Production Function Estimation for Multi-Product Firms*

Zhiyuan Chen† Moyu Liao‡

January 30, 2020

Abstract

We study a stylized model of multi-product firms with firm-product level heterogeneity in Hicks-neutral production technology. The productivity process allows flexible correlation between the production efficiency of different products. We characterize the empirical content and show that the scale and location of the production function are non-parametrically non-identified without observing the allocation of inputs and exogenous input price variations. Based on the model's empirical content, we develop an estimation strategy for any parametric family of production functions. For general production functions, we argue this procedure is subject to unclear identification and heavy computation. In the case of Cobb-Douglas production function, we show the optimal input allocation rules can be solved in closed form and the corresponding moment conditions identify the production functions. Monte Carlo evidence shows that our identification strategy performs well. We then apply our methodology to a sample of agricultural goods manufacturing firms and show that multi-product firms' production technologies differ from single-product firms even for the same product. We also

*We thank comments and suggestions from Paul Grieco, Daniel Grodzicki, Marc Henry, Karl Schurter, and other seminar participants at the Industrial Organization Brownbag at Penn State University. Two authors contribute equally to this paper.

†Department of Economics, Pennsylvania State University. Email: zxc5109@psu.edu.

‡Department of Economics, Pennsylvania State University. Email: mml5543@psu.edu.

find that multi-product firms produce its core (peripheral) products at higher (lower) technical efficiency than single-product firms. Lastly, we find a strong positive comovement between the productivity shocks of different product lines.

JEL: C51, L2, L6

Keywords: multi-product firms; production function; non-identification; input allocations

1 Introduction

With the increasing availability of product-level production data, the existence of multi-product firms has become one of the most important features of micro-level production data.[1] There are still unanswered questions in data such as why some firms become multi-product firms and others do not. A first step to understanding a firm's choice of production set is the heterogeneity in firm-product level technical efficiency. Estimating firm-product level productivity allows researchers to directly detect the impact of interesting factors (such as competition, R&D spending, trade liberalization, and etc.) on productivity efficiencies at firm-product level. It can also provide empirical tests to existing theories on multi-product firms.[2] At an aggregate level, a good estimation of firm-product level productivity can provide product-level explanations for the evolution of firm productivity and aggregate total factor productivity (TFP).[3]

The popular productivity estimation methods proposed by Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), and Gandhi et al. (2016) are more suitable for single-product firms. Extending existing methods to multi-product firms poses several challenges. This is mainly because the researcher can only observe the total inputs, but not the inputs allocated to the production of different products. In this paper we investigate the identification of the firm-product level production function for multi-product firms.

We study a stylized model of multi-product firms with firm-product level heterogeneity in Hicks-neutral production technologies and a parameterized demand system. In the model, firms choose the allocation of capital, labor, and materials to different production lines after observing the realization of productivity for each product. We extend the productivity process in existing literature to be a first-order vector auto re-

[1] In U.S., multi-product firms account for 91 % of U.S. manufacturing sales (Bernard et al., 2010) and 98% of the value of manufacturing exports (Bernard et al., 2010).

[2] For example, recent theoretical models by Eckel and Neary (2010), Bernard et al. (2011), and Mayer et al. (2014) predict that firms produce their "core" products with higher efficiency and generate more profits. Eckel et al. (2015) consider the quality-upgrading choice in a multi-product firm model and indicate the firms will invest more in improving the quality of the product with higher production efficiency.

[3] For example, Bernard et al. 2010 show theoretically that the product switching contributes to the reallocation of resources towards more efficient use, which increase the firm's production efficiency.

gressive (VAR) process, which provides a more realistic characterization of the evolution of productivity for multi-product firms. We clarify the empirical content of the model and show that firm-product level production function is generally *non-parametrically non-identified*. First, we show that the scale of production functions cannot be separated from the scale of productivities for each product, hence the scale of production functions needs to be normalized in order to have a reasonable interpretation of productivity estimates. As a result, we can only analyze the relative productivity of different products. We also show the location of production functions for multi-product firms are not identified, i.e. we can horizontally shift production functions for each product simultaneously and the newly generated production functions still satisfy the empirical content. This non-identification comes from the unobserved input allocation to different production lines and only presents in multi-product firms. As a consequence, we cannot recover the output elasticity of inputs.

Due to the non-identification result for non-parametric production functions, we restrict our attention to any parametric family of production functions with Hicks-neutral production technology. Based on the empirical content of the model, we develop an estimation procedure that can be potentially applied to any parametric family of production functions. We notice, however, in general our estimation strategy suffers from an unclear interpretation of its empirical content. This is because we not only rely on the firm's optimization behavior and production functions to characterize the allocation rule of inputs, but also have to use production functions to construct unobserved productivity shocks. A general parametric form of the production function blurs the identification of parameters. More importantly, when the production function does not have convenient form, one faces a severe computation burden when solving optimal input allocations using observed data and parameters. In general, in the estimation procedure, we need to solve input allocation numerically for each firm in each iteration of parameters.

To achieve interpret-able empirical content and feasible computation, we further study a special case of the parametric production function, i.e. the Cobb-Douglas production function. By obtaining a closed-form solution for unobserved input allocation as a function of product-level output quantities, we prove the identification of the production function. We also offer two special cases of CES production functions which feature transparent identification and relatively easy computation. As an important extension of the estimation strategy for Cobb-Douglas production function, we consider multi-product firms using multiple material inputs. We show that our estimation framework can be easily adjusted to accommodate such scenario by assuming that different materials are combined through a CES aggregator. We also show that when there are multiple material inputs, input price variation across firms is essential to the identification of production function.

We conduct Monte Carlo experiments to evaluate the validity of our estimation strategy. Using a simulated dataset, we first show that our method successfully identify the parameters of firm-product level production function. This implies that the distribution of estimated productivity is close to the true productivity distribution. Applying our method for Cobb-Douglas production function to the agricultural food manufacturing sector in China, we find that the production function of multi-product firms differs from that of single-product firms even for the same product. Ignoring the heterogeneity in production function results in an upward bias to the average productivity. We also find that multi-product firms produce its core (peripheral) products at higher (lower) technical efficiency than single-product firms. Most importantly, we find a strong positive comovement between the productivity shocks of different product lines. This implies positively correlated productivity shocks to different products.

This paper contributes to the literature on the estimation of production function for multi-product firms in following ways. First, different from the single product proxy method proposed by De Loecker et al. (2016), our method identifies firm-product level technical efficiency. Second, we do not impose that firms producing the same product have the same production function whether they are single-product or multi-product firms. Our method can separate the heterogeneity in production technology from productivity. This allows for possibly different production technologies that may motivate firms to choose different sets of products. Dhyne et al. (2017) (DPSW hereafter) develop a multi-product production function approach to identify the production function at firm-product level. By assuming production possibility frontier (PPF) has a separate additive form, they use the quantities of other products produced by the same firm to control for unobserved input allocations. Our method differs from DPSW in following ways. First, we rely on a parameterized demand structure and firms' first-order conditions of optimal input allocations to identify productivities. Similar to Gandhi et al. (2016) (GNR hereafter), our identification strategy does not rely on exogenous variations in input prices. Second, we point out that the approach proposed by DPSW faces a potential problem of mis-specification in controlling the unobserved input allocations even in the simple Cobb-Douglas case, which causes endogeneity issue and a systematical bias to the estimates of production function. In contrast, our method imposes functional form on the production functions at the firm-product level. We show that researchers need to adjust the control function for unobserved input allocations according to the chosen firm-product production function in order to obtain consistent estimate for the productivity.

The rest of this paper is organized as follows. Section 2 describes a general theoretical model of multi-product firms. In section 3, we explore the Cobb-Douglas parametric implementation and its connection to existing methods. Section 4 presents the Monte Carlo studies we use to establish the validity of our methodology and its comparison to DPSW. In section 5 we apply our method to a sample of firms in China's grain industry. Section 6 presents an extension to the setting of multiple material inputs. Section 7 concludes the paper.

2 The Model

In this section we describe a theoretical model of multi-product firms. We explicitly state the assumptions crucial to our estimation approach. Section 2.1 describes the setup of the model. Section 2.2 provides a characterization of the empirical content of the model.

2.1 Model Setup

Consider firm i producing a set of products $\mathcal{J} = \{1, 2, \cdots, J\}$ over time periods $t = 1, 2, \cdots$. Production function of product j has a Hicks-neutral production technology:

$$Q_{ijt} = \exp\left(\omega_{ijt}\right) F\left(K_{ijt},\, L_{ijt},\, M_{ijt};\, \boldsymbol{\beta}_j\right) \tag{1}$$

where ω_{ijt} is the log of unobserved productivity for j-th product. In each period, firm i makes decisions on inputs including labor $\{L_{ijt}\}$, capital $\{K_{ijt}\}$, and materials $\{M_{ijt}\}$. Note that we allow for both of the productivity and production function to vary at firm-product level: ω_{ijt} potentially captures other heterogeneity such as consumer tastes for the attributes of different products, while $\boldsymbol{\beta}_j$ reflects production function that may have different forms for different products. Since we do not restrict the dimension of $\boldsymbol{\beta}_j$, this formulation is fully general and $F(\cdot)$ can be non-parametric. The inverse demand is of CES form:[4]

$$P_{ijt} = \bar{P}_t \bar{Q}_t^{\frac{1}{\sigma}} Q_{ijt}^{-\frac{1}{\sigma}} \tag{2}$$

where $\bar{P}_t$ is the sectoral price level, and $\bar{Q}_t$ is the industry quantity level. In the demand equation, we leave out the possible heterogeneity in consumer tastes for the characteristics of each product j because it cannot be separately identified from the productivity without using the instruments for demand.[5] Without special notice, we will use F_j to represent the production function of production j. We require that the functional form varies across products, but not over time.

Assumption 1. *Firm-level total capital stock $\bar{K}_{it}$ and labor $\bar{L}_{it}$ are predetermined, and the choice of total material $\bar{M}_{it}$ is static.*

[4] The assumption of CES demand is not critical to all of our results. Fundamentally, we only require a parameterized demand system such as

$$P_{ijt} = P\left(Q_{ijt},\, \boldsymbol{d}_t;\, \sigma\right)$$

where $\boldsymbol{d}_t$ is the aggregate demand shifter and σ is the parameter characterizing the demand equation. We choose CES for its tractability and wide applicability in related literature.

[5] A model with firm specific productivity and heterogeneity in consumer taste for different products can be found in Bernard et al. (2010) and Bernard et al. (2011).

According to this assumption, firms allocate capital and labor to different products, while taking total capital stock and labor as given. Then firms source materials and determine allocations to different products to maximize short-run profits. We also assume that firms are price taker in the inputs market. These are summarized in the following assumption:

Assumption 2. *Firms choose the optimal allocation of inputs X_{ijt}, for $X \in \{K,\, L,\, M\}$, in each period taking input prices as given.*

Based on the assumptions stated above, the firm's value function can be written as:

$$
\begin{aligned}
V\left(\boldsymbol{\omega}_{it},\, \bar{K}_{it},\, \bar{L}_{it},\, \mathbf{s}_{it}\right) = \max_{\{X_{ijt}\}_{j\in\mathcal{J},X\in\{K,L,M\}}\, I_{it},\, \bar{L}_{it+1}} & \Big\{\sum_{j=1}^{J}\left(P_{ijt}Q_{ijt} - v_{it}M_{ijt}\right) \\
& + \mathbf{E}\left[V\left(\boldsymbol{\omega}_{it+1},\, \bar{K}_{it+1},\, \bar{L}_{it+1}\right)\right]\Big\}
\end{aligned}
$$

$$
s.t.\ \bar{K}_{it+1} = (1-\delta)\,\bar{K}_{it} + I_{it} \tag{3}
$$

$$
\boldsymbol{\omega}_{it+1} = \boldsymbol{\rho}_0 + \bar{\boldsymbol{\rho}}\boldsymbol{\omega}_{it} + \boldsymbol{\epsilon}_{it+1} \tag{4}
$$

$$
\sum_{j=1}^{J} K_{ijt} = \bar{K}_{it},\ \sum_{j=1}^{J} L_{ijt} = \bar{L}_{it} \tag{5}
$$

where $\boldsymbol{\omega}_{it} = (\omega_{i1t},\, \cdots,\, \omega_{iJt})' \in \mathbb{R}^J$ is a vector of productivity summarizing the technical efficiencies of different products; $\boldsymbol{\omega}_{it}$ is observed by the firm but not by the econometrician, and v_{it} is the material price. Material price can vary across firms and periods.

Our modeling of the production of multiple products can be best described as firms use a single type of material to produce multiple types of outputs. In this sense, input prices will only vary across firms and periods, but not over products. Introducing product-level heterogeneity in input prices is possible if the econometrician has information on quantities or prices of multiple inputs. We will consider the situation of multiple material inputs in the extension.

In our formulation, both total capital investment and labor hiring are dynamic choices. Equation (3) is the accumulation rule of capital, with δ being the capital depreciation rate, and I_{it} the investment. Total labor in period $t+1$ is determined at the end of period t. Therefore both the firm-level total labor and capital decisions face uncertainty caused by productivity shocks. Equation (4) states the rule of evolution for the productivity, which follows a linear VAR(1) process,[6] $\boldsymbol{\rho}_0 \in \mathbb{R}^J$ is a vector of constants, $\bar{\boldsymbol{\rho}} \in \mathbb{R}^J \times \mathbb{R}^J$ is a matrix of coefficients that characterize the persistence and correlation pattern of different product's productivity. This captures the possible productivity spillovers across different production lines within the firm.[7] $\boldsymbol{\epsilon}_{it+1}$ is a J-dimension column vector of i.i.d shocks with zero mean, and $\boldsymbol{\epsilon}_{it+1}$ can have a very flexible covariance matrix. This also has important implications for studies on the relationship between R&D investment and productivity. Existing literature analyzes the impact of R&D on firm-level productivity by estimating an endogenous productivity process where R&D plays a role in stimulating productivity growth (e.g., Doraszelski and Jaumandreu (2013); Peters et al. (2017)). With firm-product level technical efficiencies, the impact of R&D on productivity can be analyzed at the product level for each firm by taking the possible productivity spillovers across different products into account. This can deepen the understanding of the channels through which R&D investment influences firm-level productivity.

Lastly, equation (5) is the resource constraint for allocations of labor and capital. The binding constraint implies that firms have to exhaust all of their available capital and labor. In contrast, they can adjust the level of total material inputs flexibly in each period. To simplify notations, we use X_{ijt} to represent inputs for $X \in \{K, L, M\}$. We use small letters to represent logged forms, i.e., $x_{ijt} = \ln(X_{ijt})$ and $f_j = \ln(F_j)$. Therefore the static optimization problem of each firm is

$$\max_{\{K_{ijt}, L_{ijt}, M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j=1}^{J} \left\{ \bar{P}_t \bar{Q}_t^{\frac{1}{\sigma}} \left[\exp(\omega_{ijt}) F_j\right]^{\frac{\sigma-1}{\sigma}} - v_{it} M_{ijt} \right\} \tag{6}$$

$$s.t. \sum_{j=1}^{J} X_{ijt} = \bar{X}_{it} \quad \text{for} \quad X \in \{K, L\} \tag{7}$$

[6]For the implementation of our approach, we do not require that the VAR(1) process is linear. Our estimation methodology is robust to non-linear VAR(1) process. We use the linear productivity process for the purpose of simple exposition.

[7]For example, consider a company producing snowboards and snowboard boots. The knowledge for producing snowboards may help the manufacturing of snowboard boots.

The following assumption states the condition ensuring that the static optimization problem is well-behaved.

Assumption 3. *The firm's revenue function of product j is concave in all of its inputs $\{K_{ijt},\, L_{ijt},\, M_{ijt}\}$.*

The assumption above is equivalent to that $\left[F\left(K_{ijt},\, L_{ijt},\, M_{ijt};\, \boldsymbol{\beta}_j\right)\right]^{\frac{\sigma-1}{\sigma}}$ is a concave function of its arguments $K_{ijt},\, L_{ijt},\, M_{ijt}$. As an econometrician, we do not observe the productivity ($\boldsymbol{\omega}_{it}$) and material price v_{it}, which may add to the difficulty of solving the optimal input allocations. We show in the following proposition that given the firm-level inputs $\left\{\bar{K}_{it},\, \bar{L}_{it},\, \bar{M}_{it}\right\}$ and the output quantities $\{Q_{ijt}\}_{j\in\mathcal{J}}$, the allocation rules of capital, labor and material are independent of the productivity and input prices.

Proposition 1. *There exist functions $\{\mathcal{K}_j, \mathcal{L}_j, \mathcal{M}_j\}_{j=1,\ldots J}$ that depend only on total capital, labor and materials, output quantities, and unknown parameters such that*

$$
\begin{aligned}
K^*_{ijt} &= \mathcal{K}_j(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it},\, Q_{i1t}, \ldots .Q_{iJt};\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J,\, \sigma) \\
L^*_{ijt} &= \mathcal{L}_j(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it},\, Q_{i1t}, \ldots .Q_{iJt};\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J,\, \sigma) \\
M^*_{ijt} &= \mathcal{M}_j(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, Q_{i1t}, \ldots .Q_{iJt};\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J,\, \sigma)
\end{aligned}
$$

is the solution to the optimization problem (6). The optimal allocation does not depend on the firm-specific input price v_{it}.

Proof. See appendix. □

The proposition above states that if we observe the final quantity, the total capital, labor and material, the allocation of X_{ijt} should be determined by the production function and observables. More importantly, the allocation rules are independent of the unobserved material prices. Given any chosen parameters for the production function and the demand system, this proposition provides us a method of using observables to compute optimal allocation of inputs.

2.2 Empirical Content of the Model

We start with the structure of data that the econometrician can observe.

Assumption 4. *(Data) For each firm-period pair (i,t), the econometrician observes total capital $\bar{K}_{it}$, total labor $\bar{L}_{it}$, total material input $\bar{M}_{it}$, and output quantity for each product Q_{ijt}. Moreover, the econometrician also observes the ratio of material input to total revenue $s_{it}^m = v_{it}\bar{M}_{it}/R_{it}$. The revenue is observed with a measurement error: $R_{it} = R_{it}^* \exp(u_{it})$ where R_{it}^* is the true revenue and u_{it} is some measurement error.*

To aid the identification of σ, we employ the relation between the total revenue and quantities.[8] Since industry-level aggregate demand Q_t is the same across firms, we take a reference firm I and express all revenue as the ratio with respect to the reference firm. This helps us get rid of the unobserved market level aggregate demand shifter. Then the logged revenue ratio can be expressed as

The first condition imposed is similar to the material input to revenue ratio in Gandhi et al. (2016). From the first order condition with respect to material, we have:

$$\frac{v_{it}M_{ijt}}{R_{ijt}} = \frac{\partial f(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*, \boldsymbol{\beta}_j)}{\partial m_{ijt}} \Rightarrow \quad \frac{v_{it}M_{ijt}}{R_{it}^*} = \frac{R_{ijt}}{R_{it}^*}\frac{\partial f(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*; \boldsymbol{\beta}_j)}{\partial m_{ijt}}$$

Because the econometrician only observes the total material input and quantity for each product, we concentrate on the firm-level material-to-revenue ratio:

$$\frac{v_{it}M_{it}}{R_{it}^*} = \sum_j \frac{v_{it}M_{ijt}}{R_{it}^*} = \sum_{j=1}^{J} \frac{Q_{ijt}^{\frac{\sigma-1}{\sigma}}}{R_{it}^*}\frac{\partial f(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*; \boldsymbol{\beta}_j)}{\partial m_{ijt}}$$

Note that the total revenue is measured with error, we obtain

$$\ln\left(s_{it}^m\right) = \ln\left[\sum_{j=1}^{J} \Gamma_{ijt}\frac{\partial f(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*; \boldsymbol{\beta}_j)}{\partial m_{ijt}}\right]$$

where

$$\Gamma_{ijt} = \frac{Q_{it}^{\frac{\sigma-1}{\sigma}}}{\sum_{j'} Q_{ij't}^{\frac{\sigma-1}{\sigma}}}$$

[8] One can also use the product-specific revenue and related quantities to directly identify σ if data are available.

is the model predicted revenue share of product j in the firm's total revenue. This leads to a moment condition:

$$\mathbf{E}\left[\ln\left(\hat{s}_{it}^{m}\right)-\ln\left(\sum_{j}\Gamma_{ijt}^{-1}\frac{\partial f(k_{ijt}^{*},l_{ijt}^{*},m_{ijt}^{*};\,\boldsymbol{\beta}_{j})}{\partial m_{ijt}}\right)|K_{it},L_{it},M_{it},Q_{i1t},...Q_{ijt}\right]=0 \qquad (8)$$

In general,we should note that the functional form of $\frac{\partial f(k_{ijt}^{*},l_{ijt}^{*},m_{ijt}^{*};\boldsymbol{\beta}_{j})}{\partial m_{ijt}}$ can be very complicated. Because $(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_J)$ appear in the expression through the optimal input choices k_{ijt}^{*}, l_{ijt}^{*} and m_{ijt}^{*}. It is possible that two sets of parameters may generate a same value for the output-to-material elasticity $\frac{\partial f(k_{ijt}^{*},l_{ijt}^{*},m_{ijt}^{*};\boldsymbol{\beta}_{j})}{\partial m_{ijt}}$. This blurs the parameter that is identified via the material-revenue share equation.

$$\Delta r_{it}=\hat{r}_{it}-\hat{r}_{It}=\ln\left(\sum_{j}Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right)-\ln\left(\sum_{j}Q_{Ijt}^{\frac{\sigma-1}{\sigma}}\right)+\Delta u_{it}$$

where $\Delta u_{it}=u_{it}-u_{It}$. This gives us the second moment condition:

$$\mathbf{E}\left(\Delta u_{it}|\{Q_{i1t},Q_{i2t},\cdots Q_{ijt}\},\{Q_{I1t},Q_{I2t},\cdots Q_{Ijt}\}\right)=0 \qquad (9)$$

Next, we move to the output equation. For each product j, using the productivity evolution equation, we can express the logged output quantity as

$$q_{ijt}=f\left(k_{ijt}^{*},\,l_{ijt}^{*},\,m_{ijt}^{*};\boldsymbol{\beta}_{j}\right)+\rho_{0}^{j}+\bar{\boldsymbol{\rho}}^{j}\boldsymbol{\omega}_{it-1}+\epsilon_{ijt},$$

where $\boldsymbol{\omega}_{it-1}=(\omega_{i1t-1},\,\omega_{i2t-1},\,\cdots,\,\omega_{iJt-1})'$ and

$$\omega_{ijt-1}=q_{ijt-1}-f\left(k_{ijt-1}^{*},\,l_{ijt-1}^{*},\,m_{ijt-1}^{*};\boldsymbol{\beta}_{j}\right),$$

and ϵ_{ijt} is the $j-th$ term of the vector of productivity shocks $\boldsymbol{\epsilon}_{it}$. Note that the first term $f\left(k_{ijt}^{*},\,l_{ijt}^{*},\,m_{ijt}^{*};\boldsymbol{\beta}_{j}\right)$ has a complicated form because parameters $\left\{\boldsymbol{\beta}_{j'}\right\}_{j'\neq j}$ also enters the expression through the optimal solution of k_{ijt}^{*}, l_{ijt}^{*} and m_{ijt}^{*}. A similar issue exists in the lagged value term $f\left(k_{ijt-1}^{*},\,l_{ijt-1}^{*},\,m_{ijt-1}^{*};\boldsymbol{\beta}_{j}\right)$ which is embedded in the productivity vector $\boldsymbol{\omega}_{it-1}$. The identification power of $\boldsymbol{\beta}_{j}$ and $\bar{\boldsymbol{\rho}}^{j}$ comes from the orthogonality condition of $\boldsymbol{\epsilon}_{it}$, as is the case in the literature. We formally state the assumption for $\boldsymbol{\epsilon}_{it}$ as below.

Assumption 5. *Define the information set* $\mathcal{I}_t = \{K_{it}, L_{it}, K_{it-1}, L_{it-1}, M_{it-1}, Q_{i1t-1}, ...Q_{iJt-1}, ...\}$, *then the following orthogonality condition holds for* $\boldsymbol{\epsilon}_{it}$*:*

$$\mathbf{E}[\boldsymbol{\epsilon}_{it}|\mathcal{I}_t] = 0$$

Now we have our third set of moment condition:

$$\mathbf{E}\left[q_{ijt} - f\left(k^*_{ijt}, l^*_{ijt}, m^*_{ijt}; \boldsymbol{\beta}_j\right) - \rho^j_0 - \bar{\boldsymbol{\rho}}^j \boldsymbol{\omega}_{it-1}|\mathcal{I}_t\right] = 0 \quad \forall j = 1, ...J \tag{10}$$

Definition 1. The parameters of interest are $(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J, \boldsymbol{\rho}_0, \bar{\boldsymbol{\rho}}, \sigma)$. The identified set $\boldsymbol{\Theta}_I$ is the collection of parameters that satisfy moment conditions (8), (9) and (10).

We see that the moment conditions (8) and (10) have all of the parameters $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$ of the production function in the expression, which complicates the identification of production function. As we will see shortly, with the Cobb-Douglas production function, the first-stage moment condition (8) only contains parameters characterizing the output-to-material elasticity, which greatly simplifies the identification of material-output elasticity by using the firm-level material-revenue equation.

2.2.1 Identification of σ

We first note that the mark-up parameter σ is identified from (9) if there are sufficient variations in the quantity data.

Assumption 6. *The random vector* $(Q_{i1t}, ..., Q_{iJt})$ *is continuously distributed and supported on the* $\mathbb{R}^{\mathbb{J}}_+$ *space.*

Theorem 1. *Moment condition (9) identifies mark-up parameter* σ*.*

Proof. Given the full support assumption, we can condition on the event $Q_{I1t} = Q_{I2t} =$

$... = Q_{IJt} = 1$. Then moment condition (9) reduces to

$$\mathbf{E}\left[\Delta r_{it} - \ln\left(\sum_j Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right) - J|\{Q_{i1t}, Q_{i2t}, \cdots Q_{ijt}\}\right] = 0$$

Or equivalently $\ln\left(\sum_j Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right) = \mathbf{E}[\Delta r_{it} - J|\{Q_{i1t}, Q_{i2t}, \cdots Q_{ijt}\}, Q_{I1t} = Q_{I2t} = ... = Q_{IJt} = 1]$. Note that the left hand side $\ln\left(\sum_j Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right)$ is an increasing function of σ on the space $[1, \infty)^J$, so σ is identified. □

According to Theorem 1, we can identify σ using data on firm-level revenues and product quantities. In the case that one observes product specific revenues, the identification of σ is similar in the sense that we can always aggregate the firm-product level revenues to the firm level. Our identification strategy relies on the exogenous variations (measurement errors) in the observed revenue data. When there are unobserved idiosyncratic demand shifters, we may rely on instruments for demand to identify σ. Moreover, the identification of σ is independent of the identification of other parameters in the production function.

2.2.2 Non-parametric Non-identification of $\{F_j\}_{j=1}^J$

Because we do not observe how firms allocate inputs towards each product, the allocation rule is jointly determinedby the production function parameters $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$ and productivity $\boldsymbol{\omega}_{it}$. In what follows, we state two non-identification results on the scale and location of the production function F_j when it is of non-parametric form.

Proposition 2. *Let $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ be in the identified set, then $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\boldsymbol{\rho}}_0, \bar{\boldsymbol{\rho}}, \sigma)$ is also in the identified set, where*

$$\tilde{\rho}_0^j = \rho_0^j + \sum_{j'=1}^J (\bar{\rho}_{jj'} \ln \alpha_{j'}) - \ln \alpha_j$$

Proof. See appendix. □

This non-identification of the scale of the production function is firstly proved in Gandhi et al. (2016); we extend it to the firm-product level production function for multi-product firms. For single-product firms, this non-identification result means the level of productivity and the scale of production function cannot be separated, hence a meaningful comparison of productivity across industries requires some normalization. In the case of multi-product firms, Proposition 2 implies that the difference of the productivity between two products can only be identified up to a constant. We further show that the location of the production function is not identified in following proposition.

Proposition 3. *Let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ *be in the identified set, then* $(\{\tilde{F}_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ *is also in the identified set, where*

$$\tilde{F}_j(K_{ijt}, L_{ijt}, M_{ijt}) = F_j(K_{ijt} - C_j^K, L_{ijt} - C_j^L, M_{ijt} - C_j^M) \quad \forall j = 1, ...J$$

and the constants $\{C_j^X | j = 1, ...J,\ X \in \{K, L, M\}\}$ *satisfies*

$$\sum_{j=1}^J C_j^X = 0 \quad \forall X \in \{K, L, M\}$$

Proof. See appendix. □

The proposition implies that it is impossible to identify the relative elasticity of inputs $\frac{\partial f_j}{\partial x_j} / \frac{\partial f_{j'}}{\partial x_{j'}}$ for $x \in \{k, l, m\}$. The intuition is that when the input allocation is unobserved, we can always relocate the production function such that the inputs allocated to different products are rationalized using the observed data. Noticing the issue of non-parametric non-identification, we turn to focus on a parametric form of the production function which has a transparent rule of input allocations.

3 Parametric Implementation

3.1 Estimation of Parametric Production Functions

Now we propose a general estimation strategy for any parametric family of production functions. For any parametric family of product $j's$ production function $F\left(K_{ijt},\, L_{ijt},\, M_{ijt},\, \boldsymbol{\beta}_j\right)$, where $\boldsymbol{\beta}_j \in \mathbb{R}^d$ and $d < \infty$, we can identify the production function directly using moment conditions (8)-(10). In principle, we can employ an estimation procedure as follows:

1. Use moment condition (8) to estimate σ.
2. Given the estimated σ,we start with some appropriate initial value of $\boldsymbol{\beta}_j^0$ and solve for the optimal allocation rules $\left\{k_{ijt}^0,\, l_{ijt}^0,\, m_{ijt}^0\right\}$ using Proposition 1.
3. Plugging the solved input allocations into the objective function of GMM estimator based on moment conditions (9) and (10), and obtains a new estimate of production function parameters $\boldsymbol{\beta}_j^1$.
4. Repeat step 2 and step 3 until $\|\boldsymbol{\beta}_j^n - \boldsymbol{\beta}_j^{n-1}\| < tol$, where tol is a chosen tolerance value. Then $\boldsymbol{\beta}_j^n$ is the estimate for the production function of product j.

Though moment conditions (8)-(10) are general and can be applied to any parametric production functions, there are two major issues. First, our proposition 1 only insures the unique existence of optimal allocation rules, but closed form may not exist. For each firm we need to solve the allocation rule numerically. When we solve the optimization problem separately, the number of optimization problems is equal to the number of firm-year observations. This can pose computational challenge to implementation. An alternative is to choose a grid of state variables and use function interpolation to find the optimal input allocations for each firm. However, the state space $(K_{it},\, L_{it},\, M_{it},\, Q_{i1t},\, \cdots,\, Q_{iJt})$ upon which we compute the optimal input allocations is of high dimensionality. Consider firms producing two products, we need to choose a 5-dimension grid in order to compute the optimal input allocations for each firm. Choosing 10 values for each state variable leads to 10^5 grids. Second, more importantly, it is not clear what is identified from equation (8) and (10). Note that in equation (8), $\{\boldsymbol{\beta}_{j'}\}_{j'\neq j}$ also enters the partial derivatives $\frac{\partial f(k^*_{ijt}, l^*_{ijt}, m^*_{ijt}; \boldsymbol{\beta}_j)}{\partial m_{ijt}}$ through the allocation rule for capital, k^*_{ijt}. This complicates the objective that is identified through the material-to-revenue ratio equation. In general, we can not determine whether moment conditions (8)-(10) give us point identification of parameters even under parametric assumptions.

In what follows, we propose several parametric examples of production functions which have clear identification and are less intensive in computation. We use Cobb-Douglas production function as our benchmark model, and compare the gain and loss from using more general parametric forms.

3.2 Cobb-Douglas Production Function

Consider a Cobb-Douglas production function as follows:

$$Q_{ijt} = \exp\left(\omega_{ijt}\right) K_{ijt}^{\beta_k^j} L_{ijt}^{\beta_l^j} M_{ijt}^{\beta_m^j} \tag{11}$$

The production technology can differ across different products, which implies that β_k^j, β_l^j, and β_m^j vary for different product j. The Cobb-Douglas production function is widely used in the productivity literature (see Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg et al. 2015, for example). In the Cobb-Douglas production function, the elasticity of substitution between any two different inputs is one; β_x^j represent the share of input x ($x \in \{k, l, m\}$) in total inputs.

3.2.1 Optimal Input Allocations

We can simplify the firm's static problem (6) as:

$$\max_{\{K_{ijt}, L_{ijt}, M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j=1}^{J} \{R_{ijt} - v_{it} M_{ijt}\}$$

subject to constraint (5), where R_{ijt} is the revenue obtained by firm i from selling product j:

$$R_{ijt} = P_t Q_t^{\frac{1}{\sigma}} \left[\exp\left(\omega_{ijt}\right) K_{ijt}^{\beta_k^j} L_{ijt}^{\beta_l^j} M_{ijt}^{\beta_m^j}\right]^{\frac{\sigma-1}{\sigma}} \tag{12}$$

Denote $\tilde{\beta}_x^j = \frac{\sigma-1}{\sigma}\beta_x^j$, $x \in \{k, l, m\}$. The first-order conditions imply that

$$K_{ijt} = \frac{\tilde{\beta}_k^j R_{ijt}}{\lambda_{kit}} \tag{13}$$

$$L_{ijt} = \frac{\tilde{\beta}_l^j R_{ijt}}{\lambda_{lit}} \tag{14}$$

$$M_{ijt} = \frac{\tilde{\beta}_m^j R_{ijt}}{v_{it}} \tag{15}$$

where λ_{kit} and λ_{lit} are the Lagrangian multipliers for the capital constraint and labor constraint, respectively. From the first-order conditions, we know the input quantity of dynamic input are influenced by the shadow prices, but the allocation rules as the ratio of total input are not influenced by the shadow prices. In other words, the marginal price of each input is adjusted such that the optimal level of this input is consistent with the pre-determined total input. In particular, the allocation rules for inputs (in logs) are

$$x_{ijt} = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j'\in\mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt}^{j'}}\right) + \bar{x}_{it},\ x \in \{k, l, m\}$$

where

$$\gamma_{ijt}^{j'} = \frac{R_{ij't}}{R_{ijt}} = \left(\frac{Q_{ij't}}{Q_{ijt}}\right)^{\frac{\sigma-1}{\sigma}} \tag{16}$$

is the product j's revenue relative to the reference product j.[9] To simplify notations, let's define the log of input share as

$$\alpha_{xit}^j = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j'\in\mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt}^{j'}}\right).$$

[9]See the appendix for the derivation.

Therefore we express the allocation rule for inputs as

$$x_{ijt} = \alpha^j_{xit} + \bar{x}_{it},\ x \in \{k,\ l,\ m\} \tag{17}$$

The key observation is that we have a closed-form solution for allocation rules of inputs corresponding to proposition (1). It follows that the quantity of product j is

$$q_{ijt} = \beta^j_k \bar{k}_{it} + \beta^j_l \bar{l}_{it} + \beta^j_m \bar{m}_{it} + \sum_{x \in \{k,l,m\}} \beta^j_x \alpha^j_{xit} + \omega_{ijt} \tag{18}$$

This logged production function links the firm-level inputs to the firm-product level output. Because input shares α^j_{xit} are not observed by the econometrician, failing to control them will cause a bias to the estimation of productivity. We have shown that the identification of σ is independent of the form of production function; we are left to discuss the identification of the parameters in the production function. To emphasize the importance of the material-revenue share equation, we first show that the production function is not identified using the moment condition employed by Levinsohn and Petrin (2003). We further demonstrate how we use the firm-level material-revenue share to help identify the material elasticity.[10] Then we employ the productivity process to identify other parameters in the production function.

3.2.2 Identification

We first display a proposition showing that the moment restriction (10) does not identify the parameters β^j_x even in the Cobb-Douglas parametric framework if there is no exogenous variation in material input prices.

Proposition 4. *Let* $\{\beta^j_k, \beta^j_l, \beta^j_m, \boldsymbol{\rho}_0, \bar{\boldsymbol{\rho}}, \sigma\}$ *satisfies moment condition (10). If there is no variation in the price of materials across firms, i.e.* $v_{it} = v_t$, *then* $\{\beta^{*j}_k, \beta^{*j}_l, \beta^{*j}_m, \boldsymbol{\rho}^*_0, \bar{\boldsymbol{\rho}}^*, \sigma\}$

[10] One may argue that GNR method can be directly applied to estimate the firm-product level production function when the input allocation is observed. However, this is not the case when the firm's optimization problem has taken into account the difference in production efficiency in producing different products. Even when we observe the input allocations, estimating the productivity at firm-product level will introduce a bias to our estimation. See the appendix for a more detailed discussion.

which are defined as

$$\beta_k^{*j} = c\beta_k^j \quad \forall j$$
$$\beta_l^{*j} = c\beta_l^j \quad \forall j$$
$$1 - \beta_m^{*j}\frac{\sigma - 1}{\sigma} = c(1 - \beta_m^j\frac{\sigma - 1}{\sigma}) \quad \forall j$$

also satisfy moment condition (10) for some value of $\bar{\boldsymbol{\rho}}_0^*$.

Proof. See the appendix. □

This is an extension of the non-identification theorem in Gandhi et al. (2016) to the multi-product context. Their result shows that the scale of the parameter of Cobb-Douglas production function is not identified. The constructive proof in the appendix, however, shows that the scale is not identified only up to a one-dimensional constant c regardless of the total number of products J. This can be seen from the allocation rule of capital and labor:

$$\alpha_{xit}^j = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j'\in\mathcal{J}}\tilde{\beta}_x^{j'}\gamma_{ijt}^{j'}}\right) \quad x \in \{k, l\}$$

The allocation rule is not changed only when the scaling of parameter is the same for all j.

Identification power of (8) For the Cobb-Douglas production function, the material to revenue ratio equation becomes

$$\mathbf{E}\left[\hat{s}_{it}^m | Q_{i1t}, ...Q_{iJt}\right] = \ln\left(\frac{\sigma - 1}{\sigma}\sum_j \beta_m^j \frac{Q_{ijt}^{\frac{\sigma-1}{\sigma}}}{\sum_{j'} Q_{ij't}^{\frac{\sigma-1}{\sigma}}}\right) \tag{19}$$

The conditioning set does not depend on the total K_{it}, L_{it}, M_{it}. Given the identification of σ, the only unknown parameter appears in the moment condition is the material elasticity β_m^j. Variations in Q_{ijt} and $\hat{s}_{it}^m$ identifies β_m^j for all j. Intuitively, conditional on the firm's aggregate output $\sum_{j'} Q_{ijt}^{\frac{\sigma-1}{\sigma}}$, β_m^j captures the contribution of output j in the firm-level material share.

Identification of capital and labor parameters

In the case of Cobb-Douglas production, we can write the moment condition (10) as

$$\mathbf{E}\left[q_{ijt} - \sum_{x\in\{k,l,m\}} \beta_x^j \alpha_{xit}^j \bar{x}_{it} - \rho_0^j - \bar{\boldsymbol{\rho}}^j \boldsymbol{\omega}_{it-1} | \mathcal{I}_t\right] = 0, \tag{20}$$

where the productivity vector is

$$\boldsymbol{\omega}_{it-1} = \begin{bmatrix} q_{i1t-1} - \sum_{x\in\{k,l,m\}} \beta_x^1 \alpha_{xit-1}^1 \bar{x}_{it-1} \\ q_{i2t-1} - \sum_{x\in\{k,l,m\}} \beta_x^2 \alpha_{xit-1}^2 \bar{x}_{it-1} \\ \vdots \\ q_{i2t-1} - \sum_{x\in\{k,l,m\}} \beta_x^J \alpha_{xit-1}^J \bar{x}_{it-1} \end{bmatrix} \tag{21}$$

Choosing appropriate instruments for k_{it} and l_{it}, for example lagged logged capital and labor k_{it-1} , l_{it-1},we can identify the parameters for capital and labor. We provide a formal discussion of the identification of $\boldsymbol{\beta}_k = \left(\beta_k^1, \cdots, \beta_k^J\right)$ and $\boldsymbol{\beta}_l = \left(\beta_l^1, \cdots, \beta_l^J\right)$ in the math appendix.

Note that in the estimation we can easily compute the expected input allocation shares α_{xit}^j using (17). Because the allocation rule summarizes the relative revenue shares, the productivity of other products relative to product j should matter for it. Heuristically, in the revenue equation, k_{it-1} and α_{kit-1}^j help us identify β_k^j, l_{it-1} and α_{lit-1}^j contribute to the identification of β_l^j. We can separate $\bar{\boldsymbol{\rho}}^j$ from other parameters through the variations in $\bar{\boldsymbol{m}}_{it-1}$ and q_{ijt-1}. When estimating the revenue equation, failing to control the unobserved input allocations would lead to an endogeneity problem.

3.3 Estimation

Following Wooldridge (2009), we propose a joint GMM estimator to estimate the parameters of interest. Note that from (9), (19) and (20) , we have three groups of moment conditions. Let $\Theta = (\sigma, \boldsymbol{\beta}_k, \boldsymbol{\beta}_l, \boldsymbol{\beta}_m, \boldsymbol{\rho}_0, \bar{\boldsymbol{\rho}})$ be the vector of the interested parameters. We define a $(J+2) \times 1$ residual function as

$$\boldsymbol{\xi}_{it}(\Theta) = \begin{bmatrix} u_{it} \\ \Delta u_{it} \\ \boldsymbol{\epsilon}_{it} \end{bmatrix} \tag{22}$$
$$= \begin{bmatrix} \ln\left(\hat{s}_{it}^{m}\right) - \ln\left(\sum_j \tilde{\beta}_m^j \Gamma_{ijt}^{-1}\right) \\ \Delta \hat{r}_{it} - \ln\left(\sum_j Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right) + \ln\left(\sum_i \sum_j Q_{ijt}^{\frac{\sigma-1}{\sigma}}\right) \\ q_{i1t} - \sum_{x\in\{k,l,m\}} \beta_x^1 \alpha_{xit}^1 \bar{x}_{it} - \rho_0^1 - \bar{\boldsymbol{\rho}}^1 \boldsymbol{\omega}_{it-1} \\ \vdots \\ q_{iJt} - \sum_{x\in\{k,l,m\}} \beta_x^J \alpha_{xit}^J \bar{x}_{it} - \rho_0^J - \bar{\boldsymbol{\rho}}^J \boldsymbol{\omega}_{it-1} \end{bmatrix}$$

where $\boldsymbol{\omega}_{it-1}$ is given by (21). Let $\boldsymbol{Z}_{it}$ be a matrix of instruments

$$\boldsymbol{Z}_{it} = \begin{bmatrix} \boldsymbol{z}_{it1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{z}_{it2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{z}_{it3} \end{bmatrix}, t = 2, \cdots, T \tag{23}$$

In particular, for $\boldsymbol{z}_{it1}$ we choose quantities for different products and ratios of them with respect to the reference product; $\boldsymbol{z}_{it2}$ contains product quantities and their sum. Lastly, we can use variables in $\mathcal{I}_{it}$ and their appropriate polynomials for z_{it3}. The joint GMM estimation is based on the moment condition

$$\mathbf{E}\left[\boldsymbol{Z}_{it}' \boldsymbol{\xi}_{it}(\Theta)\right] = 0,\ t = 2, \cdots, T \tag{24}$$

We can stack these moment conditions for each firm i and use the standard GMM estimation to obtain the estimates for all of the parameters. The GMM estimator is

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{i^*=1}^{N} [\boldsymbol{Z}'_{it}\boldsymbol{\xi}_{it}(\Theta)]' \hat{W} [\boldsymbol{Z}'_{it}\boldsymbol{\xi}_{it}(\Theta)]$$

where $\hat{W}$ is the estimated efficient matrix that is positive definite. The analytical expression for the limit variance matrix can be complicated. Instead, we use MCMC method (Chernozhukov and Hong, 2003) to conduct inference on parameters. Essentially, given the parameterized demand structure, our estimation methodology only requires data on product-level quantities or revenues. One can easily adjust the equation for product quantities to be that for revenues. We provide a detailed discussion in the appendix.

3.4 Other parametric forms of production function

In this subsection, we discuss two other examples to show some special cases where the identification of production function is transparent.

Nested CES production function In this section, we consider the product j's production function of nested CES form:

$$F(K_{ijt}, L_{ijt}, M_{ijt}; \boldsymbol{\beta}_j) = \left(\beta_k^j K_{ijt}^{\theta_j} + \beta_l^j L_{ijt}^{\theta_j}\right)^{\frac{1}{\theta_j}} M_{ijt}^{\beta_m^j} \tag{25}$$

Here the vector of parameters as $\boldsymbol{\beta}_j = \left(\beta_k^j,\, \beta_l^j,\, \beta_m^j,\, \theta_j\right)$. The substitution elasticity between capital and labor is $\frac{1}{1-\theta_j}$. When $\theta_j = 0$, the production function is identical to the Cobb-Douglas production function. The importance of normalizing CES functions was well recognized in the existing literature. The observation is that a family of CES production functions which differ only by different elasticities of substitution needs a common benchmark point. Because the elasticity of substitution is originally defined as point elasticity, one needs to fix baseline values for the level of production level, factor inputs, and for the marginal rate of substitution.[11] For the simplicity of notation, for now we skip the normalization in the discussion of the identification procedure.

A nice feature of this production function is that the material-output elasticity is the same as the Cobb-Douglas production function: $\partial f\left(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*, \boldsymbol{\beta}_j\right)/\partial m_{ijt} = \beta_m^j$. So moment condition (8) is the same as in the Cobb-Douglas production function, and it identifies the material elasticity β_m^j. Also, we have the same allocation of the material as to the case of Cobb-Douglas production function. The quantity equation after substituting the m_{ijt} is similar

$$\frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^j} = q_{ijt} - \frac{\ln\left(\beta_k^j K_{ijt}^{\theta_j} + \beta_l^j L_{ijt}^{\theta_j}\right)}{\theta_j(1-\tilde{\beta}_m^j)} - \rho_j \frac{\omega_j}{(1-\tilde{\beta}_m^j)} + const$$

Then the analysis of moment condition (10) can be restricted to the identification conditions on β_k^j, β_l^j and θ_j.

While the allocation rules of K_{ijt}^* and L_{ijt}^* still need to be solved numerically, the identification of $\{\beta_m^j\}_{j=1}^J$ allows us to compute M_{ijt}^* analytically and plug into the optimization problem. So instead of solving an optimization problem with $3J$ argument, we only need to solve an optimization problem with $2J$ argument. This can also significantly improves the computation burden.

CES production function with parameter constraints Now we consider another example of CES with common elasticity among three inputs:

$$F_j(K_{ijt}, L_{ijt}, M_{ijt}) = \left(\beta_k^j K_{ijt}^\theta + \beta_m^j L_{ijt}^\theta + \beta_m^j M_{ijt}^\theta\right)^{\frac{1}{\theta}} \tag{26}$$

[11]This was discovered by de La Grandville (1989), and further explored by Klump and de La Grandville (2000); de La Grandville and Solow (2006); León-Ledesma et al. (2010); Grieco et al. (2016).

with the parameter constraints[12]

$$\frac{\beta_k^j}{\beta_l^j} = C_l,\ \frac{\beta_k^j}{\beta_m^j} = C_m \quad \forall j \in \mathcal{J} \tag{27}$$

Under this restriction, we can show that the optimal allocation of inputs are[13]

$$X_{ijt}^{*} = \frac{R_{ijt}}{R_{it}} X_{it},\ \forall X \in \{K,\ L,\ M\} \tag{28}$$

This simply states that the input allocation is proportional to the revenue, and thus independent of the structural production function parameters. If we can observe the product-specific revenue, we can directly compute the inputs allocated to different products. Alternatively, if we observed product quantities, we know the input allocation up to the demand elasticity σ. This means that we can compute the input allocated to different products directly. Therefore, the moment condition for material-revenue ratio is

$$\mathbf{E}\left[\ln\left(\hat{s}_{it}^{m}\right) - \ln\left(\frac{\sigma-1}{\sigma}\sum_j \frac{M_{ijt}^{*\theta}}{C_m K_{ijt}^{*\theta} + \frac{C_m}{C_l} L_{ijt}^{*\theta} + M_{ijt}^{*\theta}} \times \frac{R_{ijt}}{R_{it}}\right) | \mathcal{I}_t\right] = 0 \tag{29}$$

This delivers the identification of material-output elasticity under certain normalization for the CES function. Then with moment conditions (9) and (10), we can identify other parameters of interest.

3.5 Connection to existing methods

Existing methods differ in their approaches dealing withthe unobserved input allocations. Our methodology differs from existing methods by allowing firm-product level technical efficiencies and the structural way of controlling the unobserved input allocations. In this subsection, we discuss the relationship between our approach and other methods.

[12] De Loecker (2011) considered a common production function for different products. In particular, he restricted that β_m^j is common to all products in addition to our restriction.

[13] See the appendix for the proof.

Single-product proxy approach De Loecker et al. (2016) assume that a multi-product firm uses the same technology as a single-product firm when they produce the same product. They rely on single-product firms to control for unobserved input allocations and recover the product-level production function. They also deal with the potential selection bias caused by the firm's choice of being a multi-product firm based on unobserved productivity. Unlike their method, we propose a structural approach to control for unobserved input allocations using the information on the outputs of different products. In the case of Cobb-Douglas production functions, the unobserved input allocations has a closed form and can be computed easily. We do not assume that single product firm and multi-product firm have the same production technology when producing the same product. If the selection bias is also caused by heterogeneity in production technology, using the single-product proxy approach can lead to biased estimates for the production function and thus the productivity.

MPP function approach Dhyne et al. (2017) propose a multi-product production (MPP) function approach to estimate the production function for multi-product firms. They assume that the unobserved productivity is at the firm-product-year level; they do not rely on single-product firms to identify the product-specific technology. They describe the production technology using production possibilities set T, and define the production function using the production possibilities frontier

$$\begin{aligned} q_{ijt} &= f_j^{DPSW}(q_{i-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \\ &\equiv \max\{q_{ijt} |\ (q_{ijt}, q_{i-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \in T\} \end{aligned} \tag{30}$$

They assume a separately additive form of F_j^{DPSW} such that

$$\begin{aligned} q_{ijt} &= f_j^{DPSW}(q_{i-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \\ &= \tilde{f}_j^{DPSW}(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) + h(q_{i-jt}) \end{aligned} \tag{31}$$

where $h(\cdot)$ is some non-parametric function. However, this high-level assumption is lack of micro-level foundation on the production function at firm-product level. In the following, we will show that the production function cannot be expressed in such an additive form when the firm-product production function is Cobb-Douglas.

Our method can also be nested into the production possibilities approach. Given production functions $(F_1, ...F_J)$, the production possibilities set T is defined as

$$T \equiv \Bigg\{ (q_{i1t}, ...q_{iJt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) |$$
$$q_{ijt} = F_j(K_{ijt}, L_{ijt}, M_{ijt})$$
$$\sum_j X_{ijt} = \bar{X},\ X = K, L, M \Bigg\}$$

From T we can similarly define the DPSW type production function using quantities of other products. Compared with DPSW, our method puts assumption on the form of production functions, and how the production possibilities set T is generated. These assumptions can be interpreted as assumptions on fundamental objects. Now we show that the separately additive assumption of F_j^{DPSW} fails when the production possibilities set T is generated in our framework under Cobb-Douglas production technology.

We consider a simple case where $J = 2$, the only input is capital K, no labor and material input, and $F_1(K_{i1t}) = K_{i1t}^{\beta}$, $F_2(K_{i2t}) = K_{i2t}^{\beta}$. By definition we want to find

$$q_{i1t} = \max\{q_{i1t} |\ (q_{i1t}, q_{i2t}, \bar{K}_{it}) \in T\}$$

It is equivalent to maximizing the quantity of q_{i1t}, subject to quantity constraint of q_{i2t}, i.e.

$$\max\,(\bar{K}_{it} - K_{i2t})^{\beta}$$
$$s.t. Q_{i2t} = K_{i2t}^{\beta}$$

First order condition implies

$$\beta(\bar{K}_{it} - K_{i2t})^{\beta-1} = \lambda\beta K_{i2t}^{\beta-1}$$

and we can solve $K^*_{i2t} = \frac{\bar{K}_{it}}{1+\lambda^{1/(\beta-1)}}$. The constraint $Q_{i2t} = K^{\beta}_{i2t}$ implies λ must satisfy

$$Q_{i2t} = \left(\frac{\bar{K}_{it}}{1+\lambda^{1/(\beta-1)}}\right)^{\beta} \tag{32}$$

Now, the maximum q_{i1t} can be calculated by plug in K^*_{i2t} so

$$\begin{aligned} q_{i1t} &= \beta \log(\bar{K}_{it} - K^*_{i2t}) \\ &= \beta \log \bar{K}_{it} + \frac{\beta}{1-\beta} \log \lambda - \beta \log\left(1 + \lambda^{1/(\beta-1)}\right) \\ &= q_{i2t} + \beta \log \left(\frac{\bar{K}_{it}}{Q^{1/\beta}_{i2t}} - 1 \right) \end{aligned}$$

where the last equality holds by plugging in (32). The key observation is that despite of the simple Cobb-Douglas production function we are using, the MPP function turns out to be much more complicated than that is assumed in Dhyne et al. (2017). The main difficulty in estimating (30) is to disentangle the vector of output $\boldsymbol{q}_{i-jt}$ from inputs and unobserved production efficiency. Assuming that $F_j(\cdot)$ is linear and separately additive in terms of $\boldsymbol{q}_{i-jt}$ and all inputs leads to a problem of mis-specification. Under this assumption, it is likely that the non-linear effects of inputs are captured in the error term, which causes an endogeneity issue.

4 Monte Carlo Study

In this section, we employ a Monte Carlo Study to evaluate the validity of the proposed estimation methodology. We apply our method to a simulated dataset to see whether it can recover the production functions and productivities correctly. Since DPSW is the closest approach to ours, we also compare the estimation results with that obtained using DPSW.

4.1 The data generating process

We consider a simple case where the evolution of the labor and capital are given exogenously. We generate N firms with two periods $t = 2$ and two products $J = 2$. The parameter of interest is $\theta = (\beta_k^1, \beta_k^2, \beta_l^1, \beta_l^2, \beta_m^1, \beta_m^2, \sigma, \boldsymbol{\rho}_0, \bar{\boldsymbol{\rho}})$. Our simulation procedure is as follows. For the first period: (a) given prices of capital, labor and materials, we generate the logged capital, logged labor from the log normal distribution $k_{ij1} \sim N\left(\bar{k}_1,\, \sigma_{k1}^2\right)$, $l_{ij1} \sim N\left(\bar{l}_1,\, \sigma_{l1}^2\right)$. We generate the logged productivity $(\omega_{ij1})_{j=1}^J \sim N(\bar{\omega}_j, \sigma_{\omega_j}^2)$; (b) for each firm, we solve its optimization problem. We record their choices of total material $\bar{M}_{i1}$, output quantities quantities Q_{ij1} and revenues R_{ij1}. Also, we document the logged material-revenue share as s_{i1}^*, and introduce a measurement error τ_i such that the observed material-revenue share is $s_{i1} = s_{i1}^* + \tau_i$. For the second period: (a) given prices of capital, labor and materials, we generate data on capital and labor. The accumulation of capital is assumed to be following:

$$k_{it+1} = \vartheta_1^k \omega_{i1t} + \vartheta_2^k \omega_{i2t} + (1 - \delta)\, k_{it}$$

where $\vartheta_1^k \omega_{i1t} + \vartheta_2^k \omega_{i2t}$ represents the capital investment and $(1-\delta)\, k_{it}$ is the capital stock retained from previous period.[14] δ is the depreciation rate. Similarly, the accumulation of labor is given as:

$$l_{it+1} = \vartheta_1^l \omega_{i1t} + \vartheta_2^l \omega_{i2t} + l_{it}$$

The only difference is that we impose no depreciation for the labor. We choose the weight parameters $\left\{\vartheta_1^k,\, \vartheta_2^k,\, \vartheta_1^l,\, \vartheta_2^l\right\}$ between 0 and 1; (b) we generate the new productivity vector $\omega_{it+1} = \boldsymbol{\rho}_0 + \boldsymbol{\rho}\omega_{ijt} + \epsilon_{ijt}$ where $\epsilon_{ijt} \sim N(0, I)$. Without loss of generality, we choose $\boldsymbol{\rho}$ to be a diagonal matrix; (c) For each firm, we solve its optimization problem. Document their choice of total material M_{i2}, their quantities Q_{ij1} and revenues R_{ij2}. Also, we document the input of material as a share of total revenue s_{i2}^*, and add a measurement error such that $s_{i2} = s_{i2}^* + \tau_i$. See Appendix B for a summary of the parameter values for the Monte Carlo experiment.

[14] The simulation strategy reflects that the investment is increasing in firm-product technical efficiency. Note that our estimation strategy does not require the data on investment.

4.2 Estimation results

The simulation result is displayed in Table 1. Our simulation result shows that estimates are very close to the true value, so the selected instruments identify the parameters of interest. In most cases, the 95% confidence interval covers the true value. As sample size increases, the confidence sets become narrower.

Table 1: Estimation Result with 95% bootrapped confidence interval

	True Value	$N = 1000$	$N = 5000$
β_k^1	0.1	0.1402 [0.0972,0.2225]	0.1045 [0.0860,0.1356]
β_k^2	0.3	0.2451 [0.1700,0.4062,]	0.2913 [0.2480,0.3415]
β_l^1	0.3	0.2938 [0.3085,0.2689]	0.2992 [0.2888,0.3051]
β_l^2	0.1	0.1024 [0.0845,0.1373]	0.1006 [0.0932,0.1118]
β_m^1	0.65	0.6502 [0.6452,0.6555,]	0.6490 [0.6468,0.6514]
β_m^2	0.6	0.6047 [0.5980,0.6104]	0.6012 [0.5984,0.6038]
ρ_1^1	0.8	0.7867 [0.7289,0.8117]	0.7953 [0.7768,0.8113]
ρ_2^1	0	0.0182 [-0.0122,0.0826]	0.0102 [-0.0093,0.0327]
ρ_2^2	0.9	0.9150 [0.8804,1.0071]	0.9039 [0.8866,0.9251]
ρ_1^2	0	-0.0158 [-0.1237,0.0157]	-0.0014 [-0.0263,0.0169]
σ	3	2.9813 [2.9620,3.0006]	3.0007 [2.9918,3.0079]

4.3 Comparison to DPSW

In this section, we compare the result with Dhyne et al. (2017) using our Cobb-Douglas Setting. As discussed above, the separative additive form (31) fails for Cobb-Douglas production function. The comparison here aims to detect the degree of mis-specification error. We employ the following estimation strategy:

$$\omega_{ijt} + u_{it} = q_{ijt} - \beta_k^j \bar{k}_{it} - \beta_l^j \bar{l}_{it} - \beta_m^j \bar{m}_{it} - \kappa_j q_{i,-jt} \tag{33}$$

where ω_{ijt} is the productivity. Their idea is to use the evolution equation of productivity

$$\omega_{ijt} = \rho_0^j + \bar{\rho}_1^j \omega_{ijt-1} + \xi_{ijt}$$

and then write ω_{ijt-1} as a non-parametric function of instruments that helps identify the parameters. Substituting ω_{ijt} and keep error terms on the left-hand side yields

$$u_{it} + \xi_{ijt} = q_{ijt} - \beta_k^j \bar{k}_{it} - \beta_l^j \bar{l}_{it} - \beta_m^j \bar{m}_{it} - \kappa_j q_{i,-jt} - \rho_0^j - \bar{\rho}_1^j \omega_{ijt-1}$$

We use the first-stage identified value of $\beta_m^1, \beta_m^2, \sigma$ from (9) and (8), and treat them as knowns in the second stage . In order to solve the problem that DPSW employs some instruments that we don't have in simulation, we replace the lagged value of productivity with the true allocation rule

$$\alpha_{xit-1}^j = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j' \in \mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt-1}^{j'}}\right)$$

and

$$\omega_{ijt-1} = q_{ijt-1} - \beta_k^j(\bar{k}_{it} + \alpha_{kit}^j) - \beta_l^j(\bar{l}_{it} + \alpha_{lit}^j) - \beta_m^j(\bar{m}_{it} + \alpha_{mit}^j)$$

Our estimating equation for the second stage relies on the moment condition:

$$E[u_{it} + \xi_{ijt} | \mathcal{I}_t] = 0.$$

We use the same set of instrument as we did in implementingour method.

Table 2 shows the result from the adapted version of Dhyne et al. (2017). The adapted approach does not capture the correct coefficient on capital and labor. Moreover, the estimator is not stable when we increase the sample size, which illustrates the mis-specification error can be large when the assumptions on the separative additive functional form fails.

Table 2: Estimation Result from DPSW

	True Value	$N = 1000$	$N = 5000$
β_k^1	0.1	0.0246	0.2454
β_k^2	0.3	0.2516	0.0859
β_l^1	0.3	0.3300	0.3226
β_l^2	0.1	0.0551	0.0149
β_m^1*	0.65	0.6521	0.6497
β_m^2*	0.6	0.6044	0.6032
$\kappa_1\#$	N/A	0.0450	-0.0014
$\kappa_2\#$	N/A	-0.0268	-0.0424
ρ_1	0.8	0.7053	0.7630
ρ_2	0.9	0.9834	0.9998
$\sigma*$	3	2.9709	2.9939

Note: The coefficient on β_m^1, β_m^2 and σ is estimated using GNR's first order condition approach. The extra two coefficients in DPSW's method, κ_1 and κ_2, are their control of the unobserved allocation rule.

Lastly, we compare our estimated productivity and that obtained in the spirit of DPSW. Given the estimate $\{\hat{\beta}_x^j, \hat{\sigma}, \hat{\rho}_1, \hat{\rho}_2, \hat{\kappa}\}_{\{x\in\{k,l,m\}, j=1,2\}}$, log productivity using DPSW's method is given by

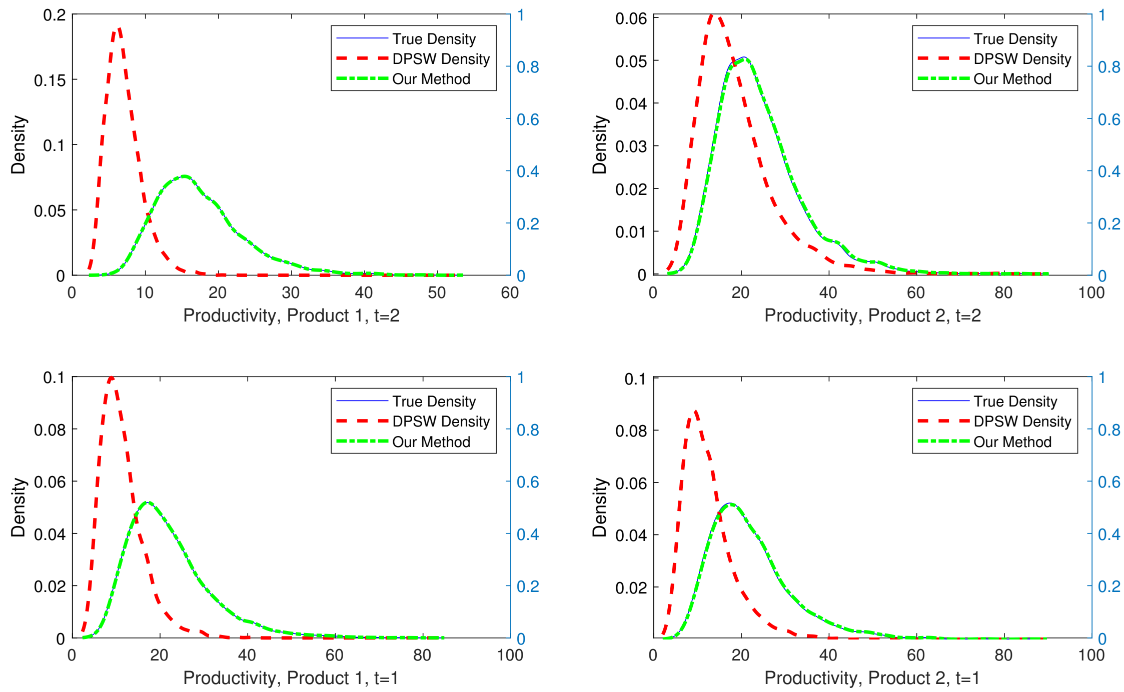
$$\hat{\omega}_{ijt}^{DPSW} = q_{ijt} - \hat{\beta}_k^j \bar{k}_{it} - \hat{\beta}_l^j \bar{l}_{it} - \hat{\beta}_m^j \bar{m}_{it} - \hat{\kappa} q_{i,-jt}$$

and the log productivity employing our method is

$$\hat{\omega}_{ijt}^{new} = q_{ijt} - \hat{\beta}_k^j (\bar{k}_{it} + \hat{\alpha}_{kit}^j) - \hat{\beta}_l^j (\bar{l}_{it} + \hat{\alpha}_{lit}^j) - \hat{\beta}_m^j (\bar{m}_{it} + \hat{\alpha}_{mit}^j)$$

The estimated marginal distributions of productivity are presented in Figure 1.

Figure 1: Comparison of productivity distribution



Note: The sample size is 5000.

Our method captures the distribution quite well, but DPSW approach significantly underestimates the productivity. The table below shows that the correlation between the technical efficiencies of these two products are estimated to be lower than our method. In contrast, our method obtains a correlation pattern that is very close to the original data.

Table 3: Estimated Productivity Correlation

	True	DPSW	Our Method
$Corr(\omega_{i12}, \omega_{i22})$	0.95	0.6784	0.9521

5 Empirical Study

In this section, we apply our proposed methodology of estimating the production function for multi-product firms to real data.

5.1 Data

We have data on the sector of food manufacturing from the Chinese Annual Industry Survey (CAIS) and the production survey between 2000 and 2006. These two data sets are complied by China's National Bureau of Statistics (NBS). In the CAIS dataset, we observe firm-level production information including sales, number of employees, capital stock, and total material inputs. In the production survey, firms are asked to report annual output quantities of their major products. Only firms with annual sales greater than 5 million RMB are considered. To implement our estimation method, we link these two datasets using the unique firm identifier–the legal code. As we mentioned before, our model of multi-product firms features a single input and multiple outputs. Considering this, we choose the sector of grain manufacturing to implement our method. There are two main products for the grain manufacturing: rice and fodder, both are made from grain. The appealing feature is that for multi-product firms operating in this sector, the prices for grain mainly vary at the firm-year level but not at the product level. This provides a good empirical counterpart to the setting of our model. Another reason is that the Chinese production survey employs a relatively aggregate classification of products. This greatly decreases the observations we can use to identify the production function for multi-product firms.

In Table 4, we display the number of observations for different types of firms in our final sample. There are three types of firms: firms only producing rice, firms only producing fodder, and firms producing both rice and fodder. The number of observations of firms only producing fodder is the greatest in each year. We also report the effective number of observations in the last row of the table. Because our estimation relies on a dynamic productivity process, we require at least two consecutive observations for each firm. We end up with 1748 observations for firms only producing rice, 4982 for firms only producing fodder, and 301 for firms producing both products.

Table 4: Sample Description

Year	Rice Only	Fodder Only	Both
2000	206	813	112
2001	247	912	119
2002	266	906	84
2003	270	902	51
2004	352	962	27
2005	717	1268	41
2006	684	1157	37
Effective	1748	4982	301

Note: Each firm-year pair is effective if the firm also exists in the previous year, e.g. if a firm is in the panel for 2000,2001 and 2002, it is counted as 2 effective observations in years 2001 and 2002.

5.2 Empirical Results

We employ a GMM estimator to estimate the production function. To avoid the problem of local minimum, we use the MCMC simulation method proposed by Chernozhukov and Hong (2003). We perform the estimation for every group of firms separately. We obtain a Markov chain based on 12000 simulations and discard the first 2000 simulations in order to minimize the impact of the chosen initial values on biasing the estimates. In the estimation, we impose the parameters for production function to be positive so that the optimization problem of the firm is well defined. The estimation results are reported in Table 5.

The demand elasticity is estimated to be 9.92, indicating a markup of 11.2%. We find differences in the production technology between bi-product firms and single-product firms even for the same product. For the production of fodder, the bi-product firms have a production function with larger values of β_m, β_k, and β_l, indicating larger output-input elasticities for the bi-product firms. We observe a similar pattern for the production of rice except that the single production firm has smaller coefficients for all of the inputs. Moreover, we notice that the coefficient of capital is smaller than that of labor and material. This may be because of the nature of the production of agricultural goods: conditional on labor and material inputs, capital stock, such as machines and warehouse buildings, does not contribute much to the quantities of outputs.

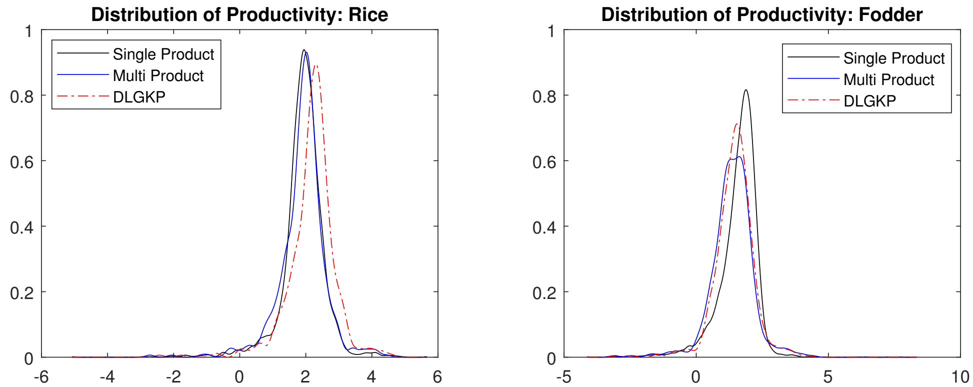
Table 5: Empirical results for the food sector

	Fodder					Rice			
	bi-product		single product			bi-product		single product	
	est.	CI	est.	CI		est.	CI	est.	CI
β_m	.836	[.836 .837]	.772	[.767 .776]	β_m	.768	[.768 .768]	.724	[.717 .733]
β_k	.038	[.002 .082]	.012	[.001 .030]	β_k	.033	[.001 0.115]	.014	[.001 .037]
β_l	.219	[.078 .364]	.026	[.003 .056]	β_l	.090	[.017 .167]	.017	[.001 .042]
ρ_0	.026	[-.061 .123]	.052	[.014 .092]	ρ_0	.113	[-.001 .178]	.520	[.370 .667]
ρ_f	.699	[.540 .840]	.938	[.909 .964]	ρ_r	.917	[.878 .958]	.685	[.605 .763]
ρ_{fr}	.003	[-.086 .100]	n.a	n.a	ρ_{rf}	-.199	[-.218 -.178]		
σ	9.920	[9.862 10.021]							

Note: Estimates for coefficients and confidence intervals are obtained using MCMC method.

Our model allows for correlation between productivity of products produced by the same firm. For bi-product firms, the matrix of correlation coefficients for the productivity process shows a high-level persistence. For the productivity of producing rice, the persistence coefficient is 0.917. The persistence coefficient is only 0.699 for the productivity of producing rice. The correlation coefficient ρ_{rf} is estimated to be -0.199, significant at 1% significance level. In contrast, ρ_{fr} is found to be insignificant, though it is . As far as we know, this paper is the first to show the productivity correlation across different product lines. One possible explanation is the span of control within the bi-product firms. When the production line of fodder is highly productive today, it is more likely that managers will exert more effort on producing fodder, which further decreases the productivity for rice tomorrow.

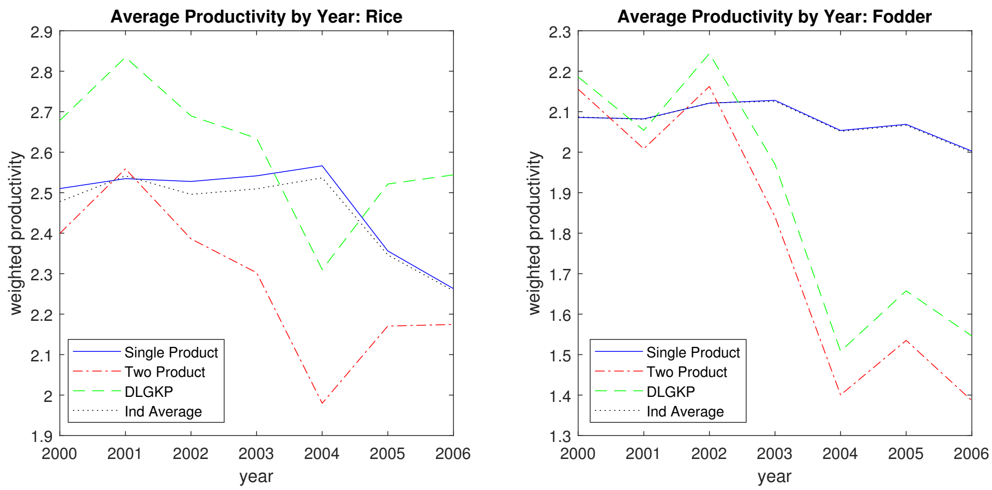
Figure 2: Density of logged productivity



Note: The density of logged productivity is estimated by pooling all firm-year obervations.

After we obtain the estimates for production function parameters, we calculate the productivity by subtracting inputs from the observed quantities. In Figure 2, we display the density of estimated productivity for single-product firms and multiple-product firms. Panel A describes the productivity distribution for rice. Bi-product firms tend to have slightly higher productivity than single-product firms. To show the importance of controlling for heterogeneity in production technology, we also compute the productivity by imposing a same production function of single-product and bi-product firms as in DLGKP. We see that the productivity estimates obtained using DLGKP are slightly higher, implying an upward bias. In Panel B, we show the density function of the productivity estimates for fodder. We see that bi-product firms producing fodder at lower technical efficiencies than firms only producing fodder. Still, ignoring the heterogeneity in the production function would bias the productivity estimates upward. The results show that, compared with single-product firms, bi-product firms do better when producing its core product (rice) but worse producing their peripheral products (fodder). As supporting evidence, we find that, on average, fodder only accounts 30% of the total outputs by the bi-product firm. Our findings are consistent with recent theoretical models by Bernard et al. (2010), Bernard et al. (2011), and Mayer et al. (2014). All of these models show that, in equilibrium, multi-product firms produce their “core” products more efficiently and generate higher revenues.

Figure 3: Evolution of average productivity



Note:The average productivity is computed as the weighted average of firm-product productivity using capital stock as the weight.

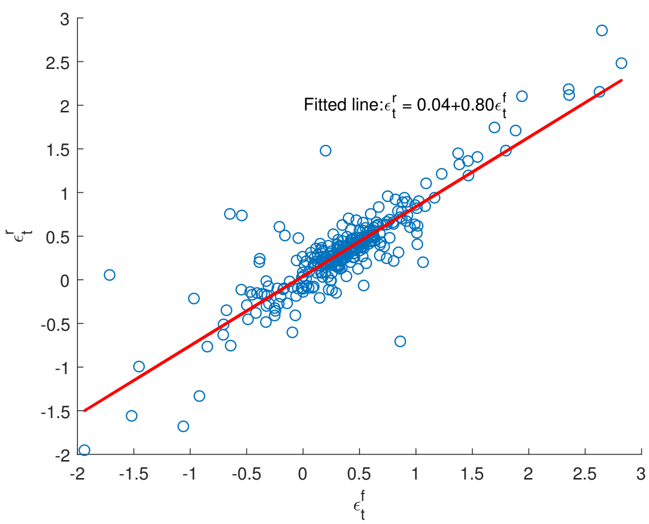
In Figure 3, we show the evolution of the average production efficiency. We compute the average productivity for different products and different types of firms using firm's capital stock as the weights.[15] We notice that, on average, bi-product firms are less productive than single product firms in the production of rice or fodder. More importantly, if we impose a common production technology for single-product and bi-product firms when they produce the same product, we tend to have an upward bias in evaluating the average productivity. By separating the heterogeneity in production technology and technical efficiencies, our method can provide a more reliable estimate of the firm-product productivity for multi-product firms.

Because our model permits productivity shocks at the firm-product level, we can also investigate the co-movement between two different products of bi-product firms.

[15]Our resutls are qualitatively robust if we choose to use total employment or sales as the weights.

We back out the productivity shocks by constructing the error term in the productivity process after we plug in the estimated productivity. In Figure 4, we present a scatter plot of the productivity shocks of rice and fodder for bi-product firms. We see a strong positive correlation between the productivity shocks. In particular, the estimated linear model has a slope of 0.80. The productivity shocks that hit the production of rice are likely to influence fodder in the same direction. More importantly, this co-movement will be reflected by the co-movement of productivity and production of different products for multi-product firms.

Figure 4: Correlation pattern of productivity shocks



Note:ϵ_t^f is the productivity shocks to the fodder's productivity; ϵ_t^r is the productivity shocks to rice's productivity.

6 Extension: Multiple Material Inputs

In this section, we extend production function to allow for multiple material inputs. Our allocation proposition 1 still holds, so moment conditions (8),(10) and (9) can still be implemented. We still suffer from the non-identification issue, so we look at a nested CES type parametric implementation. Let s denote the variety of material input, $M_{ijt}(s)$ the amount of material variety s allocated to product j. In particular, we consider following production function:

$$Q_{ijt} = \exp\left(\omega_{ijt}\right) K_{ijt}^{\beta_k^j} L_{ijt}^{\beta_l^j} \left(\sum_{s=1}^{S_m} \alpha_m^j\left(s\right) M_{ijt}^{\frac{\theta-1}{\theta}}\left(s\right) \right)^{\frac{\beta_m^j \theta}{\theta-1}} \tag{34}$$

where $\alpha_m^j\left(s\right)$ is the share of material variety s , θ is the substitution elasticity between different any two material varieties, and β_m^j is the share of aggregate material inputs. Define the averaged material input as

$$M_{ijt} = \left[\sum_{s=1}^{S_m} \alpha_m^j\left(s\right) M_{ijt}^{\frac{\theta-1}{\theta}}\left(s\right) \right]^{\frac{\theta}{\theta-1}}, \tag{35}$$

we immediately obtain the Cobb-Douglass production function as discussed before. Let $v_{it}\left(s\right)$ be the price of material input $M_{ijt}\left(s\right)$, solving the optimization problem of $M_{it}\left(s\right)$ delivers

$$\frac{v_{it}\left(s\right) M_{ijt}\left(s\right)}{R_{ijt}} = \tilde{\beta}_m^j \frac{\partial m_{ijt}}{\partial m_{ijt}\left(s\right)} \tag{36}$$

where $m_{ijt} = \ln\left(M_{ijt}\right)$ and $m_{ijt}\left(s\right) = \ln\left(M_{ijt}\left(s\right)\right)$. Using functional form (35), we have

$$\begin{aligned} \frac{\partial m_{ijt}}{\partial m_{ijt}\left(s\right)} &= \frac{\alpha_m^j\left(s\right) M_{ijt}^{\frac{\theta-1}{\theta}}\left(s\right)}{\sum_{s'=1}^{S_m} \alpha_m^j\left(s'\right) M_{ijt}^{\frac{\theta-1}{\theta}}\left(s'\right)}, \\ &= \left[\sum_{s'=1}^{S_m} \frac{\alpha_m^j\left(s'\right)}{\alpha_m^j\left(s\right)} \left(\frac{M_{ijt}\left(s'\right)}{M_{ijt}\left(s\right)} \right)^{\frac{\theta-1}{\theta}} \right]^{-1} \end{aligned} \tag{37}$$

We assume that the econometrician cannot observe the allocation of material inputs to different products. However, he knows the total expenditures on material input variety s. Summing the expenditures over different products for input variety s and divide it by the firm's total revenue:

$$
\begin{aligned}
\frac{v_{it}(s)\,M_{it}(s)}{R_{it}} &= \frac{v_{it}(s)\sum_j M_{ijt}(s)}{R_{it}} \\
&= \sum_{j:\alpha_m^j(s)\neq 0} \tilde{\beta}_m^j \frac{R_{ijt}}{R_{it}} \left(\sum_{s'=1}^{S_m} \frac{\alpha_m^j(s')}{\alpha_m^j(s)} \left(\frac{M_{ijt}(s')}{M_{ijt}(s)} \right)^{\frac{\theta-1}{\theta}} \right)^{-1} && (38) \\
&= \sum_{j:\alpha_m^j(s)\neq 0} \tilde{\beta}_m^j \frac{R_{ijt}}{R_{it}} \left(\sum_{s'=1}^{S_m} \left(\frac{v_{it}(s')}{v_{it}(s)} \right)^{1-\theta} \left(\frac{\alpha_m^j(s')}{\alpha_m^j(s)} \right)^{\theta} \right)^{-1} && (39)
\end{aligned}
$$

where we use first order condition $\frac{M_{ijt}(s')}{M_{ijt}(s)} = \left(\frac{v_{it}(s')\alpha_m^j(s)}{v_{it}(s)\alpha_m^j(s')}\right)^{-\theta}$ to obtain the last equality. The first thing we should note is that if there is no variation in input price, we cannot separately identify $\tilde{\beta}_m^j$ and the confounding ratio $\sum_{s'=1}^{S_m}\left(\frac{v_{it}(s')}{v_{it}(s)}\right)^{1-\theta}\left(\frac{\alpha_m^j(s')}{\alpha_m^j(s)}\right)^{\theta}$. This is different from the single input case, where $\sum_{s'=1}^{S_m}\left(\frac{v_{it}(s')}{v_{it}(s)}\right)^{1-\theta}\left(\frac{\alpha_m^j(s')}{\alpha_m^j(s)}\right)^{\theta} = 1$. In this case, there is no confounding issue even if material input price is the same across firms. As is well known in the existing literature, the CES production function faces non-identification problem without an appropriate normalization. This is because the material-to-revenue equation is homogeneous of degree zero with respect to $\{\alpha_m^j(s)\}_{s=1}^{S_m}$. Then the variation of $\frac{R_{ijt}}{R_{it}}$ and $\frac{v_{it}(s')}{v_{it}(s)}$ will help identify $(\tilde{\beta}_m^j, \theta, \frac{\alpha_m^j(s')}{\alpha_m^j(s)})$. The identification power of θ comes from the non-linearity in the power term $\left(\frac{v_{it}(s')}{v_{it}(s)}\right)^{1-\theta}$. After some normalization of the location and imposing that $\sum_{s=1}^{S_m}\alpha_m^j(s) = 1$, we can estimate $\{\beta_m^j\}$, $\{\alpha_m^j(s)\}$, and θ using the moment condition from the equation of material-to-revenue ratio. Note that the identification of σ is the same as we discussed before. To identify β_k^j and β_l^j, we can form a moment condition by using the productivity process as (20).

7 Conclusion

This paper studies the identification of production function for multi-product firms. Starting with a stylized model of multi-product firms, we first show that the firm-product level production function is non-parametrically non-identified without observing the allocation of inputs and exogenous variations in input prices. This result extends the finding by Gandhi et al. (2016) to the context of multi-product firms. We then propose an estimation strategy by employing moment equality for any parametric family of production functions with Hicks-neutral productivity. We point out that in general this method is computation-intensive and suffers from an unclear interpretation of the empirical content.

Given the difficulty in identifying general parametric forms of the production function of multi-product firms, we turn to discuss some practical examples that feature a clear identification and easy computation. We find that with Cobb-Douglas production function, the unobserved input allocations can be controlled by simply using information on product-level output quantities or revenues. Moreover, with the aid of material-to-revenue ratio, our methodology identifies the production function in the absence of exogenous variations in input prices.

We have discussed possible generalizations of our method and offered two parametric examples: the nested CES production function and the CES production function with parameter constraints. To show the validity of our methodology, we have conducted several Monte Carlo studies. Our Monte Carlo study shows that our estimation strategy performs well in backing out the production function parameters as well as the productivity distribution.

We apply our method to a sample of Chinese firms manufacturing rice and/or fodder between 2000 and 2006. Our empirical results show that the production technology of multi-product firms differs from the single-product firms even for the same product. After taking the heterogeneity in production technology into consideration, on average, multi-product firms are found to be more productive in its core product than single-product firms producing the same product. Moreover, we detect a strong co-movement in the productivity shocks of different products of bi-product firms.

In the current paper, we do not model the firm's decision on choosing the set of products. As the productivity of producing different products may also interact with firm's choice of products to produce, we need to adjust for firm's endogenous choice of product set in estimating the firm's productivity. We find this to be an important avenue for future research.

Appendice

A Math Appendix

A.1 Cobb-Douglas production function

A.1.1 Input allocations for C-D production function

From first-order conditions (13)(14)(15), we know that

$$\frac{X_{ijt}}{X_{ij't}} = \frac{\tilde{\beta}_x^j R_{ijt}}{\tilde{\beta}_x^{j'} R_{ij't}}, \forall x \in \{k, l, m\} \tag{A.1}$$

This implies that

$$\begin{aligned} X_{ijt} &= \frac{X_{ijt}}{X_{it}} X_{it} \\ &= \frac{\tilde{\beta}_x^j R_{ijt}}{\sum_{j'} \tilde{\beta}_x^j R_{ij't}} X_{it} \end{aligned} \tag{A.2}$$

If we define $\gamma_{ijt}^{j'} = R_{ij't}/R_{ijt}$ and write the input allocation in logged form, we obtain

$$x_{ijt} = \ln \left(\frac{\tilde{\beta}_x^j}{\sum_{j' \in \mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt}^{j'}} \right) + x_{it}. \tag{A.3}$$

A.1.2 Material-revenue share equation

For firm i and product j, the material-revenue share is

$$S_{ijt}^m = \frac{v_{jt} M_{ijt}}{R_{ijt}} = \tilde{\beta}_m^j \tag{A.4}$$

Note that the firm-level material-revenue share can be expressed as

$$
\begin{aligned}
S_{it}^{m} &= \frac{\sum_j v_{jt} M_{ijt}}{\sum_{j'} R_{ij't}} \\
&= \sum_j \frac{v_{jt} M_{ijt}}{R_{ijt}} \frac{R_{ijt}}{\sum_{j'} R_{ij't}} \\
&= \sum_j \frac{\tilde{\beta}_m^j}{\Gamma_{ijt}}
\end{aligned} \tag{A.5}
$$

A.1.3 Firm-level revenue equation

Note that the firm's total revenue is just a summation of revenues of all the products:

$$
R_{it} = \sum_{j=1}^{J} R_{ijt} = P_t Q_t^{\frac{1}{\sigma}} \sum_{j'} \gamma_{ijt}^{j'} \exp\left(\tilde{\omega}_{ijt}\right) K_{ijt}^{\tilde{\beta}_k^j} L_{ijt}^{\tilde{\beta}_l^j} M_{ijt}^{\tilde{\beta}_m^j} \tag{A.6}
$$

Using the optimal input allocations, the log of deflated revenue of firm i can be expressed as

$$
\begin{aligned}
\tilde{r}_{it} = {} & \tilde{\beta}_k^j k_{it} + \tilde{\beta}_l^j l_{it} + \tilde{\beta}_m^j m_{it} + \sum_{x \in \{k,l,m\}} \tilde{\beta}_x^j \alpha_{xit}^j \\
& + \frac{1}{\sigma} q_t + \ln\left(\sum_{j'} \gamma_{ijt}^{j'}\right) + \tilde{\omega}_{ijt}
\end{aligned} \tag{A.7}
$$

A.1.4 Only observing product-level revenues

Assumption 4 requires us to observe the total revenues and product specific quantities. In principle, one can also back out the output prices using this information. However, in many cases, researchers can only observe product revenues. If the researcher can only observe the information on product revenues, we cannot employ the moment condition (9) to identify the demand elasticity σ. Let $\tilde{\Theta}$ be the set of all the interested parameters. For any $\theta \in \Theta \backslash \{\sigma\}$, we have $\frac{(\sigma-1)\theta}{\sigma} \in \tilde{\Theta}$. We can still employ the moment condition for the material-revenue share and the the revenues. Now the residual function can be written as

$$\boldsymbol{\xi}_{it}^{R}\left(\tilde{\Theta}\right)=\begin{bmatrix} u_{it} \\ \tilde{\boldsymbol{\epsilon}}_{it} \end{bmatrix} \tag{A.8}$$

where $\tilde{\boldsymbol{\epsilon}}_{it}=\frac{\sigma-1}{\sigma}\boldsymbol{\epsilon_{it}}$ is the scaled productivity shocks. Recall that logged product revenue can be represented as

$$r_{ijt}=\tilde{\beta}_{k}^{j}k_{ijt}+\tilde{\beta}_{l}^{j}l_{ijt}+\tilde{\beta}_{m}^{j}m_{ijt}+p_{t}+\frac{1}{\sigma}q_{t}+\tilde{\omega}_{ijt} \tag{A.9}$$

where $\tilde{\beta}_{x}^{j}=\frac{\sigma-1}{\sigma}\beta_{x}^{j}$ for $x\in\{k,\, l,\, m\}$ and $\tilde{\omega}_{ijt}=\frac{\sigma-1}{\sigma}\omega_{ijt}$. We then can solve for $\tilde{\omega}_{ijt}$ as

$$\tilde{\omega}_{ijt}=r_{ijt}-\frac{\sigma-1}{\sigma}f\left(k_{ijt},\, l_{ijt},\, m_{ijt};\boldsymbol{\beta}_{j}\right)-p_{t}-\frac{1}{\sigma}q_{t} \tag{A.10}$$

Therefore the scaled error term is

$$\tilde{\epsilon}_{it}=r_{ijt}-\frac{\sigma-1}{\sigma}f\left(k_{ijt},\, l_{ijt},\, m_{ijt};\boldsymbol{\beta}_{j}\right)-p_{t}-\frac{1}{\sigma}q_{t}-\tilde{\rho}_{0}^{j}-\bar{\boldsymbol{\rho}}^{j}\tilde{\boldsymbol{\omega}}_{it-1} \tag{A.11}$$

where the productivity vector is

$$\tilde{\boldsymbol{\omega}}_{it-1}=\begin{bmatrix} \tilde{\omega}_{i1t-1} \\ \tilde{\omega}_{i2t-1} \\ \vdots \\ \tilde{\omega}_{iJt-1} \end{bmatrix}$$

According to the allocation rule (17), we can replace x_{ijt} with x_{it} and α_{xit}. Note that we can construct the revenue ratio between product j' and product j , $\gamma_{ijt}^{j'}$ using the observed revenue data. Therefore we can construct moment conditions similar to (24) and apply a joint GMM estimator. As a result, we obtain estimate for the scaled parameters. Using the scaled parameters, we can back out the productivity up to a scale of the inverse of markup. When the demand elasticity is common to all firms, the obtained productivity can serve the purpose of empirical analysis on the determinants of productivity change. However, if the demand elasticity differs for different sectors, the productivity differences will also reflect the differences in markups.[16]

A.2 Proofs for Lemmas and Propositions

A.2.1 Proofs in Section 2

Proposition 1

Proposition. *There exist functions* $\{\mathcal{K}_j, \mathcal{L}_j, \mathcal{M}_j\}_{j=1,\dots,J}$ *that depend only on total capital, labor and material, production function parameters and output quantities, such that*

$$\begin{aligned}
K^*_{ijt} &= \mathcal{K}_j(K_{it}, L_{it}, M_{it},\, Q_{i1t}, ...Q_{iJt};\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J, \sigma)\\
L^*_{ijt} &= \mathcal{L}_j(K_{it}, L_{it}, M_{it},\, Q_{i1t}, ...Q_{iJt};\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J, \sigma)\\
M^*_{ijt} &= \mathcal{M}_j(K_{it}, L_{it}, M_{it}, Q_{i1t}, ...Q_{iJt},\, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J, \sigma)
\end{aligned}$$

is the solution to the optimization problem (6). The optimal allocation does not depend on the firm-specific input price ω_{it}*,* r_{it} *and* v_{it}*.*

Proof. By taking derivatives with respect to K_{ijt}, L_{ijt}, and M_{ijt}, we obtain the first-order conditions

$$\begin{aligned}
K^*_{ijt} &= \frac{\sigma-1}{\sigma}\frac{P_t Q_t^{\frac{1}{\sigma}} Q_{ijt}^{\frac{\sigma-1}{\sigma}} \frac{\partial \ln(F_j)}{\partial \ln(K_{ijt})}}{r_{it}+\lambda_{kit}}\\
L^*_{ijt} &= \frac{\sigma-1}{\sigma}\frac{P_t Q_t^{\frac{1}{\sigma}} Q_{ijt}^{\frac{\sigma-1}{\sigma}} \frac{\partial \ln(F_j)}{\partial \ln(L_{ijt})}}{w_{it}+\lambda_{lit}}\\
M^*_{ijt} &= \frac{\sigma-1}{\sigma}\frac{P_t Q_t^{\frac{1}{\sigma}} Q_{ijt}^{\frac{\sigma-1}{\sigma}} \frac{\partial \ln(F_j)}{\partial \ln(M_{ijt})}}{v_{it}}
\end{aligned}$$

where λ_{kit} and λ_{lit} are the Lagrangian multipliers for capital constraint $\sum_j K^*_{ijt} = K_{it}$ and labor constraint $\sum_j L^*_{ijt} = L_{it}$, respectively. Note that we have $3J$ the first-order conditions along with the 3 resource constraints, $\sum_j X^*_{ijt} = X_{it}$, where $X \in \{K, L, M\}$.[17] There are $3J$ unknown allocation of input, along with 3 unknown prices:

[16] See De Loecker (2011) for a similar discussion in the estimation of physical productivity.

[17] This is because as an econometrician we observe the total input of material.

$\{\frac{P_t Q_t^{\frac{1}{\sigma}}}{r_{it}+\lambda_{kit}}, \frac{P_t Q_t^{\frac{1}{\sigma}}}{w_{it}+\lambda_{lit}}, \frac{P_t Q_t^{\frac{1}{\sigma}}}{r_{it}}\}$. Since we impose the optimization problem to be a convex problem, the first order condition along with the resource constrain is sufficient and necessary for the optimal allocation, and it has a solution. Since we have $3J + 3$ unknowns and $3J + 3$ equations, we have the unique solution.

In the system of equations, $K_{it}, L_{it}, M_{it}, Q_{ijt}$ are observed, $\boldsymbol{\beta}$ and σ are the parameters that are treated as known, and $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}, \frac{P_t Q_t^{\frac{1}{\sigma}}}{r_{it}+\lambda_{kit}}, \frac{P_t Q_t^{\frac{1}{\sigma}}}{w_{it}+\lambda_{lit}}, \frac{P_t Q_t^{\frac{1}{\sigma}}}{r_{it}}\}$ are treated as unknown to be solved. There fore the unknowns can be written as the function of observables. □

Proposition (2)

Proposition. *Let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ *be in the identified set, then* $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\boldsymbol{\rho}}_0, \bar{\boldsymbol{\rho}}, \sigma)$ *is also in the identified set, where*

$$\tilde{\rho}_0^j = \rho_0^j + \sum_{j'=1}^{J} (\rho_{jj'} \ln \alpha_{j'}) - \ln \alpha_j$$

Proof. (9), (8) are not influenced by changing from $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ to $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\boldsymbol{\rho}}_0, \bar{\boldsymbol{\rho}}, \sigma)$ because $f_j = \ln F_j$, and we take the partial derivatives with respect to m_{ijt}. It suffice to check (10) holds. Note that

$$\begin{aligned} q_{ijt} - f_j - \rho_0^j - \boldsymbol{\rho}^j \boldsymbol{\omega} &= q_{ijt} - f_j - \rho_0^j - \sum_{j'} \bar{\rho}_{jj'} (q_{ij't} - f_{j'}) \\ &= q_{ijt} - \ln \alpha_j - f_j - \sum_{j'} \bar{\rho}_{jj'} (q_{ij't} - \ln \alpha_{j'} - f_{j'}) \\ &\quad - (\rho_0^j - \ln \alpha_j + \sum_{j'} \bar{\rho}_{jj'} \ln \alpha_{j'}) \\ &= q_{ijt} - \tilde{f}_j - \bar{\boldsymbol{\rho}}^j \boldsymbol{\omega} - \tilde{\rho}_0^j \end{aligned}$$

So the moment condition ((10)) still holds. □

Proposition (3)

Proposition. *Let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ *be in the identified set, then* $(\{\tilde{F}_j\}_{j=1}^J, \boldsymbol{\rho_0}, \bar{\boldsymbol{\rho}}, \sigma)$ *is also in the identified set, where*

$$\tilde{F}_j(K_{ijt}, L_{ijt}, M_{ijt}) = F_j(K_{ijt} - C_j^K, L_{ijt} - C_j^L, M_{ijt} - C_j^M) \quad \forall j = 1, ...J$$

and the constants $\{C_j^X | j = 1, ...J,\ X \in \{K, L, M\}\}$ *satisfies*

$$\sum_{j=1}^J C_j^X = 0 \quad \forall X \in \{K, L, M\}$$

Proof. By proposition (1), let $(K_{ijt}^*, L_{ijt}^*, M_{ijt}^*)_{j=1}^J$ be the optimal allocation rule under $\{F_j\}_{j=1}^J$. Let

$$\tilde{X}_{ijt}^* = X_{ijt}^* + C_j^X \quad \forall X \in \{K, L, M\},$$

then we claim that $\tilde{X}_{ij}^*$ is the optimal allocation rule under $\tilde{F}_j$. This is because the static profit optimization problem (6) is a convex optimization, it suffice to check the first order condition and the constraint. The resource constraint holds by the construction $\sum_{j=1}^J C_j^X = 0$. So it suffices to check the first order condition. The first order condition with respect to K_{ijt} gives

$$F_j^{-1/\sigma} \frac{\partial F_j}{\partial K_{ijt}}(K_{ijt}^*, L_{ijt}^*, M_{ijt}^*) = F_{j'}^{-1/\sigma} \frac{\partial F_{j'}}{\partial K_{ij't}}(K_{ij't}^*, L_{ij't}^*, M_{ij't}^*)$$

by our construction, it is easy to see

$$\tilde{F}_j^{-1/\sigma} \frac{\partial \tilde{F}_j}{\partial K_{ijt}}(\tilde{K}_{ijt}^*, \tilde{L}_{ijt}^*, \tilde{M}_{ijt}^*) = \tilde{F}_{j'}^{-1/\sigma} \frac{\partial \tilde{F}_{j'}}{\partial K_{ij't}}(\tilde{K}_{ij't}^*, \tilde{L}_{ij't}^*, \tilde{M}_{ij't}^*)$$

so $\tilde{X}_{ij}^*$ is the optimal allocation rule under $\tilde{F}_j$. Since we do not observe the allocation rule, these two sets of production functions both generate the same observation of K_{it}, L_{it}, M_{it} and Q_{ijt}. $\square$

A.2.2 Proofs in Section 3

Proposition (4)

Proposition. *Let* $\{\beta_k^j, \beta_l^j, \beta_m^j, \boldsymbol{\rho}_0, \bar{\boldsymbol{\rho}}, \sigma\}$ *satisfies moment condition (10). If there is no variation in the price of materials across firms, i.e.* $v_{it} = v_t$*, then* $\{\beta_k^{*j}, \beta_l^{*j}, \beta_m^{*j}, \boldsymbol{\rho}_0^{*}, \bar{\boldsymbol{\rho}}^{*}, \sigma\}$ *that satisfies*

$$
\begin{aligned}
\beta_k^{*j} &= c\beta_k^j \quad \forall j \\
\beta_l^{*j} &= c\beta_l^j \quad \forall j \\
1 - \beta_m^{*j}\frac{\sigma-1}{\sigma} &= c(1 - \beta_m^j\frac{\sigma-1}{\sigma}) \quad \forall j
\end{aligned}
$$

also satisfies moment condition (10) for some value of $\boldsymbol{\rho}_0^*$ *.*

Proof. Using (11) (2) and (17), the optimal choice of material is given as

$$
m_{ijt} = \frac{1}{1-\tilde{\beta}_m^j}\left(\tilde{\beta}_k^j k_{ijt} + \tilde{\beta}_l^j l_{ijt} + \tilde{\omega}_{ijt} + \mu_{ijt}\right) \tag{A.12}
$$

where $\mu_{ijt} \equiv \ln\left(\tilde{\beta}_m^j P_t Q_t^{\frac{1}{\sigma}}/v_{it}\right)$ is a product-year fixed effects. Recall the information is

$$
\mathcal{I}_{it} = \{k_{it},\, l_{it},\, k_{it-1},\, m_{it-1},\, l_{it-1}, q_{i1t}, ...q_{iJt}\cdots\}
$$

Plug (A.12) and (17) into (18):

$$
\begin{aligned}
q_{ijt} = &\frac{\beta_k^j}{1-\tilde{\beta}_m^j}k_{it} + \frac{\beta_l^j}{1-\tilde{\beta}_m^j}l_{it} + \frac{1}{1-\tilde{\beta}_m^j}\sum_{x\in\{k,l\}}\beta_x^j\alpha_{xit}^j \\
&+ \frac{\beta_m^j}{1-\tilde{\beta}_m^j}\mu_{ijt} + \frac{\omega_{ijt}}{1-\tilde{\beta}_m^j}
\end{aligned} \tag{A.13}
$$

Using the productivity process specified in (4) and (A.12), we can write current productivity ω_{ijt} as

$$
\omega_{ijt} = \rho_0^j + \bar{\boldsymbol{\rho}}^j\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{t-1}\right) + \epsilon_{ijt} \tag{A.14}
$$

where ρ_0^j is the jth element of $\boldsymbol{\rho}_0$ and $\bar{\boldsymbol{\rho}}^j$ is the jth row of $\bar{\boldsymbol{\rho}}$; the vector of productivity is given by

$$\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1}, \boldsymbol{\beta}\right) = \begin{bmatrix} \left(1-\tilde{\beta}_m^1\right) m_{i1t-1} - \tilde{\beta}_k^1 k_{i1t-1} - \tilde{\beta}_l^1 l_{i1t-1} - \mu_{i1t-1} \\ \vdots \\ \left(1-\tilde{\beta}_m^j\right) m_{ijt-1} - \tilde{\beta}_k^j k_{ijt-1} - \tilde{\beta}_l^j l_{ijt-1} - \mu_{ijt-1} \\ \vdots \\ \left(1-\tilde{\beta}_m^J\right) m_{iJt-1} - \tilde{\beta}_k^J k_{iJt-1} - \tilde{\beta}_l^J l_{iJt-1} - \mu_{iJt-1} \end{bmatrix} \tag{A.15}$$

Combine it with (A.13) we can express the re-scaled error term for product j as

$$\begin{aligned} \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^j} = q_{ijt} - \frac{\beta_k^j}{1-\tilde{\beta}_m^j} k_{it} - \frac{\beta_l^j}{1-\tilde{\beta}_m^j} l_{it} - \frac{1}{1-\tilde{\beta}_m^j} \sum_{x\in\{k,l\}} \beta_x^j \alpha_{xit}^j \\ - \frac{\beta_m^j}{1-\tilde{\beta}_m^j}\mu_{ijt} - \frac{\rho_0^j}{1-\tilde{\beta}_m^j} - \frac{\bar{\boldsymbol{\rho}}^j}{1-\tilde{\beta}_m^j} \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1},\, \boldsymbol{\beta}\right) \end{aligned} \tag{A.16}$$

By construction, we have $\frac{\beta_k^{*j}}{1-\tilde{\beta}_m^{*j}} = \frac{\beta_k^j}{1-\tilde{\beta}_m^j}$, and recall $\alpha_{xit}^j(\boldsymbol{\beta}) = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j'\in\mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt}^{j'}}\right) =$

$\alpha^j_{xit}(\boldsymbol{\beta^*})$so we derive from (A.16)

$$\begin{aligned}
\frac{\epsilon^j_{it}}{1-\tilde{\beta}^j_m} &= q_{ijt} - \frac{\beta^{*j}_k}{1-\tilde{\beta}^{*j}_m} k_{it} - \frac{\beta^{*j}_l}{1-\tilde{\beta}^{*j}_m} l_{it} - \frac{1}{1-\tilde{\beta}^{*j}_m} \sum_{x\in\{k,l\}} \beta^{*j}_x \alpha^j_{xit}(\boldsymbol{\beta}^*) \\
&\quad - \frac{\beta^j_m}{1-\tilde{\beta}^j_m}\mu_{ijt} - \frac{\rho^j_0}{1-\tilde{\beta}^j_m} - \frac{\bar{\boldsymbol{\rho}}^j}{1-\tilde{\beta}^j_m}\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1},\, \boldsymbol{\beta}\right) \\
{\scriptstyle (i)} &= q_{ijt} - \beta^{*j}_k k_{it} - \beta^{*j}_l l_{it} - \beta^{*j}_m m_{it} - \sum_{x\in\{k,l,m\}} \beta^{*j}_x \alpha^j_{xit}(\boldsymbol{\beta}^*) + \frac{\epsilon^j_{it}}{1-\tilde{\beta}^{*j}_m} \\
&\quad + \frac{\beta^{*j}_m}{1-\tilde{\beta}^{*j}_m}\mu^*_{ijt} - \frac{\beta^j_m}{1-\tilde{\beta}^j_m}\mu_{ijt} \\
&\quad + \tilde{\beta}^{*j}_m \frac{\rho^{*j}_0 + \bar{\boldsymbol{\rho}}^{*j}\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}^*_{it-1},\, \boldsymbol{\beta^*}\right)}{1-\tilde{\beta}^{*j}_m} \\
&\quad - \frac{\rho^j_0 + \bar{\boldsymbol{\rho}}\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}^*_{it-1},\, \boldsymbol{\beta}\right)}{1-\tilde{\beta}^j_m} \qquad \text{(A.17)}
\end{aligned}$$

where in the equality (i) we use the first order condition of m_{ijt} under $\boldsymbol{\beta}^*, \rho^*_0, \bar{\boldsymbol{\rho}}^*$

$$m_{ijt} = \frac{1}{1-\tilde{\beta}^{*j}_m}\left(\tilde{\beta}^{*j}_k k_{ijt} + \tilde{\beta}^{*j}_l l_{ijt} + \rho^{*j}_0 + \bar{\boldsymbol{\rho}}^{*j}\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu^*}_{it-1},\, \boldsymbol{\beta^*}\right) + \mu^*_{ijt}\right)$$

where $\mu^*_{ijt} = \ln\left(\tilde{\beta}^{*j}_m P_t Q_t^{\frac{1}{\sigma}}/v_{it}\right)$.

Now we consider the moment condition term under parametrization $(\boldsymbol{\beta}^*, \rho^*_0, \bar{\boldsymbol{\rho}}^*)$. Define

$$\begin{aligned}
\Delta_{ijt} = q_{ijt} - \beta^{*j}_k k_{it} - \beta^{*j}_l l_{it} - \beta^{*j}_m m_{it} - \sum_{x\in\{k,l,m\}} \beta^{*j}_x \alpha^j_{xit}(\boldsymbol{\beta}^*) \\
- \rho^{*j}_0 - \boldsymbol{\rho}^{*j}\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1},\, \boldsymbol{\beta}^*\right)
\end{aligned}$$

By using equation(A.17) we have

$$\Delta_{ijt} = \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^j} - \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^{*j}} + \underbrace{\frac{\beta_m^{*j}}{1-\tilde{\beta}_m^{*j}}\mu_{ijt}^* - \frac{\beta_m^j}{1-\tilde{\beta}_m^j}\mu_{ijt}}_{A}$$

$$+ \underbrace{\frac{\rho_0^j + \bar{\boldsymbol{\rho}}^j \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{t-1},\, \boldsymbol{\beta}\right)}{1-\tilde{\beta}_m^j} - \frac{\rho_0^{*j} + \bar{\boldsymbol{\rho}}^{*j} \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{t-1},\, \boldsymbol{\beta}^*\right)}{1-\tilde{\beta}_m^{*j}}}_{B}$$

To prove the claim in the lemma, it suffice to show $E[\Delta_{ijt}|\mathcal{I}_t] = 0$.

$$A = \frac{\beta_m^{*j}\ln\tilde{\beta}_m^{*j}}{1-\tilde{\beta}_m^{*j}} - \frac{\beta_m^{*j}\ln\tilde{\beta}_m^{*j}}{1-\tilde{\beta}_m^{*j}} + \left[\frac{\beta_m^{*j}}{1-\tilde{\beta}_m^{*j}} - \frac{\beta_m^j}{1-\tilde{\beta}_m^j}\right]\ln P_t Q_t^{1/\sigma} - \left[\frac{\beta_m^{*j}}{1-\tilde{\beta}_m^{*j}} - \frac{\beta_m^j}{1-\tilde{\beta}_m^j}\right]\ln v_{it}$$

Also recall that we choose $\bar{\boldsymbol{\rho}}^j = \bar{\boldsymbol{\rho}}^{*j}$, so we have

$$\begin{aligned}
B &= \frac{\rho_0^j}{1-\tilde{\beta}_m^j} - \frac{\rho_0^{*j}}{1-\tilde{\beta}_m^{*j}} + \frac{\bar{\boldsymbol{\rho}}^j\boldsymbol{\omega}}{1-\tilde{\beta}_m^j} - \frac{\bar{\boldsymbol{\rho}}^j\boldsymbol{\omega}^*}{1-\tilde{\beta}_m^{*j}} \\
&= \frac{\rho_0^j}{1-\tilde{\beta}_m^j} - \frac{\rho_0^{*j}}{1-\tilde{\beta}_m^{*j}} + \sum_{s=1}^J \frac{\bar{\rho}_s^j\mu_{ist-1}}{1-\tilde{\beta}_m^j} - \frac{\bar{\rho}_s^j\mu_{ist-1}^*}{1-\tilde{\beta}_m^{*j}} \\
&= \frac{\rho_0^j}{1-\tilde{\beta}_m^j} - \frac{\rho_0^{*j}}{1-\tilde{\beta}_m^{*j}} + \sum_{s=1}^J \frac{\bar{\rho}_s^j\ln\tilde{\beta}_m^j}{1-\tilde{\beta}_m^j} - \frac{\bar{\rho}_s^j\ln\tilde{\beta}_m^{*j}}{1-\tilde{\beta}_m^{*j}} \\
&\quad + \sum_{s=1}^J \frac{\bar{\rho}_s^j\ln P_{t-1}Q_{t-1}^{1/\sigma}}{1-\tilde{\beta}_m^j} - \frac{\bar{\rho}_s^j\ln P_{t-1}Q_{t-1}^{1/\sigma}}{1-\tilde{\beta}_m^{*j}} + \sum_{s=1}^J \frac{\bar{\rho}_s^j\ln v_{it-1}}{1-\tilde{\beta}_m^j} - \frac{\bar{\rho}_s^j\ln v_{it-1}}{1-\tilde{\beta}_m^{*j}}
\end{aligned}$$

So our term $A + B$ satisfies

$$\begin{aligned}A + B = & \frac{\rho_0^j}{1 - \tilde{\beta}_m^j} - \frac{\rho_0^{*j}}{1 - \tilde{\beta}_m^{*j}} + \sum_{s=1}^{J} \frac{\rho_s^j \ln \tilde{\beta}_m^j}{1 - \tilde{\beta}_m^j} - \frac{\rho_s^j \ln \tilde{\beta}_m^{*j}}{1 - \tilde{\beta}_m^{*j}} + \\ & + \sum_{s=1}^{J} \frac{\rho_s^j \ln P_{t-1} Q_{t-1}^{1/\sigma}}{1 - \tilde{\beta}_m^j} - \frac{\rho_s^j \ln P_{t-1} Q_{t-1}^{1/\sigma}}{1 - \tilde{\beta}_m^{*j}} + \left[\frac{\beta_m^{*j}}{1 - \tilde{\beta}_m^{*j}} - \frac{\beta_m^j}{1 - \tilde{\beta}_m^j} \right] \ln P_t Q_t^{1/\sigma} \\ & + \left[\frac{\beta_m^{*j}}{1 - \tilde{\beta}_m^{*j}} - \frac{\beta_m^j}{1 - \tilde{\beta}_m^j} \right] \ln v_{it} + \sum_{s=1}^{J} \frac{\rho_s^j \ln v_{it-1}}{1 - \tilde{\beta}_m^j} - \frac{\rho_s^j \ln v_{it-1}}{1 - \tilde{\beta}_m^{*j}}\end{aligned}$$

When there is no variation in material input price, $A + B$ is just a constant, so we can choose ρ_0^{*j} to make $A + B = 0$. So the moment condition

$$E[\Delta | \mathcal{I}_t] = E\left[\frac{\epsilon_{it}^j}{1 - \tilde{\beta}_m^j} - \frac{\epsilon_{it}^j}{1 - \tilde{\beta}_m^{*j}} \middle| \mathcal{I}_t \right] = 0$$

holds for the set of value $(\boldsymbol{\beta}^*, \sigma, \boldsymbol{\rho}_0^*, \boldsymbol{\rho}^*)$. □

A.2.3 Local and global identification of β_k and β_l

Now, for the Cobb-Douglas production function, we use method similar to Gandhi et al. 2016 to identify β_m^j. Since m_{ijt} is a function of all other variables, we replace it by the first order condition to help understand the identification conditions. When $\boldsymbol{\rho}$ is diagonal, the revenue equation

$$\begin{aligned}\frac{\epsilon_{it}^j}{1 - \tilde{\beta}_m^j} = & q_{ijt} - \frac{\beta_k^j}{1 - \tilde{\beta}_m^j} k_{it} - \frac{\beta_l^j}{1 - \tilde{\beta}_m^j} l_{it} - \frac{1}{1 - \tilde{\beta}_m^j} \sum_{x \in \{k,l\}} \beta_x^j \alpha_{xit}^j \\ & - \frac{\bar{\rho}_j^j}{1 - \tilde{\beta}_m^j} \left[\left(1 - \tilde{\beta}_m^j\right) (m_{it-1} + \alpha_{mit-1}^j) - \tilde{\beta}_k^j k_{it-1} - \tilde{\beta}_l^j l_{ijt-1} - \sum_{x \in \{k,l\}} \tilde{\beta}_x^j \alpha_{xit-1}^j \right] \\ & + \frac{\bar{\rho}_j^j}{1 - \tilde{\beta}} \mu_{jt-1} + const\end{aligned} \tag{A.18}$$

The complication comes from the input allocation rule $\alpha_{xit}^{j} = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j'\in\mathcal{J}}\tilde{\beta}_x^{j'}\gamma_{ijt}^{j'}}\right)$. The term $\gamma_{ijt}^{j'} = \left(\frac{Q_{ij't}}{Q_{ijt}}\right)^{1-1/\sigma}$ is endogeneous, and depends on the productivity vector $\boldsymbol{\omega}_{it}$. This causes α_{xit}^{j} to be dependent on lagged values of total capital, total labor and product quantity. So by taking expectation of $\mathbf{E}\left(q_{ijt}|\mathcal{I}_{it}\right)$, we typically do not get a clean separable additive form of (k_{it}, l_{it}) and $(k_{it-1}, l_{it-1}, q_{ijt})$. Instead, we provide sufficient conditions to ensure the local and global identification of capital and labor parameters. First we define the function

$$\tilde{\alpha}_{xt}^{j} = \mathbf{E}\left(\alpha_{xit}^{j}|\mathcal{I}_{it}\right)$$

Lemma 1. *$\tilde{\alpha}_{xit}^{j}$ is a function of the information set $\mathcal{I}_{it}$ and $\tilde{\beta}_k^j$ only. That is, we can write*

$$\tilde{\alpha}_{xt}^{j} = h_x(\tilde{\beta}_x^1, ..., \tilde{\beta}_x^J, \mathcal{I}_{it}),\ x \in \{k, l\}$$

Moreover, this function is identified from the data.

Assumption 7. *(Local Identification of capital and labor input parameter) For any parameter value $\boldsymbol{\beta}_k$=$\left(\beta_k^1, \cdots, \beta_k^J\right)$ and $\boldsymbol{\beta}_l = \left(\beta_l^1, \cdots, \beta_l^J\right)$ near the true parameter,*

1. Define matrix $A^k = [A_{ij}^k]$, where the $(i,j)-th$ entry

$$A_{ij}^k = \begin{cases} \frac{1}{1-\tilde{\beta}_m^j}\left(1 - \frac{\partial\tilde{\alpha}_{kt}^j}{\partial k_{it}} - \tilde{\beta}_k^j\frac{\partial^2\tilde{\alpha}_{kt}^j}{\partial k_{it}\partial\tilde{\beta}_k^j}\right) & \text{if } i = j \\ -\frac{\tilde{\beta}_k^j}{1-\tilde{\beta}_m^j}\frac{\partial^2\tilde{\alpha}_{kt}^j}{\partial k_t\partial\tilde{\beta}_k^j} & \text{if } i \neq j \end{cases}$$

   For some value $\tilde{I} \in supp(\mathcal{I}_{it})$, the matrix A^k is invertible near the true value.

2. Define matrix $A^l = [A_{ij}^l]$, where the $(i,j)-th$ entry

$$A_{ij}^l = \begin{cases} \frac{1}{1-\tilde{\beta}_m^j}\left(1 - \frac{\partial\tilde{\alpha}_{lt}^j}{\partial l_{it}} - \tilde{\beta}_l^j\frac{\partial^2\tilde{\alpha}_{lt}^j}{\partial l_{it}\partial\tilde{\beta}_l^j}\right) & \text{if } i = j \\ -\frac{\tilde{\beta}_l^j}{1-\tilde{\beta}_m^j}\frac{\partial^2\tilde{\alpha}_{lt}^j}{\partial l_{it}\partial\tilde{\beta}_l^j} & \text{if } i \neq j \end{cases}$$

   For some value $\tilde{I} \in \text{supp}(\mathcal{I}_{it})$, the matrix A^k is invertible near the true value.

Proposition 5. *Under assumption 6, the parameters* $\boldsymbol{\beta}_k = (\beta_k^1, \cdots, \beta_k^J)$ *and* $\boldsymbol{\beta}_l = (\beta_l^1, \cdots, \beta_l^J)$ *are locally identified.*

Proposition (5) utilizes the standard local identification condition. Partial derivatives with respect to firm-level total capital helps us to get rid of the lagged terms and makes the conditions easier to verify. Given that $\tilde{\alpha}_{xit}^j = h_x(\tilde{\beta}_x^1, ..., \tilde{\beta}_x^J, \mathcal{I}_{it})$ are identified from the data, assumption (7) is testable. Given that we have a complicated functional form of $\tilde{\alpha}_{kit}^j$, it is harder to find the global identification condition. The assumption below states a set of sufficient conditions that ensure the global identification.

Assumption 8. *(Global Identification of capital and labor input parameter) For a set of vector values* $\{\tilde{I}_1, ...\tilde{I}_S\} \subset supp\{\mathcal{I}_{it}\}$*, define matrix function* B

$$B(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l) = \begin{bmatrix} B^k \\ B^l \end{bmatrix}$$

where

$$B_{js}^k = \tilde{\beta}_k^j (1 - \frac{\partial \tilde{\alpha}_{kt}^j}{\partial k_{it}}(\tilde{\beta}_k^1, ..., \tilde{\beta}_k^J, \tilde{i}_s)) - \frac{\partial \tilde{\alpha}_{lt}^j}{\partial k_{it}}(\tilde{\beta}_k^1, ..., \tilde{\beta}_k^J, \tilde{I}_s)$$
$$B_{js}^l = \tilde{\beta}_l^j (1 - \frac{\partial \tilde{\alpha}_{lt}^j}{\partial l_{it}}(\tilde{\beta}_k^1, ..., \tilde{\beta}_k^J, \tilde{i}_s)) - \frac{\partial \tilde{\alpha}_{kt}^j}{\partial l_{it}}(\tilde{\beta}_k^1, ..., \tilde{\beta}_k^J, \tilde{I}_s)$$

Also define the matrix C

$$C = \begin{bmatrix} C^k \\ C^l \end{bmatrix}$$

where $C_{js}^k = \mathbf{E}\left(\frac{\partial q_{ijt}}{\partial k_{it}} | \mathcal{I}_t = \tilde{I}_s\right)$, $C_{js}^l = \mathbf{E}\left(\frac{\partial q_{ijt}}{\partial l_{it}} | \mathcal{I}_t = \tilde{I}_s\right)$. *The system* $B(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l) = C$ *has a unique solution of* $(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l)$.

Proposition 6. *Under assumption 7, we can globally identify capital and labor input parameter* $(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l)$.

The conditions imposed in assumption 7 are very similar to the identification argument in Gandhi et al. 2016, where partial derivatives with respect to k_t or l_t help to get rid of the lagged terms in the revenue equation. Assumption 7 identifies the parameters because it directly impose that the system $\mathbf{E}\left(\frac{\partial q_{ijt}}{\partial k_{it}}|\mathcal{I}_t\right) = \frac{\tilde{\beta}_k^j - \tilde{\beta}_k^j \frac{\partial \tilde{\alpha}_{kt}^j}{\partial k_{it}} - \tilde{\beta}_l^j \frac{\partial \tilde{\alpha}_{lt}^j}{\partial k_{it}}}{1-\tilde{\beta}_m^j}$ has a unique solution when we have sufficient variation in the support of $\mathcal{I}_t$.

Proof of (5)

Proof. Note that by differentiate the moment condition, from equation A.16 we have

$$\mathbf{E}\left(\frac{\partial q_{ijt}}{\partial k_{it}}|\mathcal{I}_t\right) = \frac{\tilde{\beta}_k^j - \tilde{\beta}_k^j \frac{\partial \tilde{\alpha}_{kit}^j}{\partial k_{it}} - \tilde{\beta}_l^j \frac{\partial \tilde{\alpha}_{lit}^j}{\partial k_{it}}}{1-\tilde{\beta}_m^j}.$$

.This is because the remaining term $\boldsymbol{\omega}\left(\boldsymbol{m}_{it-1}, \boldsymbol{k}_{it-1}, \boldsymbol{l}_{it-1}, \boldsymbol{\mu}_{t-1}\right)$ only involves lagged value in the information set. Further note that $\tilde{\beta}_l^j \frac{\partial \tilde{\alpha}_{lit}^j}{\partial k_{it}}$ only includes $\boldsymbol{\beta}_l$, so their derivatives with respect to $\boldsymbol{\beta_k}$ will disappear. Assumption 6 states the Jacobian matrix of the RHS with respect to $\tilde{\beta}_k^j$ is invertible for some value in the support of the information set, which is the local identification condition of $\boldsymbol{\beta}_k$=$\left(\beta_k^1, \cdots, \beta_k^J\right)$, given the value of σ. Similar is the condition for labor input parameter. □

A.3 Optimal inputs for CES production function with parameter constraints

The optimization problem using CES production function is

$$\max_{\{K_{ijt}, L_{ijt}, M_{ijt}\}} \sum_{j=1}^{J} P_t Q_t^{\frac{1}{\sigma}} \left(\beta_k^j K_{ijt}^\theta + \beta_l^j L_{ijt}^\theta + \beta_m^j M_{ijt}^\theta\right)^{\frac{1}{\theta}} - r_t \sum_j K_{ijt} - w_t \sum_j L_{ijt} - v_t \sum_j M_{ijt}$$
$$s.t. \sum_j K_{ijt} = K_{it}, \sum_j L_{ijt} = L_{it}$$

The first-order conditions are

$$K_{ijt} = \frac{\sigma-1}{\sigma}\frac{R_{ijt}}{r_t+\lambda_{kit}}\frac{\beta_k^j K_{ijt}^\theta}{\beta_k^j K_{ijt}^\theta+\beta_l^j L_{ijt}^\theta+\beta_m^j M_{ijt}^\theta} \tag{A.19}$$

$$L_{ijt} = \frac{\sigma-1}{\sigma}\frac{R_{ijt}}{w_t+\lambda_{lit}}\frac{\beta_l^j L_{ijt}^\theta}{\beta_k^j K_{ijt}^\theta+\beta_l^j L_{ijt}^\theta+\beta_m^j M_{ijt}^\theta} \tag{A.20}$$

$$M_{ijt} = \frac{\sigma-1}{\sigma}\frac{R_{ijt}}{w_t+v_t}\frac{\beta_l^j M_{ijt}^\theta}{\beta_k^j K_{ijt}^\theta+\beta_l^j L_{ijt}^\theta+\beta_m^j M_{ijt}^\theta} \tag{A.21}$$

This implies that

$$L_{ijt} = \left[\frac{\beta_k^j\left(r_t+\lambda_{kit}\right)}{\beta_l^j\left(w_t+\lambda_{lit}\right)}\right]^{\frac{1}{\theta-1}} K_{ijt} = \left(C_l\frac{r_t+\lambda_{kit}}{w_t+\lambda_{lit}}\right)^{\frac{1}{\theta-1}} K_{ijt} \tag{A.22}$$

$$M_{ijt} = \left[\frac{\beta_k^j\left(r_t+\lambda_{kit}\right)}{\beta_m^j v_t}\right]^{\frac{1}{\theta-1}} K_{ijt} = \left(C_m\frac{r_t+\lambda_{kit}}{v_t}\right)^{\frac{1}{\theta-1}} K_{ijt} \tag{A.23}$$

Define the relative prices as $p_{lit} \equiv \frac{r_t+\lambda_{kit}}{w_t+\lambda_{lit}}$ and $p_{kit} \equiv \frac{r_t+\lambda_{kit}}{v_t}$. Plug the above two equations back to (A.19) and consider two different products for firm i, we obtain that

$$\begin{aligned}\frac{R_{ijt}}{R_{ij't}} &= \frac{K_{ijt}}{K_{ij't}}\frac{1+C_l^{\frac{1}{\theta-1}}p_{lit}^{\frac{\theta}{\theta-1}}+C_m^{\frac{1}{\theta-1}}p_{kit}^{\frac{\theta}{\theta-1}}}{1+C_l^{\frac{1}{\theta-1}}p_{lit}^{\frac{\theta}{\theta-1}}+C_m^{\frac{1}{\theta-1}}p_{kit}^{\frac{\theta}{\theta-1}}} \\ &= \frac{K_{ijt}}{K_{ij't}}\end{aligned}$$

This implies that $K_{ijt} = \frac{R_{ijt}}{R_{it}}K_{it}$. Then by (A.22) and (A.23), we also have $L_{ijt} = \frac{R_{ijt}}{R_{it}}L_{it}$ and $M_{ijt} = \frac{R_{ijt}}{R_{it}}M_{it}$.

B Monte Carlo Description

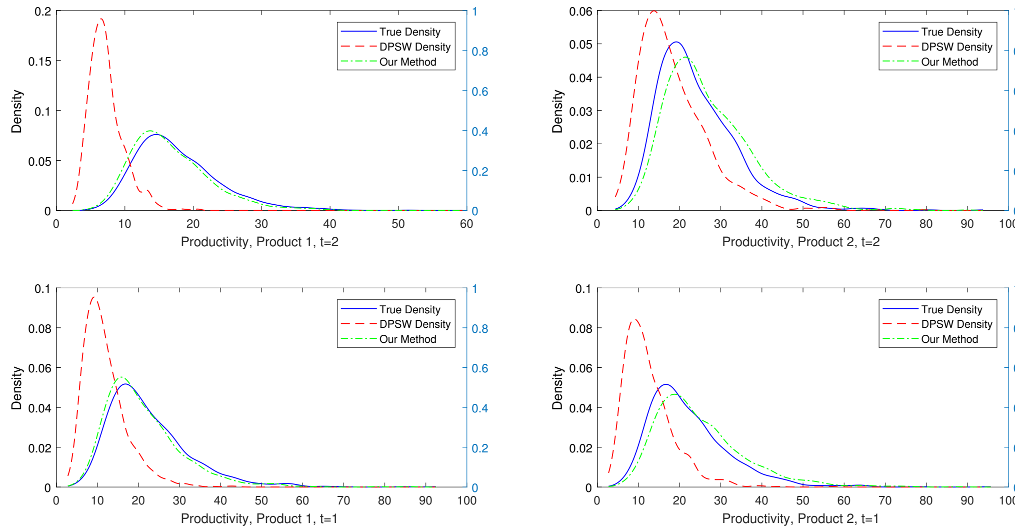
The parameters used in our Monte Carlo study is summarized in following table:

Table B.1: Parameter Values for Monte Carlo Study

Parameters	Description	Values
σ	Demand elasticity	3
$\left(\bar{k}_1, \bar{l}_1\right)$	Initial mean of capital and labor	(10, 10)
(σ_k^2, σ_l^2)	Initial variances of capital and labor	$(1, 1)$
δ	Capital depreciation rate	0.1
r_t	Capital price	1
w_t	Labor price	1
v_{jt}	Material price of product j	1
P_t	Aggregate price index	1
$(\beta_k^1, \beta_l^1, \beta_m^1)$	Production function parameters for product 1	(0.1, 0.3, 0.65)
$(\beta_k^2, \beta_l^2, \beta_m^2)$	Production function parameters for product 2	(0.3, 0.1, 0.6)
$(\bar{\omega}_1, \bar{\omega}_2)$	Initial mean of the productivity vector	$(3, 3)$
$\left(\sigma_{\omega_1}^2, \sigma_{\omega_2}^2\right)$	Initial variance of the productivity vector	(0.5, 0.5)
(ρ_0^1, ρ_0^2)	Intercept vector in productivity process	(0.4, 0.4)
$\begin{bmatrix} \rho_1^1 & \rho_2^1 \\ \rho_1^2 & \rho_2^2 \end{bmatrix}$	Persistence matrix in productivity process	$\begin{bmatrix} 0.8 & 0 \\ 0 & 0.9 \end{bmatrix}$
$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$	Variance-covariance matrix for the productivity shock	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\left(\vartheta_1^k, \vartheta_2^k\right)$	Weights in the capital accumulation rule	(0.02, 0.03)
$\left(\vartheta_1^l, \vartheta_2^l\right)$	Weights in the labor accumulation rule	(0.03, 0.02)
T	Number of periods	2
N	Number of firms	1000/5000

Productivity comparison: Sample size = 1000

Figure B.1: Productivity comparison for simulated data: sample = 1000



References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-product firms and product switching," *American Economic Review*, 100, 70–97.

——— (2011): "Multiproduct firms and trade liberalization," *The Quarterly Journal of Economics*, 126, 1271–1318.

Chernozhukov, V. and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115, 293–346.

de La Grandville, O. (1989): "In quest of the Slutsky diamond," *The American Economic Review*, 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, 1–3.

De Loecker, J. (2011): "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 79, 1407–1451.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, markups, and trade reform," *Econometrica*, 84, 445–510.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2017): "Multi Product Firms, Import Competition, and the Evolution of Firm-product Technical Efficiencies," Working Paper 23637, National Bureau of Economic Research.

Doraszelski, U. and J. Jaumandreu (2013): "R&D and productivity: Estimating endogenous productivity," *Review of Economic Studies*, 80, 1338–1383.

Eckel, C., L. Iacovone, B. Javorcik, and J. P. Neary (2015): "Multi-product firms at home and away: Cost-versus quality-based competence," *Journal of International Economics*, 95, 216–232.

Eckel, C. and J. P. Neary (2010): "Multi-product firms and flexible manufacturing in the global economy," *The Review of Economic Studies*, 77, 188–217.

Gandhi, A., S. Navarro, and D. Rivers (2016): "On the identification of production functions: how heterogenous is productivity," *manuscript*.

Grieco, P. L., S. Li, and H. Zhang (2016): "Production function estimation with unobserved input price dispersion," *International Economic Review*, 57, 665–690.

Klump, R. and O. de La Grandville (2000): "Economic growth and the elasticity of substitution: Two theorems and some suggestions," *American Economic Review*, 90, 282–291.

León-Ledesma, M. A., P. McAdam, and A. Willman (2010): "Identifying the elasticity of substitution with biased technical change," *American Economic Review*, 100, 1330–57.

Levinsohn, J. and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 70, 317–341.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2014): "Market size, competition, and the product mix of exporters," *American Economic Review*, 104, 495–536.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Peters, B., M. J. Roberts, V. A. Vuong, and H. Fryges (2017): "Estimating dynamic R&D choice: an analysis of costs and longrun benefits," *The RAND Journal of Economics*, 48, 409–437.

Wooldridge, J. M. (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 104, 112–114.