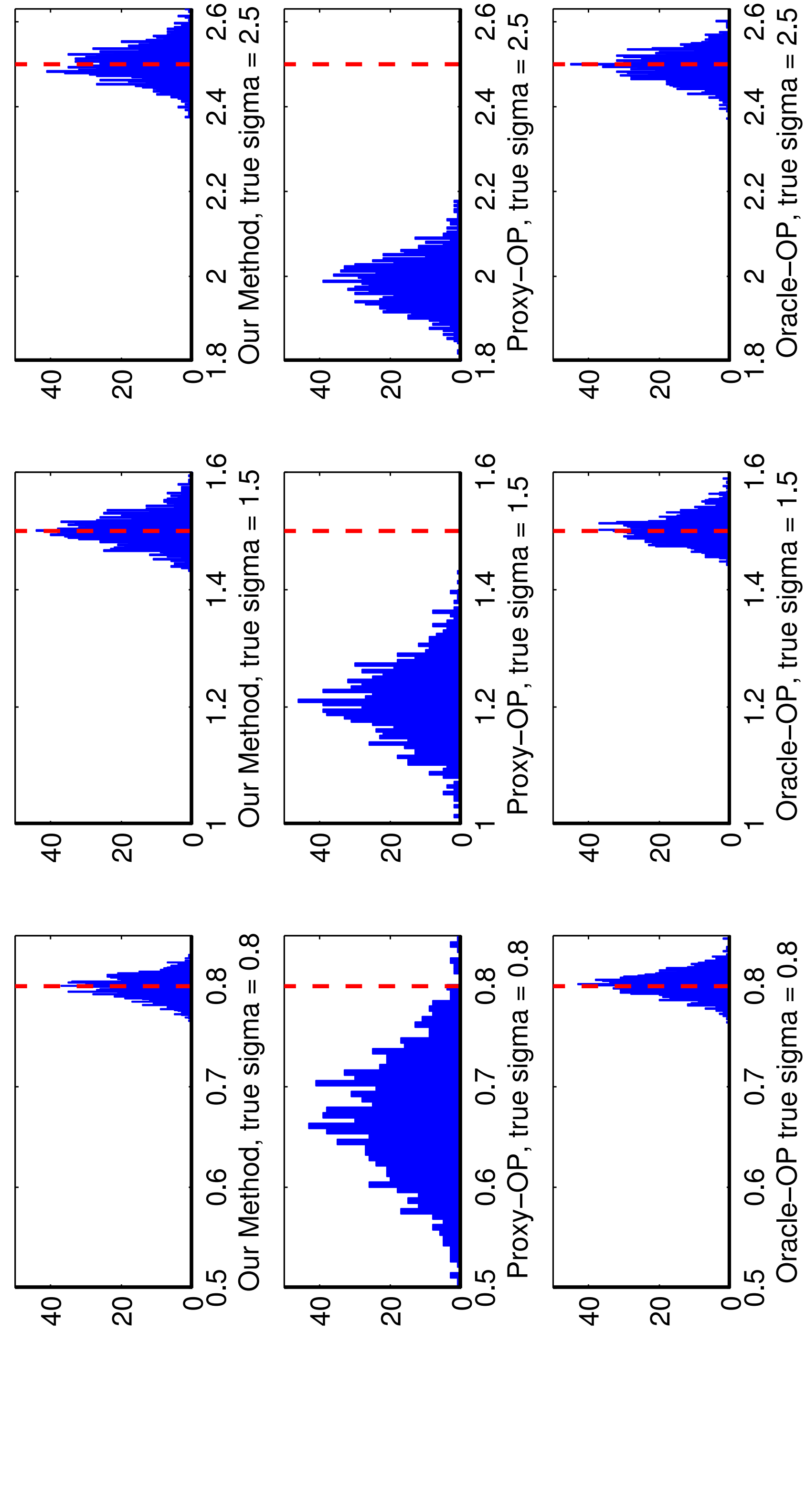

Figure 2: Kernel density estimation: recovered material prices v.s. true material prices

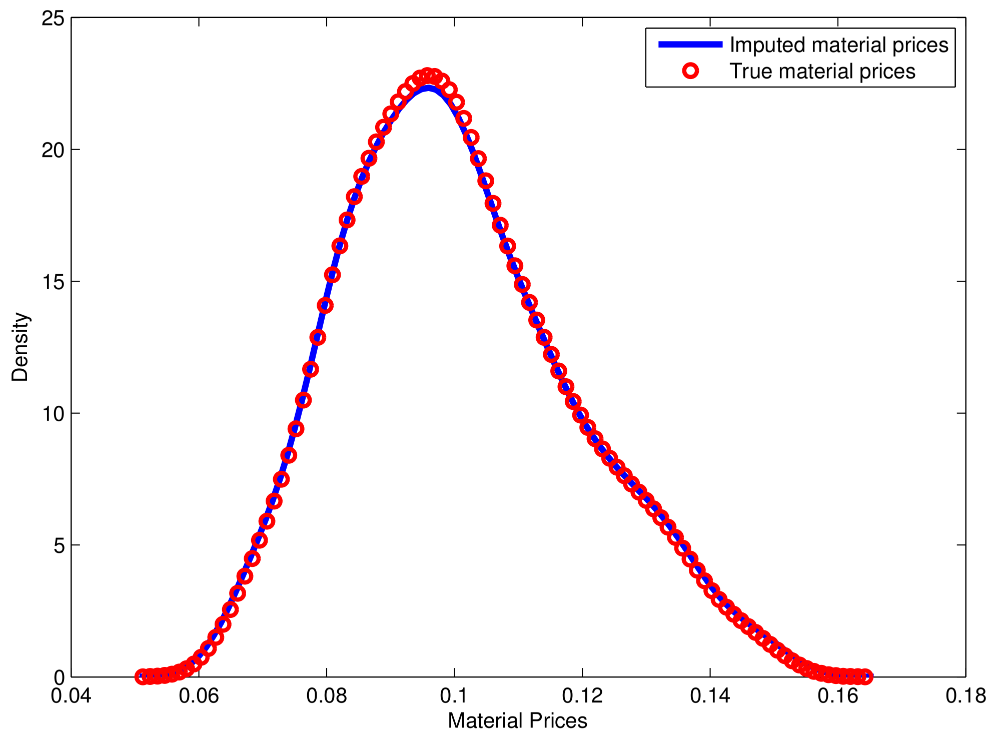

Table 3: Estimated results for Colombian industries

	Clothing		Bakery Products		Printing & Publishing		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	

Table 4: Persistence of recovered 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 5: Correlations between recovered productivity and input prices in logarithm

	$\text{corr}(\hat{\omega}, \log(\hat{P}_M))$	$\text{corr}(\hat{\omega}, \log(P_L))$	$\text{corr}(\log(\hat{P}_M), \log(P_L))$
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 3: Kernel density estimation of productivity $(\widehat{\omega_{jt} + u_{jt}})$

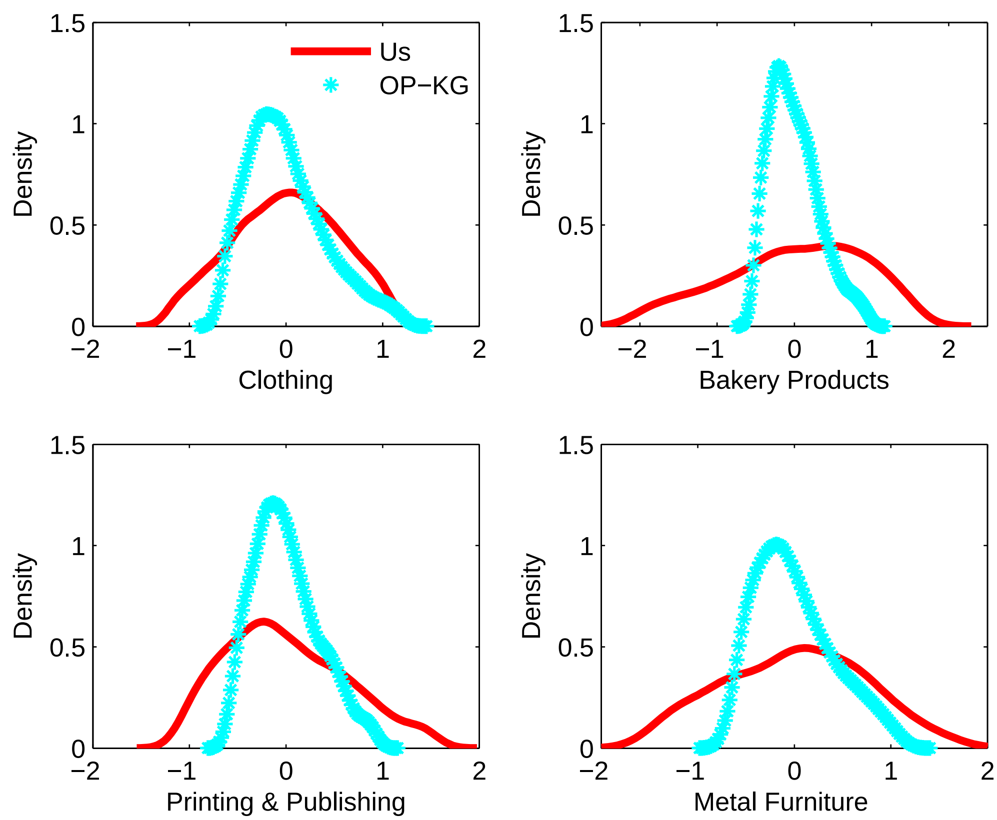

Figure 4: Kernel density estimation of recovered material prices in logarithm