Identification and Estimation of Production Functions for Multiproduct Firms*

Zhiyuan Chen† Moyu Liao‡

January 21, 2022

Abstract

This paper studies the identification and estimation of a multiproduct firm model with firm-product level heterogeneity in production functions and productivity. We characterize a set of moment conditions using the multiproduct firm's first-order conditions and the productivity's evolution process. We first show that the production functions are non-parametrically non-identified without observing the input allocations. For parametric production functions, we formally prove that the Cobb-Douglas production functions are point identified under mild conditions and the standard GMM method can be applied. For more general parametric production functions, we propose using the partial identification method. We provide Monte Carlo evidence suggesting that the CES production functions are partially identified. Applying our method to a sample of grain manufacturing firms, we find that the production functions for single-product and multiproduct firms are different for the same product.

JEL: C51, L2, L6

Keywords: multiproduct firms; production functions; productivity; input allocations; identification; economies of scope

*We thank comments from Yuan Gao, Paul Grieco, Shengyu Li, Rong Luo, Yao Luo, Scott Orr, Robin Sickles, Valerie Smeets, and Li Su. Special thanks go to Ulig Doraszelski, Mark Roberts, Marc Henry, and Hongsong Zhang who read our paper and provided insightful suggestions. We also thank seminar participants at Peking University(SOE), Renmin University(SOB), the 2021 Econometric Society China Meeting, and the Industrial Organization Brownbag at Penn State University.

†School of Business, Renmin University of China. Email: `chenzhiyuan@rmbs.ruc.edu.cn`.

‡School of Business, Nanjing University. Email: `liaomoyu@nju.edu.cn`.

1 Introduction

Multiproduct firms are dominant in product-level production data.[1] Empirical research on the determinants of productivity (such as R&D spending, trade liberalization, managerial efficiency, etc.) and the choice of product set calls for good firm-product level productivity estimates for multiproduct firms. The popular production function estimation methods proposed by Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), and Gandhi et al. (2020) are developed for single-product firms. Extending these methods to multiproduct firms leads to a problem: Usually, the researcher can only observe the total inputs, but not the inputs allocated to different products.[2] In this paper, we study the identification of multiproduct firms' production functions and propose estimation methods to recover the product-specific production functions.

We adopt a model of multiproduct firms with product-specific production functions and firm-product level productivity. The production technology is Hicks-neutral, and the demand function is parametric. In the model, capital is predetermined, while labor and materials are static choices. Firms optimally allocate the dynamic input (capital) to different product lines and choose static inputs (labor and materials) for different products after observing the realized productivity. We assume that the evolution of the productivity vector follows a Markov process so that productivities can be correlated across products and periods.

We assume that the econometrician can observe total inputs and product-specific output quantities, but not the inputs allocated to different product lines—a common feature of the available data on multiproduct firms. We prove that, under some regularity conditions, the optimal allocation of inputs is a function of the total amount of inputs, output quantities, and production function parameters. This result enables us to use observed information to control for the unobserved input allocations.

We derive and characterize moment conditions induced by the model. The first moment condition is the quantity-revenue relationship, which helps identify the demand elasticity for the parametric demand. We then derive moment conditions from the firm's profit maximization and productivity evolution. These moment conditions include: (1) the firm-level labor-to-revenue shares, (2) the firm-level material-to-revenue shares, and (3) the product-specific productivity evolution process. All of these moment conditions contain input allocations that are unobservable to the econometrician. It turns out that the identification of production function parameters depends on whether the parameters of the production function can be separatedly identified from the unobserved input allocations.

We first show a new non-identification result. We argue that the production functions of multiproduct firms are non-parametrically non-identified. First, the scale of production functions

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

[2] There are rare instances when information on cost shares by product lines within a plant is available. See Lamorgese et al. (2015), Marin and Voigtländer (2018), and Garcia-Marin and Voigtländer (2019) for using the unique dataset from Chile.

cannot be separated from the scale of productivity for each product. Therefore we need to normalize the scale of production functions to obtain meaningful interpretations of the estimated productivity. In this context, we can only analyze the relative productivity of different products. Second, we also show that the output elasticity of any input is non-identified because two sets of horizontally shifted production functions are observationally equivalent to each other. The non-identification problem only occurs in the case of multiproduct firms. This is because we cannot separate the unobserved product-specific input allocations from the production functions.

Having established that the non-parametric production functions are non-identified, we study the identification of parametric production functions. We first study the Cobb-Douglas production functions. We show that the allocation of any input is a closed-form function of data and parameters. We prove that the Cobb-Douglas production functions are point identified under mild conditions that can be easily verified. The conditions ensuring the identification are non-trivial. While it is a common practice to assume identification of parameters under the non-linear GMM setting (Newey and McFadden, 1994), this practice can be misleading in our case: it is not clear what sources of variation can help identify output elasticities from input allocations.[3] To the best of our knowledge, we are the first to show this identification result. Under the guidance of this identification result, the classical joint GMM estimator can be employed to estimate the Cobb-Douglas production functions.

We extend the analysis to Constant-Elasticity-of-Substitution (CES) production functions. In this case, there is no closed-form solution to the input allocations: the production-side moment condition for a product depends not only on its own production function's parameters but also on all other products' production function parameters. Because the unknown optimal allocations and production function parameters change simultaneously, the production functions are potentially non-identified. Given any set of parameters, computing the optimal allocations for the CES production functions is challenging. It involves solving the firm's maximization problem numerically for each observation. We prove that, under some restrictions on the substitution elasticity, the system of equations on the optimal input allocations corresponds to a contraction mapping. For parameters satisfying this condition, we can stack equations for all observations together and use iteration to solve the optimal allocations efficiently. Our proposed methodology can also be applied in empirical studies on multiproduct firms with the CES production functions. In these studies, researchers often need to solve optimal input allocations for numerous firms to calibrate structural parameters and perform counterfactual analyses.

Our method requires solving input allocations based on the product-specific production functions. Given parametric production functions, we can use data and parameters to infer the unobserved optimal input allocations. For Transcendental Logarithmic (Translog) productions functions and more general parametric families, two problems arise: (1) The regularity conditions on production function parameters to ensure the existence of the optimal allocations can be complicated; (2) Computing the optimal input allocations is difficult.

[3] As Orr (2021) pointed out: The GMM based approach used in Valmari (2016), which employes a set of moment conditions different from ours, also faces a similar identification issue.

We conduct Monte Carlo simulations to investigate the identification property of the CES production functions. We generate datasets randomly from an identical set of parameters and apply the GMM point estimator for each dataset. We then compare the estimates with the corresponding true values. It turns out that the GMM estimator acts poorly: the density of the estimates has double spikes with one hump centering away from the true value. In contrast, the distribution of the GMM estimator for the Cobb-Douglas production functions has only one hump centering around the parameter's true value. Considering the potential non-identification problem for the CES production functions, we propose using partial identification methods to obtain the identified set and conduct inference.

We apply our methodology for the Cobb-Douglas production functions to a sample of grain and mixed feed manufacturing firms in China's agricultural sector. We find that the production functions of multiproduct firms differ from that of single-product firms even for the same product. This result is robust when we assume the production functions are CES. Therefore, ignoring the difference in production functions between single-product and multiproduct firms can lead to biased productivity estimates. This is because the production function heterogeneity will be mistakenly attributed to the productivity differences. By assuming that bi-product firms and single-product firms have the same production functions, we show in our empirical application that we over estimate the productivity for the rice producers and underestimate the productivity for the mixed feed producer.

Focusing on the estimated productivity from the Cobb-Douglas production functions, we find that, compared to single-product firms, bi-product firms are more efficient in producing mixed feed but less so in producing rice. We also detect a strong positive contemporaneous correlation between productivities of different product lines. This suggests that firm-level common factors are the main determinant of the productivity of different product lines. This is consistent with the recent finding by Abito (2021) that the time-invariant component accounts for most of the variation in the firm-level productivity. Lastly, we show that both the differences in productivity and shape of the production functions between single-product and multi-product firms contribute to economies of scope for the bi-product firms' production.

Our paper is closely related to the literature on empirical methods for estimating the multiproduct firms' production functions.[4] Based on non-joint production functions, De Loecker et al. (2016), Valmari (2016), Gong and Sickles (2021), and Orr (2021) provide empirical methods to uncover unobserved input allocations and the multiproduct firm's productivity. This paper contributes to this strand of literature in the following ways. First, we consider a general framework with both the product-level production functions and the firm-product-level productivity. Our setting generalizes De Loecker et al. (2016) who only identifies firm-level productivity, and Orr (2021) who abstracts away the heterogeneity in the shape of production functions. Since we allow that the single-product firms and multiproduct firms are different in both production functions and productivity, our model can allow for the (dis)economies of scope originating from

[4]See Section 5.4 in De Loecker and Syverson (2021) for an excellent review of recent developments.

the production side in a more flexible way than the existing literature based on product-level production functions (De Loecker et al., 2016; Orr, 2021). On the other hand, Valmari (2016) only considers a model with the Cobb-Douglas production technologies. In contrast, we provide a comprehensive study on the identification and estimation of non-parametric and parametric production functions for multiproduct firms.

Second, based on a new set of moment conditions containing output elasticities and input allocations, we show a bunch of identification results that are new to existing literature. Both Valmari (2016) and Gong and Sickles (2021) confront the complications that originate from estimating the input allocations and production functions simultaneously. However, in their models, it is not clear whether the output elasticities can be identified from the input allocations. In contrast, we provide proofs for the identification property of production functions both for non-parametric and parametric functional forms. We show that the production functions cannot be identified non-parametrically, while the Cobb-Douglas production functions can be identified under mild conditions. For general parametric production functions like the CES production functions, we point out the issue of partial identification. Under the guidance of the identification results, we develop estimation and computation methods corresponding to different production functions.

Third, unlike most of existing production function estimation methods that rely on exogenous variations in input prices (see De Loecker et al. (2016), Valmari (2016), Orr (2021)), we consider first-order conditions of optimal input allocations to identify firm-level production functions. As Gandhi et al. (2020) point out, the instrument strategy relying on past material prices may face a non-identification problem when there is a lack of variation in material prices. In comparison, our method does not require the researchers to know the input prices and can allow for rich heterogeneity in input prices. In this sense, we make the first attempt of extending the approaches by Grieco et al. (2016) and Gandhi et al. (2020) (GNR hereafter) to the multiproduct setting.

Dhyne et al. (2017) and Dhyne et al. (2021) (DPSW hereafter) develop theories and empirical methods for estimating firm-product productivity for multiproduct firms. Instead of building on non-joint product-level production functions, their framework is based on the transformation function defined for multiproduct firms. Their method does not require the researchers to solve the input allocations and can be applied to more general settings of joint-production where some inputs simultaneously contribute to the production of different product lines. Our paper complements their study by investigating the identification and estimation of well-defined firm-product production functions in the non-joint setting. We also show that we can derive a subset of the transformation functions considered in DPSW from firm-product production functions. However, the simple Cobb-Douglas production functions at the firm-product level can result in a transformation function more complicated than the log-linear case required in Dhyne et al. (2017) and Dhyne et al. (2021). Therefore, their methods are potentially subject to a misspecification issue when outputs are not jointly produced.

The rest of the paper is organized as follows. Section 2 describes a model of multiproduct firms. In Section 3, we first lay out the moment conditions induced by the model. Then we show the non-identification result for non-parametric production functions and the implementation of the Cobb-Douglas and the CES families of production functions. We briefly summarize the estimation methods in Section 4. In Section 5, we present a Monte Carlo study on the CES production functions. In Section 6, we apply our method to a sample of firms in China's grain industry. Section 7 concludes the paper.

2 The Model

In this section, we describe a model of multiproduct firms. In the model, firms organize production by allocating resources to different product lines or different operating divisions.

2.1 Production Functions and Demand

Production Functions We consider a firm i belonging to a group of multiproduct firms producing a set of products $\mathcal{J} = \{1, 2, \cdots, J\}$ over periods $t = 1, 2, \cdots$.[5] Firm i produces product j using a Hicks-neutral production technology:

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

where Q_{ijt} is the total physical output of product j, K_{ijt}, L_{ijt}, and M_{ijt} refer to capital, labor, and materials respectively; ω_{ijt} is the log of unobserved productivity for j-th product, and $\boldsymbol{\beta}_j$ is the product-specific parameters characterizing the production technology. The production technology we consider deserves some further discussion, as it departs from related work in important ways. First, the productivity ω_{ijt} varies at firm-product-period level. This flexibility captures the possibility that multiproduct firms can produce some goods more efficiently than other goods, which cannot be captured by frameworks with firm-level production efficiency (see Bernard et al. (2010), Bernard et al. (2011), De Loecker (2011), De Loecker et al. (2016), among others).

Other than the productivity differences, we allow that the shape of the production technology, $F(\cdot)$, differs across products. When $\boldsymbol{\beta}_j$ is infinite-dimensional, the function $F(\cdot)$ is non-parametric. Our formulation, therefore, generalizes existing frameworks based on non-joint firm-product level production functions by considering firm-product level productivity with the product-level production technology in the production function. To see this, imposing that $\omega_{ijt} = \omega_{it}$, the production function degenerates to be the case of firm-level technical efficiency and product-level production technology considered by De Loecker et al. (2016). If we remove the product-level production technology by requiring that $\boldsymbol{\beta}_j = \boldsymbol{\beta}$, the production function with

[5] Our model allows that different firms produce different sets of products. That is, we can denote the set of products as $\mathcal{J}_i$ and put firms into different groups based on their sets of products. We find focusing on a particular group of firms producing the same set of products for the ease of exposition.

only firm-product productivity differences resembles that considered in Orr (2021). On the other hand, Valmari (2016) only considers the case of Cobb-Douglas production functions.

Dhyne et al. (2021) also allow for firm-product productivity and product-specific production technology. They incorporate unobserved productivity to the transformation function of multiproduct firms defined by Diewert (1973) and Lau (1976). As it will be clear shortly, we can derive a transformation function similar to theirs based on the product-level production functions and the firm's optimal input allocations (See Section 3.5.1). Therefore the family of production functions we consider can be viewed as a subset of production functions defined using the transformation function.

Demand We consider a parameterized demand system with an inverse demand function:

$$P_{ijt} = e^{u_{it}} P_d(Q_{ijt}, \delta_{jt}; \sigma_j),$$

where u_{it} summarizes demand shocks that are not anticipated by firms when they choose inputs, δ_{jt} is the product-level demand shifter shared by firms producing product j, $P_d(Q_{ijt}, \delta_{jt}; \sigma_j)$ is the deterministic component known to firms, and σ_j is a demand parameter that can differ across products. The current formulation nests the single-parameter demand model used by Klette and Griliches (1996) and Levinsohn and Melitz (2002), and the segment-specific demand model in De Loecker (2011) because we do not impose a specific functional form on the demand function.

A Firm i's revenue in period t is:

$$R_{it}^{obs} = \sum_{j \in \mathcal{J}} P_{ijt} Q_{ijt} = \sum_{j \in \mathcal{J}} e^{u_{it}} Q_{ijt} P_d(Q_{ijt}, \delta_{jt}; \sigma_j). \tag{2}$$

Note that the firm's revenue depends on the demand factors u_{it} and δ_{jt}, as well as the production efficiency ω_{ijt} through Q_{ijt}.

2.2 The Firm's Information and Timing of Decisions

When firm i chooses input allocations in period t, it knows its productivity vector $\boldsymbol{\omega}_{it} \equiv (\omega_{ijt})_{j \in \mathcal{J}}$, available total capital $\bar{K}_{it}$, and all variables in the past periods. We make the following assumption on the firm's information set $\mathcal{D}_{it}$ in period t.

Assumption 1. *(Information) The firm's information set*

$$\mathcal{D}_{it} = \{(\boldsymbol{\omega}_{is}, \bar{K}_{is}, K_{ijs-1}, L_{ijs-1}, M_{ijs-1})_{j \in \mathcal{J}, s \leq t}\}$$

contains the productivity vector $\boldsymbol{\omega}_{it} \equiv (\omega_{ijt})_{j \in \mathcal{J}}$ but not the demand shock u_{it}. Moreover:

1. *$\boldsymbol{\omega}_{it}$ follows a first-order Markov process:*

$$\boldsymbol{\omega}_{it} = \boldsymbol{h}(\boldsymbol{\omega}_{it-1}; \boldsymbol{\rho}) + \boldsymbol{\epsilon}_{it}, \tag{3}$$

where $\boldsymbol{\rho}$ is the vector of parameters, and $\boldsymbol{\epsilon}_{it}$ is mean independent of last period's information set: $\mathbf{E}(\boldsymbol{\epsilon}_{it}|\mathcal{D}_{it-1}) = 0$.

2. *u_{it} is mean independent of the variables in $\mathcal{D}_{it}$: $\mathbf{E}(u_{it}|\mathcal{D}_{it}) = 0$.*

Our assumption on the productivity evolution is more general than the exogenous AR(1) process considered in De Loecker et al. (2016), Valmari (2016), and Orr (2021). In their frameworks, the future productivity of a certain product line will only depend on its current productivity and exogenous productivity shocks. We depart from their assumption to allow intertemporal correlation between productivities of different product lines. This is a natural extension when we confront the productivity vector of multiproduct firms. The generalized productivity process allows us to detect potential productivity spillovers across product lines within multiproduct firms.[6]

Assumption 2. *(Timing) Following the definitions of predetermined and static variables in Gandhi et al. (2020), we assume firm-level total capital stock $\bar{K}_{it}$ is predetermined, and the choices of labor $\{L_{ijt}\}_{j\in\mathcal{J}}$ and materials $\{M_{ijt}\}_{j\in\mathcal{J}}$ are static.*[7]

According to Assumptions 1 and 2, at the beginning of period t, the demand shock u_{it} is unanticipated and firm i cannot make input choices contingent on the value of u_{it}. Firm i allocates capital to different products given the total capital stock, while it chooses total labor and materials directly for different product lines.

2.3 The Firm's Optimal Allocation Problem

For the dynamic input, we further assume firms exhaust all capital that can be transferred freely across product lines.

Assumption 3. *Capital $\bar{K}_{it}$ can be costlessly transferred between product lines, and a firm uses all its capital stock in the production.*

We do not make an explicit assumption of zero transferring costs of static inputs as it is immediately implied by Assumption 2. Assumption 3 states there is no adjustment cost of capital between production lines. If there are product-level capital adjustment costs, solving firms' optimal allocation rule becomes infeasible since we neither observe the capital allocation in the last period nor the adjustment cost. For the tractability of our model, we assume there is no product-level capital adjustment cost. However, Assumption 3 does not rule out the adjustment cost at the firm level.

[6]Though we do not explicitly model the network effects of productivity within multiproduct firms, i.e., one product line's productivity is affected by other product lines, our framework allows that productivities of different product lines are correlated.

[7]The definitions of predetermined, dynamic, and static variables are given in Section II.A in Gandhi et al. (2020). We briefly summarize the definitions here: a variable X_t is predetermined if the firm chooses X_t's quantity before the realization of the current period's productivity shocks. A variable X_t is called dynamic if: (1) X_t is also a state variable; (2) the firm chooses X_t's quantity after the productivity shocks realize. Otherwise, X_t is a static variable.

We also assume that firms are price-takers in the inputs market:

Assumption 4. *Firms take material prices and labor prices as given, and material price* v_{it} *and labor price* w_{it} *vary across firms and periods.*

Our setting accommodates the unobservable firm-level input price heterogeneity considered by Grieco et al. (2016). The above assumption on firm-level input prices is more appropriate when a firm uses one material to produce multiple outputs. We denote the mean of the exponential of the demand shock as $\mathcal{U} \equiv \mathbf{E}(e^{u_{it}})$. We use function $R^{exp}(K_{ijt}, L_{ijt}, M_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j)$ and $R^{real}(K_{ijt}, L_{ijt}, M_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j)$ to denote firm i's *expected* and *realized* revenue from selling product j in period t, respectively:

$$
\begin{aligned}
R^{exp}_{ijt} &\equiv R^{exp}(K_{ijt}, L_{ijt}, M_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j) = e^{\delta_{jt}} \mathcal{U} P_d(Q_{ijt}, \delta_{jt}; \sigma_j) Q_{ijt}, \\
R^{real}_{ijt} &\equiv R^{real}(K_{ijt}, L_{ijt}, M_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j) = e^{u_{it}+\delta_{jt}} P_d(Q_{ijt}, \delta_{jt}; \sigma_j) Q_{ijt}.
\end{aligned} \tag{4}
$$

Firm i chooses $\{K_{ijt}, L_{ijt}, M_{ijt}\}_{j\in\mathcal{J}}$ to maximize its expected short-run profits (5):[8]

$$
\begin{aligned}
\max_{\{K_{ijt}, L_{ijt}, M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j\in\mathcal{J}} &\Big\{ R^{exp}(K_{ijt}, L_{ijt}, M_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j) - w_{it} L_{ijt} - v_{it} M_{ijt} \Big\} \\
s.t. \sum_{j\in\mathcal{J}} &K_{ijt} = \bar{K}_{it}.
\end{aligned} \tag{5}
$$

The following assumption states conditions ensuring that the firm's static optimization problem of input allocations is well-behaved and the solution is unique.

Assumption 5. *The firm's expected revenue function* R^{exp} *of product* j *is strictly increasing and strictly concave in its first three arguments* $\{K_{ijt}, L_{ijt}, M_{ijt}\}$*. Moreover, for all values of* $\{K_{ijt}\}_{j\in\mathcal{J}}$*,* $\boldsymbol{\omega}_{it}$*, and* δ_t*, the following Inada condition holds for any* $L^d_{ijt}, M^d_{ijt} \geq 0$*,* $j \in \mathcal{J}$*:*

$$
\lim_{\lambda\to\infty} \sum_{j\in\mathcal{J}} \frac{\partial}{\partial\lambda} R^{exp}(K_{ijt}, L_{ijt} + \lambda L^d_{ijt}, M_{ijt} + \lambda M^d_{ijt}, \boldsymbol{\omega}_{ijt}, \delta_{jt}, \sigma_j) = 0,
$$

where $(L^d_{i1t}, ..., L^d_{iJt}, M^d_{i1t}, ..., M^d_{iJt}) \neq \mathbf{0}$ *is a non-zero direction to increase labor and materials.*

The Inada condition in Assumption 5 requires that, ceteris paribus, when labor or materials increase to infinity, their marginal contributions to the revenue should decrease to zero. Given any parametric demand and production functions, this assumption can be easily verified.

To simplify notation, we use X_{ijt} to represent inputs allocated to product j by firm i in period t, where $X \in \{K, L, M\}$. For any variable Y, we also use lower case letters to represent logged forms, i.e., $y = \log(Y)$. We define the elasticity of output j with respect to input $X \in \{K, L, M\}$ and evaluated at X_{ijt} as γ^X_{ijt}, and the inverse of demand elasticity as η_{ijt}:

$$
\gamma^X_{ijt} \equiv \frac{\partial f(K_{ijt}, L_{ijt}, M_{ijt}; \boldsymbol{\beta}_j)}{\partial x_{ijt}}, \qquad \eta_{ijt} \equiv \frac{\partial p_d(Q_{ijt}, \delta_{jt}; \sigma_j)}{\partial q_{ijt}}. \tag{6}
$$

[8] See our Online Appendix A for the full formulation of the firm's dynamic optimization problem.

Accordingly, the elasticity of the deterministic component of the revenue for input X is given by:

$$\tilde{\gamma}_{ijt}^{X} = \gamma_{ijt}^{X}\left(1 + \frac{1}{\eta_{ijt}}\right).$$

The econometrician does not observe the productivity vector $\boldsymbol{\omega}_{it}$, the material price v_{it}, and the labor price w_{it}. However, we show in the following proposition that given the total inputs $\{\bar{K}_{it},\, \bar{L}_{it},\, \bar{M}_{it}\}$ and the output quantities $\{Q_{ijt}\}_{j\in\mathcal{J}}$, solving the input allocation rules does not require us to know the productivity and input prices.

Proposition 1. *Given Assumption 1-5, there exists a unique solution $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ to optimization problem (5). The optimal allocation of inputs is a function of observables and parameters:*

$$X^*_{ijt} = \mathcal{A}^X_j(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j\in\mathcal{J}}; \{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}), \quad \textit{for} \quad X \in \{K, L, M\}. \tag{7}$$

Moreover, $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ is the unique solution to the following system of equations:

$$\frac{X^*_{ijt}}{\bar{X}_{it}} = \frac{R^{exp}_{ijt}\tilde{\gamma}^{X^*}_{ijt}}{\sum_{j'=1}^{J} R^{exp}_{ij't}\tilde{\gamma}^{X^*}_{ij't}} = \frac{R^{real}_{ijt}\tilde{\gamma}^{X^*}_{ijt}}{\sum_{j'=1}^{J} R^{real}_{ij't}\tilde{\gamma}^{X^*}_{ij't}},\ \textit{for}\ X \in \{K, L, M\}, j \in \mathcal{J}, \tag{8}$$

where $\tilde{\gamma}^{X^*}_{ijt}$ is the elasticity of revenue of product j to input X evaluated at $X_{ijt} = X^*_{ijt}$.

Conversely, if Assumption 5 fails, then there is no solution to optimization problem (5).

Proof. We delegate the full proof to Appendix A.1. Here we only derive Equation (8). We show that the solution to the optimization problem (5) is also the solution to the optimization problem (A.1). Note that the first-order conditions of optimization problem (A.1) deliver that:

$$X_{ijt} = \frac{R^{exp}_{ijt}}{\lambda^X_{it}}\frac{\partial f_j}{\partial x_{ijt}}\left(1 + \frac{1}{\eta_{ijt}}\right) \equiv \frac{\tilde{\gamma}^X_{ijt}R^{exp}_{ijt}}{\lambda^X_{it}} \quad \forall X \in \{K, L, M\},$$

where λ^X_{it} is the Lagrangian multiplier for the constraint $\sum_j X_{ijt} = \bar{X}_{it}$ in the alternative optimization problem (A.1). Moreover, $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ is the unique solution to the first-order conditions, since the optimization problem is differentiable and strictly convex. Using $\sum_{j\in\mathcal{J}} X^*_{ijt} = \bar{X}_{it}$, we have

$$\lambda^X_{it} = \frac{\sum_{j\in\mathcal{J}} \tilde{\gamma}^X_{ijt}R^{exp}_{ijt}}{\bar{X}_{it}}, \quad \forall X \in \{K, L, M\}.$$

Substitute this back to the first-order condition, we obtain the first equality in (8). Note that from (4), we have $R^{exp}_{ijt} = \mathcal{U}e^{-u_{it}}R^{real}_{ijt}$, which implies the second equality in (8). □

The allocation rule (8) states that we can infer the input shares using the revenue shares adjusted by revenue elasticities. First, notice that the optimal allocation rules do not depend on the firm-specific input prices and $\mathcal{U}$. Also, productivity enters the allocation rule through the output quantities $\{Q_{ijt}\}_{j\in\mathcal{J}}$, which enter the revenues $\{P_d(Q_{ijt}, \delta_{jt}; \sigma_j)Q_{ijt}\}_{j\in\mathcal{J}}$ and revenue elasticities $\{\tilde{\gamma}_{ijt}^{X^*}\}_{j\in\mathcal{J}}$. Given parameters in the production functions and the demand equation, Proposition 1 provides us with a method to compute the optimal allocation of inputs from observables and model parameters. When the revenue elasticity is assumed to be common across products, the inputs are allocated in proportion to the revenue shares of each product. This is the shortcut approach employed by Foster et al. (2008), De Loecker (2011), and Blum et al. (2021). In general, however, the allocation rule (8) does not deliver a closed-form expression for the optimal input allocations, and we have to solve them numerically.

Gong and Sickles (2021) and Orr (2021) have obtained a similar expression for input allocations.[9] However, they do not establish the existence and uniqueness of the input allocations. Therefore, it is not clear how the solved input allocations can help estimate the production functions. In contrast, Proposition 1 states the existence and uniqueness to the firm's static optimization problem. We argue that, in order to be able to uncover the input allocations, we have to impose certain restrictions on the shape of the revenue functions (See Assumption 5).

3 Identification

We are interested in the identification of parameters $\left(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}\right)$. The formal definition of the identified set of the parameters of interest requires a full characterization of the firm's dynamic optimization problem. To avoid the technical discussion of the dynamic optimization problem and maintain a clean presentation, we instead derive several implementable moment conditions and focus on the identification of parameters of interests from these moment conditions.[10] Some of our non-identification results still hold even when we consider the model with the firm's dynamic optimization behavior (See Online Appendix A). Throughout the rest of the paper, we maintain all assumptions in Section 2. We start with the data that the econometrician can observe.

Assumption 6. *(Data) For each firm-period pair (i, t), the econometrician observes total capital $\bar{K}_{it}$, total labor $\bar{L}_{it}$, total materials $\bar{M}_{it}$, output quantities for each product $\{Q_{ijt}\}_{j\in\mathcal{J}}$ and total revenue $R_{it}^{obs} = \sum_{j\in\mathcal{J}} R_{ijt}^{real}$. Econometricians also observe the ratio of labor to the total revenue ($S_{it}^L = w_{it}\bar{L}_{it}/R_{it}^{obs}$) and the ratio of materials to the total revenue ($S_{it}^M = v_{it}\bar{M}_{it}/R_{it}^{obs}$). Lastly, $(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j\in\mathcal{J}}, R_{it}^{obs}, S_{it}^L, S_{it}^M)_{t=1}^T$ is independent and identically sampled across i.*

Assumption 6 is aligned with the data feature that the firm-level inputs are observable, but not inputs allocated to each product or each division. We assume that the econometrician observes the product-specific quantities but not the product-specific revenues. The main reason

[9] Orr (2021) considers the firm's cost minimization problem with production constraints for different products. As a result, the input allocations contain each product's marginal costs. Whereas the allocation rules in this paper include each product's revenue elasticity.

[10] It is possible that some of the identification results can be overturned if different moment conditions are considered. However, we emphasize that the commonly used moment conditions may lead to problems.

is that in our empirical setting, the product quantities are recorded in a separate survey. However, if the econometrician can observe the product-specific revenues, our method can be easily adapted to identify the revenue functions and the revenue productivity (See Section 3.5.3).

3.1 The Moment Conditions

Revenue-Quantity Equation We first exploit the relationship between the realized total revenue and expected total revenue to identify the demand equation parameter $\{\sigma_j\}_{j\in\mathcal{J}}$. The revenue-quantity relationship imposes a moment constraint on the inverse demand equation $P_d(Q_{ijt}, \delta_{jt}; \sigma_j)$:

$$\mathbf{E}\left[r_{it}^{obs} - \log\left(\sum_{j\in\mathcal{J}} P_d(Q_{ijt}, \delta_{jt}; \sigma_j) Q_{ijt} \right) \middle| \{Q_{ijt}\}_{j\in\mathcal{J}} \right] = 0, \tag{9}$$

where r_{it}^{obs} is the logged realized revenue.

Labor and Material Share Equations We start by considering the first-order condition for materials from optimization problem (5), which indicates that:

$$\frac{v_{it} M_{ijt}^*}{R_{ijt}^{exp}} = \tilde{\gamma}_{ijt}^{M^*}, \tag{10}$$

where R_{ijt}^{exp} is the firm's expected revenue from selling product j defined in (4). This condition is similar to the material share equation in Gandhi et al. (2020). It also has been widely used to estimate markups with the cost-minimization method (De Loecker and Warzynski, 2012). However, by Assumption 6, we do not observe materials allocated to different products. To employ condition (10), we aggregate the product-level material shares (10) to the firm level:

$$\begin{aligned} S_{it}^M \equiv \frac{v_{it}\bar{M}_{it}}{R_{it}^{obs}} &= \frac{\sum_{j'} R_{ij't}^{exp}}{\sum_{j'} R_{ij't}^{real}} \sum_j \frac{R_{ijt}^{exp}}{\sum_{j'} R_{ij't}^{exp}} \tilde{\gamma}_{ijt}^{M^*} \\ &= \frac{\mathcal{U}}{e^{u_{it}}} \sum_j \underbrace{\frac{R_{ijt}^{exp}}{\sum_{j'} R_{ij't}^{exp}}}_{product\, j's\, revenue\, share} \tilde{\gamma}_{ijt}^{M^*}, \end{aligned}$$

where in the last equality we use the definitions of expected revenue and observed total revenue in (4). Since $\mathbf{E}[u_{it}|\mathcal{D}_{it}] = 0$, where $\mathcal{D}_{it}$ is the information set defined in Assumption 1, the material share equation leads to a moment condition:

$$\mathbf{E}\left[s_{it}^M - \log(\mathcal{U}) - \log\left(\sum_j \frac{P_d(Q_{ijt}, \delta_{jt}; \sigma_j) Q_{ijt}}{\sum_{j'} P_d(Q_{ij't}, \delta_{j't}; \sigma_{j'}) Q_{ij't}} \tilde{\gamma}_{ijt}^{M^*} \right) \middle| \mathcal{D}_{it} \right] = 0. \tag{11}$$

Similarly for labor, we can derive:

$$\mathbf{E}\left[s_{it}^{L}-\log(\mathcal{U})-\log\left(\sum_{j}\left[\frac{P_d(Q_{ijt},\delta_{jt};\sigma_j)Q_{ijt}}{\sum_{j'}P_d(Q_{ij't},\delta_{j't};\sigma_{j'})Q_{ij't}}\tilde{\gamma}_{ijt}^{L^*}\right]\right)\middle|\mathcal{D}_{it}\right]=0. \tag{12}$$

In addition, the constant $\mathcal{U}$ satisfies the following moment condition:

$$\mathcal{U}=\mathbf{E}\left[S_{it}^{X}\left(\sum_{j}\frac{P_d(Q_{ijt},\delta_{jt};\sigma_j)Q_{ijt}}{\sum_{j'}P_d(Q_{ij't},\delta_{j't};\sigma_{j'})Q_{ij't}}\tilde{\gamma}_{ijt}^{X^*}\right)^{-1}\right],\ X\in\{L,M\}. \tag{13}$$

Productivity Evolution Equation Plugging the productivity evolution equation (3) into product j's production function (1) and taking logs, we obtain:

$$q_{ijt}=f\left(k_{ijt},\, l_{ijt},\, m_{ijt};\boldsymbol{\beta}_j\right)+h_j(\boldsymbol{\omega}_{it-1};\boldsymbol{\rho})+\epsilon_{ijt},$$

where $h_j(\boldsymbol{\omega}_{it-1};\boldsymbol{\rho})$ and ϵ_{ijt} are the j-th element of $\boldsymbol{h}(\boldsymbol{\omega_{it-1}};\boldsymbol{\rho})$ and $\boldsymbol{\epsilon}_{it}$, respectively. We define the instrument set for firm i in period t as

$$\mathcal{I}_{it}=\{\bar{K}_{it},\bar{K}_{it-1},\bar{L}_{it-1},\bar{M}_{it-1},\{Q_{ijt-1}\}_{j\in\mathcal{J}},...\}.$$

Note that $\mathcal{I}_{it}$ does not contain input prices because our identification strategy does not rely on their exogenous variations. Assumptions 1 and 2 imply $\mathbf{E}[\boldsymbol{\epsilon}_{it}|\mathcal{I}_{it}]=0$, which leads to the third set of moment conditions for all $j\in\mathcal{J}$:

$$\begin{aligned}\mathbf{E}\Big\{&q_{ijt}-f\left(k_{ijt},\, l_{ijt},\, m_{ijt};\boldsymbol{\beta}_j\right)\\&-h_j\big[\underbrace{(q_{ij't-1}-f(k_{ij't-1},l_{ij't-1},m_{ij't-1};\boldsymbol{\beta}'_j))_{j'\in\mathcal{J}}}_{\boldsymbol{\omega}_{it-1}=(\omega_{ij't-1})_{j'\in\mathcal{J}}};\boldsymbol{\rho}\big]|\mathcal{I}_{it}\Big\}=0\end{aligned} \tag{14}$$

We define the identified set implied by the moment equality models above as the following:

Definition 1. The parameters of interest are $\left(\{\boldsymbol{\beta}_j,\sigma_j,\{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}},\boldsymbol{\rho},\mathcal{U}\right)$. The identified set $\boldsymbol{\Theta}_I$ is the collection of parameters that satisfy moment conditions (9), (11), (12), (13) and (14).

Given the model-implied moment conditions, we first show that the demand parameters $\{\sigma_j\}_{j\in\mathcal{J}}$ and $\{\delta_{jt}\}_{j\in\mathcal{J},t=1,...,T}$ are point identified using moment condition (9). Then we discuss the identification of other parameters.

3.2 Identification of a Dixit-Stiglitz Demand

Throughout the rest of this section, we consider a constant-elasticity demand such that:

$$P_d(Q_{ijt},\delta_{jt};\sigma_j)=e^{\delta_{jt}}Q_{ijt}^{-1/\sigma_j}, \tag{15}$$

where $e^{\delta_{jt}}$ is a product-time demand shifter.[11] This demand system follows the specification in De Loecker (2011). We choose this demand function mainly for the convenience of illustrating the identification property of the model. However, our methodology can be extended to a large class of parameterized demand systems that can be separately identified using market-level information.[12] We first show that the demand elasticity $\{\sigma_j\}_{j\in\mathcal{J}}$ and demand shifters $\{\delta_{jt}\}_{j\in\mathcal{J},t=1,\dots,T}$ are identified from (9) if there are sufficient variations in the quantity data.

Assumption 7. *The random vector $(Q_{ijt})_{j\in\mathcal{J}}$ is supported on $\mathbb{R}^J_+$.*

Assumption 7 is satisfied if the productivity vector $\boldsymbol{\omega}_{it}$ is supported on $\mathbb{R}^J$. To see this, consider $\omega_{ijt} \to -\infty$. In this case, firm i will reduce its inputs allocated to product j to zero, and the quantity Q_{ijt} will also shrink towards zero. Similarly, when $\omega_{ijt} \to +\infty$, the quantity Q_{ijt} will increase to infinity.

Theorem 1. *Under Assumption 7, moment condition (9) identifies $\{\sigma_j\}_{j\in\mathcal{J}}$ and $\{\delta_{jt}\}_{j\in\mathcal{J},t=1,\dots,T}$.*

Proof. See the Appendix A.2. □

Note that the moment equation (9) does not depend on production function parameters. In what follows, we fix the demand to be the Dixit-Stiglitz and focus on identifying the production function parameters.

3.3 Non-identification of Non-parametric Production Functions

Because we do not observe how firms allocate inputs towards each product, the allocation rule is jointly determined by the production function parameters $(\boldsymbol{\beta}_j)_{j\in J}$ and the unobserved productivity $\boldsymbol{\omega}_{it}$. This causes a problem for identifying the non-parametric production functions. We show two non-identification results on the scale and location of the production functions. For notational simplicity, we omit inputs and express the production function for product j as $F(\cdot\,;\,\boldsymbol{\beta}_j)$.

Proposition 2. *(Scale non-identification) Let $(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ be in the identified set Θ_I, then $(\{\tilde{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \tilde{\boldsymbol{\rho}}, \mathcal{U})$ is also in Θ_I, where*

$$F(\cdot\,;\,\tilde{\boldsymbol{\beta}}_j) = e^{a_j} F(\cdot\,;\,\boldsymbol{\beta}_j),$$
$$\boldsymbol{h}(\boldsymbol{\omega};\tilde{\boldsymbol{\rho}}) = \boldsymbol{h}(\boldsymbol{\omega}+\boldsymbol{a};\,\boldsymbol{\rho}) - \boldsymbol{a},\ j\in\mathcal{J}$$

where $\boldsymbol{\omega}\in\mathbb{R}^J$ and $\boldsymbol{a}\equiv(a_1, a_2, \cdots, a_J)\in\mathbb{R}^J$.

Proof. See the Appendix A.2. □

[11] For example, in the case of Dixit-Stiglitz demand, we have $e^{\delta_{jt}} = \bar{P}_{jt}\bar{Q}_{jt}^{1/\sigma_j}$, $\bar{P}_{jt}$ is the aggregate price index for product j, $\bar{Q}_{jt}$ is the aggregate demand index and σ_j is the demand elasticity.

[12] Incorporating an extensive discussion on applying various demand estimation methods to our model is out of the scope of the paper, see Orr (2021) for using demand estimation strategy based on the discrete choice model to infer the input allocations of multiproduct firms.

The argument for non-identification of the scale of the single-product firm's production function has already been illustrated in Gandhi et al. (2020). Our contribution is to extend it to the product-specific production function for multiproduct firms. In the case of multiproduct firms, Proposition 2 implies that the ratio of the productivity between two products can only be identified up to a constant.

More importantly, we proceed to show that the location of the production function is also not identified.

Proposition 3. *(Location non-identification) Let* $(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ *be in the identified set* Θ_I*, then* $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ *is also in* Θ_I*, where*

$$F(K_{ijt}, L_{ijt}, M_{ijt}; \breve{\boldsymbol{\beta}}_j) \equiv F(K_{ijt} - C_j^K, L_{ijt} - C_j^L, M_{ijt} - C_j^M; \boldsymbol{\beta}_j), \quad j \in \mathcal{J},$$

and the constants $\left\{C_j^X \middle| j \in \mathcal{J},\ X \in \{K, L, M\}\right\}$ *satisfy*

$$\sum_{j\in\mathcal{J}} C_j^X = 0, \quad for \quad X \in \{K, L, M\}.$$

Proof. See the Appendix A.2. □

When $J = 2$, the intuition of this proposition is simple: when we horizontally shift the production functions symmetrically, the optimal total input quantities remain the same. Here is a simple example: consider a firm chooses only materials. Suppose the true objective function $10-(2-M_1)^2-(2-M_2)^2$, for which the optimal choices of materials are $M_1 = M_2 = 2$. In this case, we observe the total materials to be $\bar{M} = M_1 + M_2 = 4$. If the objective function becomes $10-(1-M_1)^2-(3-M_2)^2$, the optimal solution is $M_1 = 1$, $M_2 = 3$, but we still observe $\bar{M} = 4$. Only observing $\bar{M}$, we cannot tell the difference between these two production functions. The non-identification of location is new to the existing literature on multiproduct firms' productivity estimation and has strong implications for the identification and estimation of multiproduct firms' production functions. It implies that the econometrician can infer limited information on the shape of the production function from the data for non-parametric production functions. While the results in Proposition 3 are stated for the identified set defined through moment conditions (Definition 1), we prove the same results under the firm's dynamic optimization model. That is, for non-parametric production functions, we can find two sets of parameters that satisfy the dynamic optimization problem. See our Online Appendix A.1 for details.

The following corollary states that the relative output-to-input elasticities between different products and between different inputs are non-identified.

Corollary 1. *The ratio of output-to-input elasticities* $\frac{\partial f_j}{\partial x_j} / \frac{\partial f_{j'}}{\partial x_{j'}}$ *(for* $j \neq j'$ *and* $x \in \{k, l, m\}$*) and* $\frac{\partial f_j}{\partial x_j} / \frac{\partial f_j}{\partial x'_j}$ *(for* $j = 1, ..., J$ *and* $x \neq x'$ *) are not identified.*

Non-parametric production functions have another two problems. First, to evaluate the moment equations, we need to solve the allocation rules using the production functions. Without knowing the functional form, it is hard to solve the allocation rules. Second, the identified set of non-parametric production function parameters can be so large that empiricists cannot give any interesting interpretations. We thus have to consider parametric production functions.

3.4 Parametric Production Functions

We now consider parametric production functions $F(K_{ijt}, L_{ijt}, M_{ijt}; \boldsymbol{\beta}_j)$, where $\boldsymbol{\beta}_j$ is finite-dimensional. In Proposition 1, we have shown that the optimal allocation rule can be computed as a function of observables and parameters. Under the Dixit-Stiglitz demand, the revenue elasticity is $\tilde{\gamma}_{ijt}^X = \gamma_{ijt}^X(1 - \frac{1}{\sigma_j})$. Therefore we can write the optimal allocation rule for input X as:

$$X_{ijt}^* = \frac{e^{\delta_{jt}} Q_{ijt}^{\frac{\sigma_j - 1}{\sigma_j}} \tilde{\gamma}_{ijt}^X(K_{ijt}^*, L_{ijt}^*, M_{ijt}^*; \boldsymbol{\beta}_j)}{\sum_{j'} e^{\delta_{j't}} Q_{ij't}^{\frac{\sigma_{j'} - 1}{\sigma_{j'}}} \tilde{\gamma}_{ij't}^X(K_{ij't}^*, L_{ij't}^*, M_{ij't}^*; \boldsymbol{\beta}_{j'})} \bar{X}_{it}, \quad X \in \{K, L, M\}. \tag{16}$$

Note first that, in general, the output elasticity of input X, i.e., $\tilde{\gamma}_{ijt}^X$, depends on production function parameters $\{\boldsymbol{\beta}_j\}_{j \in \mathcal{J}}$ and optimal inputs $(K_{ijt}^*, L_{ijt}^*, M_{ijt}^*)_{j \in \mathcal{J}}$. Because the optimal inputs themselves also depend on outputs and some unknown parameters, we may not be able to sort out a single parameter that determines the output elasticity. As a result, it's hard to tell what parameters are identified from the material-to-revenue equation (11) and labor-to-revenue (12). Moreover, the optimal allocation of inputs can be hard to compute from (16) for a large number of observations simultaneously.

We then look at the parametric family such that: First, we can compute the optimal allocation rules easily; Second, we can identify some parameters from (11) and (12). Consider the output elasticity satisfies $\gamma_{ijt}^X = b_j(\{Q_{ijt}\}_{j \in \mathcal{J}}) \beta_j^X$, where $b_j(\{Q_{ijt}\}_{j \in \mathcal{J}})$ is a known function of observed product quantities and is free of unknown parameters, and β_j^X is the parameter of interest. In this case, we can compute the optimal allocation rules using the closed-form expression:[13]

$$X_{ijt}^* = \frac{\frac{\sigma_j - 1}{\sigma_j} e^{\delta_{jt}} Q_{ijt}^{\frac{\sigma_j - 1}{\sigma_j}} b_j(\{Q_{ijt}\}_{j \in \mathcal{J}}) \beta_j^X}{\sum_{j'} \frac{\sigma_{j'} - 1}{\sigma_{j'}} e^{\delta_{j't}} Q_{ij't}^{\frac{\sigma_{j'} - 1}{\sigma_{j'}}} b_j(\{Q_{ijt}\}_{j \in \mathcal{J}}) \beta_{j'}^X} \bar{X}_{it}, \quad X \in \{K, L, M\}. \tag{17}$$

Moreover, we can show that moment condition (11) (resp. (12)) identifies $\{\beta_j^M\}_{j \in \mathcal{J}}$ (resp. $\{\beta_j^L\}_{j \in \mathcal{J}}$). The intuition is that conditional on output quantities, variations in input-to-revenue shares can help us identify the corresponding output elasticity. The Cobb-Douglas production function satisfies these two requirements, thus serving as a good benchmark case.

We also consider the CES production function, for which the output elasticities of inputs

[13] This is only a sufficient condition that simplifies the identification analysis. There are probably other sufficient conditions, but we find characterizing the general sufficient condition is challenging and out of the scope of this paper.

depend on all products' inputs. While we cannot show what parameters are identified from (11) and (12), we propose an efficient algorithm to solve the input allocations (see Section 4.2.1). However, even for parametric production functions, we cannot guarantee the existence and uniqueness of optimal allocations because the conditions stated in Proposition 1 can be violated. We use the Translog production function as an example to illustrate this point.

3.4.1 The Cobb-Douglas Production Function

The Cobb-Douglas production function is specified as:

$$Q_{ijt} = e^{\omega_{ijt}} K_{ijt}^{\beta_j^K} L_{ijt}^{\beta_j^L} M_{ijt}^{\beta_j^M}. \tag{18}$$

In the Cobb-Douglas production function, the elasticity of substitution between any two different inputs is one, and β_j^X is the elasticity of product j's output to its input X_j.

Identification of the Labor and Material Elasticities For static inputs (labor and materials), the moment condition for the input-to-revenue ratio becomes

$$\mathbf{E}\left[s_{it}^X | Q_{i1t}, ...Q_{iJt}\right] = \log\left(\mathcal{U}\sum_j \frac{\frac{\sigma_j-1}{\sigma_j}\beta_j^X e^{\delta_{jt}} Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'} e^{\delta_{j't}} Q_{ij't}^{\frac{\sigma_j-1}{\sigma_j}}}\right) \quad \text{for} \quad X \in \{L, M\}. \tag{19}$$

We should note that the expectation of ex-post demand shocks $\mathcal{U} \equiv \mathbf{E}(e^{u_{it}})$ does not enter the revenue-quantity equation (9) and the productivity evolution equation (14). Following Doraszelski and Jaumandreu (2013) and Gandhi et al. (2020), we first identify $\frac{\mathcal{U}\beta_j^M(\sigma_j-1)}{\sigma_j}$ and $\frac{\mathcal{U}\beta_j^L(\sigma_j-1)}{\sigma_j}$ using moment conditions (19), then identify $\mathcal{U}$ using moment condition (13). This leads us to the following proposition:

Proposition 4. *Under Assumption 7, moment conditions (13) and (19) identify* $\{\beta_j^M\}_{j\in\mathcal{J}}$, $\{\beta_j^L\}_{j\in\mathcal{J}}$ *and* $\mathcal{U}$.

Proof. See the Appendix A.2. □

Identification of the Capital Elasticity and Other Parameters Under the Cobb-Douglas specification, the allocation rule in Equation (8) can be characterized in the closed-form:

$$\alpha_{ijt}^X \equiv \frac{X_{ijt}^*}{\bar{X}_{it}} = \frac{\frac{\sigma_j-1}{\sigma_j} e^{\delta_{jt}} \beta_j^X Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'\in\mathcal{J}} \frac{\sigma_{j'}-1}{\sigma_{j'}} \beta_{j'}^X e^{\delta_{j't}} Q_{ij't}^{\frac{\sigma_{j'}-1}{\sigma_{j'}}}}, \quad \text{for} \quad X \in \{K, L, M\}, \tag{20}$$

where in the second equality we use the demand equation (15) to substitute price with quantity. From (20) the quantity of product j can be expressed as

$$Q_{ijt} = e^{\omega_{ijt}}(\alpha^K_{ijt}\bar{K}_{it})^{\beta^K_j}(\alpha^L_{ijt}\bar{L}_{it})^{\beta^L_j}(\alpha^M_{ijt}\bar{M}_{it})^{\beta^M_j}. \tag{21}$$

Equation (21) links the firm-level inputs to the firm-product level output. Plugging equation (21) into equation (14), we can write down the moment condition for productivity evolution as:

$$\begin{aligned}\mathbf{E}(\epsilon_{ijt}|\mathcal{I}_{it}) = \mathbf{E}\Big\{ & q_{ijt} - \sum_{X\in\{K,L,M\}} \beta^X_j \log(\alpha^X_{ijt}\bar{X}_{it}) \\ & - h_j[\underbrace{(q_{ij't-1} - \sum_{X\in\{K,L,M\}} \beta^X_{j'} \log(\alpha^X_{ij't-1}\bar{X}_{it-1}))_{j'\in\mathcal{J}}}_{\boldsymbol{\omega}_{it-1}}]\big|\mathcal{I}_{it}\Big\} = 0.\end{aligned} \tag{22}$$

We now provide a sufficient condition under which $\{\beta^K_j\}_{j\in\mathcal{J}}$ are identified. We first subtract the labor and materials from the output quantity and define

$$\tilde{q}_{ijt} \equiv q_{ijt} - \beta^L_j \log(\alpha^L_{ijt}\bar{L}_{it}) - \beta^M_j \log(\alpha^M_{ijt}\bar{M}_{it}),$$

where α^L_{ijt} and α^M_{ijt} are calculated by plugging the identified β^L_j and β^M_j into (20). To simplify notation, we also define $E^K_{jt}(\mathcal{I}_{it}) \equiv \mathbf{E}(\partial q_{ijt}/\partial\bar{k}_t|\mathcal{I}_{it})$, and $\tilde{E}^K_{jt}(\mathcal{I}_{it}) \equiv \mathbf{E}(\partial\tilde{q}_{ijt}/\partial\bar{k}_t|\mathcal{I}_{it})$. Since $\{\beta^L_j\}_{j\in\mathcal{J}}$ and $\{\beta^M_j\}_{j\in\mathcal{J}}$ are identified, and α^L_{ijt} and α^M_{ijt} only contain identified parameters, $\tilde{E}^K_{jt}(\mathcal{I}_{it})$ is also identified from data.

Proposition 5. *Suppose there exist two values ($\mathcal{I}_{it}$ and $\mathcal{I}'_{it}$) in the support of instrument set such that $\tilde{E}^K_{jt}(\mathcal{I}_{it})/\tilde{E}^K_{j't}(\mathcal{I}_{it}) \neq \tilde{E}^K_{jt}(\mathcal{I}'_{it})/\tilde{E}^K_{j't}(\mathcal{I}'_{it})$, then β^K_j and $\beta^K_{j'}$ are identified from (22).*

Proof. See Appendix A.2. □

The condition that $\tilde{E}^K_{jt}(\mathcal{I}_{it})/\tilde{E}^K_{j't}(\mathcal{I}_{it}) \neq \tilde{E}^K_{jt}(\mathcal{I}'_{it})/\tilde{E}^K_{j't}(\mathcal{I}'_{it})$ requires the ratio of partial derivatives of allocated capital with respect to total capital $\bar{k}_{it}$ is not a constant over the support of instrument set. While this condition is imposed on data rather than on exogenous variables, it is relatively mild and can be verified after β^L_j and β^M_j are estimated.[14]

3.4.2 More General Parametric Production Functions

The Cobb-Douglas production function serves as a benchmark for its simplicity and tractability. However, the Cobb-Douglas specification is quite restrictive in that the substitution elasticity between different inputs is one for any product. Moreover, the Cobb-Douglas production function

[14]For example, we can use a parametric estimation of $\tilde{\Xi}_j(\mathcal{I}_{it}) \equiv \mathbf{E}(\tilde{q}_{ijt}|\mathcal{I}_{it})$. Denote the estimator as $\hat{\tilde{\Xi}}_j$ and we want to test $\frac{\hat{\tilde{\Xi}}_j(\mathcal{I}_{it})}{\partial\bar{k}_{it}}/\frac{\hat{\tilde{\Xi}}_{j'}(\mathcal{I}_{it})}{\partial\bar{k}_{it}}$ is not a constant. We check this condition for both our simulation and empirical settings in Appendix C.

may have strong testable implications that are likely to fail in the data.[15] This calls for more general production functions.

The CES Production Function The CES production function family nests the Cobb-Douglas production function family. The CES production function for product j is given by

$$Q_{ijt} = e^{\omega_{ijt}} \left(\beta_j^K K_{ijt}^{\frac{\theta_j - 1}{\theta_j}} + \beta_j^L L_{ijt}^{\frac{\theta_j - 1}{\theta_j}} + \beta_j^M M_{ijt}^{\frac{\theta_j - 1}{\theta_j}} \right)^{\frac{\nu_j \theta_j}{\theta_j - 1}}, \quad \text{for } j \in \mathcal{J}. \tag{23}$$

The parameters β_j^K, β_j^L, and β_j^M represent the share of capital, labor, and materials in total inputs separately. We restrict that $\beta_j^K + \beta_j^L + \beta_j^M = 1$. The parameter $\nu_j > 0$ measures the degree of return to scale. θ_j is the substitution elasticity between different inputs. When $\theta_j \to 1$, this production function degenerates to be the Cobb-Douglas production function. To ensure that the revenue function is concave in all inputs, we restrict $\nu_j(\sigma_j - 1)/\sigma_j < 1$ for all j. For the CES production function, the output elasticity of product j to optimal input X (evaluated at the optimal allocations) is given by:

$$\gamma_{ijt}^{X*} = \frac{\nu_j \beta_j^X X_{ijt}^{*\frac{\theta_j - 1}{\theta_j}}}{\beta_j^K K_{ijt}^{*\frac{\theta_j - 1}{\theta_j}} + \beta_j^L L_{ijt}^{*\frac{\theta_j - 1}{\theta_j}} + \beta_j^M M_{ijt}^{*\frac{\theta_j - 1}{\theta_j}}}, \quad X \in \{K, L, M\}. \tag{24}$$

We see that the elasticity γ_{ijt}^{X*} depends on all parameters in the production function and the optimal levels of inputs. Therefore a closed-form allocation rule does not exist. As a result, we are unclear about which parameters are identified from equations (11) and (12).

The Translog Production Function The Translog production function is a second-order approximation of the CES production functions, which is specified as:

$$\begin{aligned} q_{ijt} = &\omega_{itj} + \beta_j^K k_{ijt} + \beta_j^L l_{ijt} + \beta_j^M m_{ijt} + \frac{1}{2}\beta_j^{KK} k_{ijt}^2 + \frac{1}{2}\beta_j^{LL} l_{ijt}^2 + \frac{1}{2}\beta_j^{MM} m_{ijt}^2 \\ &+ \beta_j^{KM} k_{ijt} m_{ijt} + \beta_j^{KL} k_{ijt} l_{ijt} + \beta_j^{LM} l_{ijt} m_{ijt}. \end{aligned} \tag{25}$$

The output elasticity of any input X can be written as

$$\gamma_{ijt}^X = \beta_j^X + \beta_j^{XX} x_{ijt} + \sum_{X' \neq X} \beta_j^{XX'} x'_{ijt}, \quad X \in \{K, L, M\} \tag{26}$$

where $\beta_j^{XX'} = \beta_j^{X'X}$. Similar to the CES production function, the output elasticity γ_{ijt}^X depends on all production function parameters and optimal input allocations. However, imposing that the

[15] Grieco et al. (2016) point out that the ratio between different factor inputs should be constant across firms for the Cobb-Douglas production function in the single-product case, which is against the data. We extend and illustrate this point in the multiproduct setting in a companion paper.

Tranlog production function is a concave function of all inputs involves complicated non-linear parameter restrictions. This adds to the difficulty of solving the optimal input allocations.

For general parametric production functions, they can be potentially non-identified using the moment conditions we characterized. This motivates us to use an estimation strategy that is robust to partially identified parameters.

3.5 Discussion

In this subsection, we discuss the implications of our model assumptions, which serve as a foundation for our approach. We first show that our model of firm-product-level production functions is closely related to the multiproduct production functions defined using the transformation function approach. We then show that our model allows for economies of scope manifested in the technological difference between single-product and multiproduct firms. We also discuss several extensions.

3.5.1 Product-Level Production Functions and Transformation Function Approach

Our setting of non-joint firm-product production functions follows De Loecker et al. (2016), Valmari (2016), Gong and Sickles (2021), and Orr (2021). Another strand of literature following the seminal contributions by Diewert (1973) and Lau (1976) defines the multiproduct firm's production function using the transformation function approach. Dhyne et al. (2021) extends the theory of multiproduct production function to allow for the unobserved factor of inputs, i.e., the productivity. To see the connection between our model and theirs, note that we can plug the optimal allocation rule (7) into the production function (1) so that the product j's output quantity can be written as

$$Q_{ijt} = \tilde{F}_{ijt}\left\{\left[\mathcal{A}_j^X(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j\in\mathcal{J}}; \{\boldsymbol{\beta}_j, \sigma_j\}_{j\in\mathcal{J}})\right]_{X\in\{K,L,M\}}; \boldsymbol{\beta}_j\right\}; \tag{27}$$

$$\equiv e^{\omega_{ijt}} F\left\{\left[\mathcal{A}_j^X(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j\in\mathcal{J}}; \{\boldsymbol{\beta}_j, \sigma_j\}_{j\in\mathcal{J}})\right]_{X\in\{K,L,M\}}; \boldsymbol{\beta}_j\right\}. \tag{28}$$

In Equation (27), we treat productivity as a known input factor. It is easy to show that $\tilde{F}_{ijt}$ is invertible with respect to Q_{ijt}. Solving for Q_{ijt}, we obtain a multiproduct production function similar to that specified in Diewert (1973) and Lau (1976). If we treat ω_{ijt} as an unknown input factor, we can solve Q_{ijt} using (28) as a function containing a productivity term, which corresponds to the specification considered by Dhyne et al. (2021). In our model, conditioning on total inputs and productivity, increasing one product's output must decrease other products' outputs. Like Gong and Sickles (2021), our model is more appropriate for a firm's different divisions competing for resources. However, the multiproduct production function derived from the transformation function approach is very general and can capture scenarios such as the production of byproducts.

However, starting from the widely used separable product-specific production functions, the transformation function for a product can be much more complicated than the separable log-linear form considered by Dhyne et al. (2021).[16] Therefore, using the simplified transformation function can potentially lead to a misspecification problem and result in biased productivity estimates. However, this does not mean that our method is strictly better than that proposed by Dhyne et al. (2021). The production function based on the transformation function is more general and can potentially capture production patterns richer than the non-joint firm-product production functions. We emphasize that building the transformation function from micro-founded firm-product functions can guide the choice of functional forms for estimating the transformation function for multiproduct firms when modeling the product-level production is possible.

3.5.2 Economies of Scope for Multiproduct Firms' Production

It is worth mentioning that our framework allows the (dis)economies of scope for multiproduct firms. Baumol et al. (1982) called the economies of scope exists if the multiproduct firms can produce at lower costs than the sum of the costs when produced by single-product firms separately. In our setting, the scope of economies can arise from three channels: (1) The multiproduct firms' productivity premium over single-product firms. Conditional on input prices, if multiproduct firms are more productive than single-product firms, they can produce at lower costs. (2) Different factor prices. Our assumption allows the factor prices to be different across firms. It could be that multiproduct firms have access to cheaper input prices than single-product firms. (3) Differences in the production technology. Given different production functions, economies of scope can arise if multiproduct firms can produce more given the same input bundle and productivity. Note that in both De Loecker et al. (2016) and Orr (2021), the economies of scope can only arise through the channels of productivity and input prices. Therefore, our framework can allow for economies of scope more flexibly.

However, as pointed out by Orr (2021), the non-joint firm-product production functions do restrict the role of economies of scope originating from using public inputs across product lines. He also shows that if the public component of input is a constant fraction of total inputs, and the production technology is Cobb-Douglas, the shifter of economies of scope is observationally equivalent to TFP shifters that depend on the number of products produced by the firm. This implies that one can control the number of products when estimating the production functions to deal with the problem. We can easily incorporate this result into the estimation to account for the possibility of public factors.

[16]For the Cobb-Douglas specifications, Dhyne et al. (2021) define the system of the production functions for multiproduct firms as: $q_{jt} = \beta_j^0 + \beta_j^L l_t + \beta_j^K k_t + \beta_j^M m_t + {\boldsymbol{\gamma}_{-j}^j}' \boldsymbol{q}_{-jt} + \omega_{jt}, \quad j = 1, \cdots, J$. One can verify that even when the product-level production functions are Cobb-Douglas, the corresponding transformation function cannot be written in this separable log-linear form. We provide a simple example of two products and one input factor to further illustrate this point in Online Appendix B.2.

3.5.3 Extensions

We consider several extensions to our model. We briefly summarize them here; Details of these extensions are delegated into Online Appendix D. First, as a variant of Assumption 6, we analyze the case in which researchers only observe the firm-product-level revenues. It turns out that we can identify revenue elasticities and revenue productivity. Second, considering the firing and hiring costs of labor, we explore the possibility that firm-level labor is a dynamic choice. In this case, we rely on the moment conditions for the productivity shocks to identify capital and labor parameters. Third, we also relax the assumption of a single type of materials to multiple types of materials. We assume that the bundle of materials is a CES aggregator of different material varieties. As a result, the variation in relative material prices can help identify the parameters governing the material varieties' shares.

4 Estimation Strategy and Inference

In this section, we first briefly introduce the GMM estimator for the Cobb-Douglas production functions. For more general parametric production functions, we discuss the implementation of partial identification estimation. To fix ideas, we assume that the productivity's evolution process follows a VAR(1): $\mathbf{E}(\boldsymbol{\omega}_{it}|\boldsymbol{\omega}_{it-1}) = \boldsymbol{\rho}_0 + \boldsymbol{\rho}_1\boldsymbol{\omega}_{it-1}$.

4.1 The Cobb-Douglas Production Function

We use a joint GMM estimator to estimate the parameters for the Cobb-Douglas production. Based on the conditional moment conditions (9), (13), (19), and (22), we can obtain associated unconditional moment conditions using appropriate instruments. We choose three groups of instruments: (1) For the revenue-quantity moment condition (9), we can use product quantities and their sum as the instruments; (2) For the input-to-revenue equations (19), we employ quantities for different products and ratios between different products as the instruments; (3) For the moment condition (22) related to the productivity process, we choose variables in the information set $\mathcal{I}_{it}$ and their second-order polynomials. We call this collection of instruments as the instrument set $\mathbb{I}$. We stack all of the unconditional moment conditions and use the standard GMM method to obtain parameter estimates. We follow Wooldridge (2009) to construct the confidence intervals for the parameter estimates and make statistical inferences.

4.2 Partially Identified Parametric Production Functions

For general parametric families, we cannot conveniently compute the input allocations using (20). More importantly, we may lose the point identification if we consider more general parametric production functions. In this subsection, we first discuss the problem of computing the optimal input allocations. Then we propose using a partial identification method to obtain the identified set of production function parameters and make statistical inferences.

4.2.1 Solving the Optimal Input Allocations

For general parametric production functions, we first need to solve the optimal input allocations to construct the moment equalities. To do so, we need to numerically solve the system of non-linear equations (8) for firms with different $(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j\in\mathcal{J}})$. For input vector $(\{k_{ijt}, l_{ijt}, m_{ijt}\}_{j\in\mathcal{J}})$, we first define

$$T_{it}(X,j) = \log\left(\frac{\frac{\sigma_j-1}{\sigma_j}e^{\delta_{jt}}Q_{ijt}^{1-1/\sigma_j}\gamma_{ijt}^X}{\sum_{j'=1}^{J}\frac{\sigma_{j'}-1}{\sigma_{j'}}e^{\delta_{j't}}Q_{ij't}^{1-1/\sigma_j}\gamma_{ij't}^X}\bar{X}_{it}\right), \tag{29}$$

where the (X,j) serves as an index for input (K,L,M) and product j. For example, if $(X,j) = (K,1)$, then $T_{it}(X,j)$ solves the optimal allocation of capital to the first product. Then we define that

$$\mathbf{T}_{it} \equiv [\{T_{it}(K,j)\}_{j\in\mathcal{J}}, \{T_{it}(L,j)\}_{j\in\mathcal{J}}, \{T_{it}(M,j)\}_{j\in\mathcal{J}}]'.$$

Therefore $\mathbf{T}_{it}$ is a mapping from a vector of logged input allocations $(k_{ijt}, l_{ijt}, m_{ijt})_{j\in\mathcal{J}}$ to a new vector of input allocations for each observation:

$$\mathbf{T}_{it} : (-\infty, \bar{k}_{it})^J \times (-\infty, \bar{l}_{it})^J \times (-\infty, \bar{m}_{it})^J \to (-\infty, \bar{k}_{it})^J \times (-\infty, \bar{l}_{it})^J \times (-\infty, \bar{m}_{it})^J.$$

For the CES production function, a sufficient condition ensuring that $\mathbf{T}_{it}$ is a contraction is $|\frac{\theta_j-1}{\theta_j}| < \frac{1}{2}$. An advantage of contraction mapping is that we can stack all $\mathbf{T}_{it}$'s together and iterate to obtain the optimal allocations simultaneously. We propose using an algorithm that combines direct iteration and Newton's method to solve the optimal allocations for CES production functions. We provide the details on the algorithm and proofs in Online Appendix C.

4.2.2 Dimension Reduction and Estimation Methods

While the production functions are identified under the Cobb-Douglas specification, we do not know whether more general parametric families are point identified. Now we discuss the estimation and inference of partially identified production functions. A difficulty of implementing the partial identification method is caused by the high dimension of parameters. Here we propose a method that reduces the computation burden. Heuristically, the idea is the following: given production function and demand parameters $\theta_{-\boldsymbol{\rho}} = (\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \mathcal{U})$, we can recover the productivity $\omega_{ijt}(\theta_{-\boldsymbol{\rho}}) = q_{ijt} - f(k^*_{ijt}, l^*_{ijt}, m^*_{ijt}; \theta_{-\boldsymbol{\rho}})$, where $\{k^*_{ijt}, l^*_{ijt}, m^*_{ijt}\}$ are the optimal inputs computed using data and parameters $\theta_{-\boldsymbol{\rho}}$. Regressing $\omega_{ijt}(\theta_{-\boldsymbol{\rho}})$ on $\omega_{ijt-1}(\theta_{-\boldsymbol{\rho}})$ yields $\boldsymbol{\rho}$.

Formally, we use the same instrument set $\mathbb{I}$ as in the Cobb-Douglas model (See Subsection 4.1) to get unconditional moment conditions $m_i(\theta) \equiv m(\{Q_{ijt}\}_{j\in\mathcal{J}}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}; \theta)$ from (11)-(14). For any $\theta_{-\boldsymbol{\rho}}$, moment condition (14) implies $\mathbf{E}[\omega_{ijt}(\theta_{-\boldsymbol{\rho}}) - h_j(\boldsymbol{\omega}_{it-1}(\theta_{-\boldsymbol{\rho}}); \boldsymbol{\rho}) | \mathcal{I}_{it}] = 0$. We transform this moment condition on $\omega_{ijt}(\theta_{-\boldsymbol{\rho}})$ to

$$\mathbf{E}\left[\left(\omega_{ijt}(\theta_{-\boldsymbol{\rho}}) - \rho_0^j - \boldsymbol{\rho}_{1j}\boldsymbol{\omega}_{it-1}(\theta_{-\boldsymbol{\rho}})\right) \otimes \mathbf{Z}_{it}\right] = 0, \quad \forall j \in \mathcal{J}, \tag{30}$$

where $\mathbf{Z}_{it} \in \mathbb{I}$ are instruments in the instrument set, and ρ_0^j and $\boldsymbol{\rho}_{1j}$ are the j-th row of $\boldsymbol{\rho}_0$ and $\boldsymbol{\rho}_1$, respectively. Since moment condition (30) is linear, we can estimate $\boldsymbol{\rho}$ from (30) in a closed form. We propose an estimation algorithm to reduce the dimension for parameter search:

Algorithm 1 Dimension reduction estimation method.

1. Fix $\theta_{-\boldsymbol{\rho}}$, define $\mathbf{W}_j = [\mathbf{1}, ((\omega_{it-1}(\theta_{-\boldsymbol{\rho}}))_{i=1}^N)']$, where $\mathbf{1}$ is a vector of ones, and $\mathbf{Y}_j = (\omega_{ijt}(\theta_{-\boldsymbol{\rho}}))_{i=1}^n$, and the matrix of instruments $\mathbf{Z}$. The GMM estimator of $(\rho_0^j, \boldsymbol{\rho}_{1j})'$ is:

$$\left(\tilde{\rho}_0^j(\theta_{-\boldsymbol{\rho}}), \tilde{\boldsymbol{\rho}}_{1j}(\theta_{-\boldsymbol{\rho}})\right)' = \left(S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{W}_j\mathbf{Z}}\right)^{-1}\left(S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{Y}_j\mathbf{Z}}\right), \tag{31}$$

   where $S_{\mathbf{W}_j\mathbf{Z}} = \mathbf{Z}'\mathbf{W}_j$, $S_{\mathbf{Y}_j\mathbf{Z}} = \mathbf{Z}'\mathbf{Y}_j$. Denote the vector of estimators for all j as $\tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})$.

2. Compute a hyperplane

$$\mathsf{p}_j(\theta_{-\boldsymbol{\rho}}) = \left(\tilde{\rho}_0^j(\theta_{-\boldsymbol{\rho}}), \tilde{\boldsymbol{\rho}}_{1j}(\theta_{-\boldsymbol{\rho}})\right) + \widehat{\mathrm{Null}}\left(S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{W}_j\mathbf{Z}}\right)$$

   where $\widehat{\mathrm{Null}}(\cdot)$ denotes an estimated null space. Let $\mathsf{p}(\theta_{-\boldsymbol{\rho}}) \equiv \prod_j \mathsf{p}_j(\theta_{-\boldsymbol{\rho}})$ be the estimated identified set of moment condition (30).

3. Let $\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}) = \frac{1}{N}\sum_{i=1}^N m_i(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}})$. Define the GMM criterion function at $\theta_{-\boldsymbol{\rho}}$ as:

$$M_n(\theta_{-\boldsymbol{\rho}}) = \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))' A_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))^{-1} \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})), \tag{32}$$

   where $A_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))$ is a weighting matrix. For example, we can use $A_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) = \frac{1}{N}\sum_{i=1}^N m_i(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}})' m_i(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}})$, the efficient weighting matrix.

4. We follow Chernozhukov et al. (2007) to estimate the identified set as

$$\hat{\Theta}^{ID} = \{(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}) \in \Theta : a_N M_n(\theta_{-\boldsymbol{\rho}}) \leq c b_N, \ \boldsymbol{\rho} \in \mathsf{p}(\theta_{-\boldsymbol{\rho}})\}, \tag{33}$$

   where $c > 0$ is a constant, and $\{a_N\}$ and $\{b_N\}$ are two sequences of positive numbers that diverge to infinity and $b_N/a_N \to 0$.

Given $\theta_{-\boldsymbol{\rho}}$, the $\tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})$ solved from Step 1 is the unique minimizer of $\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}})'\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}})$. In order to maintain the closed-form expression (31), we do not use a continuous updating weighting matrix in the first step. Step 2 in Algorithm 1 is robust to the case when $\left(S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{W}_j\mathbf{Z}}\right)^{-1}$ is close to singular and the identification of $\boldsymbol{\rho}$ fails. If $S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{W}_j\mathbf{Z}}$ is not close to singular, then the estimated null space contains only the zero vector and Step 2 does not influence the result. We propose the following method to estimate the null space: We first find the singular decomposition of $S'_{\mathbf{W}_j\mathbf{Z}} S_{\mathbf{W}_j\mathbf{Z}} = D'\Lambda D$. We then define $\breve{\Lambda}$ such that $\breve{\Lambda}_{ii} = \Lambda_{ii} 1(\Lambda_{ii} > c b_N/\sqrt{a_N})$ for some constant $c > 0$. The estimated null space is the null space of $D'\breve{\Lambda}D$. In Step 3 of Algorithm 1, we can use the closed-form solution (31) from Step 1 regardless of whether $S'_{\mathbf{W}_j\mathbf{Z}}S_{\mathbf{W}_j\mathbf{Z}}$ is close to singular or not. If $S'_{\mathbf{W}_j\mathbf{Z}}S_{\mathbf{W}_j\mathbf{Z}}$ is close to singular, and we use another $\boldsymbol{\rho}'$ in the estimated identified set, the difference of the GMM criterion function between using $\tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})$ and using $\boldsymbol{\rho}'$ will not affect the consistency of the identified set estimator. We provide a detailed discussion for the consistency of the estimator in Appendix A.3.

5 Monte Carlo Study

In this section, we use a Monte Carlo study to answer three questions on the identification of the CES production function: (1) If we ignore that the model is potentially partially identified, how the GMM estimator performs over repeated experiments? (2) What is the shape of an (estimated) identified set under a finite sample? (3) How long does it take to obtain the identified set or the GMM estimator? If the CES production functions are partially identified, the identified set will not cluster around the $(\log N/\sqrt{N})$-neighborhood of the true parameter, and the distribution of GMM estimates from repeated experiments is likely to be multimodal.

5.1 Data Generating Process

We generate $N = 1500$ firms with two periods $t = 2$ and two products $J = 2$. We assume a common substitution elasticity between inputs, i.e., $\theta_1 = \theta_2 = \theta$ and a common demand parameter, i.e., $\sigma_1 = \sigma_2 = \sigma$, $\delta_{jt} = \delta$ for all $j = 1, 2$ and $t = 1, 2$. The parameters of interest are $\{\beta_1^K, \beta_2^K, \beta_1^L, \beta_2^L, \beta_1^M, \beta_2^M, \nu_1, \nu_2, \theta, \sigma, \mathcal{U}, \boldsymbol{\rho}_0, \boldsymbol{\rho}_1\}$. As specified in (23), we impose that $\beta_j^K + \beta_j^L + \beta_j^M = 1$.

Our simulation procedure is the following. For each firm in the first period: We fix prices of labor and materials, and we generate the logged capital from a normal distribution $k_{ij1} \sim N\left(\bar{k}_1, \xi_{k_1}^2\right)$. We generate the logged productivity $\omega_{ij1} \sim N(\bar{\omega}_j, \xi_{\omega_j}^2)$ for all j. We introduce the ex-post demand shock $u_{i1} \sim N(0, 0.03)$, and the constant $\mathcal{U} \equiv \mathbf{E}[e^{u_{i1}}] \approx 1$. We solve the firm's optimization problem (5) and record the total labor $\bar{L}_{i1}$, total materials $\bar{M}_{i1}$, realized output quantities Q_{ij1}, and total revenues R_{i1}^{obs}. For each firm in the second period: We generate data on capital assuming that the accumulation of capital follows $k_{it+1} = \vartheta_1^k \omega_{i1t} + \vartheta_2^k \omega_{i2t} + (1-\delta)\, k_{it}$, where $\vartheta_1^k \omega_{i1t} + \vartheta_2^k \omega_{i2t}$ is the capital investment and δ is the depreciation rate;[17] We generate the new productivity vector $\boldsymbol{\omega}_{it+1} = \boldsymbol{\rho}_0 + \boldsymbol{\rho}_1 \boldsymbol{\omega}_{it} + \boldsymbol{\epsilon}_{it}$, where $\boldsymbol{\epsilon}_{it} \sim N(\mathbf{0}, \boldsymbol{I})$. We choose $\boldsymbol{\rho}_1$ to be a diagonal matrix. We introduce the demand shock $u_{i2} \sim N(0, 0.03)$ and $\mathcal{U} \approx 1$ to solve the firm's optimization problem (5). We document firms' total labor $\bar{L}_{i2}$, total materials $\bar{M}_{i2}$, output quantities Q_{ij2} and firm-level revenues R_{i2}^{obs}. See Appendix B for a summary of the parameter values for the Monte Carlo study.

[17] The simulation strategy reflects that the capital investment is increasing in firm-product-level productivity.

5.2 Simulation Results

Figure 1 shows the kernel density of GMM estimates of the CES model over 200 repeated simulations. The sample size of each simulation is $N = 1500$. The vertical dash line denotes the true value. For all production function parameters, we observe two humps in the distribution of GMM estimators: Around 50% of the estimates center around the true value while the rest are far from the truth. This pattern is unlikely caused by sampling error because we see the kernel density between the two humps is zero for some parameters (e.g., β_1^K, β_1^L, and ν_1).[18]

Two things are revealed from Figure 1: First, the GMM estimators are not consistent estimators of the truth; Second, the CES model is likely to be partially identified from the selected moments. We, therefore, recommend using the partial identification estimation approach. We use the following algorithm to calculate the estimated identified set. We first mesh the parameter space. We then search for local minimums of the partial identification criterion function near these grid points. We keep the local minimums whose function values are smaller than the partial identification threshold ($c \log(N)$, where c is a constant). Last, we generate random points near these local minimums and evaluate their function values. The identified set is the collection of all points whose function values are below the threshold. In the benchmark result, c is equal to the inverse of the logged number of production function parameters. We also set c to be different values to check the robustness of the estimated identified set. See our Online Appendix E.2 for the results.

In Figure 2 we plot the estimated identified set. The green points denote the true values used in the simulation. Consistent with the density plots in Figure 1, the estimated identified set is wide for the production function parameters. The identified set covers both of the hump values in Figure 1.

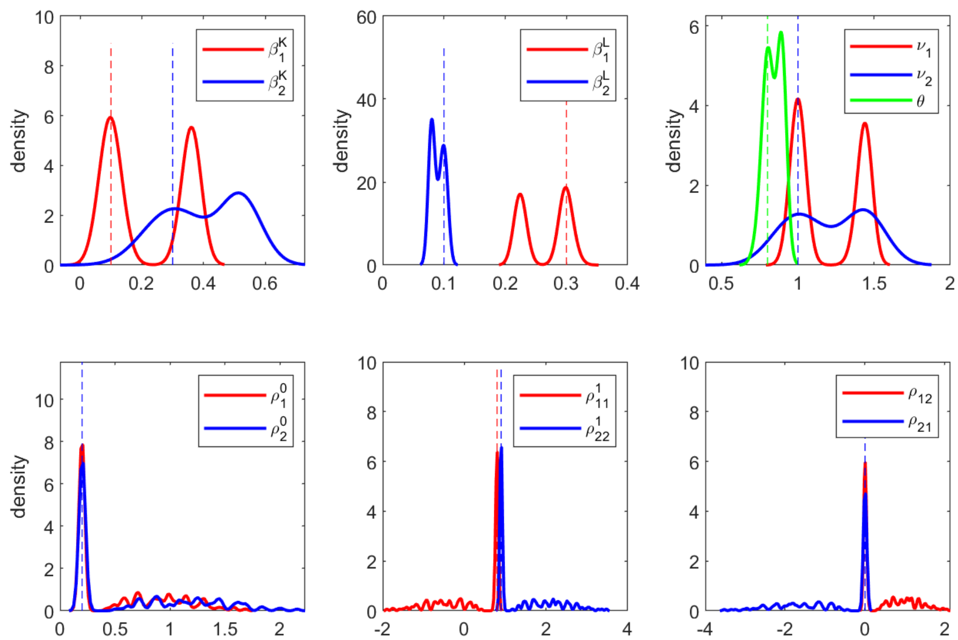
The algorithm to solve the allocation rule is fast. For each set of production function parameters, the running time to solve for the optimal input allocations ranges from 0.1 seconds to 2 seconds, depending on the value of the substitution elasticity θ_j. When $\left|\frac{\theta_j}{\theta_j - 1}\right| < 0.5$, the direct iteration based on (29) works efficiently to solve the optimal input allocations. For a 50-core workstation and a sample size $N = 1500$, it takes around 40 hours to get the estimated identified set.

6 Empirical Study

In this section, we apply our method to a sample of multiproduct firms. As applications of our methodology, we estimate both the Cobb-Douglas and the CES production functions. To emphasize that our method does not require data on single-product firms, we compare our results with those obtained using the single-product proxy method.

[18] In contrast, for the Cobb-Douglas model, we use the same set of moment conditions and find that the GMM estimates are centered around the true value. The distribution of GMM estimators for the Cobb-Douglas model is presented in Appendix B.

Figure 1: Kernel Density of the GMM estimator for the CES Production Function



Note: The GMM estimates are obtained by running 200 experiments over randomly generated samples.

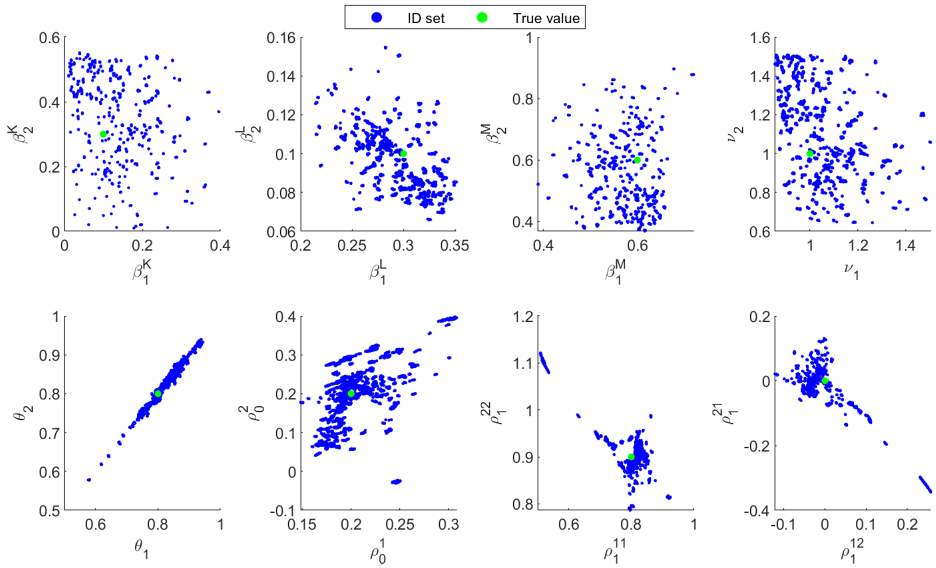
6.1 Data

We have data on the grain manufacturing industry from the Chinese Annual Industry Survey (CAIS) and the production survey between 2000 and 2006. These two datasets are compiled by China's National Bureau of Statistics (NBS). In the CAIS dataset, we observe firm-level production information on sales, the number of employees, capital stock, and materials. In the production survey, firms report their annual output quantities of the major products. Both of these datasets contain only firms with annual sales greater than 5 million RMB. We link these two datasets using the legal code of firms. One limitation of our dataset is that the Chinese production survey employs a relatively aggregate classification of products, which results in a large fraction of single-product firms compared to that reported in the literature. Because rice and mixed feed are two products that are differentiated in the production survey, we select the grain manufacturing industry to avoid losing multiproduct firms due to the survey method.

Rice and mixed feed are two main products for the grain manufacturing industry. Rice is made from rice grains, and is cooked and used for food. Generally, the process of rice manufacturing can include drying, hulling, milling, and enriching.[19] These tasks are mainly performed by machines operated by humans. Mixed feed is a primary compound feed for livestock. It is

[19]The manufacturing of brown rice includes drying and hulling. After processing, brown rice retains the outer bran layers of the rice. In comparison, the manufacturing of white rice needs further milling and enriching. As the milling process that produces white rice also removes much of the vitamins and minerals found primarily in the outer bran layers. The enriching process is to restore the nutrients to the grain.

Figure 2: Identified Set of the CES Production Function



Note: The identified set is obtained by choosing the threshold function value to be log(1500)/ log(8).

made of various feed materials (such as wheat particles, bran, middlings, etc.) through simple processing and mixing. Given the nature of production, we can treat firms producing both rice and mixed feed as multi-divisional firms, which is suitable for our empirical framework.

Table 1: Sample Description

Year	Rice-Only		Mixed Feed-Only		Rice-Feed	
	all	effective	all	effective	all	effective
2000	205	n.a.	809	n.a.	112	n.a.
2001	242	202	904	798	111	104
2002	266	167	900	697	83	61
2003	270	202	902	743	51	49
2004	352	159	962	673	27	17
2005	717	327	1268	889	41	22
2006	684	684	1157	1157	37	37
Total	2736	1741	6902	4597	462	290

Note: Each firm-year pair is an effective observation if the firm also exists in the previous year, i.e., if a firm appears in the panel for years 2000, 2001, and 2002, it is counted as useful in the years 2001 and 2002.

In Table 1, we display the number of observations for different types of firms in our final sample. There are three types of firms: rice-only firms, mixed feed-only firms, and rice-mixed feed firms. The number of observations of mixed feed-only firms is the greatest. For each group of firms, we report the yearly observations in the column called 'all' and the total number 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 also report the number of useful observations for each year in the column named 'effective'. We end up with 1,741 observations for firms only producing rice, 4,597 for firms only producing mixed feed, and 290 for firms producing both products.

6.2 The Cobb-Douglas Production Function

We first apply our method to the Cobb-Douglas specification of the multiproduct firms' production functions. The estimation results are reported in Table 2.

Table 2: Empirical results for the food sector

	Mixed Feed					Rice			
	bi-product		single-product			bi-product		single-product	
	est.	95% 95% CI	est.	95% 95% CI		est.	95% 95% CI	est.	95% 95% CI
β_1^M	.867	[.784, .950]	.891	[.792, 932]	β_2^M	.862	[.807, .918]	.777	[.755, .796]
β_1^K	2.5e-6	[0, .048]	.009	[0, 0.018]	β_2^K	1e-6	[0, .054]	.010	[0, .021]
β_1^L	.049	[.032, .066]	.023	[.021, .024]	β_2^L	.016	[.013, .020]	.017	[.016, .017]
ρ_0	.308	[-.137, .753]	.009	[-.074, .110]	ρ_0	.452	[.026, .878]	.440	[.284, .732]
ρ_{11}	.598	[.309, .887]	.836	[.610, 751]	ρ_{22}	.805	[.459,1.15]	.627	[.509, .732]
ρ_{12}	.089	[-.231, .408]			ρ_{21}	-.320	[-.534, -.106]		
$\mathcal{U}$	1.185	[1.093, 1.276]	1.052	[1.006,1.156]				1.060	[1.033, 1.092]
σ_f	8.439				σ_r	21.787			

Note: Confidence intervals are generated by the GMM bootstrap method. When a parameter estimate is close to the boundary of the parameter space, we use an one-sided confidence interval.

The estimated markup for mixed feed $\sigma_f/(\sigma_f - 1)$ is 13.4%, while the estimated markup for rice $\sigma_r/(\sigma_r - 1)$ is only 4.8%. This is consistent with our expectation given that rice products are more homogenous than mixed feed products. The estimation results show that grain manufacturing is material-intensive, with the material share β_j^M being the largest among the three inputs. For both single-product and multiproduct firms, capital parameters are close to the lower boundary of the parameter space. In this case, the two-sided confidence interval is invalid, so we only report the one-sided confidence interval. The low values of capital coefficients reflect the nature of the production process in the grain manufacturing industry: conditional on labor and materials, capital stock (such as machines and warehouse buildings) does not contribute much to producing outputs.[20]

For the output elasticity of rice concerning materials, the confidence interval of bi-product firms does not cover that of rice-only firms. Since the sample size of rice-only firms is much larger than the rice-feed firms, we can infer that the production functions for rice-only and rice-feed firms are significantly different. For other production function parameters, we also see some differences between bi-product and single-product firms. Compared to single-product firms, multi-product firms are significantly more labor-intensive. The capital coefficients vanish for both mixed feed and rice, though we cannot reject that single- and bi-product firms have the same capital coefficients. We notice that $\mathcal{U} = \mathbf{E}(e^{u_{it}})$ for bi-product firms is larger than that for single-product firms. This may indicate that bi-product firms' demand shocks are more dispersed than that of single-product firms. For single-product firms, mixed feed's productivity

[20]We also tried to reestimate the production function by excluding capital and find very similar results.

process tends to be more persistent than rice's. For bi-product firms, we obtain the transition matrix of the productivity vector. The diagonal terms capture the intertemporal self-correlation of the productivity, while the off-diagonal terms indicate the intertemporal correlation between different products' productivities. Interestingly, we find that ρ_{21} is significantly negative. This implies that the current productivity of rice has a negative impact on the future productivity of mixed feed.

Productivity Distribution We compare the distribution of productivity for single and multi-product firms in Figure 3. Single-product firms are more efficient at producing rice but less efficient at producing mixed feed. Compared to single-product firms, multiproduct firms are more balanced between rice and mixed feed in terms of productivity. This may be the reason why some firms choose to produce both products while others only focus on one output. In Figure 4, we plot the estimated productivity vector (ω_r, ω_f) for multiproduct firms. The two products' productivities are positively correlated and scattered around the 45° line. This indicates the existence of firm-level common factors (e.g., managerial ability, organization efficiency.) that simultaneously influence the productivity of all product lines. Our results are largely consistent with the finding by Abito (2021) that the firm-level fixed effect accounts for most of the variation in the estimated firm-level total factor productivity.

Figure 3: Distribution of Productivity for Single and Multiproduct Firms

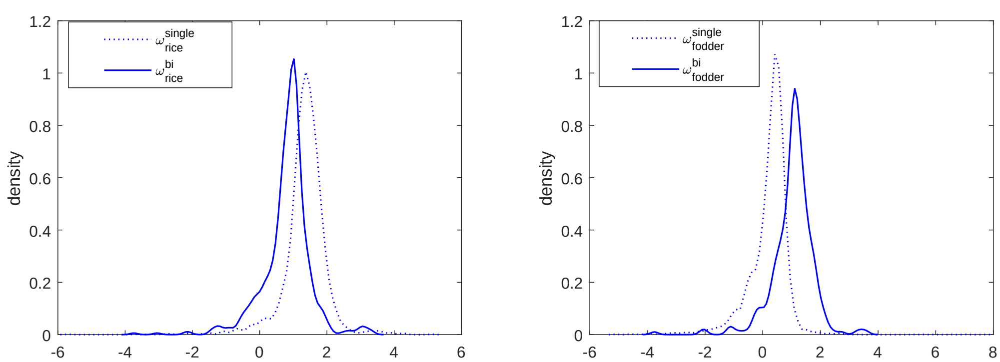
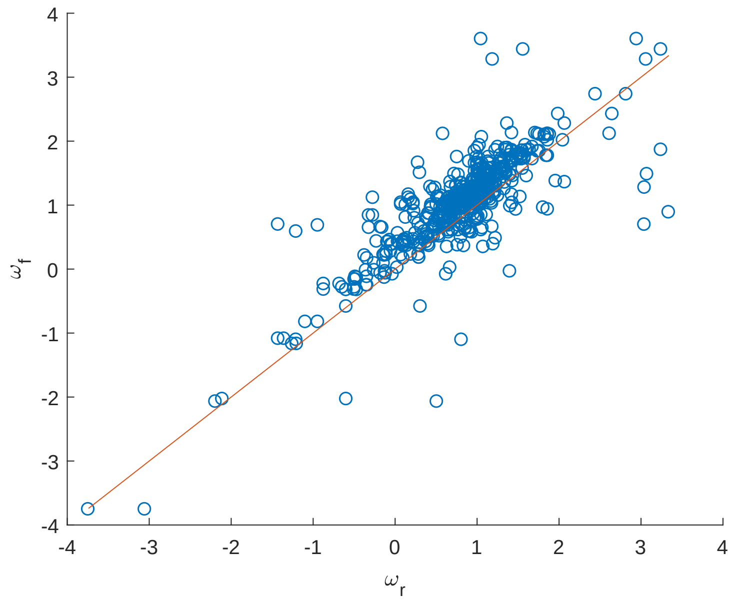


Figure 4: Scatter Plot of (ω_r, ω_f) with a 45° Line



Economies of Scope Conditional on Input Prices We now explore to what extent the multiproduct firms have (dis)economies of scope. As discussed in Section 3.5.2, in our framework, the (dis)economies of scope can arise through the productivity and the production technology differences between single-product and multiproduct firms. To examine the importance of the productivity channel and the production technology channel, we perform counterfactual exercises to merge two representative single-product firms to be a bi-product firm. We construct a representative single-product firm with the average capital, labor, materials, and productivity for rice and mixed feed, separately. Let $(\bar{K}_{sp}, \bar{L}_{sp}, \bar{M}_{sp}), sp \in \{r, f\}$ be the input vector and $(\bar{\omega}_r, \bar{\omega}_f)$ be the productivity vector. We consider a rice-mixed feed firm using the same input bundle $(\bar{K}_r + \bar{K}_f, \bar{L}_r + \bar{L}_f, \bar{M}_r + \bar{M}_f)$. We compare the outputs for different scenarios of merging two single-product firms. Since we do not observe the capital rental price and material prices, we cannot compare the costs to detect the economies of scope as in Baumol et al. (1982). Instead, we consider the dual problem of revenue maximization for the bi-product firms. We call the bi-product firm has the economies of scope if the revenue generated from $(\bar{K}_r + \bar{K}_f, \bar{L}_r + \bar{L}_f, \bar{M}_r + \bar{M}_f)$ is greater than the total revenue of single products. Therefore, our counterfactual exercise is conditional on that input prices are the same for single-product and bi-product firms.

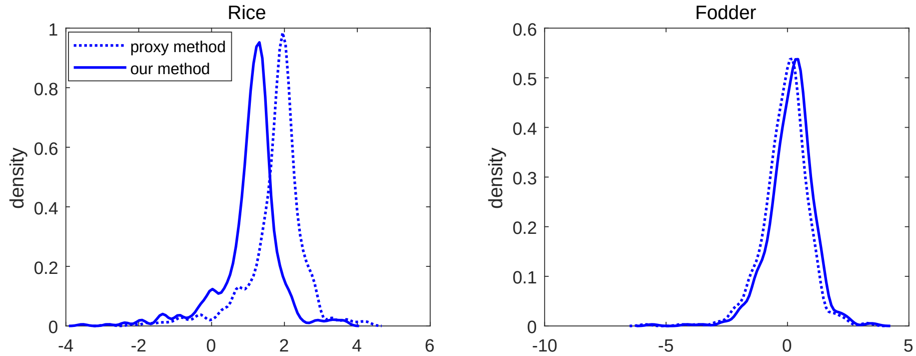
To examine the strength of the productivity channel, we set the bi-product firm's production function to be that of single-product firms while the productivity to be the sample mean of bi-product firms' estimated productivity. Similarly, to gauge the importance of the production technology channel, we set the productivity vector of the bi-product firm to be $(\bar{\omega}_r, \bar{\omega}_f)$ while the production function to be that of bi-product firms. The results are shown in Table 3. We see that both channels contribute to the overall economies of scope. In the productivity channel, more resources are allocated to mixed feed; In the technology channel, more resources are allocated to rice. Overall, compared to two representative single-product firms, the representative bi-product firm can generate around a 75% increase in revenue.

Table 3: Economies of Scope

	Rice-Only	Fodder-Only	Bi-product (rice, mixed feed)
Inputs:			
$\bar{K}$	6.80e3	1.16e4	1.84e4
$\bar{L}$ (persons)	68.89	129.98	198.87
$\bar{M}$	2.29e4	6.48e4	8.78e4
Productivity($\bar{\omega}$)	1.34	0.21	(0.82,1.04)
Outputs:			
Q(tons)	1.10e4	2.94e4	productivity: (1.62e3,8.51e4) technology: (7.62e4,11.54) overall: (4.07e4, 9.47e3)
Revenue (1,000 RMB)	7.19e3	8.70e3	productivity: 2.33e4 technology: 4.84e4 overall: 2.82e4

Comparison to the Proxy Method In comparison to De Loecker et al. (2016), our method does not use the production functions of single-product firms to infer parameters of multiproduct firms. Therefore, we do not restrict that single-product firms and multiproduct firms produce the same product using the same technology. We calculate two productivity estimates for each rice-mixed feed firm observation using the proxy method and our method, respectively. In Figure 5, we display these two productivity estimates for rice-mixed feed firms. The left panel describes the productivity distribution for rice and the right panel for mixed feed. The proxy method leads to higher productivity estimates for rice but almost unbiased productivity estimates for mixed feed. The results show that the proxy method may lead to biased estimates, thus invalidating the comparison of productivity between single-product and multiproduct firms.

Figure 5: Comparison of Our Method and Proxy Method



Note: The 'proxy' method refers using the single-product firm's production function estimates to proxy for the multiproduct firm's production function of the same product. This method is proposed by De Loecker et al. (2016).

6.3 The CES Production Function

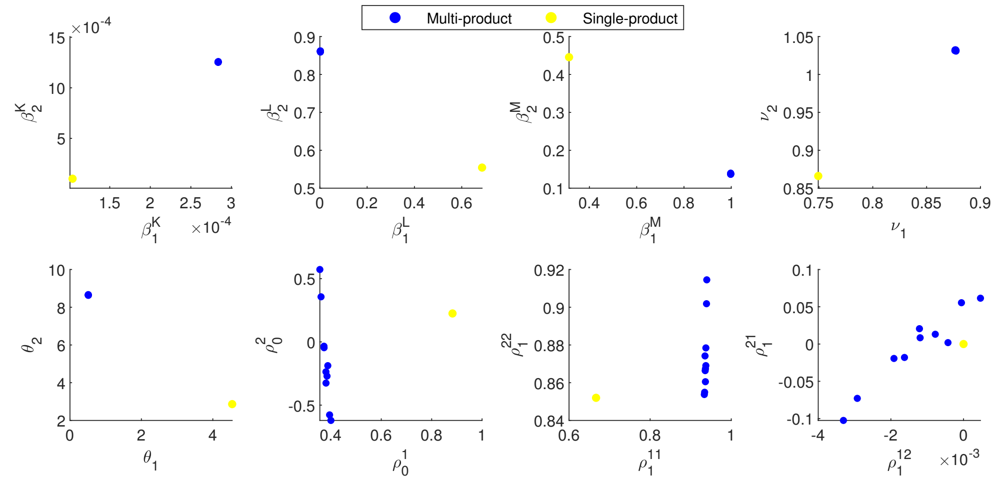
We also apply the partial identification method to estimate the CES production functions of multiproduct firms. The production functions for rice and mixed feed are assumed to follow specification (23). We apply the GMM estimator to obtain point estimates for single-product firms and the partial identification method for bi-product firms.[21] We allow the two products have different substitution elasticities between any two inputs.

We display the estimation results for the identified set in Figure 6. For factor shares' parameters β_j^K, β_j^L, and β_j^M, the parameter values in the identified set of bi-product firms are away from those of single-product firms. This again indicates that the parameter estimates of single-product firms' production function parameters are different from the multiproduct firms'. In Figure 7, we display the confidence region for the identified set of production function parameters. We see that except for the productivity evolution process parameters, the estimates for single-product firms lie outside or on the boundary of the confidence region of multiproduct firms. These results show that imposing that bi-product firms and single-product firms have the same production technology may bias the productivity estimates.

Due to the small sample size, the partial identification estimation results are sensitive to the choice of threshold constant c in (33). In this section, consistent with our Monte Carlo study, we report the results when c is the inverse of the logged number of the dimension of production function parameters. The results under other choices of c are reported in Online Appendix E.

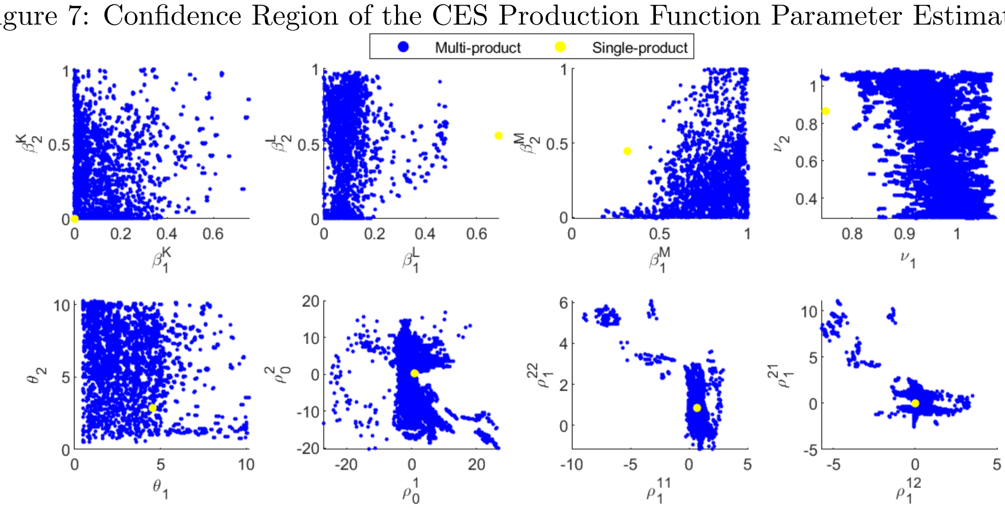
[21] Gandhi et al. (2020) show the identification of non-parametric production functions using moment conditions similar to ours. According to their results, the CES production function for single-product firms is identified.

Figure 6: Identified Set of the CES Production Function Parameters



Note: '1' and '2' refer to rice and mixed feed, respectively.

Figure 7: Confidence Region of the CES Production Function Parameter Estimates



Note: '1' and '2' refer to rice and mixed feed, respectively.

7 Conclusion

This paper studies the identification of production functions for multiproduct firms. Starting with a model of multiproduct firms, we first characterize the moment conditions implied by the model. Then, we show that the product-specific production functions are non-parametrically non-identified without observing the allocation of inputs. We also discuss the difficulty of obtaining the optimal input allocations under non-parametric production functions.

Given the difficulty of non-parametrically identifying production functions and solving the optimal input allocations, we turn to discuss commonly used parametric families. We find that with the Cobb-Douglas production functions, the unobserved input allocations can be controlled by using product-specific output quantities. Moreover, the Cobb-Douglas production functions are point identified under our proposed moment conditions. For the CES production functions, we propose an iterative algorithm to solve the optimal input allocations. We use the Monte Carlo study to show the CES production functions are potentially partially identified and the GMM estimator is not consistent.

We apply our method to a sample of Chinese firms manufacturing rice and/or mixed feed between 2000 and 2006. Our empirical results show that the production functions of multiproduct firms differ from that of single-product firms. Moreover, we show that the single-product proxy method (De Loecker et al., 2016) results in a biased distribution of productivity for multiproduct firms.

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

Appendices

A Proofs and Supplementary Discussions

A.1 Proof of Proposition 1

We break the proofs of Proposition 1 into several lemmas. In Lemma 1, we prove that the optimization problem (5) has a solution. In Lemma 2, we prove that the solution to the optimization problem (5) is also the solution to the following alternative optimization problem:

$$\begin{aligned} \max_{\{K_{ijt},\, L_{ijt},\, M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j\in\mathcal{J}} & \left\{\mathcal{U} e^{\omega_{ijt}} P_d\left[e^{\omega_{ijt}} F_j(K_{ijt},\, L_{ijt},\, M_{ijt}), \delta_{jt};\, \sigma_j\right] F_j(K_{ijt},\, L_{ijt},\, M_{ijt})\right\} \\ s.t. \sum_{j\in\mathcal{J}} & X_{ijt} = \bar{X}_{it} \quad \text{for} \quad X \in \{K, L, M\}. \end{aligned} \tag{A.1}$$

Lastly, we show that the system of equations (8) can be derived from the alternative optimization problem (A.1).

Lemma 1. *There exists a solution $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ to the optimization problem (5).*

Proof. Without loss of generality, we impose that $\mathcal{U} = 1$. Define the objective function value of a point $(K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}}$ by $V((K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}})$ and define the supremum of the value function as $\sup V$. Suppose there does not exists an optimizer of (5), by Theorem 27.1 of Rockafellar (2015), our objective function has a recession direction, i.e. for any $C < \sup V$, the level set

$$Level^C = \{(K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}} : V((K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}}) > C\}$$

has a recession direction. That means we can find a point $(K^0_{ijt}, L^0_{ijt}, M^0_{ijt})_{j\in\mathcal{J}} \in Level^C$, and a non-zero direction $(K^d_{ijt}, L^d_{ijt}, M^d_{ijt})_{j\in\mathcal{J}}$ such that $(K^0_{ijt} + \lambda K^d_{ijt}, L^0_{ijt} + \lambda L^d_{ijt}, M^0_{ijt} + \lambda M^d_{ijt})_{j\in\mathcal{J}} \in Level^C$ for all $\lambda > 0$. [22]

Since the constraints for capital must be binding, the recession direction must satisfy $K^d_{ijt} = 0$, $L^d_{ijt}, M^d_{ijt} \geq 0$ for all $j \in \mathcal{J}$, so the above requirement reduces to $(K^0_{ijt}, L^0_{ijt} + \lambda L^d_{ijt}, M^0_{ijt} + \lambda M^d_{ijt})_{j\in\mathcal{J}} \in Level^C$ for all $\lambda > 0$. Then we note that

$$\begin{aligned} & V((K^0_{ijt}, L^0_{ijt} + \lambda L^d_{ijt}, M^0_{ijt} + \lambda M^d_{ijt})_{j\in\mathcal{J}}) \\ =& P_d\left[e^{\omega_{ijt}} F_j(K_{ijt},\, L_{ijt} + \lambda L^d_{ijt},\, M_{ijt} + \lambda M^d_{ijt}), \delta_{jt};\, \sigma_j\right] F_j(K_{ijt},\, L_{ijt} + \lambda L^d_{ijt},\, M_{ijt} + \lambda M^d_{ijt}) \\ & - w_{it} \sum_{j\in\mathcal{J}} (L^0_{ijt} + \lambda L^d_{ijt}) - v_{it} \sum_{j\in\mathcal{J}} (M^0_{ijt} + \lambda M^d_{ijt}) \\ \equiv& \bar{V}(\lambda). \end{aligned}$$

[22] For example, $h(x) = 1 - e^x$ is a concave function and $\sup_{x\in R} h(x) = 0$. The supremum is not achieved by any real x. For any $C < 0$, the level set $Level^C = \{x : h(x) > C\} = (-\infty, \log(1 - C)]$ has a recession direction at $y = -1$, since for any $x_0 \in Level^C$, and any $\lambda > 0$, $x_0 + \lambda y \in Level^C$.

If we can show $\bar{V}(\lambda) < C$ when $\lambda \to \infty$, then we get the contradiction that $(0, L^d_{ijt}, M^d_{ijt})_{j\in\mathcal{J}}$ is a recession direction. By the Inada condition, we can find a $\lambda_0 > 0$ such that

$$\begin{aligned}&\sum_{j\in\mathcal{J}} e^{\omega_{ijt}} \frac{\partial}{\partial\lambda} P_d\left[e^{\omega_{ijt}} F_j(K_{ijt},\, L_{ijt}+\lambda L^d_{ijt},\, M_{ijt}+\lambda M^d_{ijt}, \delta_{jt};\, \sigma_j\right] F_j(K_{ijt},\, L_{ijt}+\lambda L^d_{ijt},\, M_{ijt}+\lambda M^d_{ijt})\\ &< \frac{v_{it}}{2}\sum_{j\in\mathcal{J}} M^d_{ijt} + \frac{w_{it}}{2}\sum_{j\in\mathcal{J}} L^d_{ijt}\end{aligned}$$

whenever $\lambda > \lambda_0$. We now evaluate $\bar{V}(\lambda)$ using the fundamental theorem of calculus:

$$\begin{aligned}\bar{V}(\lambda) &= \bar{V}(0) + \int_0^\lambda \frac{d\bar{V}(t)}{dt} dt\\ &= \int_0^\lambda \left[\sum_{j\in\mathcal{J}} e^{\omega_{ijt}} \frac{\partial}{\partial\lambda} P_d(e^{\omega_{ijt}} F_j(X_{ijt}+sX^d_{ijt}), \delta_{jt};\, \sigma_j) F_j(X_{ijt}+sX^d_{ijt}) - v_{it}\sum_{j\in\mathcal{J}} M^d_{ijt} - w_{it}\sum_{j\in\mathcal{J}} L^d_{ijt}\right] ds\\ &= \bar{V}(0) + \int_0^{\lambda_0} \left[\sum_{j\in\mathcal{J}} e^{\omega_{ijt}} \frac{\partial}{\partial\lambda} P_d(e^{\omega_{ijt}} F_j(X_{ijt}+sX^d_{ijt}), \delta_{jt};\, \sigma_j) F_j(X_{ijt}+sX^d_{ijt}) - v_{it}\sum_{j\in\mathcal{J}} M^d_{ijt} - w_{it}\sum_{j\in\mathcal{J}} L^d_{ijt}\right] ds\\ &+ \int_{\lambda_0}^\lambda \left[\sum_{j\in\mathcal{J}} e^{\omega_{ijt}} \frac{\partial}{\partial\lambda} P_d(e^{\omega_{ijt}} F_j(X_{ijt}+sX^d_{ijt}), \delta_{jt};\, \sigma_j) F_j(X_{ijt}+sX^d_{ijt}) - v_{it}\sum_{j\in\mathcal{J}} M^d_{ijt} - w_{it}\sum_{j\in\mathcal{J}} L^d_{ijt}\right] ds\\ &\le \bar{V}(0) + \underbrace{\int_0^{\lambda_0} \left[\sum_{j\in\mathcal{J}} e^{\omega_{ijt}} \frac{\partial}{\partial\lambda} P_d(e^{\omega_{ijt}} F_j(X_{ijt}+sX^d_{ijt}), \delta_{jt};\, \sigma_j) F_j(X_{ijt}+sX^d_{ijt}) - v_{it}\sum_{j\in\mathcal{J}} M^d_{ijt} - w_{it}\sum_{j\in\mathcal{J}} L^d_{ijt}\right] ds}_{Term \quad A}\\ &- \frac{\lambda-\lambda_0}{2}\left(v_{it}\sum_{j\in\mathcal{J}} M^d_{ijt} + w_{it}\sum_{j\in\mathcal{J}} L^d_{ijt}\right),\end{aligned}$$

where we use $F_j(X_{ijt}+sX^d_{ijt})$ to denote $F_j(K_{ijt},\, L_{ijt}+sL^d_{ijt},\, M_{ijt}+sM^d_{ijt})$. Since Term A is finite and $\sum_{j\in\mathcal{J}} M^d_{ijt} + \sum_{j\in\mathcal{J}} L^d_{ijt} > 0$, as we push $\lambda \to \infty$, $\bar{V}(\lambda) < C$ must hold. This implies that $V((K^0_{ijt}, L^0_{ijt}+\lambda L^d_{ijt}, M^0_{ijt}+\lambda M^d_{ijt})_{j\in\mathcal{J}}) \le \bar{V}(\lambda) < C$ must hold when λ is sufficiently large. This contradicts that $(K^0_{ijt}, L^0_{ijt}, M^0_{ijt}+\lambda M^d_{ijt})_{j\in\mathcal{J}}$ is a recession direction of $V((K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}})$. Therefore, there cannot exist any recession direction of $V((K_{ijt}, L_{ijt}, M_{ijt})_{j\in\mathcal{J}})$. By Theorem 27.1 of Rockafellar (2015), there exists a solution $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ to optimization problem (5). □

Solution to optimization (5) is a function of unobserved labor and materials prices. In contrast, solution to optimization problem (A.1) is a function of observables.

Lemma 2. *The optimization problem (5) has a unique solution $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$, and this solution is also the solution to the alternative optimization (A.1).*

Proof. By Lemma 1, a solution to (5) exists. By Assumption 5, the optimization problem is strictly concave, so the solution is unique. We consider the alternative optimization problem (A.1).

We claim that $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ is also the unique optimizer to the alternative optimization problem. Suppose not, we can find $\{K^\#_{ijt}, L^\#_{ijt}, M^\#_{ijt}\}_{j\in\mathcal{J}}$ that generates a weakly higher functional value in the alternative optimization problem (A.1) than $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ can generate, i.e.,

$$\sum_{j\in\mathcal{J}}\{P_d(e^{\omega_{ijt}}F_j^*,\delta_{jt};\sigma_j)e^{\omega_{ijt}}F_j(K_{ijt}^*,L_{ijt}^*,M_{ijt}^*)\}\leq\sum_{j\in\mathcal{J}}\left\{P_d(e^{\omega_{ijt}}F_j^{\#},\delta_{jt};\sigma_j)e^{\omega_{ijt}}F_j(K_{ijt}^{\#},L_{ijt}^{\#},M_{ijt}^{\#})\right\},$$

where we use F_j^* and $F_j^{\#}$ to denote the functional value with the corresponding values of capital, labor and materials. Note that $\{K_{ijt}^{\#},L_{ijt}^{\#},M_{ijt}^{\#}\}_{j\in\mathcal{J}}$ satisfies resource constraint (A.1), $\sum_{j\in\mathcal{J}}X_{ijt}^{\#}=\sum_{j\in\mathcal{J}}X_{ijt}^*=\bar{X}_{it}$ for $X\in\{L,M\}$. As a result, we can show

$$\begin{aligned}&\sum_{j\in\mathcal{J}}\left\{e^{\omega_{ijt}}P_d(e^{\omega_{ijt}}F_j^{\#},\delta_{jt};\sigma_j)F_j(K_{ijt}^{\#},L_{ijt}^{\#},M_{ijt}^{\#})\right\}-v_{it}\sum_{j\in\mathcal{J}}M_{ijt}^{\#}-w_{it}\sum_{j\in\mathcal{J}}L_{ijt}^{\#}\\&\geq\sum_{j\in\mathcal{J}}\{e^{\omega_{ijt}}P_d(e^{\omega_{ijt}}F_j^*,\delta_{jt};\sigma_j)F_j(K_{ijt}^*,L_{ijt}^*,M_{ijt}^*)\}-v_{it}\sum_{j\in\mathcal{J}}M_{ijt}^*-w_{it}\sum_{j\in\mathcal{J}}L_{ijt}^*.\end{aligned}\tag{A.2}$$

Moreover, $\{K_{ijt}^{\#},L_{ijt}^{\#},M_{ijt}^{\#}\}_{j\in\mathcal{J}}$ also satisfies resource constraint in optimization problem (5). The inequality (A.2) contradicts that $\{K_{ijt}^*,L_{ijt}^*,M_{ijt}^*\}_{j\in\mathcal{J}}$ is the unique optimizer of optimization problem (5). □

Main Proof of Proposition 1

Proof. Lemma 2 shows that finding $\{K_{ijt}^*,L_{ijt}^*,M_{ijt}^*\}_{j\in\mathcal{J}}$ in optimization problem (5) is equivalent to finding the optimizer of optimization problem (A.1). The alternative optimization problem (A.1) is a convex optimization problem since both the objective function and the constraint set are convex, so the solution can be equivalently characterized by the first-order conditions:

$$X_{ijt}=\frac{R_{ijt}^{exp}}{\lambda_{it}^X}\frac{\partial f_j}{\partial x_{ijt}}\left(1+\frac{1}{\eta_{ijt}}\right),\quad\forall X\in\{K,L,M\}\tag{A.3}$$

where λ_{it}^X is the Lagrangian multiplier for the constraint $\sum_j X_{ijt}=\bar{X}_{it}$ in the alternative optimization problem, and η_{ijt} is the demand elasticity. Moreover, $\{K_{ijt}^*,L_{ijt}^*,M_{ijt}^*\}_{j\in\mathcal{J}}$ is the unique solution to the first-order conditions, since the optimization problem is differentiable and strictly convex. Using $\sum_{j\in\mathcal{J}}X_{ijt}^*=\bar{X}_{it}$, we have

$$\lambda_{it}^X=\frac{\sum_{j\in\mathcal{J}}\frac{\partial f_j}{\partial x_j}(1+1/\eta_{ijt})R_{ijt}^{exp}}{\bar{X}_{it}},\quad for\ X\in\{K,L,M\}.\tag{A.4}$$

Substitute this back to the first-order condition, we have the following system of functions implied by (A.3):

$$\bar{X}_{it}\frac{R_{ijt}^{exp}\frac{\partial f_j}{\partial x_{ijt}}(1+1/\eta_{ijt})}{\sum_{j\in\mathcal{J}}R_{ijt}^{exp}\frac{\partial f_j}{\partial x_j}(1+1/\eta_{ijt})}-X_{ijt}^*=0,\quad for\ X\in\{K,L,M\}.\tag{A.5}$$

We now claim that if $\{K_{ijt}^*,L_{ijt}^*,M_{ijt}^*\}_{j\in\mathcal{J}}$ is the unique solution to the optimization problem (A.1), it is also the unique solution to the system of equations (A.5). Suppose not, we can find

$\{\tilde{K}_{ijt}, \tilde{L}_{ijt}, \tilde{M}_{ijt}\}_{j\in\mathcal{J}}$ as another solution to (A.5), then by choosing

$$\frac{1}{\tilde{\lambda}^X_{it}} = \frac{\sum_{j\in\mathcal{J}} \frac{\partial f_j}{\partial \tilde{x}_{ijt}}(1+1/\eta_{ijt})R^{exp}_{ijt}}{\bar{X}_{it}} \quad for\ X \in \{K, L, M\}, \tag{A.6}$$

then $\{\{\tilde{K}_{ijt}, \tilde{L}_{ijt}, \tilde{M}_{ijt}\}_{j\in\mathcal{J}}, \tilde{\lambda}^K_{it}, \tilde{\lambda}^L_{it}, \tilde{\lambda}^M_{it}\}$ is a solution to (A.4), which implies that $\{\tilde{K}_{ijt}, \tilde{L}_{ijt}, \tilde{M}_{ijt}\}_{j\in\mathcal{J}}$ is also a solution to the alternative optimization problem (A.1). This contradicts that $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ is the unique optimizer of (A.1).

Then, we establish that the optimal solution $\{\tilde{K}_{ijt}, \tilde{L}_{ijt}, \tilde{M}_{ijt}\}_{j\in\mathcal{J}}$ is a function of observables and parameters. Note that the system of equations (A.6) contains

$$(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, Q_{i1t}, ...Q_{iJt}, \delta_{jt}, \sigma_j, F_1, ...F_J)$$

and $\{X_{ijt} : X \in \{K, L, M\}, j \in \mathcal{J}\}$. The mapping

$$H : (\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}, Q_{i1t}, ...Q_{iJt}, \{\{\delta_{jt}\}_{t\le T}, \sigma_j, \boldsymbol{\beta}_j\}_{j\in\mathcal{J}},) \to \{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$$

is a function since the solution $\{K^*_{ijt}, L^*_{ijt}, M^*_{ijt}\}_{j\in\mathcal{J}}$ is unique. The components of the mapping H above correspond to the allocation functions $\mathcal{A}^K_j, \mathcal{A}^L_j, \mathcal{A}^M_j$ in Proposition 1.

Lastly, we show that if Assumption 5 fails, optimization problem (5) has no solution. If Assumption 5 fails, we can find $\{K_{ijt}\}_{j\in\mathcal{J}}$ and a positive direction $(L^d_{i1t}, ..., L^d_{iJt}, M^d_{i1t}, ..., M^d_{iJt}) \neq \vec{0}$ and $L^d_{ijt}, M^d_{ijt} \ge 0$ for all $j \in \mathcal{J}$, such that $\liminf_{\lambda\to\infty} V((K_{ijt}, \lambda L^d_{ijt}, \lambda M^d_{ijt})_{j\in\mathcal{J}} = \infty$. Therefore, there is no solution to optimization problem (5). □

A.2 Proofs in Section 3

Theorem 1

Proof. Under the Dixit-Stiglitz demand, moment condition (9) becomes

$$E^r_{it}\left(\{Q_{ijt}\}_{j\in\mathcal{J}}\right) \equiv \mathbf{E}\left[r^{obs}_{it} \middle| \{Q_{ijt}\}_{j\in\mathcal{J}}\right] = \log\left(\sum_{j\in\mathcal{J}} e^{\delta_{jt}} Q_{ijt}^{\frac{\sigma_j - 1}{\sigma_j}}\right),$$

and we can identify the following partial derivatives:

$$\frac{\partial exp(E^r_{it}\left(\{Q_{ijt}\}_{j\in\mathcal{J}}\right))}{\partial Q_{ijt}} = e^{\delta_{jt}} \frac{\sigma_j - 1}{\sigma_j} Q_{ijt}^{-1/\sigma_j}.$$

Now fix a j and any positive constant $c > 0$, we consider $\{\tilde{Q}_{ijt}\}_{j\in\mathcal{J}}$ such that $\tilde{Q}_{ij't} = Q_{ij't}$ for all $j' \neq j$ and $\tilde{Q}_{ij't} = cQ_{ijt}$. We have

$$\frac{\partial exp(E^r_{it}\left(\{Q_{ijt}\}_{j\in\mathcal{J}}\right))}{\partial Q_{ijt}} \Bigg/ \frac{\partial exp(E^r_{it}(\{\tilde{Q}_{ijt}\}_{j\in\mathcal{J}}))}{\partial \tilde{Q}_{ijt}} = c^{-1/\sigma_j}.$$

Therefore, σ_j is identified. By repeating this procedure for all $j \in \mathcal{J}$, we can show the demand parameter $\{\sigma_j\}_{j\in\mathcal{J}}$ are identified. Lastly, given the identified $\{\sigma_j\}_{j\in\mathcal{J}}$, δ_{jt} is identified as $\log\left[\frac{\partial exp(E^r_{it}(\{Q_{ijt}\}_{j\in\mathcal{J}}))}{\partial Q_{ijt}}\right] - \log\frac{\sigma_j - 1}{\sigma_j} - \log Q_{ijt}^{-1/\sigma_j}$.

□

Proposition 2

Proof. To simplify notation, we denote $F_j = F(\cdot, \boldsymbol{\beta}_j)$, and $\tilde{F}_j = F(\cdot, \tilde{\boldsymbol{\beta}}_j)$. Similarly, we denote $h = h(\cdot, \boldsymbol{\rho})$ and $\tilde{h} = h(\cdot, \tilde{\boldsymbol{\rho}})$. We aim to show that all moment conditions (9), (11), (12), (13) and (14) hold under the new vector of parameters $(\{\tilde{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \tilde{\boldsymbol{\rho}}, \mathcal{U})$.

The moment conditions in (11), (12) and (13) are not influenced if we change the parameter value from $(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ to $(\{\tilde{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \tilde{h}, \mathcal{U})$ because $\gamma^X_{ijt} = \frac{\partial f_j}{\partial x_{ijt}} = \frac{\partial(a_j + f_j)}{\partial x_{ijt}}$ for $X \in \{L, M\}$. Equation (9) does not involve the production function $\{F_j\}_{j\in\mathcal{J}}$, so it is not influenced either. It suffices to check (14) holds under the new parameter. Note that

$$\begin{aligned} q_{ijt} - & f_j(k_{ijt}, l_{ijt}, m_{ijt}) - h_j\left((q_{ij't-1} - f(k_{ij't-1}, l_{ij't-1}, m_{ij't-1}))_{j'\in\mathcal{J}}\right) \\ =& q_{ijt} - a_j - f_j(k_{ijt}, l_{ijt}, m_{ijt}) + a_j - h_j\left((q_{ij't-1} - a_j - f(k_{ij't-1}, l_{ij't-1}, m_{ij't-1}))_{j'\in\mathcal{J}} + \boldsymbol{a}\right) \\ =& q_{ijt} - \tilde{f}_j(k_{ijt}, l_{ijt}, m_{ijt}) - \tilde{h}_j\left((q_{ij't-1} - \tilde{f}(k_{ij't-1}, l_{ij't-1}, m_{ij't-1}))_{j'\in\mathcal{J}} + \boldsymbol{a}\right), \end{aligned} \tag{A.7}$$

where the first line of (A.7) is the expression of (14) under $(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$, while the last line of (A.7) is the expression of (14) under $(\{\tilde{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \tilde{\boldsymbol{\rho}}, \mathcal{U})$.

So if the moment condition (14) holds for $(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$, it must also hold for $(\{\tilde{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \tilde{\boldsymbol{\rho}}, \mathcal{U})$. □

Proposition 3

Proof. To simplify notation, we denote $F_j = F(\cdot, \boldsymbol{\beta}_j)$, and $\breve{F}_j = F(\cdot, \breve{\boldsymbol{\beta}}_j)$. By Proposition 1, we denote $(K^*_{ijt}, L^*_{ijt}, M^*_{ijt})_{j\in\mathcal{J}}$ as the optimal allocation rule under $\{F_j\}_{j\in\mathcal{J}}$. Our proof of Proposition 3 consists of two steps:

1. We show that we can find a unique optimal allocation $\{\breve{K}_{ijt}, \breve{L}_{ijt}, \breve{M}_{ijt}\}_{j\in\mathcal{J}}$ under $\{\breve{F}_j\}_{j\in\mathcal{J}}$.
2. We show that moment conditions (9), (11), (12), (13) and (14) hold under $\{\breve{K}^*_{ijt}, \breve{L}^*_{ijt}, \breve{M}^*_{ijt}\}_{j\in\mathcal{J}}$ and $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$.

Step 1 Given the constants C^X_j in Proposition 3, let

$$\breve{X}^*_{ijt} = X^*_{ijt} + C^X_j \quad \forall X \in \{K, L, M\},$$

then we claim that $\breve{X}^*_{ijt}$ is the optimal allocation under $\{\breve{F}_j\}_{j\in\mathcal{J}}$. The resource constraint holds for $\breve{X}^*_{ij}$ by the construction $\sum_{j\in\mathcal{J}} C^X_j = 0$. Since the static profit optimization problem (5) is a strictly convex optimization optimization problem under $\breve{F}_j$, it suffices to check the first-order condition and the resource constraint to ensure $\{\breve{X}^*_{ijt}\}_{j\in\mathcal{J},X\in\{K,L,M\}}$ is the optimal allocation under $\{\breve{F}_j\}_{j\in\mathcal{J}}$. That is, we want to show there exist a Lagrangian multiplier $\breve{\lambda}^K_{it}$ such that $(\breve{K}^*_{ijt}, \breve{L}^*_{ijt}, \breve{M}^*_{ijt})_{j\in\mathcal{J}}$ satisfies the first-order condition

$$\frac{\partial P_d(Q_{ijt};\sigma_j)}{\partial Q_{ijt}}\frac{\partial \breve{F}_j(\breve{K}^*_{ijt},\breve{L}^*_{ijt},\breve{M}^*_{ijt})}{\partial X_{ijt}}e^{\omega_{ijt}}\breve{F}_j + P_d\frac{\partial \breve{F}_j(\breve{K}^*_{ijt},\breve{L}^*_{ijt},\breve{M}^*_{ijt})}{\partial X_{ijt}}e^{\omega_{ijt}} = \breve{\lambda}^X_{it}, \quad for\, X\in\{K,L,M\}, \tag{A.8}$$

where $\breve{\lambda}^L_{it} = w_{it}$ and $\breve{\lambda}^M_{it} = v_{it}$.

To show (A.8) holds, we first look at the first-order condition with respect to X_{ijt} under F_j, which is satisfied by the optimal allcation rule $(K^*_{ijt}, L^*_{ijt}, M^*_{ijt})_{j\in\mathcal{J}}$:

$$\frac{\partial P_d(Q_{ijt};\sigma_j)}{\partial Q_{ijt}}\frac{\partial F_j(K^*_{ijt},L^*_{ijt},M^*_{ijt})}{\partial X_{ijt}}e^{\omega_{ijt}}F_j + P_d\frac{\partial F_j(K^*_{ijt},L^*_{ijt},M^*_{ijt})}{\partial X_{ijt}}e^{\omega_{ijt}} = \lambda^X_{it} \quad X\in\{K,L,M\}, \tag{A.9}$$

where λ^K_{it} is the Lagrangian multiplier for the resource constraint and $\lambda^L_{it} = w_{it}$ and $\lambda^M_{it} = v_{it}$.

By construction of $\breve{F}_j$ and $\breve{X}^*_{ij}$, we have $\frac{\partial \breve{F}_j(\breve{K}^*_{ijt},\breve{L}^*_{ijt},\breve{M}^*_{ijt})}{\partial \breve{X}_{ijt}} = \frac{\partial F_j(K^*_{ijt},L^*_{ijt},M^*_{ijt})}{\partial X_{ijt}}$ for all $X \in \{K,L,M\}$. By taking $\breve{\lambda}^K_{it} = \lambda^K_{it}$, equation (A.9) implies (A.8) holds. Therefore, $\breve{X}^*_{ijt}$ is the optimal allocation under $\breve{F}_j$.

Step 2 It remains to show that the moment conditions (9), (11), (12), (13) and (14) hold under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$. First note that (9) does not depend on the production function, so it holds under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ since we do not change σ_j.

We then observe that by the construction of Step 1, $f_j\left(k^*_{ijt},\, l^*_{ijt},\, m^*_{ijt}\right) = \breve{f}_j\left(\breve{k}^*_{ijt},\, \breve{l}^*_{ijt},\, \breve{m}^*_{ijt}\right)$. As a result, moment condition (14) holds under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$.

Last, we show moment conditions (11), (12) and (13) hold under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$. Note that since $\{\breve{K}^*_{ijt}, \breve{L}^*_{ijt}, \breve{M}^*_{ijt}\}_{j\in\mathcal{J}}$ is the optimal solution to (8) under $\{\breve{F}_j\}_{j\in\mathcal{J}}$, we can write the first-order condition (A.8) for $X = M$:

$$\frac{\partial P_d(Q_{ijt};\sigma_j)/P_d}{\partial Q_{ijt}/Q_{ijt}}\frac{\partial \breve{F}_j(\breve{K}^*_{ijt},\breve{L}^*_{ijt},\breve{M}^*_{ijt})/\breve{F}_j}{\partial M^*_{ijt}/\breve{M}^*_{ijt}} + \frac{\partial \breve{F}_j(\breve{K}^*_{ijt},\breve{L}^*_{ijt},\breve{M}^*_{ijt})/\breve{F}_j}{\partial M^*_{ijt}/\breve{M}^*_{ijt}} = \frac{v_{it}\breve{M}^*_{ijt}}{P_d Q_{ijt}},$$

which is the same as the equation of material-to-revenue share in (10) with F_j replaced by $\breve{F}_j$ and $(K^*_{ijt}, L^*_{ijt}, M^*_{ijt})_{j\in\mathcal{J}}$ replaced by $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$ Therefore moment condition (11) is also satisfied under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$. Moment condition (12) for labor-to-revenue share can be verified similarly. Since moment condition (13) is simply an exponential transformation of (11) and (12), moment condition (13) also holds under $(\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\le T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U})$. □

Proposition 4

Proof. We only show the proof for the identification of β^M_j and $\mathcal{U}$; The proof for the identification of β_j^L is similar to that for β_j^M. We first take exponent on both side of (19) to get

$$\exp\left\{\mathbf{E}\left(s_{it}^M|Q_{i1t},...Q_{iJt}\right)\right\} = \sum_{j\in\mathcal{J}} \frac{(\sigma_j-1)\mathcal{U}\beta_j^M}{\sigma_j}\frac{Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'} Q_{ij't}^{\frac{\sigma_{j'}-1}{\sigma_{j'}}}}.$$

Define $\beta_j^{M,\mathcal{U},\sigma} \equiv \frac{\sigma_j-1}{\sigma_j}e^{\delta_{jt}}\mathcal{U}\beta_j^M$. We consider a $J\times J$ matrix $\mathbb{Q}$, whose (j,ι) element is $\frac{\mathcal{Q}_{j,\iota}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'}\mathcal{Q}_{j',\iota}^{\frac{\sigma_j-1}{\sigma_j}}}$ such that $\mathbb{Q}$ is of full rank, where $\mathcal{Q}_{j,\iota}$ are real numbers. Fix an ι, conditioned on $Q_{ijt} = \mathcal{Q}_{j,\iota}$ for all j, we have

$$e^{s_\iota} \equiv \exp\left\{\mathbf{E}\left(s_{it}^M|Q_{ijt} = \mathcal{Q}_{j,\iota},\ for\ j\in\mathcal{J}\right)\right\} = \sum_j \beta_j^{M,\mathcal{U},\sigma}\frac{\mathcal{Q}_{j,\iota}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'}\mathcal{Q}_{j',\iota}^{\frac{\sigma_j-1}{\sigma_j}}}.$$

which can be written in the following equation of matrix:

$$[e^{s_1},...e^{s_J}] = [\beta_1^{M,\mathcal{U},\sigma},...,\beta_J^{M,\mathcal{U},\sigma}]\mathbb{Q}.$$

Therefore, $[\beta_1^{M,\mathcal{U},\sigma},...,\beta_J^{M,\mathcal{U},\sigma}] = [e^{s_1},...e^{s_J}]\mathbb{Q}^{-1}$ is identified. Observing that $u_{it} = s_{it}^M - \log\left(\sum_j \frac{\beta_j^{M,\mathcal{U},\sigma}Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'}Q_{ij't}^{\frac{\sigma_{j'}-1}{\sigma_{j'}}}}\right)$, we can identify $\mathcal{U}$ as

$$\mathcal{U} = \mathbf{E}\left\{\exp\left[s_{it}^M - \log\left(\sum_j \frac{\beta_j^{M,\mathcal{U},\sigma}Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}}}{\sum_{j'}Q_{ij't}^{\frac{\sigma_{j'}-1}{\sigma_{j'}}}}\right)\right]\right\}. \tag{A.10}$$

Lastly, β_j^M is identified as the ratio between $\beta_j^{M,\mathcal{U},\sigma}$ and $\frac{\mathcal{U}(\sigma_j-1)}{\sigma_j}e^{\delta_{jt}}$. □

Proposition 5

Proof. Throughout this proof, we use the following notation: $\tilde{\beta}_j^X \equiv \frac{\sigma_j-1}{\sigma_j}e^{\delta_{jt}}\beta_j^X$. We treat β_j^L and β_j^M as known since they are identified from the labor share and material share equations, respectively. The moment condition (22) can be rewritten as

$$\mathbf{E}\left\{\tilde{q}_{ijt} - \beta_j^K\log(\alpha_{ijt}^K) - \beta_j^K\bar{k}_t - h_j\left((\tilde{q}_{ijt-1} - \beta_j^K[\log(\alpha_{ijt-1}^K) + \bar{k}_{it}])_{j\in\mathcal{J}};\boldsymbol{\rho}\right)\middle|\mathcal{I}_{it}\right\} = 0. \tag{A.11}$$

Take the partial derivative of (A.11) with respect to $\bar{k}_{it}$, we obtain

$$\mathbf{E}\left(\frac{\partial \tilde{q}_{ijt}}{\partial \bar{k}_{it}}\Big|\mathcal{I}_{it}\right) = \beta_j^K + \beta_j^K \mathbf{E}\left(\frac{\partial \log(\alpha_{ijt}^K)}{\partial \bar{k}_{it}}\Big|\mathcal{I}_{it}\right). \tag{A.12}$$

Recall that $\mathbf{E}\left(\frac{\partial \tilde{q}_{jt}}{\partial \bar{k}_t}|\mathcal{I}_t\right)$ is identified from data. Also note that the allocation rule (20) implies that

$$\begin{aligned}
\mathbf{E}\Big(\frac{\partial \log(\alpha_{ijt}^K)}{\partial \bar{k}_{it}}|\mathcal{I}_{it}\Big) &= \mathbf{E}\Big(\frac{\partial \log(\tilde{\beta}_j^K Q_{ijt}^{\frac{\sigma_j-1}{\sigma_j}})}{\partial \bar{k}_{it}} - \frac{\partial \log(\sum_{j'} \tilde{\beta}_{j'}^K Q_{j't}^{\frac{\sigma_j-1}{\sigma_j}})}{\partial \bar{k}_{it}}|\mathcal{I}_{it}\Big) \\
&= \mathbf{E}\Big(\frac{\sigma_j - 1}{\sigma_j}\frac{\partial q_{ijt}}{\partial \bar{k}_{it}} - \underbrace{\frac{\partial \log(\sum_{j'} \tilde{\beta}_{j'}^K Q_{j't}^{\frac{\sigma_j-1}{\sigma_j}})}{\partial \bar{k}_{it}}}_{Term\ B}|\mathcal{I}_{it}\Big).
\end{aligned} \tag{A.13}$$

Note that $Term\, B$ on the right-hand side of (A.13) is independent of the index j on the left-hand side. Take the difference between j and j', we obtain

$$\mathbf{E}\Big(\frac{\partial \log(\alpha_{jt}^K)}{\partial \bar{k}_t} - \frac{\partial \log(\alpha_{j't}^K)}{\partial \bar{k}_t}|\mathcal{I}_t\Big) = \mathbf{E}\Big(\frac{\sigma_j - 1}{\sigma_j}\frac{\partial q_{jt}}{\partial \bar{k}_{it}} - \frac{\sigma_{j'} - 1}{\sigma_{j'}}\frac{\partial q_{j't}}{\partial \bar{k}_t}|\mathcal{I}_t\Big), \tag{A.14}$$

where the right-hand side of (A.14) is identified.

We then solve $\mathbf{E}\left(\frac{\partial \log(\alpha_{ijt}^K)}{\partial \bar{k}_{it}}\Big|\mathcal{I}_{it}\right)$ from (A.12), and plug it into (A.14). We obtain an equation characterizing the relationship between β_j and β_j':

$$\frac{\tilde{E}_{jt}^K(\mathcal{I}_{it})}{\beta_j^K} - \frac{\tilde{E}_{j't}^K(\mathcal{I}_{it})}{\beta_{j'}^K} = \left[\frac{\sigma_j - 1}{\sigma_j}E_{jt}^K(\mathcal{I}_{it}) - \frac{\sigma_{j'} - 1}{\sigma_{j'}}E_{j't}^K(\mathcal{I}_{it})\right]. \tag{A.15}$$

We can also get another equation when we condition on the $\mathcal{I}_{it}'$:

$$\frac{\tilde{E}_{jt}^K(\mathcal{I}_{it}')}{\beta_j^K} - \frac{\tilde{E}_{j't}^K(\mathcal{I}_{it}')}{\beta_{j'}^K} = \left[\frac{\sigma_j - 1}{\sigma_j}E_{jt}^K(\mathcal{I}_{it}') - \frac{\sigma_{j'} - 1}{\sigma_{j'}}E_{j't}^K(\mathcal{I}_{it}')\right]. \tag{A.16}$$

From the system of equations (A.15) and (A.16), we treat β_j^K and $\beta_j'^K$ as unknown parameters to be solved.

Note that the following system of equation has a unique solution of (x_1, x_2) if $a_1/a_2 \neq b_1/b_2$:

$$\frac{a_1}{x_1} - \frac{a_2}{x_2} = c_1, \qquad \frac{b_1}{x_1} - \frac{b_2}{x_2} = c_2.$$

Therefore the condition $\tilde{E}_{jt}^K(\mathcal{I}_{it})/\tilde{E}_{j't}^K(\mathcal{I}_{it}) \neq \tilde{E}_{jt}^K(\mathcal{I}_{it}')/\tilde{E}_{j't}^K(\mathcal{I}_{it}')$ ensures that (A.15) and (A.16) identify β_j^K and $\beta_{j'}^K$. □

A.3 Appendix to Section 4: Dimension Reduction for Partial Identification Method

We now state sufficient conditions so that we have a consistent estimator of the identified set.

Assumption 8. *Fix $\theta_{-\boldsymbol{\rho}}$, let $\boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}})$ be a minimizer of $\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho})]'\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho})]$. The closed-form estimator $\tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})$ solved from (31) satisfies*

$$\inf_{\theta_{-\boldsymbol{\rho}}} \left\| \frac{1}{N} \sum_{i=1}^{N} m_i(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))] \right\| = O_p(1/\sqrt{a_N}).$$

Assumption 8 requires the GMM criterion function (32) to be $O_p(1/a_N)$-uniformly consistent when we plug in the estimator $\tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})$. This condition can be satisfied even if there are many $\boldsymbol{\rho}$'s that minimize $\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho})]'\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho})]$. We also need standard partial identification regularity conditions in Chernozhukov et al. (2007), which are summarized in the following assumption:

Assumption 9. *The following conditions hold for the moment function $m_i(\theta)$: (1) The parameter space Θ is compact and the identified set Θ^{ID} lies in the interior of the parameter space Θ; (2) The class $\{m_i(\theta) : \theta \in \Theta\}$ is $P-$Donsker; (3) There exist constants C and δ such that for all $\theta \in \Theta$, $||E_P[m_i(\theta)]|| \geq C(d(\theta, \Theta^{ID}) \wedge \delta)$ and the Jacobian matrix $G(\theta) = \nabla E_P[m_i(\theta)]$ is continuous in θ; (4) The weighting matrix $A_n(\theta)$ converges uniformly to a positive definite matrix $A(\theta)$, and the eigenvalues of $A(\theta)$ is uniformly bounded away from zero and infinity: $\inf_\theta \min eig(A(\theta)) > 0$ and $\sup_\theta \max eig(A(\theta)) < \infty$; (5) The partial derivatives are uniformly bounded by some constant $C_{\boldsymbol{\rho}}$: $\nabla_{\boldsymbol{\rho}} |m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho})| < C_{\boldsymbol{\rho}}$.*

In Assumption 9: Condition (1) is similar to the requirement that the true parameter is in the interior of the parameters space in a point identified model; Conditions (2) and (4) are regularity conditions and can be verified if the form of moment function is known; Condition (3) requires the expectation of the moment functions to be bounded linearly by a distance between the parameter in $(\Theta^{ID})^c$ and the identified set Θ^{ID}.

Condition (3) in Assumption 9 involves the expectation under the true data generating process P. This creates a difficulty for verifying condition (3) analytically. This is because when we construct moment condition (22), the input ratios $\{\alpha_{ijt}^X\}_{X \in \{K,L,M\}}$ depend on the endogenous variables $\{Q_{ijt}\}_{j \in \mathcal{J}}$. As a result, we cannot get a closed-form expression for $E[m_i(\theta)]$ at a given value of θ. We also note that if $||E_P[m_i(\theta)]|| \lesssim \frac{1}{n^{1/2-\epsilon}}(d(\theta, \Theta^{ID}) \wedge \delta)$ for some $\epsilon > 0$, then it is not possible to test condition (3) using data. We will maintain this high level condition in our simulation for the Cobb-Douglas case and our empirical application.

Proposition 6. *Let $\Theta_{-\boldsymbol{\rho}}^{ID} \equiv \{\theta_{-\boldsymbol{\rho}} : \exists \boldsymbol{\rho} \quad s.t. \ (\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}) \in \Theta^{ID}\}$ be the projected identified set. Suppose Assumptions 8 and 9 hold and $a_N/b_N \to \infty$. Then $\hat{\Theta}_{-\boldsymbol{\rho}}^{ID} \equiv \{\theta_{-\boldsymbol{\rho}} : \exists \boldsymbol{\rho} \quad s.t. \ (\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}) \in \hat{\Theta}^{ID}\}$ is a consistent estimator of $\Theta_{-\boldsymbol{\rho}}^{ID}$ in the Hausdorff metric.*

Proof. We use $A_n(\theta)$ to denote $A_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))$ and its eigonvalue is uniformly bounded away from 0. We first write down the following expansion:

$$
\begin{aligned}
M_n(\theta_{-\boldsymbol{\rho}}) &= \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))' A_n(\theta)^{-1} \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) \\
&= \underbrace{\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))]' A_n(\theta)^{-1} \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))]}_{\text{Term 1}} \\
&\quad - \underbrace{2(\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])' A_n(\theta)^{-1} \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^0(\theta_{-\boldsymbol{\rho}}))}_{\text{Term 2}} \\
&\quad + \underbrace{(\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])' A_n(\theta)^{-1} (\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])}_{\text{Term 3}}.
\end{aligned}
\tag{A.17}
$$

We first show that $\liminf_{n\to\infty} \Pr(\Theta^{ID}_{-\boldsymbol{\rho}} \subseteq \hat{\Theta}^{ID}_{-\boldsymbol{\rho}}) = 1$. For any $\theta_{-\boldsymbol{\rho}} \in \Theta^{ID}_{-\boldsymbol{\rho}}$, by definition, $\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))] = 0$. Therefore Term 1 in (A.17) equals zero, and by Assumption 8,

$$
\begin{aligned}
&\sup_{\theta_{-\rho}} ||\text{Term 2+ Term 3}|| \\
=& \sup_{\theta_{-\rho}} (\bar{m}_n(\theta_{-\rho}, \tilde{\boldsymbol{\rho}}(\theta_{-\rho})) - \mathbf{E}[m_i(\theta_{-\rho}, \boldsymbol{\rho}^*(\theta_{-\rho}))])' A_n(\theta)^{-1} (\bar{m}_n(\theta_{-\rho}, \tilde{\boldsymbol{\rho}}(\theta_{-\rho})) - \mathbf{E}[m_i(\theta_{-\rho}, \boldsymbol{\rho}^*(\theta_{-\rho}))]) \\
=& \inf_{\theta} \left[\min(eig(A_n(\theta))^{-1})\right] O_p(1/a_N) = O_p(1/a_N),
\end{aligned}
$$

where the last equality holds because we assume $A_n(\theta)$ converges uniformly to $A(\theta)$ and $\inf_\theta \left[\min(eig(A(\theta))^{-1})\right]$ is bounded. As a result, $\sup_{\theta_{-\boldsymbol{\rho}}} ||a_N M_n(\theta_{-\boldsymbol{\rho}})|| = O_p(1)$ and

$$
\Pr\left(\Theta^{ID}_{-\boldsymbol{\rho}} \subseteq \hat{\Theta}^{ID}_{-\boldsymbol{\rho}}\right) \geq \Pr\left(\sup_{\theta_{-\boldsymbol{\rho}}} ||a_N M_n(\theta_{-\boldsymbol{\rho}})|| \leq c b_N\right) = \Pr(O_p(1) \leq c b_N) \to 1.
$$

We then show that for any $\epsilon > 0$, $\liminf_{n\to\infty} \Pr\left([\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c \subseteq [\hat{\Theta}^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c\right) = 1$, where $\hat{\Theta}^{ID,\epsilon}_{-\boldsymbol{\rho}} = \{\theta_{-\rho} : d(\theta_{-\rho}, \hat{\Theta}^{ID}_{-\boldsymbol{\rho}}) < \epsilon\}$, and $\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}} = \{\theta_{-\rho} : d(\theta_{-\rho}, \Theta^{ID}_{-\boldsymbol{\rho}}) < \epsilon\}$. For any $\theta_{-\boldsymbol{\rho}} \in [\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c$, by condition (3) in Assumption 9, we know $\inf_{\theta_{-\boldsymbol{\rho}}} ||\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))]|| \geq C(\epsilon \wedge \delta)$. In the expansion of (A.17), Term 1 is a constant, Term 2 is $O_p(1/\sqrt{a_N})$ and Term 3 is $O_p(1/a_N)$. As a result, $\inf_{\theta_{-\boldsymbol{\rho}} \in [\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c} ||M_n(\theta_{-\boldsymbol{\rho}})|| = O_p(1)$, and

$$
\Pr\left([\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c \subseteq [\hat{\Theta}^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c\right) \geq \Pr\left(\inf_{\theta_{-\boldsymbol{\rho}} \in [\Theta^{ID,\epsilon}_{-\boldsymbol{\rho}}]^c} ||a_N M_n(\theta_{-\boldsymbol{\rho}})|| > c b_N\right) = \Pr(O_p(1) \geq c b_N / a_N) \to 1.
$$

The result follows. □

We can choose $a_N = N$ and $b_N = \log N$. For the confidence region, we aim to find a set Θ^α such that $\liminf_{n\to\infty} \Pr(\Theta^{ID} \subseteq \Theta^\alpha) \geq 1 - \alpha$. We follow Remark 4.2 in Chernozhukov et al. (2007) to use the multiplier bootstrap to construct a confidence region in the following way. For the b^{th} bootstrapped sample, we generate a random vector $(z_i^{*b})_{i=1}^n \sim N(0, I_{n\times n})$ and calculate the random variable $\mathcal{C}^{*b} = \sup_{\theta\in\hat{\Theta}^{ID}} N\left[\bar{m}_n^{*b}(\theta)' A_n(\theta) \bar{m}_n^{*b}(\theta)\right]$ where $\bar{m}_n^{*b}(\theta) = \frac{1}{N}\sum_{i=1}^N m_i(\theta) z_i^{*b}$. Repeating the bootstrap process B times, we get the $(1-\alpha)^{th}$ quantile of the bootstrapped sample $\{\mathcal{C}^{*1}, ..., \mathcal{C}^{*B}\}$, denoted by $\mathcal{C}^*_\alpha$. The $(1-\alpha)$-confidence region is constructed as:

$$\Theta^\alpha = \{\theta \in \Theta : NM_n(\theta) \leq \mathcal{C}^*_\alpha + c\log(N)\},$$

where we use the same constant c as in (33).

Proposition 7. *Under Assumption 9, Θ^α is a valid α-confidence region for the identified set.*

Proof. By Lemma 4.1 in Chernozhukov et al. (2007), Assumption 9 implies the conditions in Theorem 4.1 of Chernozhukov et al. (2007). The result follows. □

B Monte Carlo Study

B.1 Description

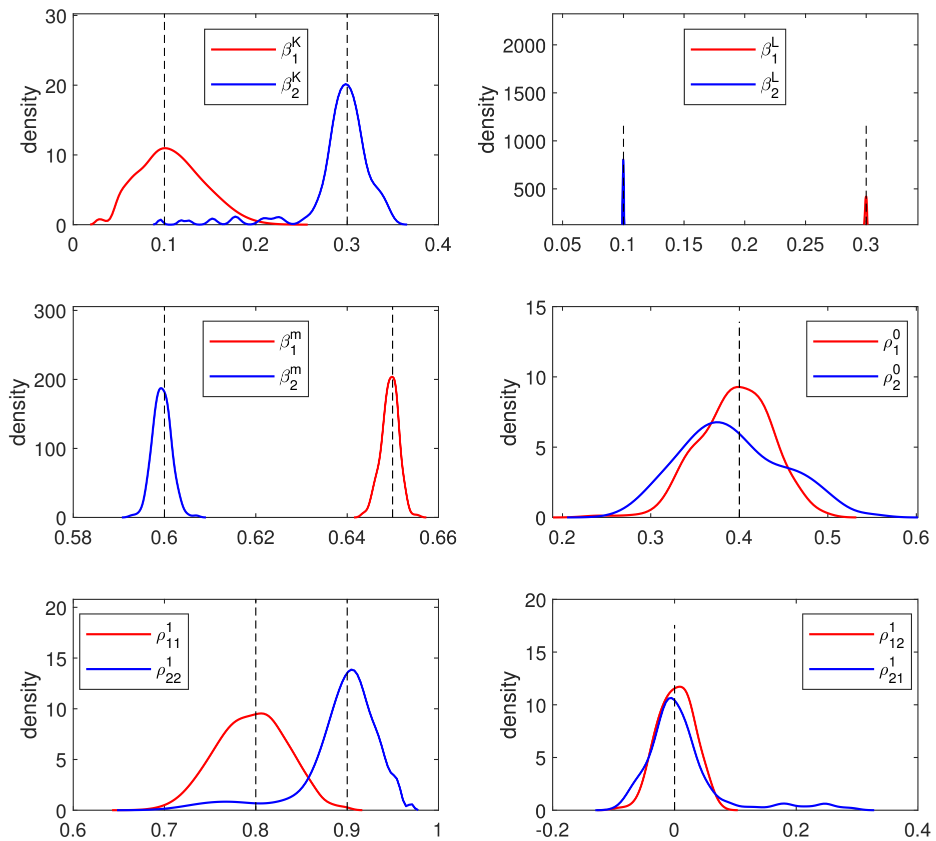
The parameters used in our Monte Carlo study is summarized in the following table:

Table B.1: Parameter Values for Monte Carlo Study

Parameters	Description	Values
$(\bar{k}_1, \bar{l}_1)$	Initial means of capital and labor	(10, 10)
(ξ_k^2, ξ_l^2)	Initial variances of capital and labor	$(1, 1)$
δ	Capital depreciation rate	0.1
(γ_1^K, γ_l^1)	Production function parameters for product 1	(0.1, 0.3)
(γ_k^2, γ_l^2)	Production function parameters for product 2	(0.3, 0.1)
(ν_1, ν_2)	Degree of the return to scale	(1, 1)
σ	Demand elasticity	3
(θ_1, θ_2)	Substitution elasticity	(0.8, 0.8)
(ρ_0^1, ρ_0^2)	Constants in productivity process	(0.2, 0.2)
$\begin{bmatrix} \rho_1^1 & \rho_2^1 \\ \rho_1^2 & \rho_2^2 \end{bmatrix}$	Transition 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)
$(\bar{\omega}_1, \bar{\omega}_2)$	Initial mean of the productivity vector	$(3, 3)$
T	Number of periods	2
N	Number of firms	1500

B.2 Distribution of the GMM Estimator of the Cobb-Douglas Production Function

Figure B.1: Kernel Density of the GMM Estimator for the Cobb-Douglas Production Function



Note: The GMM estimates are obtained by running 200 experiments over randomly generated samples.

C Testing the Identification Conditions in Proposition 5

We now use a parametric regression to check whether $\tilde{E}^K_{jt}(\mathcal{I}_{it})/\tilde{E}^K_{j't}(\mathcal{I}_{it}) \neq \tilde{E}^K_{jt}(\mathcal{I}'_{it})/\tilde{E}^K_{j't}(\mathcal{I}'_{it})$ holds in our simulation and empirical settings. A sufficient condition for $\tilde{E}^K_{jt}(\mathcal{I}_{it})/\tilde{E}^K_{j't}(\mathcal{I}_{it}) \neq \tilde{E}^K_{jt}(\mathcal{I}'_{it})/\tilde{E}^K_{j't}(\mathcal{I}'_{it})$ is

$$\tilde{E}^K_{jt}(\bar{k}_{it}, \bar{m}_{it})/\tilde{E}^K_{j't}(\bar{k}_{it}, \bar{m}_{it}) \neq \tilde{E}^K_{jt}(\bar{k}'_{it}, \bar{m}_{it})/\tilde{E}^K_{j't}(\bar{k}'_{it}, \bar{m}_{it}), \tag{C.1}$$

where $\tilde{E}^K_{jt}(\bar{k}_{it}, \bar{m}_{it}) = \partial \mathbf{E}[\tilde{q}_{ijt}|\bar{k}_{it}, \bar{m}_{it}]/\partial \bar{k}_{it}$. To check sufficient condition (C.1), we consider a second-order polynomial:

$$\tilde{\Xi}_j(\bar{k}_{it}, \bar{m}_{it}) = \tau_{0j} + \tau_{1j}\bar{k}_{it} + \tau_{2j}\bar{k}^2_{it} + \tau_{3j}\bar{k}_{it}\bar{m}_{it} + \tau_{4j}\bar{m}_{it} + \tau_{5j}\bar{m}^2_{it}, \tag{C.2}$$

and hence

$$\frac{\tilde{E}_{jt}^{K}(\bar{k}_{it})}{\tilde{E}_{j't}^{K}(\bar{k}_{it})} = \frac{\tau_{1j} + 2\tau_{2j}\bar{k}_{it} + \tau_{3j}\bar{m}_{it}}{\tau_{1j'} + 2\tau_{2j'}\bar{k}_{it} + \tau_{3j'}\bar{m}_{it}}. \tag{C.3}$$

The ratio (C.3) is not a constant if and only if $(\tau_{1j}, \tau_{2j}, \tau_{3j}) \neq c(\tau_{1j'}, \tau_{2j'}, \tau_{3j'})$ for all $c \in \mathbb{R}$. We estimate (C.2) using OLS regression, and test the following joint hypothesis using the delta method:

$$\mathcal{H}_0 : \tau_{1j}/\tau_{1j'} - \tau_{2j}/\tau_{2j'} = 0 \quad \text{and} \quad \tau_{1j}/\tau_{1j'} - \tau_{3j}/\tau_{3j'} = 0.$$

We test the hypothesis above using the chi-squared test:

$$T_n = n\left(\frac{\hat{\tau}_{1j}}{\hat{\tau}_{1j'}} - \frac{\hat{\tau}_{2j}}{\hat{\tau}_{2j'}}, \frac{\hat{\tau}_{1j}}{\hat{\tau}_{1j'}} - \frac{\hat{\tau}_{3j}}{\hat{\tau}_{3j'}}\right)\hat{\Sigma}\left(\frac{\hat{\tau}_{1j}}{\hat{\tau}_{1j'}} - \frac{\hat{\tau}_{2j}}{\hat{\tau}_{2j'}}, \frac{\hat{\tau}_{1j}}{\hat{\tau}_{1j'}} - \frac{\hat{\tau}_{3j}}{\hat{\tau}_{3j'}}\right)',$$

where $\hat{\Sigma}$ is estimated using the delta method. The results are reported in Table C.1.

Table C.1: Chi-squared Test of the Identification Condition

	Simulation	Empirical
Test statistics	6.15	0.154
p value	0.046	0.921

We can reject the null hypothesis in the simulation setting test, but we cannot reject the null hypothesis in the empirical application. There are two reasons why we fail to reject the null hypothesis in the empirical application: First, the sample size is small ($N = 290$); Second, capital plays almost no role in determining the output quantities in our empirical setting. As a result, the variation of $\tilde{\Xi}_j(\mathcal{I}_{it})$ with respect to $\bar{k}_{it}$ is weak and hard to detect. The sufficient condition in Proposition 5 may not hold. However, we note that the condition in Proposition 5 is only sufficient but not necessary. Identification may still hold when the condition in Proposition 5 fails.

References

Abito, J. M. (2021). Estimating Production Functions with Unobserved Time-invariant Heterogeneity and Time-varying Productivity Shocks.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Baumol, W., J. Panzar, and R. Willig (1982). *Contestable markets and the theory of industry structure*. New York: Harcourt Brace Jovanovich.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010). Multiple-product firms and product switching. *American Economic Review 100*(1), 70–97.

Bernard, A. B., S. J. Redding, and P. K. Schott (2011). Multiproduct firms and trade liberalization. *The Quarterly Journal of Economics 126*(3), 1271–1318.

Blum, B. S., S. Claro, I. Horstmann, and D. A. Rivers (2021). The ABCs of Firm Heterogeneity when Firms Sort into Markets: The Case of Exporters. *Working Paper*.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

De Loecker, J. and C. Syverson (2021, September). An industrial organization perspective on productivity. Working Paper 29229, National Bureau of Economic Research.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *The American Economic Review 102*(6), 2437–2471.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2017). Multi product firms, import competition, and the evolution of firm-product technical efficiencies. Technical report, National Bureau of Economic Research.

Dhyne, E., A. Petrin, V. Smeets, and F. Warzynski (2021). Theory for extending single-product production function estimation to multi-product settings. Unpublished manuscript.

Diewert, W. (1973). Functional forms for profit and transformation functions. *Journal of Economic Theory 6*(3), 284 – 316.

Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies 80*(4), 1338–1383.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review 98*(1), 394–425.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Garcia-Marin, A. and N. Voigtländer (2019). Exporting and Plant-Level Efficiency Gains: It's in the Measure. *Journal of Political Economy 127*(4), 1777–1825.

Gong, B. and R. C. Sickles (2021, April). Resource allocation in multi-divisional multi-product firms. *Journal of Productivity Analysis 55*(2), 47–70.

Grieco, P. L., S. Li, and H. Zhang (2016). Production function estimation with unobserved input price dispersion. *International Economic Review 57*(2), 665–690.

Klette, T. J. and Z. Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics 11*(4), 343–361.

Lamorgese, A. R., A. Linarello, and F. Warzynski (2015). Free Trade Agreements and Firm-Product Markups in Chilean Manufacturing. *Working Paper*.

Lau, L. J. (1976). A characterization of the normalized restricted profit function. *Journal of Economic Theory 12*(1), 131–163.

Levinsohn, J. and M. Melitz (2002). Productivity in a differentiated products market equilibrium. *Unpublished manuscript 9*, 12–25.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Marin, A. G. and N. Voigtländer (2018). Product-Level Efficiency and Core Competence in Multi-Product Plants. 2018 Meeting Papers 737, Society for Economic Dynamics.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbooks of Econometrics 4*, 2111–2245.

Olley, G. S. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Orr, S. (2021). Within-firm productivity dispersion: Estimates and implications. *University of British Columbia Working Paper*.

Rockafellar, R. (2015). *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press.

Valmari, N. (2016). Estimating production functions of multiproduct firms. *ETLA-Working Paper*.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.