Production Function Estimation for Multi-Product Firms*

Zhiyuan Chen† Moyu Liao‡

Preliminary draft, please do not cite without permission

November 4, 2019

Abstract

We study a stylized model of multi-product firm with firm-product level heterogeneity in Hicks-neutral production technology. We characterize the empirical content of the model 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. We develop the model's empirical content to be an estimation strategy for any parametric family of production functions. This procedure, however, suffers from unclear identification and the problem of computing optimal input allocations. In the case of Cobb-Douglas production function, we show the identification of the production function by obtaining a closed-form solution for unobserved input allocations as a function of product-level output quantities or revenues. Monte Carlo evidence shows that our identification strategy performs well. We then apply our methodology to the sector of agricultural goods manufacturing and find multi-product firms' production technologies differs from single-product firms even for the same product. We also find that multi-product firms produce its core (peripheral) products at higher (lower) technical efficiency than single-product firms. Lastly, we document a negative productivity spillover across different products within the multi-product firm.

JEL: C51, L2, L6

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

*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.

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

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

1 Introduction

With the increasing availability of product-level production data, the existence of multi-product firms have become one of the most important features of micro-level production data.[1] A key to understanding the behavior of multi-product firms is the firm-product level technical efficiency. Firm-product level productivity allows researchers to directly detect the impact of interested 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] Also, a good estimation of the 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 suitable for single-product firms. Identifying the firm-product level production function and the productivity has posed challenges to existing estimation methodologies. This is mainly because usually 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 production technologies that are Hicks-neutral and a parameterized demand system. In this model, firms choose the allocation of capital, labor, and material across different production lines. We first clarify the empirical content of the model and show that the firm-product level production function is *non-parametrically non-identified* without observing the allocation of inputs towards different products and exogenous input price variations. Due to the non-identification result, 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

[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.

any parametric family of production functions. We notice, however, in general our estimation strategy suffers from an unclear interpretation of its empirical content. More importantly, one faces the curse of dimensionality when computing the optimal input allocations given the observed data on total inputs and product-specific outputs. We further study a special case of the parametric production functions, i.e., the Cobb-Douglas production. 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 provide two special cases of CES production functions that features transparent identification and relatively easy computation. As an important extension of the estimation framework 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.

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 technical efficiency. Second, we do not impose that firms producing the same product have the same production function no matter they are single-product or multi-product firms. Therefore we can separate the impact of heterogeneous production functions from that of the production efficiency. This allows for possibly different production technologies that motivate firms to choose different sets of products. Based on the seminal contributions by Diewert (1973) and Lau (1976), Dhyne et al. (2017) (DPSW hereafter) develop a multi-product production function approach to identify the production function at firm-product level. For each product, 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 certain parameterized demand structures and firms' first-order conditions of optimal input allocations to identify the physical productivity. 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 this approach proposed by DPSW faces a potential problem of mis-specification in controlling the unobserved input allocations, which causes endogeneity issue and a systematical bias to the estimation of production function. This implies that the researchers need to adjust the control function for unobserved input allocations in order to obtain consistent estimate for the productivity according to the chosen firm-product production function.

In most of existing literature on productivity estimation, one crucial assumption in identifying the production function is the structure of the stochastic productivity process. Under this assumption, a firm's future productivity is determined by its current productivity and exogenous iid productivity shocks. This implies that there is no productivity spillovers across different firms. Considering the existence of possible productivity spillovers within a multi-product firm, we generalize this assumption to allow that the vector of firm-product productivity follows a VAR(1) process with arbitrary productivity correlations between different products produced by the firm. Therefore our approach permits rich firm-product productivity dynamics.

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 that of the true productivity. We also apply DPSW method to the simulated data and show that their method fails to correctly identify the production function parameters. Actually, it systematically under estimate the productivity. This shows that our method can provide a more reliable estimate for the production function. 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 single product. Ignoring the heterogeneity in production function results in an upward bias to the average 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 explain the estimation methodology and its connection to existing methods. Section 4 discusses the generalization of our method. Section 5 presents the Monte Carlo studies we use to establish the validity of our methodology and its comparison to DPSW. Section 6 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 a firm i produces a set of products $\mathcal{J} = \{1, 2, \cdots, J\}$ over time periods $t = 1, 2, \cdots$. For product j, consider a production function with Hicksian-neutral production technology:

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

where ω_{ijt} is the log of unobserved productivity for j-th product. In each period, firms make decisions on inputs including labor $\{L_{ijt}\}$, capital $\{K_{ijt}\}$, and materials $\{M_{ijt}\}$. Note that we allow both of the productivity and production functionto vary at firm-product level: ω_{ijt} potentially captures other heterogeneity such as consumer tastes for the attributes of different products, while $\boldsymbol{\beta}_j \in \mathbb{R}^d$ reflects production function may have different forms for different products. When d is finite, the production function has a parametric form characterized by a finite number of parameters; when d is infinite, $F(\cdot)$ is a non-parametric production function. The inverse demand is of CES form:[4]

$$P_{ijt} = P_t Q_t^{\frac{1}{\sigma}} Q_{ijt}^{-\frac{1}{\sigma}} \tag{2}$$

where P_t is the sectoral price level, and 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]

[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).

Without special notice, we will use F_j to represent the production function of production j. F_j can be in anon-parametric form. We require that the functional form varies across products, but not over time.

Assumption 1. *Firm-level capital stock K_{it} and labor L_{it} are predetermined, and the choice of material M_{it} is static.*

According to this assumption, firms allocate capital and labor to the production lines of different products, while taking its total capital stock and labor as given. Then firms source materials and determine its allocations to different products to maximize its 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} in each period taking input prices as given.*

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

$$V\left(\boldsymbol{\omega}_{it},\, K_{it},\, L_{it},\, \mathbf{s}_{it}\right) = \max_{\{X_{ijt}\}_{j\in\mathcal{J}},\, I_{it},\, L_{it+1}} \{\sum_{j=1}^{J}\left(P_{ijt}Q_{ijt} - w_{it}L_{ijt} - v_{it}M_{ijt} - r_{it}K_{ijt}\right) + \mathbf{E}\left[V\left(\boldsymbol{\omega}_{it+1},\, K_{it+1},\, L_{it+1},\, \mathbf{s}_{it+1}\right)\right]\}$$

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

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

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

where w_{it} is the price of labor, v_{it} is the material price, and r_{it} is the rental price of capital. These prices assumed to be varying across time and firm.$\mathbf{s}_{it}$ represents the exogenous state variables which are not affected by the firm's decision. $\mathbf{s}_{it}$ may include the age of the firm, the aggregate price index, as well as input prices. $\boldsymbol{\omega}_{it} = (\omega_{i1t},\, \cdots,\, \omega_{iJt})' \in \mathbb{R}^J$ 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.

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, the input prices will only vary across firms and periods, but not at the product level. This is a limitation mainly caused by the unobserved input allocations to different products. Introducing product-level heterogeneity in input prices is possible if the econometrician has information on quantities or prices of multiple inputs. In that case, we can easily incorporate multiple material inputs into the production function. In this case, we can write down the production function as $F\left(K_{ijt},\, L_{ijt},\, \left\{M_{ijt}^{s}\right\}_{s=1}^{S};\boldsymbol{\beta}_j\right)$, where S is the number of types of material inputs.[6]

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. The employment in the next period 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.[7] $\boldsymbol{\rho}_0 \in \mathbb{R}^J$ is a vector of constants, $\boldsymbol{\rho} \in \mathbb{R}^J \times \mathbb{R}^J$ 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.[8] $\boldsymbol{\epsilon}_{it+1}$ is J-dimension column vector of i.i.d shocks with zero mean. The identification assumption that we introduce later is that the shocks are mean independent of our information set; we do not impose any other assumption on the higher-order moments. As a result, $\boldsymbol{\epsilon}_{it+1}$ 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 with 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 should be analyzed at the product level for each firm by taking the possible productivity spillovers into account. This can provide a deeper understanding the channels through which R&D investment influences firm-level productivity.

[6]Though this is not formally discussed in the current paper, considering multiple material inputs is an interesting topic to explore when more detailed data on inputs are available.

[7]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.

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

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 material inputs flexibly in each period. To simplify notations, sometimes we use X_{ijt} to represent inputs. We use small letters to represent logged forms, i.e., $x_{ijt} = \ln(X_{ijt})$ and $f_j = \ln(F_j)$. Aggregating the firm-product level inputs to the firm-level we obtain

$$X_{it} = \sum_{j \in \mathcal{J}} X_{ijt},\ X \in \{K,\ L,\ M\} \tag{6}$$

Therefore the static optimization problem of the firm is to

$$\max_{\{K_{ijt},\, L_{ijt},\, M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j=1}^{J} \left\{ P_t Q_t^{\frac{1}{\sigma}} \left[\exp(\omega_{ijt}) F_j\right]^{\frac{\sigma-1}{\sigma}} - w_{it} L_{ijt} - v_{it} M_{ijt} - r_{it} K_{ijt} \right\} \tag{7}$$

$$s.t. \sum_{j=1}^{J} X_{ijt} = X_{it} \quad \text{for} \quad X \in \{K, L\} \tag{8}$$

Following assumption states the condition ensuring the static optimization problem is well-defined.

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 econometrician, we do not observe the productivity ($\boldsymbol{\omega}_{it}$) and input prices (ω_{it}, r_{it}, and 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 $\{K_{it},\ L_{it},\ M_{it}\}$ and the output quantities $\{Q_{ijt}\}_{j\in\mathcal{J}}$, the allocation rules of capital, labor and material is independent of the productivity and input prices.

Proposition 1. *There exist functions $\{\mathbb{K}_j, \mathbb{L}_j, \mathbb{M}_j\}_{j=1,\dots J}$ that depend only on total capital, labor and* *material,output quantities, amd some unknown parameterssuch that*

$$K^*_{ijt} = \mathbb{K}_j(K_{it}, L_{it}, M_{it},\, Q_{i1t}, ...Q_{iJt};\, \boldsymbol{\beta}_j,\, \sigma)$$

$$L^*_{ijt} = \mathbb{L}_j(K_{it}, L_{it}, M_{it},\, Q_{i1t}, ...Q_{iJt};\, \boldsymbol{\beta}_j,\, \sigma)$$

$$M^*_{ijt} = \mathbb{M}_j(K_{it}, L_{it}, M_{it}, Q_{i1t}, ...Q_{iJt};\, \boldsymbol{\beta}_j,\, \sigma)$$

is the solution to the optimization problem (7). The optimal allocation does not depend on the firm-specific input price ω_{it}, r_{it} and 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 and independent of the unobserved capital rental price, material price and wage. This proposition provides us a method of using observables to compute optimal allocation of inputs given any chosen parameters for the production function and the demand system.

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 (i, t) pair, the econometrician observes total capital K_{it}, total labor L_{it}, total material input M_{it}, and output quantity for each product Q_{ijt}. Moreover, the econometrician also observes the ratio of material input to total revenue $s^m_{it} = v_{it}M_{it}/R_{it}$. 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.*

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 look at 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

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

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

$$\mathbf{E}\left[\ln\left(\hat{s}_{it}^m\right) - \ln\left(\sum_j \Gamma_{ijt}^{-1} \frac{\partial f(k_{ijt}^*, l_{ijt}^*, m_{ijt}^*; \boldsymbol{\beta}_j)}{\partial m_{ijt}}\right) | K_{it}, L_{it}, M_{it}, Q_{i1t}, ...Q_{ijt}\right] = 0 \tag{9}$$

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}^*, as we vary the parameters $\{\boldsymbol{\beta}_j\}$, the optimal input allocations $\{k_{ijt}^*,\ l_{ijt}^*,\ m_{ijt}^*\}$ change accordingly. 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 object that is identified via the material-revenue share equation.

To aid the identification of σ, we employ the relation between the total revenue and quantities.[9] 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

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

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

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

$$\mathbf{E}\left(\Delta u_{it} | \{Q_{i1t}, Q_{i2t}, \cdots Q_{ijt}\}, \{Q_{I1t}, Q_{I2t}, \cdots Q_{Ijt}\}\right) = 0 \tag{10}$$

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

$$q_{ijt} = f\left(k^*_{ijt}, l^*_{ijt}, m^*_{ijt}; \boldsymbol{\beta}_j\right) + \rho^j_0 + \boldsymbol{\rho}^j \boldsymbol{\omega}_{it-1} + \epsilon_{ijt},$$

where $\boldsymbol{\omega}_{it-1} = \left(\omega_{i1t-1}, \omega_{i2t-1}, \cdots, \omega_{iJt-1}\right)'$ 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 existsin 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 $\boldsymbol{\rho}^j$ comes from the orthogonality condition of $\boldsymbol{\epsilon}_{it}$, as is the casein the literature. We formally state the assumption for $\boldsymbol{\epsilon}_{it}$ as below.

Assumption 5. *Define the information set* $\mathcal{I}_t = \{K_{it}, L_{it}, K_{it-1}, L_{it-1}, M_{it-1}, Q_{i1t-1}, ...Q_{iJt-1}, ...\}$, *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:

$$\mathbf{E}\left[q_{ijt} - f\left(k_{ijt}^{*},\, l_{ijt}^{*},\, m_{ijt}^{*}; \boldsymbol{\beta}_j\right) - \rho_0^j - \boldsymbol{\rho}^j \boldsymbol{\omega}_{it-1} | \mathcal{I}_t\right] = 0 \quad \forall j = 1, ...J \tag{11}$$

Definition 1. The parameters of interest are $(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J, \boldsymbol{\rho}_0, \boldsymbol{\rho}, \sigma)$. The identified set $\boldsymbol{\Theta}_I$ is the collection of parameters that satisfy moment conditions (9), (10) and (11).

We see that the moment conditions (9) and (11) 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 (9) only containsparameters characterizing the output-to-material elasticity, which greatly simplifies the identification of material-output elasticity by simplyusing the firm-level material-revenue equation.

2.2.1 Identification of σ

We first note that the mark-up parameter σ is identified from (10) 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 conditions (10) identify 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 (10) 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. $\square$

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 capital, labor and material towards each product, the allocation rule can be driven by the production function parameters $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$ as well as the 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* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ *be in the identified set, then* $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\boldsymbol{\rho}}_0, \boldsymbol{\rho}, \sigma)$ *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. See appendix. □

This non-identification of the scale of the production function is firstly proved in Gandhi et al. 2016, but 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 thatthe 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. *When* $\boldsymbol{\beta}_j \in \mathbb{R}^\infty$,*let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ *be in the identified set, then* $(\{\tilde{F}_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$

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. $\square$

The proposition says that it is impossible to identify the relative elasticity of capital $\frac{\partial f_j}{\partial k_j} / \frac{\partial f_{j'}}{\partial k_{j'}}$ at any data points of (k_j, l_j, m_j) and $(k_{j'}, l_{j'}, m_{j'})$. 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 Strategy for 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 potentially identify the production function through following steps:

1. Use moment condition (9) 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 (10) and (11), 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 this estimation algorithm is general and can be applied to any parametric production functions, there are two major issues. First, 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 use. 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. Suppose we only choose 10 points for each state variable, the number of grid points would be 10^5! Second, more importantly, it is not clear what is identified from equation (9) and (11). Note that in equation (9), $\boldsymbol{\beta}_j$ 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 being identified through the material-to-revenue ratio equation. In general, we can not determine whether the estimation procedure we have just stated converges or not.

In what follows, we propose several parametric examples of production functions which have clear identification and are relatively less intensive in computing. Because of the problems of identification and computation, we rely on evidence from Monte Carlo experiments to determine the identification for general parametric forms such as constant elasticity of substitution (CES) and translog.

3.2 Cobb-Douglas Production Function

Consider production functions is following the Cobb-Douglas form:[10]

$$Q_{ijt} = \exp(\omega_{ijt})\, K_{ijt}^{\beta_k^j} L_{ijt}^{\beta_l^j} M_{ijt}^{\beta_m^j} \tag{12}$$

[10]We show later that our approach can be extended to other production functional forms such as Constant Elasticity of Substitution (CES) and translog production functions.

The production technology can also be differing 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 has a wide applicability 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 kinds of 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 describe the firm's static problem as

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

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

$$R_{ijt} = P_t Q_t^{\frac{1}{\sigma}} \left[\exp(\omega_{ijt}) K_{ijt}^{\beta_k^j} L_{ijt}^{\beta_l^j} M_{ijt}^{\beta_m^j} \right]^{\frac{\sigma-1}{\sigma}} \tag{13}$$

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

$$K_{ijt} = \frac{\tilde{\beta}_k^j R_{ijt}}{r_{it} + \lambda_{kit}} \tag{14}$$

$$L_{ijt} = \frac{\tilde{\beta}_l^j R_{ijt}}{w_{it} + \lambda_{lit}} \tag{15}$$

$$M_{ijt} = \frac{\tilde{\beta}_m^j R_{ijt}}{v_{it}} \tag{16}$$

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 that whether the input is a dynamic input does not affect its allocation rule. However, whether the firm is allocationg a dynamic inputwill be reflected by the shadow prices of inputs. In other words, the marginal price will be adjusted such that the optimal level of this input is consistent with the predetermined total input. In particular, the allocation rule for inputs (in logs) are

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

where

$$\gamma_{ijt}^{j'} = \frac{R_{ij't}}{R_{ijt}} = \left(\frac{Q_{ij't}}{Q_{ijt}}\right)^{\frac{\sigma-1}{\sigma}} \tag{17}$$

is the product j's revenue relative to the reference product j.[11] To simplify notations, define the log of input share as

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

Therefore we express the allocation rule for inputs as

$$x_{ijt} = \alpha_{xit}^j + x_{it},\, x \in \{k,\, l,\, m\} \tag{18}$$

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

$$q_{ijt} = \beta_k^j k_{it} + \beta_l^j l_{it} + \beta_m^j m_{it} + \sum_{x \in \{k,l,m\}} \beta_x^j \alpha_{xit}^j + \omega_{ijt} \tag{19}$$

This logged production function links the firm-level inputs to the firm-product level output. Because input shares α_{xit}^j 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.[12] Then we employ the productivity process to identify other parameters in the

[11] See the appendix for the derivation.

[12] 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

production function.

3.2.2 Identification

We first display a lemma showing that the moment restriction (11) alone does not identify the parameters β_x^j even in the Cobb-Douglas parametric framework if there is no exogenous variation in material input prices.

Lemma 1. *Let* $\{\beta_k^j, \beta_l^j, \beta_m^j, \boldsymbol{\rho}_0, \boldsymbol{\rho}, \sigma\}$ *satisfies moment condition (11). If there is no variation in the price of materials across firms, i.e.* $v_{it} = v_t$*, then* $\{\beta_k^{*j}, \beta_l^{*j}, \beta_m^{*j}, \boldsymbol{\rho}_0^*, \boldsymbol{\rho}^*, \sigma\}$ *which are defined as*

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

also satisfies moment condition (11) for some value of $\boldsymbol{\rho}_0^*$ *.*

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. This can be seen from the 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 allocation rule is not changed only when the scaling of parameter is proportional for all j.

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.

Identification power of (9) Under the Cobb-Douglas production function, the material to revenue ratio equation becomes

$$\mathbf{E}\left[\hat{s}_{it}^{m}|Q_{i1t},...Q_{iJt}\right]=\ln\left(\frac{\sigma-1}{\sigma}\sum_{j}\beta_{m}^{j}\frac{Q_{ijt}^{\frac{\sigma-1}{\sigma}}}{\sum_{j'}Q_{ij't}^{\frac{\sigma-1}{\sigma}}}\right) \tag{20}$$

The conditioning set now does not depend on the total K_{it}, L_{it}, M_{it}. Given the identification of σ, the only unknownparameter 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 (11) as

$$\mathbf{E}\left[q_{ijt}-\sum_{x\in\{k,l,m\}}\beta_x^j\alpha_{xit}^j x_{it}-\rho_0^j-\boldsymbol{\rho}^j\boldsymbol{\omega}_{it-1}|\mathcal{I}_t\right]=0, \tag{21}$$

where the productivity vector is

$$\boldsymbol{\omega}_{it-1}=\begin{bmatrix} q_{i1t-1}-\sum_{x\in\{k,l,m\}}\beta_x^1\alpha_{xit-1}^1 x_{it-1} \\ q_{i2t-1}-\sum_{x\in\{k,l,m\}}\beta_x^2\alpha_{xit-1}^2 x_{it-1} \\ \vdots \\ q_{i2t-1}-\sum_{x\in\{k,l,m\}}\beta_x^J\alpha_{xit-1}^J x_{it-1} \end{bmatrix} \tag{22}$$

Choosing appropriate instruments for k_{it} and l_{it}, 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 (18). 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} and α_{kit}^j help us identify β_k^j, l_{it} and α_{lit}^j contribute to the identification of β_l^j. We can separate $\boldsymbol{\rho}^j$ from other parameters through the variations in $\boldsymbol{m}_{it-1}$ 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

We propose a joint GMM estimator to estimate the parameters of interest. As suggested by Wooldridge (2009), this leads to simple inference and more efficient estimators. Note that from (10), (20) and (21) , we have three groups of moment conditions. Let $\Theta = (\sigma, \boldsymbol{\beta}_k, \boldsymbol{\beta}_l, \boldsymbol{\beta}_m, \boldsymbol{\rho}_0, \boldsymbol{\rho})$ be the vector of all ofthe interested parameters. We define a $(J+2) \times 1$ residual function as

$$\boldsymbol{\xi}_{it}(\Theta) = \begin{bmatrix} u_{it} \\ \Delta u_{it} \\ \boldsymbol{\epsilon}_{it} \end{bmatrix} \tag{23}$$

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

where $\boldsymbol{\omega}_{it-1}$ is given by (22). Let $\boldsymbol{Z}_{it}$ be a matrix of instruments

$$\boldsymbol{Z}_{it} = \begin{bmatrix} \boldsymbol{z}_{it1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{z}_{it2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{z}_{it3} \end{bmatrix}, t = 2, \cdots, T \tag{24}$$

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 canuse variables in $\mathcal{I}_{it}$ and their appropriate polynomials for z_{it3}. The joint GMM estimation is based on the moment condition

$$\mathbf{E}\left[\boldsymbol{Z}_{it}'\boldsymbol{\xi}_{it}(\Theta)\right] = 0,\, t = 2, \cdots, T \tag{25}$$

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 the bootstrap to do inference. Given the GMM estimator, the bootstrapped estimator is

$$\Theta^b = \arg\min_{\Theta} \sum_{i^b=1}^{N} [\boldsymbol{Z}'_{i^b t}\boldsymbol{\xi}_{i^b t}(\Theta) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}'_{it}\boldsymbol{\xi}_{it}(\hat{\Theta})]'\hat{W}[\boldsymbol{Z}'_{i^b t}\boldsymbol{\xi}_{i^b t}(\Theta) - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}'_{it}\boldsymbol{\xi}_{it}(\hat{\Theta})]$$

where $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}'_{it}\boldsymbol{\xi}_{it}(\hat{\Theta})$ serves as the re-centering of the bootstrap process, and i^b is the bootstrap index used to obtain the standard errors.

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 motivating 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:

$$F(K_{ijt}, L_{ijt}, M_{ijt}; \boldsymbol{\beta}_j) = \left(\beta_k^j K_{ijt}^{\theta_j} + \beta_l^j L_{ijt}^{\theta_j}\right)^{\frac{1}{\theta_j}} M_{ijt}^{\beta_m^j} \tag{26}$$

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 $\theta_j = 1$, 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.[13] 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 (9) 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 (11) 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:

$$F_j(K_{ijt}, L_{ijt}, M_{ijt}) = \left(\beta_k^j K_{ijt}^\theta + \beta_m^j L_{ijt}^\theta + \beta_m^j M_{ijt}^\theta\right)^{\frac{1}{\theta}} \tag{27}$$

[13] 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[14]

$$\frac{\beta_k^j}{\beta_l^j} = C_l,\ \frac{\beta_k^j}{\beta_m^j} = C_m \quad \forall j \in \mathcal{J} \tag{28}$$

Under this restriction, we can show that the optimal allocation of inputs are[15]

$$X_{ijt}^* = \frac{R_{ijt}}{R_{it}} X_{it}, \forall X \in \{K,\ L,\ M\} \tag{29}$$

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 we can compute the input allocated to different products directly. Therefore, the moment condition for the material-revenue ratio is

$$\mathbf{E}\left[\ln\left(\hat{s}_{it}^m\right) - \ln\left(\frac{\sigma-1}{\sigma}\sum_j \frac{M_{ijt}^{*\theta}}{C_m K_{ijt}^{*\theta} + \frac{C_m}{C_l} L_{ijt}^{*\theta} + M_{ijt}^{*\theta}} \times \frac{R_{ijt}}{R_{it}}\right) | \mathcal{I}_t\right] = 0 \tag{30}$$

This delivers the identification the identification of material-output elasticity under certain normalization for the CES function. Then with moment conditions 10 and 11 we can identify other parameters of interest.

3.5 Connection to existing methods

Existing methods differ in their approaches to control for the 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 thi subsection, we discuss the relationship between our approach and exisiting methods.

[14] 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.

[15] 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 unobserved productivity due to the firm's decision on becoming a multi-product firm. 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 leadto 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; theydo not rely on single-product firms to identify the product-specific technology. In their method, the estimating equations link each product's quantity with respect to total inputs and the output of other products.[16] However, according to our theoretical framework, their estimation strategyfaces a potential problem of mis-specification. Note that the CES demand implies that

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

After taking logs, it indicates that

$$\gamma_{ijt}^{j'} = \exp\left[\frac{\sigma-1}{\sigma}\left(q_{ij't} - q_{ijt}\right)\right] \tag{31}$$

[16]Refer to equation (9) on page 16 in Dhyne, Petrin, Smeets, and Warzynski (2017).

Plugging it into the expression of α_{xit}^{j}, we obtain that quantity of product j is

$$\begin{aligned} q_{ijt} &= \beta_k^j k_{it} + \beta_l^j l_{it} + \beta_m^j m_{it} + \sum_{x \in \{k,l,m\}} \tilde{\beta}_x^j \alpha_{xit}^j + \omega_{ijt} \\ &= \beta_k^j k_{it} + \beta_l^j l_{it} + \beta_m^j m_{it} + \sum_{x \in \{k,l,m\}} \tilde{\beta}_x^j \ln \left\{ \tilde{\beta}_x^j \left[\sum_{j' \in \mathcal{J}} \tilde{\beta}_x^{j'} \exp \left(\frac{\sigma - 1}{\sigma} \left(q_{ij't} - q_{ijt} \right) \right) \right]^{-1} \right\} + \omega_{ijt} \end{aligned} \tag{32}$$

It is obvious that there exists a unique solution for q_{ijt}. But since the complicated way in which q_{ijt} enters into the allocation rules, we cannot find a closed-form solution for it. Suppose that solving for q_{ijt} gives us the production function

$$q_{ijt} = F_j \left(\boldsymbol{q}_{i,-jt}, k_{it}, l_{it}, m_{it}, \omega_{ijt} \right), \tag{33}$$

where $\boldsymbol{q}_{i,-jt}$ is a vector of the quantities of products other than j, which is the multi-product production function is first defined by Diewert (1973). One observation is that despite of the simple Cobb-Douglas production functionwe 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 (33) 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 of the inputs leads to a problem of mis-specification. Under this assumption, it is likely that the the non-linear effects of inputs exist in the error term and cause an endogeneity issue.

4 Monte Carlo Study

To evaluate the validity of the proposed estimation methodology, we provide a Monte Carlo Study. 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 $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \sigma, \rho)$. Our simulation procedure is as follows. For the first period: (a) given prices of capital, labor and materials, generate the logged capital, logged labor from the log normal distribution $k_{ij1} \sim N\left(\bar{k}_1, \sigma_{k1}^2\right)$, $l_{ij1} \sim N\left(\bar{l}_1, \sigma_{l1}^2\right)$. Generate the logged productivity $(\omega_{ij1})_{j=1}^J \sim N(\bar{\omega}_j, \sigma_{\omega_j}^2)$; (b) For each firm, solve its optimization problem. Record their choice of total material M_{i1}, their quantities Q_{ij1} and revenues R_{ij1}. Also, document the logged material-revenue share as s_{i1}^*, introducing 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, generate 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.[17] 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) 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, solve its optimization problem. Document their choice of total material M_{i2}, , their quantities Q_{ij1} and their revenues R_{ij2}. Also, document the input of material as a share of total revenue total revenue s_{i2}^*, create the noisy version $s_{i2} = s_{i2}^* + \tau_i$. See Appendix B for a summary of the parameter values for the Monte Carlo experiment.

[17] 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 given in Table 1. Our simulation result shows that the estimator is very close to the true value, so the selected instrument identifies the parameter of interest and in most cases, the 95% confidence interval covers the true value. As sample size increases, the confidence sets get 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). The following estimation strategy is suggested:

$$\omega_{ijt} + u_{it} = q_{ijt} - \beta_k^j k_{it} - \beta_l^j l_{it} - \beta_m^j m_{it} - \kappa_j q_{i,-jt} \tag{34}$$

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

$$\omega_{ijt} = \rho_0 + \rho_1^j \omega_{ijt-1} + \xi_{ijt}$$

and then write ω_{ijt-1} as a non-parametric function of instruments that helps identify the parameters. A simple substitution gives

$$u_{it} + \xi_{ijt} = q_{ijt} - \beta_k^j k_{it} - \beta_l^j l_{it} - \beta_m^j m_{it} - \kappa_j q_{i,-jt} - \rho_0 - \rho_1^j \omega_{ijt-1}$$

Their method is similar to the non-parametric two stage estimation proposed by Levinsohn and Petrin (2003), which also subject to the non-identification problem by GNR. Instead, we use a combination of first order condition method in GNR and their quantity approximation method to see how it works in simulation. We use the first stage identified value of $\beta_m^1, \beta_m^2, \sigma$ from (10) and (9), and treat them as known in the second stage . In order to solve the problem that DPSW have some instrument that we don't have in simulation, we replace the lagged value of productivity with the true allocation rule

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

and

$$\omega_{ijt-1} = q_{ijt-1} - \beta_k^j (k_{it} + \alpha_{kit}^j) - \beta_l^j (l_{it} + \alpha_{lit}^j) - \beta_m^j (m_{it} + \alpha_{mit}^j)$$

Our estimating equation for the second stage is to have

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

we use the same set of instrument as in our method.

Table 2 shows the result from the adapted version of approaches of Dhyne et al. (2017). The adapted approach does no manage to capture the correct coefficient on the capital and labor. Moreover, the estimator is not stable when we increase the sample size, which illustrates there is fundamental identification issue with their approach.

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
κ_1#	N/A	0.0450	-0.0014
κ_2#	N/A	-0.0268	-0.0424
ρ_1	0.8	0.7053	0.7630
ρ_2	0.9	0.9834	0.9998
σ*	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 the estimated productivity from our model and DPSW Given the estimator $\{\hat{\beta}_x^j, \hat{\sigma}, \hat{\rho}_1, \hat{\rho}_2, \hat{\kappa}\}_{\{x\in\{k,l,m\},j=1,2\}}$, log productivity in DPSW's method is given by

$$\hat{\omega}_{ijt}^{DPSW} = q_{ijt} - \hat{\beta}_k^j k_{it} - \hat{\beta}_l^j l_{it} - \hat{\beta}_m^j m_{it} - \hat{\kappa} q_{i,-jt}$$

and the log productivity in our method is given by

$$\hat{\omega}_{ijt}^{new} = q_{ijt} - \hat{\beta}_k^j (k_{it} + \hat{\alpha}_{kit}^j) - \hat{\beta}_l^j (l_{it} + \hat{\alpha}_{lit}^j) - \hat{\beta}_m^j (m_{it} + \hat{\alpha}_{mit}^j)$$

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

Our method captures the distribution quite well, but DPSW significantly underestimate 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, which is 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

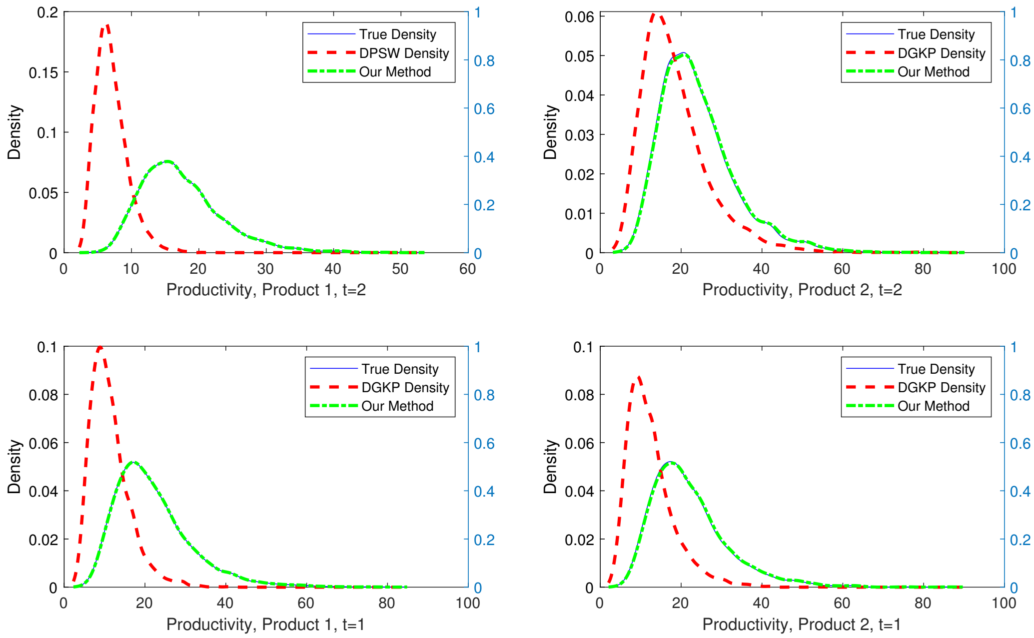


Figure 1: Comparison of productivity distribution

Note: The sample size is 5000.

5 Empirical Study

In the empirical study, we provide an applicationof our proposed methodology of estimating the production function for multi-product firms.

5.1 Data

We have obtained 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 yearly output quantities of their major products. 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 only vary at the firm-year level but not 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.

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.

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.

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 are positive so that the optimization problem of the firm is well defined. The estimation results are reported in Table 5.

Table 5: Empirical results for the food sector

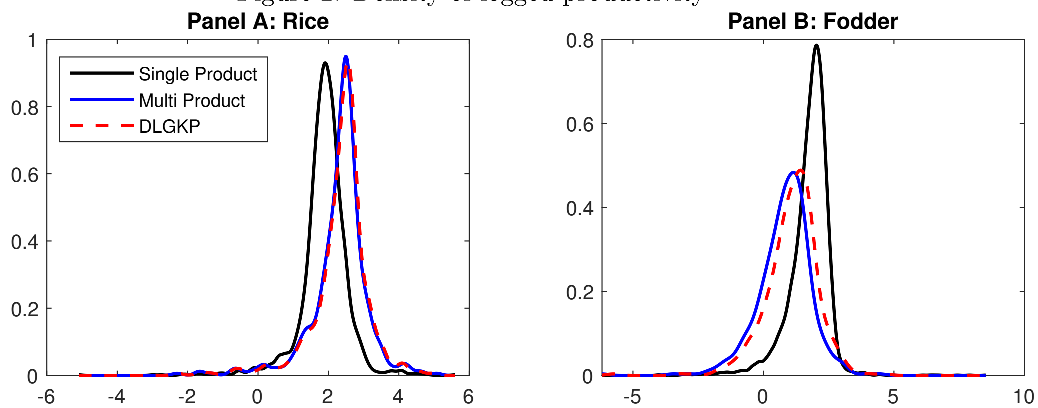
	Fodder					Rice			
	bi-product		single product			bi-product		single product	
	est.	CI	est.	CI		est.	CI	est.	CI
β_m	0.84	[0.84 0.84]	0.77	[0.77 0.77]	β_m	0.77	[0.77 0.77]	0.72	[0.72 0.73]
β_k	0.03	[0.00 0.07]	0.02	[0.00 0.04]	β_k	0.01	[0.00 0.04]	0.03	[0.00 0.07]
β_l	0.18	[0.09 0.28]	0.08	[0.04 0.12]	β_l	0.11	[0.05 0.15]	0.04	[0.00 0.09]
ρ_0	1.27	[1.08 1.45]	0.12	[0.07 0.17]	ρ_0	1.08	[0.94 1.27]	0.35	[0.11 0.60]
ρ_f	0.93	[0.78 0.99]	0.89	[0.85 0.92]	ρ_r	0.28	[0.21 0.32]	0.76	[0.62 0.89]
ρ_{fr}	-0.86	[-0.98 -0.68]	n.a	n.a	ρ_{rf}	-0.16	[-0.45 0.05]		
σ	9.92	[9.86 10.02]							

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

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 a larger coefficient for capital. Moreover, we notice that the coefficient of capital is much smaller 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.

Our model allows for productivity spillovers across products produced by the same firm. Focusing on bi-product firms, the matrix of correlation coefficients for the productivity process shows a positive self-correlation. For the productivity of producing rice, the persistence coefficient is 0.93. The persistence coefficient is only 0.28 for the productivity of producing rice. In the estimation results, we do find some productivity spillovers from the production of rice to that of fodder. The correlation coefficient ρ_{fr} is estimated to be -0.86, significant at 1% significance level. In contrast, ρ_{rf} is found to be insignificant, though it is also negative. As far as we know, this paper is the first to detect the productivity spillovers across different product lines. One possible explanation is the span of control within the bi-product firms. When the production line of rice is highly productive today, it is more likely that managers will exert more effort on producing rice, which further decreases the productivity for fodder tomorrow.

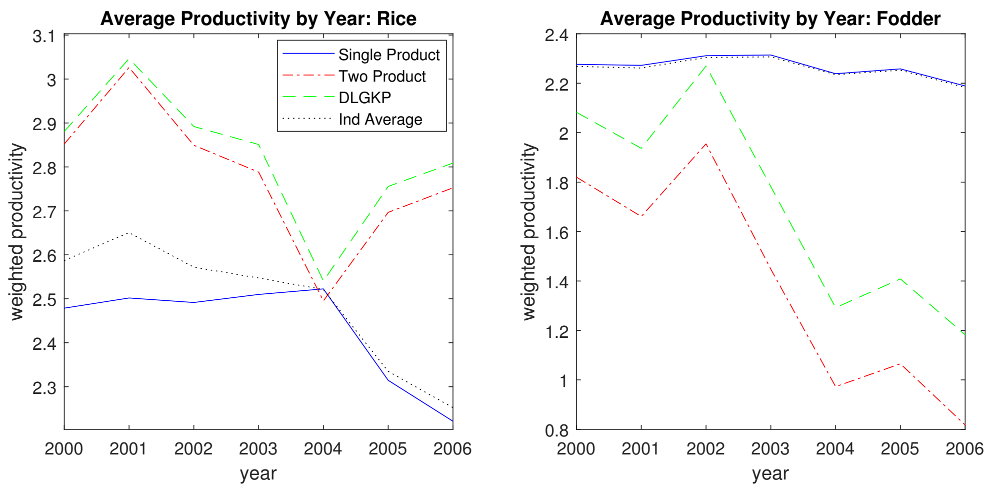
Figure 2: Density of logged productivity



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. It is clear that bi-product firms tend to have 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 larger, 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.[18] We notice that, on average, bi-product firms are more (less) productive than single product firms in the production of rice (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

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

heterogeneity in production technology and technical efficiencies, our method can provide a more reliable estimate of the firm-product productivity for multi-product firms.

6 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 features 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 the equation of material-to-revenue ratio, our methodology can identify the production function in the absence of exogenous variations in input prices. Unlike Dhyne et al. (2017) who use a reduced-form approach to control the unobserved inputs, we rely on the firm's optimization problem from a structural model. We show that, even under the setting of simple Cobb-Douglas production function, their method faces a potential problem of mis-specification.

We have discussed possible generalizations of our method and provide 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. Applying the DPSW method to the simulated data, we find a systematic bias in the estimated production function parameters and the distribution of the productivity.

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 negative productivity spillovers from the productivity of rice to the productivity of fodder.

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.

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.

Diewert, W. (1973): "Functional forms for profit and transformation functions," *Journal of Economic Theory*, 6, 284 – 316.

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.

Lau, L. J. (1976): "A characterization of the normalized restricted profit function," *Journal of Economic Theory*, 12, 131–163.

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.

Appendice

A Math Appendix

A.1 Cobb-Douglas production function

A.1.1 Input allocations for C-D production function

From first-order conditions (14)(15)(16), 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 at the firm level

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

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 (10) 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

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

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 (18), 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 (25) 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.[19]

A.2 Proofs for Lemmas and Propositions

A.2.1 Proofs in Section 2

Proposition 1

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

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

is the solution to the optimization problem (7). 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

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

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 $X \in \{K, L, M\}$.[20] There are $3J$ unknown allocation

[19] See De Loecker (2011) for a similar discussion in the estimation of physical productivity.
[20] This is because as an econometrician we observe the total input of material.

of input, along with 3 unknown prices: $\{\frac{P_tQ_t^{\frac{1}{\sigma}}}{r_{it}+\lambda_{kit}}, \frac{P_tQ_t^{\frac{1}{\sigma}}}{w_{it}+\lambda_{lit}}, \frac{P_tQ_t^{\frac{1}{\sigma}}}{r_{it}}\}$. 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 $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}, \frac{P_tQ_t^{\frac{1}{\sigma}}}{r_{it}+\lambda_{kit}}, \frac{P_tQ_t^{\frac{1}{\sigma}}}{w_{it}+\lambda_{lit}}, \frac{P_tQ_t^{\frac{1}{\sigma}}}{r_{it}}\}$ are treated as unknown to be solved. There fore the unknowns can be written as the function of observables. □

Proposition (2)

Proposition. *Let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ *be in the identified set, then* $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\boldsymbol{\rho}}_0, \boldsymbol{\rho}, \sigma)$ *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. (10), (9) is not influenced by changing from $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ to $(\{\alpha_j F_j\}_{j=1}^J, \tilde{\rho}_0, \boldsymbol{\rho}, \sigma)$ because $f_j = \ln F_j$, and we take the partial derivatives with respect to m_{ijt}.It suffice to check (11) holds. Note that

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

So the moment condition ((11)) still holds. □

Proposition (3)

Proposition. *Let* $(\{F_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ *be in the identified set, then* $(\{\tilde{F}_j\}_{j=1}^J, \boldsymbol{\rho_0}, \boldsymbol{\rho}, \sigma)$ *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 (7) 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 suffice 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}. □

A.2.2 Proofs in Section 3

Lemma (1)

Lemma. *Let* $\{\beta_k^j, \beta_l^j, \beta_m^j, \boldsymbol{\rho}_0, \boldsymbol{\rho}, \sigma\}$ *satisfies moment condition (11). If there is no variation in the* *price of material across firms, i.e.* $v_{it} = v_t$,*then* $\{\beta_k^{*j}, \beta_l^{*j}, \beta_m^{*j}, \boldsymbol{\rho}_0^*, \boldsymbol{\rho}^*, \sigma\}$ *that satisfies*

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

also satisfies moment condition (11) for some value of $\boldsymbol{\rho}_0^*$ *.*

Proof. Using (12) (2) and (18), 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 (18) into (19):

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

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

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

where ρ_0^j is the jth element of $\boldsymbol{\rho}_0$ and $\boldsymbol{\rho}^j$ is the jth row of $\boldsymbol{\rho}$; 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

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

By construction, we have $\frac{\beta_k^{*j}}{1-\tilde{\beta}_m^{*j}} = \frac{\beta_k^j}{1-\tilde{\beta}_m^j}$, and recall $\alpha_{xit}^j(\boldsymbol{\beta}) = \ln\left(\frac{\tilde{\beta}_x^j}{\sum_{j' \in \mathcal{J}} \tilde{\beta}_x^{j'} \gamma_{ijt}^{j'}}\right) = \alpha_{xit}^j(\boldsymbol{\beta^*})$so we derive from (A.16)

$$\begin{aligned} \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^j} &= q_{ijt} - \frac{\beta_k^{*j}}{1-\tilde{\beta}_m^{*j}} k_{it} - \frac{\beta_l^{*j}}{1-\tilde{\beta}_m^{*j}} l_{it} - \frac{1}{1-\tilde{\beta}_m^{*j}} \sum_{x \in \{k,l\}} \beta_x^{*j} \alpha_{xit}^j(\boldsymbol{\beta}^*) \\ &\quad - \frac{\beta_m^j}{1-\tilde{\beta}_m^j} \mu_{ijt} - \frac{\rho_0^j}{1-\tilde{\beta}_m^j} - \frac{\boldsymbol{\rho}^j}{1-\tilde{\beta}_m^j} \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1},\, \boldsymbol{\beta}\right) \\ {\scriptstyle (i)} &= q_{ijt} - \beta_k^{*j} k_{it} - \beta_l^{*j} l_{it} - \beta_m^{*j} m_{it} - \sum_{x \in \{k,l,m\}} \beta_x^{*j} \alpha_{xit}^j(\boldsymbol{\beta}^*) + \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^{*j}} \\ &\quad + \frac{\beta_m^{*j}}{1-\tilde{\beta}_m^{*j}} \mu_{ijt}^* - \frac{\beta_m^j}{1-\tilde{\beta}_m^j} \mu_{ijt} \\ &\quad + \tilde{\beta}_m^{*j} \frac{\rho_0^{*j} + \boldsymbol{\rho}^{*j} \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1}^*,\, \boldsymbol{\beta^*}\right)}{1-\tilde{\beta}_m^{*j}} \\ &\quad - \frac{\rho_0^j + \boldsymbol{\rho}^j \boldsymbol{\omega}\left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1}^*,\, \boldsymbol{\beta}\right)}{1-\tilde{\beta}_m^j} \end{aligned} \tag{A.17}$$

where in the equality (i) we use the first order condition of m_{ijt} under $\boldsymbol{\beta}^*, \rho_0^*, \boldsymbol{\rho}^*$

$$m_{ijt} = \frac{1}{1-\tilde{\beta}_m^{*j}} \left(\tilde{\beta}_k^{*j} k_{ijt} + \tilde{\beta}_l^{*j} l_{ijt} + \rho_0^{*j} + \boldsymbol{\rho}^{*j} \boldsymbol{\omega} \left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu^*}_{it-1},\, \boldsymbol{\beta^*} \right) + \mu_{ijt}^* \right)$$

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

Now we consider the moment condition term under parametrization $(\boldsymbol{\beta}^*, \rho_0^*, \boldsymbol{\rho}^*)$. Define

$$\begin{aligned} \Delta_{ijt} = q_{ijt} - \beta_k^{*j} k_{it} - \beta_l^{*j} l_{it} - \beta_m^{*j} m_{it} - \sum_{x \in \{k,l,m\}} \beta_x^{*j} \alpha_{xit}^j(\boldsymbol{\beta}^*) \\ - \rho_0^{*j} - \boldsymbol{\rho}^{*j} \boldsymbol{\omega} \left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{it-1},\, \boldsymbol{\beta}^* \right) \end{aligned}$$

By using equation(A.17) we have

$$\begin{aligned} \Delta_{ijt} = \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^j} - \frac{\epsilon_{it}^j}{1-\tilde{\beta}_m^{*j}} + \underbrace{\frac{\beta_m^{*j}}{1-\tilde{\beta}_m^{*j}} \mu_{ijt}^* - \frac{\beta_m^j}{1-\tilde{\beta}_m^j} \mu_{ijt}}_{A} \\ + \underbrace{\frac{\rho_0^j + \boldsymbol{\rho}^j \boldsymbol{\omega} \left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{t-1},\, \boldsymbol{\beta} \right)}{1-\tilde{\beta}_m^j} - \frac{\rho_0^{*j} + \boldsymbol{\rho}^{*j} \boldsymbol{\omega} \left(\boldsymbol{m}_{it-1},\, \boldsymbol{k}_{it-1},\, \boldsymbol{l}_{it-1},\, \boldsymbol{\mu}_{t-1},\, \boldsymbol{\beta}^* \right)}{1-\tilde{\beta}_m^{*j}}}_{B} \end{aligned}$$

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 $\boldsymbol{\rho}^j = \boldsymbol{\rho}^{*j}$, so we have

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

So our term $A+B$ satisfies

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

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 equalszero. 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 $(\boldsymbol{\beta}_k, \boldsymbol{\beta}_l)$

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

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

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 2. *$\tilde{\alpha}_{xit}^j$ 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 $A^l = [A^l_{ij}]$, where the $(i,j)-th$ entry

$$A^l_{ij} = \begin{cases} \frac{1}{1-\tilde{\beta}^j_m}\left(1 - \frac{\partial \tilde{\alpha}^j_{lt}}{\partial l_{it}} - \tilde{\beta}^j_l \frac{\partial^2 \tilde{\alpha}^j_{lt}}{\partial l_{it} \partial \tilde{\beta}^j_l}\right) & \text{if } i = j \\ -\frac{\tilde{\beta}^j_l}{1-\tilde{\beta}^j_m} \frac{\partial^2 \tilde{\alpha}^j_{lt}}{\partial l_{it} \partial \tilde{\beta}^j_l} & \text{if } i \neq j \end{cases}$$

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

Proposition 4. *Under assumption 6, the parameter* $\boldsymbol{\beta}_k = \left(\beta^1_k, \cdots, \beta^J_k\right)$ *and* $\boldsymbol{\beta}_l = \left(\beta^1_l, \cdots, \beta^J_l\right)$ *are locally identified.*

Proposition (4) 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 $\tilde{\alpha}^j_{xit} = h_x(\tilde{\beta}^1_x, ..., \tilde{\beta}^J_x, \mathcal{I}_{it})$ are identified from the data, assumption (7) is testable. Given that we have a complicated functional form of $\tilde{\alpha}^j_{kit}$, 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

$$B^k_{js} = \tilde{\beta}^j_k(1 - \frac{\partial \tilde{\alpha}^j_{kt}}{\partial k_{it}}(\tilde{\beta}^1_k, ..., \tilde{\beta}^J_k, \tilde{i}_s)) - \frac{\partial \tilde{\alpha}^j_{lt}}{\partial k_{it}}(\tilde{\beta}^1_k, ..., \tilde{\beta}^J_k, \tilde{I}_s)$$
$$B^l_{js} = \tilde{\beta}^j_l(1 - \frac{\partial \tilde{\alpha}^j_{lt}}{\partial l_{it}}(\tilde{\beta}^1_k, ..., \tilde{\beta}^J_k, \tilde{i}_s)) - \frac{\partial \tilde{\alpha}^j_{kt}}{\partial l_{it}}(\tilde{\beta}^1_k, ..., \tilde{\beta}^J_k, \tilde{I}_s)$$

Also define the matrix C

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

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

Proposition 5. *Under assumption 7, we have 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 $\mathbf{E}\left(\frac{\partial q_{ijt}}{\partial k_{it}}|\mathcal{I}_t\right) = \frac{\tilde{\beta}^j_k - \tilde{\beta}^j_k \frac{\partial \tilde{\alpha}^j_{kt}}{\partial k_{it}} - \tilde{\beta}^j_l \frac{\partial \tilde{\alpha}^j_{lt}}{\partial k_{it}}}{1 - \tilde{\beta}^j_m}$ has a unique solution when we have sufficient variation in the support of $\mathcal{I}_t$.

Proof of (4)

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

$$\mathbf{E}\left(\frac{\partial q_{ijt}}{\partial k_{it}}|\mathcal{I}_t\right) = \frac{\tilde{\beta}^j_k - \tilde{\beta}^j_k \frac{\partial \tilde{\alpha}^j_{kit}}{\partial k_{it}} - \tilde{\beta}^j_l \frac{\partial \tilde{\alpha}^j_{lit}}{\partial k_{it}}}{1 - \tilde{\beta}^j_m}.$$

.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 $\tilde{\beta}^j_l \frac{\partial \tilde{\alpha}^j_{lit}}{\partial k_{it}}$ 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 $\tilde{\beta}^j_k$ is invertible for some value in the support of the information set, which is the local identification condition of $\boldsymbol{\beta}_k = \left(\beta^1_k, \cdots, \beta^J_k\right)$, given the value of σ. Similar is the condition for labor input parameter. $\square$

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

$$L_{ijt} = \left[\frac{\beta_k^j (r_t + \lambda_{kit})}{\beta_l^j (w_t + \lambda_{lit})}\right]^{\frac{1}{\theta-1}} K_{ijt} = \left(C_l \frac{r_t + \lambda_{kit}}{w_t + \lambda_{lit}}\right)^{\frac{1}{\theta-1}} K_{ijt} \tag{A.22}$$
$$M_{ijt} = \left[\frac{\beta_k^j (r_t + \lambda_{kit})}{\beta_m^j v_t}\right]^{\frac{1}{\theta-1}} K_{ijt} = \left(C_m \frac{r_t + \lambda_{kit}}{v_t}\right)^{\frac{1}{\theta-1}} K_{ijt} \tag{A.23}$$

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 same 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
$(\bar{k}_1, \bar{l}_1)$	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	$(3, 3)$
$(\sigma_{\omega_1}^2, \sigma_{\omega_2}^2)$	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	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$(\vartheta_1^k, \vartheta_2^k)$	Weights in the capital accumulation rule	(0.02, 0.03)
$(\vartheta_1^l, \vartheta_2^l)$	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