Online Appendix to "Identification and Estimation of Production Functions for Multiproduct Firms"

Zhiyuan Chen
Renmin University of China

Moyu Liao
Nanjing University

First Version: December 2019
This Version: July 18, 2022

A The Firm's Dynamic Problem

We assume there is no firm entry and exit. The firm owner maximizes the firm value. Moreover, we assume that there is no change in the production set $\mathcal{J}_{it} = \mathcal{J}$ for all i and t. The firm's value function is given by:

$$V\left(\boldsymbol{\omega}_{it},\, \bar{K}_{it},\, \mathbf{s}_t\right) = \max_{\{K_{ijt}, L_{ijt}, M_{ijt}\}_{j\in\mathcal{J}}, I_{it}} \Bigg\{ \sum_{j=1}^{J} \left(\mathcal{U}P_d(\{Q_{ij't}\}_{j'\in\mathcal{J}})Q_{ijt} - v_{it}M_{ijt} - w_{it}L_{ijt}\right) - C(I_{it}) \tag{1}$$

$$+ \gamma \mathbf{E}\left[V\left(\boldsymbol{\omega}_{it+1},\, \bar{K}_{it+1},\, \mathbf{s}_{t+1}\right)\right] \Bigg\},$$

$$s.t.\ \bar{K}_{it+1} = (1-\delta)\,\bar{K}_{it} + I_{it}, \tag{2}$$

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

$$\sum_{j=1}^{J} K_{ijt} = \bar{K}_{it}, \tag{4}$$

where $\boldsymbol{\omega}_{it} = (\omega_{ijt})'_{j\in\mathcal{J}} \in \mathbb{R}^J$ is a vector of productivity, $\boldsymbol{s}_t$ represents the aggregate state variable that can be perfectly foreseen by the firm (such as the aggregate industrial trend and expected macroeconomic conditions), v_{it} is the material price, γ is the discounting factor, and w_{it} is the labor price. We use I_{it} to represent the firm-level investment, and $C(I_{it})$ to represent the costs of investment.

In our formulation, the firm-level capital investment is the only dynamic choice. Equation (2) is the accumulation rule of capital, with δ being the capital depreciation rate.

Equation (3) states that the productivity process is Markovian, and $\boldsymbol{\epsilon}_{it}$ is a J-dimension column vector of productivity shocks with zero mean. Equation (4) is the resource constraint for capital.

Based on the firm's dynamic problem, we can write down the optimization problem for the current period's profits as:

$$\max_{\{K_{ijt},L_{ijt},M_{ijt}\}_{j\in\mathcal{J}}} \sum_{j=1}^{J} \{\mathcal{U}P_{ijt}Q_{ijt} - v_{it}M_{ijt} - w_{it}L_{ijt}\}, \tag{5}$$

$$s.t. \sum_{j=1}^{J} K_{ijt} = \bar{K}_{it}, \tag{6}$$

which is the static optimization problem in our main text.

A.1 Empirical Content of the Dynamic Model

We now characterize the empirical content of the dynamic model. We will show that our nonparametric non-identification results still hold when we consider the full dynamic model. Let $\mathcal{O}_{it} \equiv (\bar{K}_{it}, \bar{K}_{it-1}, \bar{L}_{it-1}, \bar{M}_{it-1}, \{P_{ijt-1}, Q_{ijt-1}\}_{j=1}^{J})$ denote the a of variables exogenous to the t-period productivity shocks. We omit the aggregate state variable $\boldsymbol{s}_{it}$ for simplicity.

We first note that the left-hand side of the Bellman equation (1) depends on $\bar{K}_{it}$ and $\boldsymbol{\omega}_{it}$, and $\boldsymbol{\omega}_{it}$ depends on the observables $(\bar{K}_{it-1}, \bar{L}_{it-1}, \bar{M}_{it-1}, \{P_{ijt-1}, Q_{ijt-1}\}_{j=1}^{J})$ through $\boldsymbol{\omega}_{it-1}$. This shows that the observables $\mathcal{O}_{it+1}$ only depend on the past observables through $\mathcal{O}_{it}$, i.e. we have a Markov structure for $\mathcal{O}_{it}$. Since the empirical content of the dynamic are stated as the constrains on the distribution on observables, we only look at the data distribution of $\mathcal{O}_{it+1}$ conditional on $\mathcal{O}_{it}$. Let $G^0(\bar{K}_{it+1}, \bar{L}_{it}, \bar{M}_{it}, \{Q_{ijt}\}_{j=1}^{J}, R_{it}^{obs}|\mathcal{O}_{it})$ be the identified data distribution conditional on $\mathcal{O}_{it}$.

Definition A.1. (Policy Function) Given $\left(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}, C(\cdot)\right)$, we denote the policy functions solved from (1)-(4) by $L_{ijt}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$, $M_{ijt}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$, $K_{ijt}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$, and $\bar{K}_{it+1}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$. Under the optimal policy rule, we get the optimal quantity $Q_{ijt}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$, the corresponding prices P_{ijt}^* and the observed revenue $R_{it}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})$.

Definition A.2. The sharp identified set of the dynamic model, denoted by Θ_I^{EC}, is the collection of $\left(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}, C(\cdot)\right)$ such that we can find joint distributions of $H_t(\boldsymbol{\epsilon}_{it}, u_{it}, \mathcal{O}_{it}, \{P_{ijt}^*, Q_{ijt}^*(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}_{j=1}^{J})$ for $t = 1, 2, ...T$ so that the following holds:

1. $E[\boldsymbol{\epsilon}_{it}|\mathcal{O}_{it}] = 0$;

2. $E[u_{it}|\mathcal{O}_{it}, \{P^*_{ijt}, Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}_{j=1}^J] = 0$;

3. Under $H_t(\cdot)$, the distribution of

$$\left(\bar{K}^*_{it+1}(\bar{K}_{it}, \boldsymbol{\omega}_{it}), \sum_{j=1}^J L^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}), \sum_{j=1}^J M^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}), \{P^*_{ijt}, Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}_{j=1}^J, R^*_{it}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\right)$$

conditional on $\mathcal{O}_{it}$ is the same as $G^0(\cdot|\mathcal{O}_{it})$.

We then say that $H_t(\cdot)$ for $t = 1, 2, ...T$ rationalize $\big(\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}, C(\cdot)\big)$.

Conditions 1 and 2 in Definition A.2 impose the identification Assumption 1 in the main text. Condition 3 requires that the model predicted distribution of observables agree with the observed data distribution. We now generalize the location non-identification result to Θ_I^{EC}.

Proposition A.1. *Let* $\theta \equiv (\{\boldsymbol{\beta}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}, C(\cdot))$ *be in* Θ_I^{EC}*, then* $\breve{\theta} \equiv (\{\breve{\boldsymbol{\beta}}_j, \sigma_j, \{\delta_{jt}\}_{t\leq T}\}_{j\in\mathcal{J}}, \boldsymbol{\rho}, \mathcal{U}, C(\cdot))$ *is also in* Θ_I^{EC}*, 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 $\{C_j^X \big| j \in \mathcal{J},\ X \in \{K, L, M\}\}$ *satisfy*

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

Proof. The proof is separated into three steps. In each step, we first state goal and then provide the proof.

Step 1. Let $H_t^\theta(\boldsymbol{\epsilon}_{it}, u_{it}, \mathcal{O}_{it}, \{P^*_{ijt}, Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}_{j=1}^J)$, $t = 1, ..., T$ be the distributions in Definition A.2 that rationalize θ. Let $H_t^{\breve{\theta}}$ be a distribution such that $H_t^{\breve{\theta}} = H_t^\theta$ for all $t = 1, ..., T$ almost surely, and let $\omega_{it}^\theta = h(\boldsymbol{\omega}_{it-1}^\theta; \boldsymbol{\rho}) + \boldsymbol{\epsilon}_{it}^\theta$ and $\boldsymbol{\omega}_{it-1}^\theta = q_{ijt-1} - f(k^*_{ijt-1}, l^*_{ijt-1}, m^*_{ijt-1}; \boldsymbol{\beta}_j)$. *Goal*: $\boldsymbol{\omega}_{it}^\theta \equiv \boldsymbol{\omega}_{it}^{\breve{\theta}}$ almost surely.

Proof: By construction, $H_t^{\breve{\theta}} = H_t^\theta$ almost surely, so $\boldsymbol{\epsilon}_{it}^\theta = \boldsymbol{\epsilon}_{it}^{\breve{\theta}}$ almost surely. Moreover, by assumption, the evolution process is the same under θ and $\breve{\theta}$ (note that they only differ in the production function parameters), so it suffices to show that $f(k^*_{ijt-1}, l^*_{ijt-1}, m^*_{ijt-1}; \boldsymbol{\beta}_j) = f(\breve{k}^*_{ijt-1}, \breve{l}^*_{ijt-1}, \breve{m}^*_{ijt-1}; \breve{\boldsymbol{\beta}}_j)$. In the proof of Proposition 3 in the main text, we have shown that $\breve{X}^*_{ijt-1} = X^*_{ijt-1} - C_j^X$ for $X \in \{K, L, M\}$. So by construction of $\breve{\boldsymbol{\beta}}_j$, $f(k^*_{ijt-1}, l^*_{ijt-1}, m^*_{ijt-1}; \boldsymbol{\beta}_j) = f(\breve{k}^*_{ijt-1}, \breve{l}^*_{ijt-1}, \breve{m}^*_{ijt-1}; \breve{\boldsymbol{\beta}}_j)$ must hold.

Step 2. Fix $\boldsymbol{\omega}_{it}$. *Goal*: Suppose $\boldsymbol{\omega}^{\theta}_{it+1} = \boldsymbol{\omega}^{\breve{\theta}}_{it+1}$ almost surely, then the policy functions and quantity functions satisfy:

$$\begin{aligned} X^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) &= X^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) - C^X_j \quad \forall\ X \in \{K, L, M\}, \\ Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) &= Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta), \\ \bar{K}^*_{it+1}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) &= \bar{K}^*_{it+1}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta). \end{aligned} \tag{7}$$

Proof: Note that for $X \in \{K, L, M\}$, X_{ijt} are static choices that do not enter the value function in the future. Therefore, X_{ijt} satisfies the static optimization problem (5). As a result, the first two lines of (7) are direct corollary of the proof in Proposition 3 in the main text. It remains to show that the optimal investment is the same.

We denote the static revenue under θ and $\breve{\theta}$ by

$$\pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) \equiv \sum_{j=1}^{J} \left(\mathcal{U}P_d(\{Q^*_{ij't}(\theta)\}_{j\in\mathcal{J}})Q^*_{ijt}(\theta) - v_{it}M^*_{ijt}(\theta) - w_{it}L^*_{ijt}(\theta)\right),$$

$$\pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) \equiv \sum_{j=1}^{J} \left(\mathcal{U}P_d(\{Q^*_{ij't}(\breve{\theta})\}_{j'\in\mathcal{J}})Q^*_{ijt}(\breve{\theta}) - v_{it}M^*_{ijt}(\breve{\theta}) - w_{it}L^*_{ijt}(\breve{\theta})\right),$$

where we omit $(\bar{K}_{it}, \boldsymbol{\omega}_{it})$ in the expression of L^*_{ijt} and M^*_{ijt}. Using the same argument as in Step 1, we can show that $f(k^*_{ijt}, l^*_{ijt}, m^*_{ijt}; \boldsymbol{\beta}_j) = f(\breve{k}^*_{ijt}, \breve{l}^*_{ijt}, \breve{m}^*_{ijt}; \breve{\boldsymbol{\beta}}_j)$ and thus $\pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) = \pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta})$. Then the Bellman equations under θ and $\breve{\theta}$ become

$$V(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) = \max_{I_{it}} \pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) - C(I_{it}) + \beta E[V(\bar{K}_{it+1}, \boldsymbol{\omega}^{\theta}_{it+1}; \theta)], \tag{8}$$

$$V(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) = \max_{I_{it}} \pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta}) - C(I_{it}) + \beta E[V(\bar{K}_{it+1}, \boldsymbol{\omega}^{\breve{\theta}}_{it+1}; \breve{\theta})]. \tag{9}$$

Since $\pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta) = \pi(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \breve{\theta})$ and we impose $\boldsymbol{\omega}^{\theta}_{it+1} = \boldsymbol{\omega}^{\breve{\theta}}_{it+1}$ almost surely, (8) and (9) are the same dynamic programming problem. As a result, the investment level should be the same. This shows that if $\bar{K}^*_{it+1}(\bar{K}_{it}, \boldsymbol{\omega}_{it}; \theta)$ is the policy function under θ, it is also the policy function under $\breve{\theta}$.

Step 3. *Goal*: Let $H^{\theta}_t(\boldsymbol{\epsilon}_{it}, u_{it}, \mathcal{O}_{it}, \{P^*_{ijt}, Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}^J_{j=1})$, $t = 1, ..., T$ be the distributions in Definition A.2 that rationalize θ. Then $H^{\theta}_t(\boldsymbol{\epsilon}_{it}, u_{it}, \mathcal{O}_{it}, \{P^*_{ijt}, Q^*_{ijt}(\bar{K}_{it}, \boldsymbol{\omega}_{it})\}^J_{j=1})$ for $t = 1, 2, ..., T$ also rationalize $\breve{\theta}$.

Proof: By constructing a replication distribution $H^{\breve{\theta}}_t(\cdot)$ such that $H^{\breve{\theta}}_t(\cdot) = H^{\theta}_t(\cdot)$ almost surely, we have $\boldsymbol{\omega}^{\breve{\theta}}_{it} = \boldsymbol{\omega}^{\theta}_{it}$ almost surely by the argument in Step 1. Because $\boldsymbol{\omega}^{\breve{\theta}}_{it} = \boldsymbol{\omega}^{\theta}_{it}$ almost surely, we can use the result in Step 2. Since $\sum_{j=1}^{J} C^X_j = 0$, the following expression holds almost surely

$$
\begin{aligned}
&(\bar{K}^*_{it+1}(\theta), \sum_{j=1}^{J} L^*_{ijt}(\theta), \sum_{j=1}^{J} M^*_{ijt}(\theta), \{Q^*_{ijt}(\theta)\}_{j=1}^{J}) \\
&= (\bar{K}^*_{it+1}(\breve{\theta}), \sum_{j=1}^{J} L^*_{ijt}(\breve{\theta}), \sum_{j=1}^{J} M^*_{ijt}(\breve{\theta}), \{Q^*_{ijt}(\breve{\theta})\}_{j=1}^{J}),
\end{aligned}
$$

where we omit the $(\bar{K}_{it}, \boldsymbol{\omega}_{it})$ in the policy and quantity function. Then, since prices are functions of the quantities, the prices generated under two sets of parameters must also equal to each other. $R^{\theta}_{it} = P^*_{ijt}(\theta))Q^*_{ijt}(\theta)e^{u^{\theta}_{it}}$ and $u^{\theta}_{it} = u^{\breve{\theta}}_{it}$ almost surely by the construction of $H^{\breve{\theta}}_t$, we have $R^{\breve{\theta}}_{it} = R^{\theta}_{it}$ almost surely. Therefore, we have shown that the $\breve{\theta}$-model predicted data distribution is the same as the θ-model predicted data distribution. This means that if θ can be rationalized by H^{θ}_t, $\breve{\theta}$ can also be rationalized by H^{θ}_t. By Definition A.2, $\breve{\theta}$ is also in the identified set Θ^{EC}_I. $\square$

B Additional Proofs

B.1 Proof of Proposition 6

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:

$$
\begin{aligned}
M_n(\theta_{-\boldsymbol{\rho}}) &= \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}}))' A_n(\theta)^{-1} \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) \\
&= \underbrace{\mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))]' A_n(\theta)^{-1} \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))]}_{\text{Term 1}} \\
&- \underbrace{2(\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])' A_n(\theta)^{-1} \bar{m}_n(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^0(\theta_{-\boldsymbol{\rho}}))}_{\text{Term 2}} \\
&+ \underbrace{(\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])' A_n(\theta)^{-1} (\bar{m}_n(\theta_{-\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}(\theta_{-\boldsymbol{\rho}})) - \mathbf{E}[m_i(\theta_{-\boldsymbol{\rho}}, \boldsymbol{\rho}^*(\theta_{-\boldsymbol{\rho}}))])}_{\text{Term 3}}.
\end{aligned}
\tag{10}
$$

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 (10) equals zero, and by Assumption 8,

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

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

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

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 (10), 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

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

The result follows. $\square$

C Connection to Existing Methods

Existing methods differ in how to deal with the unobserved input allocations. Recall that we employ firms' optimization behavior to control the unobserved input allocations. In this section, we discuss the relationship between our approach and other methods.

MPP function approach: An Example

Dhyne et al. (2021) (DPSW hereafter) propose a multiproduct production (MPP) function approach to estimate the production function for multiproduct firms. They assume firm-product-year level productivity; They do not rely on single-product firms to identify the product-specific technology. They describe the production technology using a production possibilities set $\mathcal{T}$, and define the production function using the production possibilities frontier (PPF) as f_j^{MPP}:

$$\begin{aligned} q_{ijt} &= f_j^{MPP}(\boldsymbol{q}_{i,-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \\ &\equiv \max\{q_{ijt}|\ (q_{ijt}, \boldsymbol{q}_{i,-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \in T\}, \end{aligned} \tag{11}$$

where $\boldsymbol{q}_{i,-jt}$ stands for the vector of quantities of other products. DPSW assume f_j^{MPP} has a separately additive form such that

$$\begin{aligned} q_{ijt} &= f_j^{MPP}(\boldsymbol{q}_{i,-jt}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) \\ &= \tilde{f}_j^{MPP}(\bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it}) + g(\boldsymbol{q}_{i,-jt}), \end{aligned} \tag{12}$$

where $\tilde{f}_j^{MPP}$ is only related to firm-level inputs and $g(\cdot)$ is a function of other products' quantities. However, this high-level assumption may be inconsistent with the product-specific technology. In the following, we will show that the production function cannot be expressed in such an additive form when the firm-product production functions are the Cobb-Douglas. We first show that our method can nest the production possibilities approach. Given logged production functions $\{f_j\}_{j\in\mathcal{J}}$, the production possibilities set $\mathcal{T}$ is defined as

$$\mathcal{T} \equiv \left\{ (\{q_{ijt}\}_{j\in\mathcal{J}}, \bar{K}_{it}, \bar{L}_{it}, \bar{M}_{it})|,\, q_{ijt} = f_j(K_{ijt}, L_{ijt}, M_{ijt}),\ \sum_j X_{ijt} = \bar{X}_{it},\ X \in \{K, L, M\} \right\}.$$

From $\mathcal{T}$ we can similarly define the DPSW-type PPF using quantities of other products. Compared with DPSW, our method imposes assumptions directly on the product-specific production functions. In what follows, we use an example to show that the separately additive assumption of f_j^{MPP} fails when the production possibilities set $\mathcal{T}$ is constructed under the Cobb-Douglas production function.

An Example We consider a simple case where $J = 2$, and the only input is capital: $F_1(K_{i1t}) = K_{i1t}^{\beta}$, $F_2(K_{i2t}) = K_{i2t}^{\beta}$. By definition we want to find

$$q_{i1t} = f_1^{MPP}(\tilde{q}_{i2t}, \bar{K}_{it}) = \max_{\tilde{q}_{i1t}}\{\tilde{q}_{i1t}|\ (\tilde{q}_{i1t},\ \tilde{q}_{i2t}, \bar{K}_{it}) \in \mathcal{T}\}. \tag{13}$$

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

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

The first-order condition implies $\beta(\bar{K}_{it} - K_{i2t})^{\beta-1} = \lambda\beta K_{i2t}^{\beta-1}$, where λ is the Lagrangian multiplier for the constraint $Q_{i2t} = K_{i2t}^{\beta}$. We can solve for the optimal input of the second product as: $K^*_{i2t} = \frac{\bar{K}_{it}}{1+\lambda^{1/(\beta-1)}}$. The constraint $Q_{i2t} = K^{\beta}_{i2t}$ implies λ must satisfy

$$Q_{i2t} = \left(\frac{\bar{K}_{it}}{1+\lambda^{1/(\beta-1)}} \right)^{\beta}. \tag{14}$$

We can construct f_1^{MPP} by plugging in K^*_{i2t} into the production function of product 1:

$$\begin{aligned} f_1^{MPP} &= \beta \log(\bar{K}_{it} - K^*_{i2t}) = \beta \log \bar{K}_{it} + \frac{\beta}{1-\beta} \log \lambda - \beta \log \left(1 + \lambda^{1/(\beta-1)}\right) \\ &= q_{i2t} + \beta \log \left(\frac{\bar{K}_{it}}{Q_{i2t}^{1/\beta}} - 1 \right), \end{aligned}$$

where the last equality holds by (14). Despite of the simple Cobb-Douglas production function, the MPP function turns out to be non-separably additive between total inputs $\bar{K}_{it}$ and the logged quantity of the second product q_{i2t}. This is contradictive with the assumption in Dhyne et al. (2021). The main difficulty in estimating (11) is to disentangle the vector of output $\boldsymbol{q}_{i,-jt}$ from inputs and unobserved production efficiency. Assuming that $f_j^{MPP}(\cdot)$ is separately additive in terms of $\boldsymbol{q}_{i,-jt}$ and all inputs leads to a potential problem of mis-specification.

D Computing the Optimal Allocations for the CES Production Function

D.1 Proof of Contraction Mapping

We first note that the mapping defined in Section 4 always has a fixed point. This is because the fixed point must be the unique solution to the firm's optimization problem, and by Proposition 1, such a solution always exists. Throughout this section, we assume the Dixit-Stiglitz demand: $P_{ijt} = \bar{P}\bar{Q}^{1/\sigma_j} Q_{ijt}^{-1/\sigma_j}$, and we use the following notation $\tilde{\gamma}^X_{ijt} \equiv \frac{\sigma_j - 1}{\sigma_j} \gamma^X_{ijt}$.

To show that the $\mathbf{T}_{it}$ is a contraction mapping, we first prove the following Lemma.

Lemma D.1. *For any $j, j' = 1, 2, \cdots, J$ and $X, X' \in \{K, L, M\}$, then* $\left| \frac{\partial T_{it}(X,j)}{\partial x'_{ij't}} \right| < \left| \frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x'_{ij't}} \right|$.

Proof. 1. Diagonal terms in $\nabla \mathbf{T}_{it}$. For element (X, j):

$$\begin{aligned}\frac{\partial T_{it}(X,j)}{\partial x_{ijt}} &= \frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x_{ijt}} - \frac{Q^{1-1/\sigma}_{ijt}\tilde{\gamma}^X_{ijt}}{\sum_{j'=1}^J Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}\frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x_{ijt}} \\ &= \frac{\sum_{j'\neq j}^J Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{j'=1}^J Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}\frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x_{ijt}}\end{aligned}$$

Note that the first term $\frac{\sum_{j'\neq j}^J Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{j'=1}^J Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}} < 1$, therefore $|\frac{\partial T_{it}(X,j)}{\partial x_{ijt}}| < |\frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x_{ijt}}|$.

2. Off-diagonal terms in $\nabla \mathbf{T}_{it}$. Note that for any $X, X' \in \{K, L, M\}$, $\partial \ln \tilde{\gamma}^X_{ijt}/\partial x'_{ij't} = 0$ for $j \neq j'$:

   (a) For input X and different products $j \neq j'$:

$$\begin{aligned}\frac{\partial T_{it}(X,j)}{\partial x_{ij't}} &= \frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x_{ij't}} - \frac{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\frac{\partial \tilde{\gamma}^X_{i\tilde{j}t}}{\partial x_{ij't}}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}} \\ &= -\frac{Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}\frac{\partial \ln \tilde{\gamma}^X_{ij't}}{\partial x_{ij't}}\end{aligned}$$

   Note that $0 < \frac{Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}} < 1$, therefore $|\frac{\partial T_{it}(X,j)}{\partial x_{ij't}}| < |\frac{\partial \ln \tilde{\gamma}^X_{ij't}}{\partial x_{ij't}}|$.

   (b) For distinct inputs $X \neq X'$ but the same product j:

$$\begin{aligned}\frac{\partial T_{it}(X,j)}{\partial x'_{ijt}} &= \frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x'_{ijt}} - \frac{Q^{1-1/\sigma}_{ijt}\tilde{\gamma}^X_{ijt}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}\frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x'_{ijt}} \\ &= \frac{\sum_{\tilde{j}\neq j} Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}\frac{\partial \ln \tilde{\gamma}^X_{ijt}}{\partial x'_{ijt}}\end{aligned}$$

   Note that $0 < \frac{\sum_{\tilde{j}\neq j} Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}} < 1$, therefore $|\frac{\partial T_{it}(X,j)}{\partial x'_{ijt}}| < |\frac{\partial \ln \tilde{\gamma}^X_{ij't}}{\partial x_{ij't}}|$.

   (c) For distinct inputs $X \neq X'$ and distinct products $j \neq j'$:

$$\frac{\partial T_{it}(X,j)}{\partial x'_{ij't}} = -\frac{Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}}\frac{\partial \ln \tilde{\gamma}^X_{ij't}}{\partial x'_{ijt}}$$

   Note that $0 < \frac{Q^{1-1/\sigma}_{ij't}\tilde{\gamma}^X_{ij't}}{\sum_{\tilde{j}=1}^J Q^{1-1/\sigma}_{i\tilde{j}t}\tilde{\gamma}^X_{i\tilde{j}t}} < 1$, therefore $|\frac{\partial T_{it}(X,j)}{\partial x'_{ij't}}| < |\frac{\partial \ln \tilde{\gamma}^X_{ij't}}{\partial x'_{ijt}}|$.

$\square$

Proposition D.1. *When* $|\frac{\theta_j-1}{\theta_j}| < \frac{1}{2}$, *the mapping* $\mathbf{T}_{it}$ *defined for CES production function is a contraction mapping.*

Proof. To show that $\mathbf{T}_{it}$ is a contraction mapping, by Theorem 1 in Berry et al. (1995), it suffices to show that $\|\nabla \mathbf{T}_{it}\|_r < 1$, where $\nabla \mathbf{T}_{it}$ is the Jacobian of $\mathbf{T}_{it}$ with respect to X_{ijt}'s and $\|\cdot\|_r$ is the max absolute row sum: $||M||_r = \max_i \sum_{j=1}^J |M_{ij}|$.

The output elasticity of input X is:

$$\gamma_{ijt}^X = \frac{\upsilon_j \beta_j^X X_{ijt}^{\frac{\theta_j-1}{\theta_j}}}{\beta_j^K K_{ijt}^{\frac{\theta_j-1}{\theta_j}} + \beta_j^L L_{ijt}^{\frac{\theta_j-1}{\theta_j}} + \beta_j^M M_{ijt}^{\frac{\theta_j-1}{\theta_j}}}, \quad \text{for } X \in \{K, L, M\}.$$

Note that γ_{ijt}^X is homogeneous of degree zero with respect to $(K_{ij't}, L_{ij't}, M_{ij't})$ for $j = 1, \cdots, J$, so is $\ln \gamma_{ijt}^X$. The Euler Theorem implies that

$$\frac{\partial \ln \gamma_{ijt}^X}{\partial k_{ij't}} + \frac{\partial \ln \gamma_{ijt}^X}{\partial l_{ij't}} + \frac{\partial \ln \gamma_{ijt}^X}{\partial m_{ij't}} = 0, \quad for \quad j' = 1, \cdots, J.$$

For any three real number a, b, c, if $a + b + c = 0$, then $|a| + |b| + |c| = 2\max\{|a|, |b|, |c|\}$. It follows from Lemma D.1 that

$$\begin{aligned}
\left|\frac{\partial T_{it}(X,j)}{\partial k_{ij't}}\right| + \left|\frac{\partial T_{it}(X,j)}{\partial l_{ij't}}\right| + \left|\frac{\partial T_{it}(X,j)}{\partial m_{ij't}}\right| &\leq \left|\frac{\partial \ln \tilde{\gamma}_{ijt}^X}{\partial k_{ij't}}\right| + \left|\frac{\partial \ln \tilde{\gamma}_{ijt}^X}{\partial l_{ij't}}\right| + \left|\frac{\partial \ln \tilde{\gamma}_{ijt}^X}{\partial m_{ij't}}\right| \\
&= \left|\frac{\partial \ln \gamma_{ijt}^X}{\partial k_{ij't}}\right| + \left|\frac{\partial \ln \gamma_{ijt}^X}{\partial l_{ij't}}\right| + \left|\frac{\partial \ln \gamma_{ijt}^X}{\partial m_{ij't}}\right| \\
&\leq 2 \max_{X' \in \{K,L,M\}\}} \left|\frac{\partial \ln \gamma_{ijt}^X}{\partial x'_{ij't}}\right|.
\end{aligned}$$

Note that:

$$\begin{aligned}
\frac{\partial \ln \gamma_{ijt}^X}{\partial x_{ijt}} &= \frac{X_{ijt}}{\gamma_{ijt}^X} \frac{\nu_j \frac{\theta_j - 1}{\theta_j} \beta_j^X X_{ijt}^{-1/\theta_j} \left(\sum_{X' \neq X} \beta_j^{X'} X_{ijt}^{'(\theta_j-1)/\theta_j}\right)}{\beta_j^K K_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^L L_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^M M_{ijt}^{(\theta_j-1)/\theta_j}} \\
&= \frac{\theta_j - 1}{\theta_j} \frac{\sum_{X' \neq X} \beta_j^{X'} X_{ijt}^{'\frac{\theta_j-1}{\theta_j}}}{\beta_j^K K_{ijt}^{\frac{\theta_j-1}{\theta_j}} + \beta_j^L L_{ijt}^{\frac{\theta_j-1}{\theta_j}} + \beta_j^M M_{ijt}^{\frac{\theta_j-1}{\theta_j}}}, \\
\Rightarrow \quad \left|\frac{\partial \ln \gamma_{ijt}^X}{\partial x_{ijt}}\right| &< \left|\frac{\theta_j - 1}{\theta_j}\right|,
\end{aligned}$$

and $\frac{\partial \ln \gamma^X_{ijt}}{\partial x_{ijt}} = 0$ for $j' \neq j$. Similarly, for $X' \neq X$:

$$\Rightarrow \quad \left|\frac{\partial \ln \gamma^X_{ijt}}{\partial x'_{ijt}}\right| = \left|\frac{\theta_j - 1}{\theta_j}\right| \frac{\beta^{X'}_j X'^{\frac{\theta_j-1}{\theta_j}}_{ijt}}{\beta^K_j K^{\frac{\theta_j-1}{\theta_j}}_{ijt} + \beta^L_j L^{\frac{\theta_j-1}{\theta_j}}_{ijt} + \beta^M_j M^{\frac{\theta_j-1}{\theta_j}}_{ijt}} < \left|\frac{\theta_j - 1}{\theta_j}\right|$$

$$\frac{\partial \ln \gamma^X_{ijt}}{\partial x'_{ij't}} = 0 \quad \text{for } j' \neq j.$$

For row (X, j), the sum of this row is

$$\begin{aligned}
\sum_{X' \in \{K,L,M\}, j' \in \mathcal{J}} \left|\frac{\partial T_{it}(X,j)}{\partial x'_{ij't}}\right| &\leq \sum_{X' \in \{K,L,M\}, j' \in \mathcal{J}} \left|\frac{\partial \ln \gamma^X_{ijt}}{\partial x'_{ij't}}\right| \\
&\leq 2 \max_{X' \in \{K,L,M\}} \left|\frac{\partial \gamma^X_{ijt}}{\partial x'_{ijt}}\right| \\
&< 2 \left|\frac{\theta_j - 1}{\theta_j}\right|.
\end{aligned}$$

This implies that $\|\nabla \mathbf{T}_{it}\|_r < 1$ if $|\frac{\theta_j - 1}{\theta_j}| < 1/2$. Condition (1) of Theorem 1 in Berry et al. (1995) is satisfied. The result follows. $\square$

We do not need to check conditions (2) and (3) of Theorem 1 in Berry et al. (1995) which are required to ensure the existence of a unique fixed point. The uniqueness of the fixed point is guaranteed by Proposition 1.

Proposition D.2. *When $|\frac{\theta_j - 1}{\theta_j}| < \frac{1}{2}$, any Cauchy sequence $\{(\{x^n_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})\}_{n=1,2,\ldots}$ generated by $\mathbf{T}_{it}$ converge to the unique solution $(\{x^*_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})$ characterized by Proposition 1.*

Proof. We consider the complete space $\mathcal{D} \equiv (-\infty, \bar{k}_{it}]^J \times (-\infty, \bar{l}_{it}]^J \times (-\infty, \bar{m}_{it}]^J$. Any Cauchy sequence $\{(\{x^n_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})\}_{n=1,2,\ldots}$ generated by $\mathbf{T}_{it}$ must converges in $\mathcal{D}$.

We first show that the limit cannot lie at the boundary of $\mathcal{D}$. Suppose not, let $(\{\tilde{x}_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})$ be the limit that lies on the boudary. For some $x \in \{k, l, m\}$ and some $j \in \mathcal{J}$, $\tilde{x}_{ijt} = \bar{x}_{it}$ must hold. For this x, $\sum_{j\in\mathcal{J}} e^{\tilde{X}_{ijt}} > \bar{X}_{it}$ hold. However, along the sequence $\{(\{x^n_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})\}_{n=1,2,\ldots}$ generated by $\mathbf{T}_{it}$, we know that $\sum_{j\in\mathcal{J}} \exp[x^n_{ijt}] = \bar{X}_{it}$ for all $n \geq 2$. This generates a contradiction.

Therefore this sequence converges in $\mathcal{D}^{int}$, and the limit must be $(\{x^*_{ijt}\}_{x\in\{k,l,m\}, j\in\mathcal{J}})$ by the uniqueness of the solution to the optimization problem of input allocations. $\square$

D.2 Jacobian Matrix of Newton's Method

We first substitute $X_{iJt} = \bar{X}_{it} - \sum_{j=1}^{J-1} X_{ijt}$ and find the partial derivatives:

$$\frac{\partial \gamma_{ijt}^X}{\partial X_{ijt}} = \frac{\nu_j \frac{\theta_j - 1}{\theta_j} \beta_j^X X_{ijt}^{-1/\theta_j} \left(\sum_{X' \neq X} \beta_j^{X'} {X'}_{ijt}^{(\theta_j-1)/\theta_j} \right)}{\left(\beta_j^K K_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^L L_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^M M_{ijt}^{(\theta_j-1)/\theta_j} \right)^2} \tag{15}$$

$$= \frac{\theta_j - 1}{X_{ijt}\theta_j} \gamma_{ijt}^X (1 - \gamma_{ijt}^X / \nu_j), \tag{16}$$

$$\tag{17}$$

and for $X' \neq X$:

$$\frac{\partial \gamma_{ijt}^X}{\partial X'_{ijt}} = -\frac{\nu_j \frac{\theta_j - 1}{\theta_j} \beta_j^X X_{ijt}^{1-1/\theta_j} \beta_j^{X'} {X'}_{ijt}^{-1/\theta_j}}{\left(\beta_j^K K_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^L L_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^M M_{ijt}^{(\theta_j-1)/\theta_j} \right)^2}$$
$$= -\frac{\gamma_{ijt}^X(\theta_j - 1)}{\theta_j X'_{ijt}} \frac{\beta_j^{X'} {X'}_{ijt}^{1-1/\theta_j}}{\beta_j^K K_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^L L_{ijt}^{(\theta_j-1)/\theta_j} + \beta_j^M M_{ijt}^{(\theta_j-1)/\theta_j}},$$
$$\frac{\partial \gamma_{ij't}^X}{\partial X'_{ijt}} = 0 \quad \text{for } j' \notin \{j, J\},$$
$$\frac{\partial \gamma_{iJt}^X}{\partial X'_{ijt}} = \frac{\nu_J \frac{\theta_J - 1}{\theta_J} \beta_J^X X_{iJt}^{1-1/\theta_J} \beta_J^{X'} {X'}_{iJt}^{-1/\theta_J}}{\left(\beta_J^K K_{iJt}^{(\theta_J-1)/\theta_J} + \beta_J^L L_{iJt}^{(\theta_J-1)/\theta_J} + \beta_J^M M_{iJt}^{(\theta_J-1)/\theta_J} \right)^2}$$
$$= \frac{\gamma_{iJt}^X(\theta_J - 1)}{\theta_J X'_{iJt}} \frac{\beta_J^{X'} {X'}_{iJt}^{1-1/\theta_J}}{\beta_J^K K_{iJt}^{(\theta_J-1)/\theta_J} + \beta_J^L L_{iJt}^{(\theta_J-1)/\theta_J} + \beta_J^M M_{iJt}^{(\theta_J-1)/\theta_J}}.$$

We can then show:

$$\frac{\partial H_{it}(X,j)}{X_{ijt}} = \bar{X}_{it} \frac{Q_{ijt}^{1-1/\sigma_j} \frac{\partial \tilde{\gamma}_{ijt}^X}{\partial X_{ijt}} \left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X \right) - \left(Q_{ijt}^{1-1/\sigma_j} \frac{\partial \tilde{\gamma}_{ijt}^X}{\partial X_{ijt}} + Q_{iJt}^{1-1/\sigma_J} \frac{\partial \tilde{\gamma}_{iJt}^X}{\partial X_{ijt}} \right) Q_{ijt}^{1-1/\sigma_j} \tilde{\gamma}_{ijt}^X}{\left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X \right)^2} - 1. \tag{18}$$

Similarly, we can show for $\tilde{j} \notin \{j, J\}$:

$$\frac{\partial H_{it}(X,j)}{X_{i\tilde{j}t}} = -\bar{X}_{it} \frac{\left(Q_{i\tilde{j}t}^{1-1/\sigma_{\tilde{j}}} \frac{\partial \tilde{\gamma}_{i\tilde{j}t}^X}{\partial X_{i\tilde{j}t}} + Q_{iJt}^{1-1/\sigma_J} \frac{\partial \tilde{\gamma}_{iJt}^X}{\partial X_{i\tilde{j}t}} \right) Q_{ijt}^{1-1/\sigma_j} \tilde{\gamma}_{ijt}^X}{\left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X \right)^2}. \tag{19}$$

And for $X' \neq X$, we have

$$\frac{\partial H_{it}(X,j)}{X'_{ijt}} = \bar{X}_{it} \frac{Q_{ijt}^{1-1/\sigma_j} \frac{\partial \tilde{\gamma}_{ijt}^X}{\partial X'_{ijt}} \left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X\right) - \left(Q_{ijt}^{1-1/\sigma_j} \frac{\partial \tilde{\gamma}_{ijt}^X}{\partial X'_{ijt}} + Q_{iJt}^{1-1/\sigma} \frac{\partial \tilde{\gamma}_{iJt}^X}{\partial X'_{ijt}}\right) Q_{ijt}^{1-1/\sigma_j} \tilde{\gamma}_{ijt}^X}{\left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X\right)^2}. \tag{20}$$

Similarly, we can show for $\tilde{j} \neq j$:

$$\frac{\partial H_{it}(X,j)}{X'_{i\tilde{j}t}} = -\bar{X}_{it} \frac{\left(Q_{i\tilde{j}t}^{1-1/\sigma_{\tilde{j}}} \frac{\partial \tilde{\gamma}_{i\tilde{j}t}^X}{\partial X'_{i\tilde{j}t}} + Q_{iJt}^{1-1/\sigma_J} \frac{\partial \tilde{\gamma}_{iJt}^X}{\partial X'_{i\tilde{j}t}}\right) Q_{ijt}^{1-1/\sigma_j} \tilde{\gamma}_{ijt}^X}{\left(\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{j'}} \tilde{\gamma}_{ij't}^X\right)^2}. \tag{21}$$

D.3 Algorithm

We propose the following algorithm to solve for the optimal allocation rules:

1. Set a tolerance level and max iteration number. Start with an initial guess of the allocation rules $\log \mathbf{X}^{(0)} \equiv (k_{ijt}^{(0)}, l_{ijt}^{(0)}, m_{ijt}^{(0)})_{j=1}^J$.

2. Iterate over $\mathbf{T}_{it}$: $\log X_{ijt}^{(n+1)} = T_{it}(X^{(n)}, j)$ and define $\log \mathbf{X}^{(n+1)} = (k_{ijt}^{(n+1)}, l_{ijt}^{(n+1)}, m_{ijt}^{(n+1)})_{j=1}^J$.

3. Check $||\log \mathbf{X}^{(n+1)} - \log \mathbf{X}^{(n)}||$. Stop iteration if the difference is less than the tolerance level or if the iteration count exceeds the max iteration number.

4. If the difference does not fall below the tolerance level within finite steps, or if the solver generates an allocation quantity near 0, we proceed to solve the following nonlinear system with Newton's method:

$$H(X,j) \equiv \frac{Q_{ijt}^{1-1/\sigma_j} \tilde{\gamma}_{ijt}^X}{\sum_{j'=1}^J Q_{ij't}^{1-1/\sigma_{\tilde{j}}} \tilde{\gamma}_{ij't}^X} \bar{X}_{it} - X_{ijt} = 0 \quad \forall j = 1, .., J. \tag{22}$$

In practice, we can stack all firms and run the iteration mapping together. The iteration algorithm runs fast and works well for most firms. We then pick out firms where the iteration does not generate a converging sequence and run Newton's method on these firms. For the CES production functions, we employ the Jacobian matrix for (22) in the previous subsection.

E Extensions

E.1 Only Observing Firm-Product-Level Revenues

Assumption 5 requires the econometrician to observe the total revenues and product-specific quantities. In some datasets, researchers can observe firm-product-level revenues. Though consistent with our empirical setting, one may wonder how important this assumption is to our methods. In this subsection, we analyze the situation when the researcher can only observe firm-product-level revenues.[1] To fix ideas, we maintain the assumption of the Dixit-Stiglitz demand. We show that in this case, we can estimate the revenue elasticities and revenue productivity. We use the Cobb-Douglas function as an illustration. For the general parametric production function, the logic is similar.

First, note that we cannot employ the moment condition (8) for revenue-quantity relationship to identify the demand parameter σ. Instead, we focus on the revenue elasticities with respect to inputs: $\tilde{\beta}_j^X = \frac{\sigma_j - 1}{\sigma_j}\beta_j^X$, for $X \in \{K, L, M\}$. The moment condition (20) for the input shares now help us identify $\{\mathcal{U}\tilde{\beta}_j^M\}_{j\in\mathcal{J}}$ and $\{\mathcal{U}\tilde{\beta}_j^L\}_{j\in\mathcal{J}}$. Then we can still use moment condition (13) to separate $\mathcal{U}$ from $\tilde{\beta}_j^M$ and $\tilde{\beta}_j^L$.

To employ the productivity evolution equation, we first define the *revenue productivity* to be $\tilde{\omega}_{ijt} \equiv \frac{\sigma_j - 1}{\sigma_j}\omega_{ijt} + u_{it}$ We then employ the product-specific revenues to solve the revenue produtivity as:

$$\tilde{\omega}_{ijt} = r_{ijt} - \frac{\sigma_j - 1}{\sigma_j} f\left(k_{ijt},\, l_{ijt},\, m_{ijt}; \boldsymbol{\beta}_j\right) - p_t - \frac{1}{\sigma_j} q_t - \delta_t. \tag{23}$$

We impose the following assumption on the evolution of the vector of revenue productivity:

Assumption 1. *The revenue productivity vector $\tilde{\boldsymbol{\omega}}_{it} \equiv (\tilde{\omega}_{ijt})_{j\in\mathcal{J}}$ evolution process is Markovian:*

$$\tilde{\boldsymbol{\omega}}_{it} = \tilde{h}(\tilde{\boldsymbol{\omega}}_{it-1};\, \tilde{\boldsymbol{\rho}}) + \tilde{\boldsymbol{\epsilon}}_{it} \tag{24}$$

This assumption is similar to that in Peters et al. (2017). Based on Assumption 1, the

[1] When the researcher observes the firm-product-level quantities and revenues, our method applies naturally.

moment condition for the error term $\tilde{\epsilon}_{ijt}$ is

$$\mathbf{E}(\tilde{\epsilon}_{ijt}|\mathcal{I}_{it}) = \mathbf{E}\Big\{ r_{ijt} - \sum_{X\in\{K,L,M\}} \tilde{\beta}_j^X \log(\alpha_{ijt}^X \bar{X}_{it}) - u_t \tag{25}$$

$$- h_j \underbrace{[(r_{ij't-1} - \sum_{X\in\{K,L,M\}} \tilde{\beta}_{j'}^X \log(\alpha_{ij't-1}^X \bar{X}_{it-1}) - u_{t-1})_{j'\in\mathcal{J}}]}_{\tilde{\boldsymbol{\omega}}_{it-1}} \big| \mathcal{I}_{it} \Big\} = 0,$$

where $u_t = p_t + \frac{1}{\sigma_j} q_t + \delta_t$ is the fixed effect for the aggregate demand factors. Note that we can still use the observed revenue data and parameters to compute the input allocations $\{\alpha_{ijt}^X\}$, $X \in \{K, L, M\}$:

$$\alpha_{ijt}^X \equiv \frac{X_{ijt}^*}{\bar{X}_{it}} = \frac{\tilde{\beta}_j^X R_{ijt}}{\sum_{j'\in\mathcal{J}} \tilde{\beta}_{j'}^X R_{ijt}}, \quad \text{for} \quad X \in \{K, L, M\}.$$

Similar to the case of observing product-level quantities, we can show that moment condition (25) identifies the revenue elasticity of capital $\{\tilde{\beta}_j^K\}_{j\in\mathcal{J}}$. In this case, one can still use the GMM estimator to back out the scaled production function. The only difference is that we cannot separate the physical productivity $\boldsymbol{\omega}_{it}$ from the demand shocks u_{it}.

For general parametric production functions, we can also use the allocation rule derived in Proposition 1 to compute the input allocations. Therefore we can still employ the input-to-revenue share equations for labor and materials, as well as the productivity evolution equation based on revenues similar to (25). However, we face a similar problem that the input share's equation may not deliver identification to any production function parameters. We can still use the partial identification strategy to obtain the identified set and confidence region for the parameters for the revenue functions.

F Implementation of Partial Identification

F.1 OLS Regression to Obtain $\boldsymbol{\rho}$'s in the Empirical Application

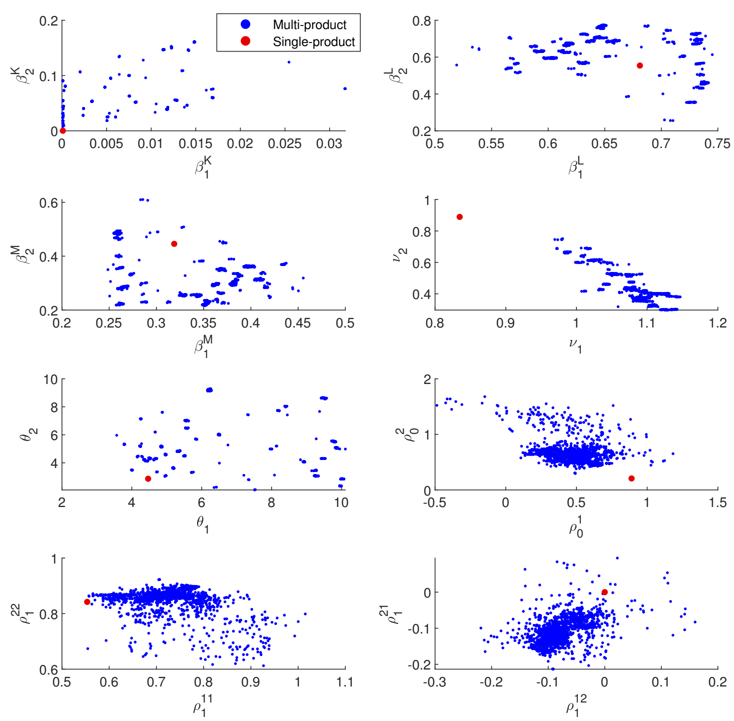
In the implementation of the dimension reduction method, we did not use an efficient weighting matrix to estimate $\boldsymbol{\rho}$ because we want to maintain a closed-form solution. When the sample size is small, the linear GMM estimation result can be irregular. Because ω_{ijt-1} can be written as a function of production function parameters and

$$(\bar{k}_{it-1}, \bar{l}_{it-1}, \bar{m}_{it-1}, \{Q_{ijt-1}\}_{j\in\mathcal{J}}, \omega_{ijt-1}) \in \mathcal{I}_{it},$$

we can also use OLS regression to recover $\boldsymbol{\rho}$ by regressing $\boldsymbol{\omega}_{it}$ on $\boldsymbol{\omega}_{it-1}$.

We report the the estimated identified set in Figure 1 and the confidence region in Figure 2. Compared to the results obtained by recovering $\boldsymbol{\rho}$ using GMM method, the estimated identified set by estimating $\boldsymbol{\rho}$ using OLS is wider. The patterns of the identified set and confidence region are similar to what we see in the main text. We see that there are points in the identified set that are far away from the single-product firms' parameter values. For θ and ν, the single-product firms' parameter values do not lie in the convex hull of the confidence region.

Figure 1: Identified Set Using OLS Regression for ρ's

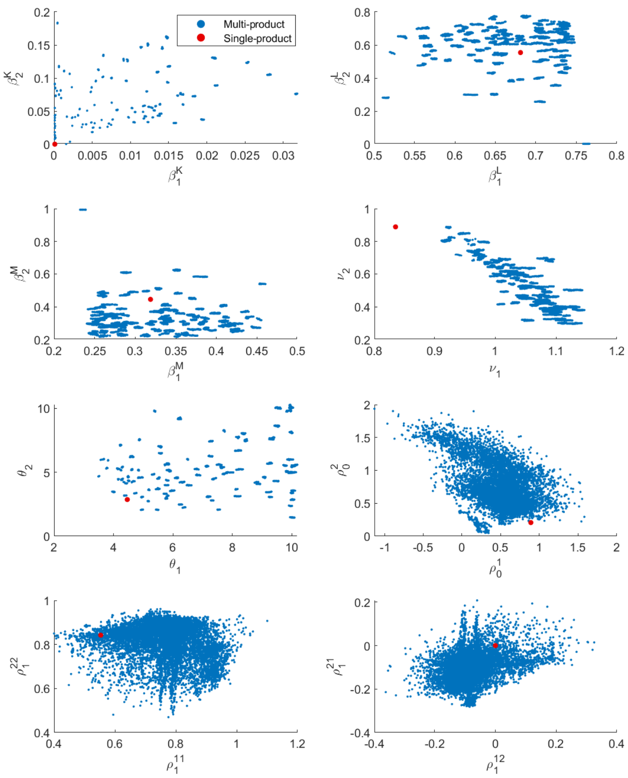


F.2 Sensitivity Analysis of Identified Sets and Confidence Regions

F.2.1 Monte Carlo Study

We now present the estimated identified set and the confidence region when setting different values for c in step 4 of the dimension-reduction algorithm. In Figure 3, we plot the estimated identified set for the simulation setting ($N = 1500$). Since the identified

Figure 2: Confidence Region Using OLS Regression for ρ's



set is larger when c increases, we overlay the points in the identified set with different c's. For example, the estimated identified set when $c = 3$ is the collection of all green, red and blue points in 3. We see that the estimated identified set is not very sensitive to the choice of the constant c.

F.2.2 Empirical Analysis

For the empirical application ($N = 300$), the estimated identified set (Figure 4) is more sensitive to change of c when c is small. We see that when we change from $c = 1/\log(8)$ (the constant chosen in the main text) to $c = 1$, the identified set becomes wider. When c increases from 1 to 3, the estimated identified sets become much wider. On the other hand, the confidence region (Figure 5) is not sensitive to the choice of c. However, we note that the convex hull of the confidence region still does not cover single-product firms' ν and θ. We, therefore, conclude that there is strong evidence that the production functions of single- and multiproduct firms are different.

Figure 3: Identified Set of the CES Production Function in the MC Simulation

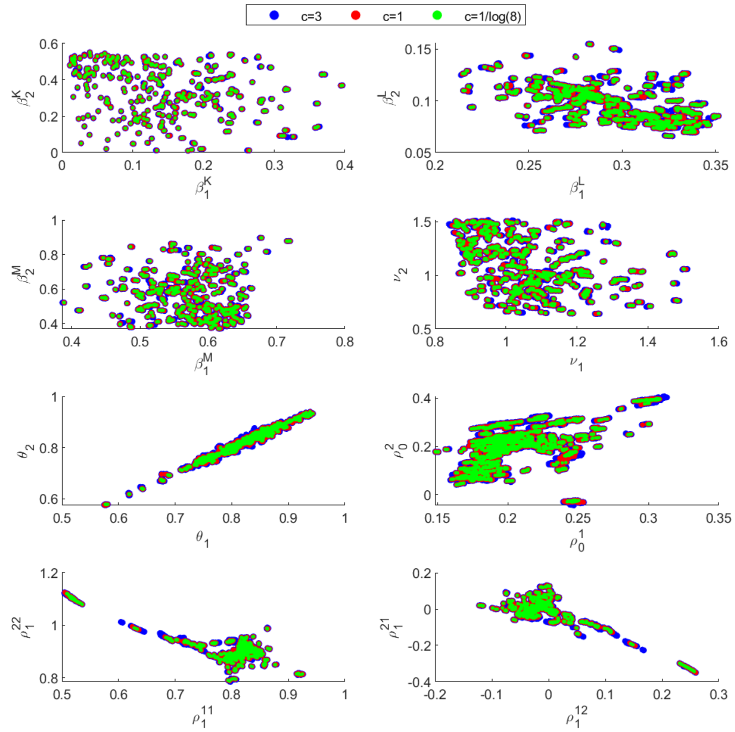


Figure 4: Identified Set of the CES Production Function

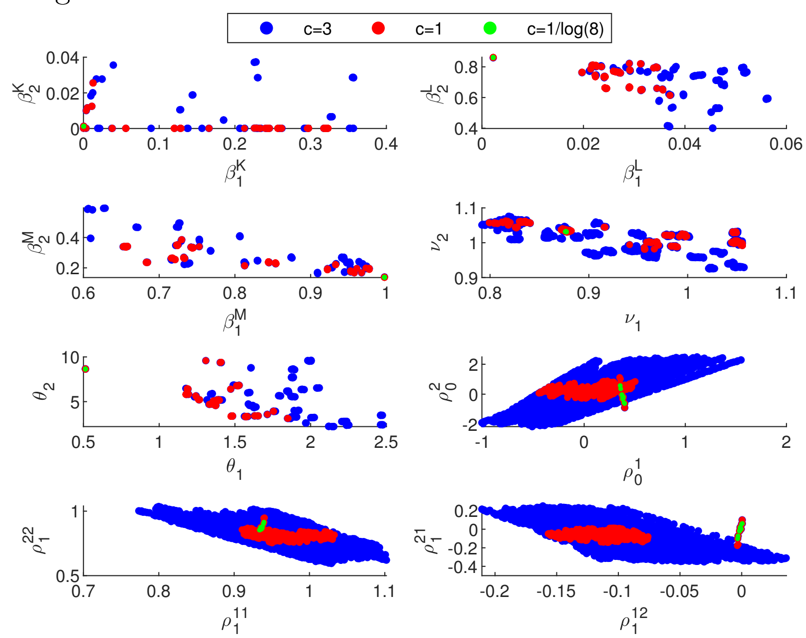
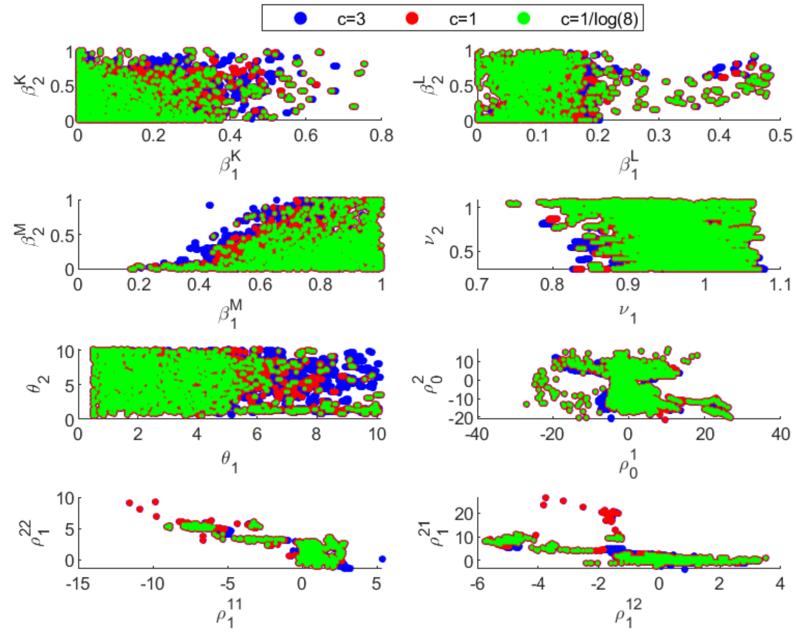


Figure 5: Confidence Region of the CES Production Function



G 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

$$\tilde{E}^K_{jt}(\bar{k}_{it}, \bar{m}_{it})/\tilde{E}^K_{j't}(\bar{k}_{it}, \bar{m}_{it}) \neq \tilde{E}^K_{jt}(\bar{k}'_{it}, \bar{m}_{it})/\tilde{E}^K_{j't}(\bar{k}'_{it}, \bar{m}_{it}), \tag{26}$$

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 (26), we consider a second-order polynomial:

$$\tilde{\Xi}_j(\bar{k}_{it}, \bar{m}_{it}) = \tau_{0j} + \tau_{1j}\bar{k}_{it} + \tau_{2j}\bar{k}^2_{it} + \tau_{3j}\bar{k}_{it}\bar{m}_{it} + \tau_{4j}\bar{m}_{it} + \tau_{5j}\bar{m}^2_{it}, \tag{27}$$

and hence

$$\frac{\tilde{E}^K_{jt}(\bar{k}_{it})}{\tilde{E}^K_{j't}(\bar{k}_{it})} = \frac{\tau_{1j} + 2\tau_{2j}\bar{k}_{it} + \tau_{3j}\bar{m}_{it}}{\tau_{1j'} + 2\tau_{2j'}\bar{k}_{it} + \tau_{3j'}\bar{m}_{it}}. \tag{28}$$

The ratio (28) 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 (27) 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 1.

Table 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 4 may not hold. However, we note that the condition in Proposition 4 is only sufficient but not necessary. Identification may still hold when the condition in Proposition 4 fails.

References

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*, 63, 841–890.

Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski (2021), "Theory for extending single-product production function estimation to multi-product settings." Unpublished manuscript.

Peters, Bettina, Mark J. Roberts, Van Anh Vuong, and Helmut Fryges (2017), "Estimating dynamic R&D choice: an analysis of costs and longrun benefits." *The RAND Journal of Economics*, 48, 409–437.

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