Identification and Estimation of Production Function with Unobserved Heterogeneity

Hiroyuki Kasahara
University of British Columbia

Paul Schrimpf
University of British Columbia

Michio Suzuki
Tohoku University

March 2023

Motivation

- Estimation of production function and TFP is important in empirical applications (Empirical IO, Trade, Macro).

- In most empirical applications, the coefficients of Cobb-Douglas production function are assumed to be common across firms

$$y_{it} = \beta_0 + \beta_m m_{it} + \beta_\ell \ell_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it}$$

$\rightarrow (\beta_m, \beta_\ell, \beta_k)$ is assumed to be common across i's within "narrowly defined" industry

Is $(\beta_m, \beta_\ell, \beta_k)$ really common across firms?

Implications of Cobb-Douglas Production Function:

$$\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}} \approx \beta_m$$

$$\frac{P_{M,t}M_{it}}{P_{M,t}M_{it} + W_t L_{it}} \approx \frac{\beta_m}{\beta_m + \beta_\ell}$$

$\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)_i}$ over 25 yrs in Japanese Concrete Product and Electric Audio Industries, 30+ Workers

Figure: Concrete Product (2223)

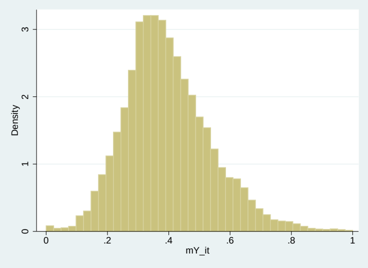
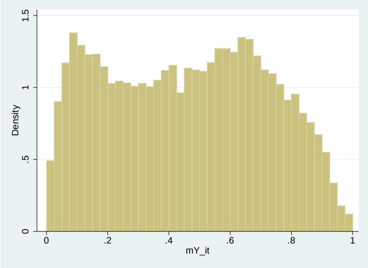


Figure: Electric Audio (2814)



The share of materials is **heterogenous** across firms and **persistent** over time within firm.

$\overline{\left(\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}\right)}_i$ over 25 years

Figure: Concrete Product (2223)

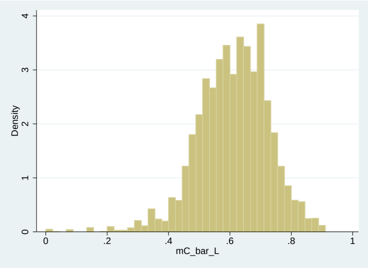
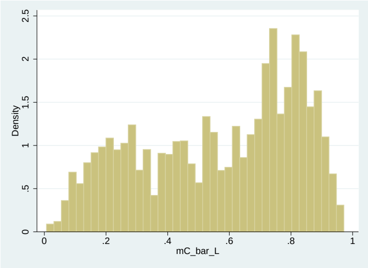


Figure: Electric Audio (2814)



$\Rightarrow$ The difference in markups is not the main reason for the heterogeneity in $\frac{PM_{it}}{PY_{it}}$.

$\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)}_i$: 2-digit vs. 3-digit vs. 4-digit

Figure: 2-digit Electric Parts, Device, Circuit (28)

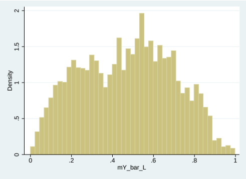


Figure: 3-digit Electric Device (281)

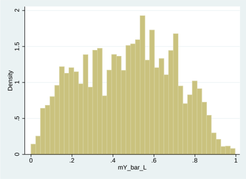
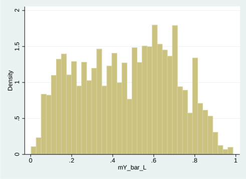


Figure: 4-digit Electric Audio (2814)



Industry Code : Name	No. of Obs.	90-10 diff in $\left(\frac{PM_{it}}{PQ_{it}}\right)_i$	90-10 diff in $\left(\frac{PM_{it}}{PM_{it}+WL_{it}}\right)_i$
28: Electric Parts/Devise/Circuit	33,467	0.62	0.66
281: Electric Device	21,712	0.62	0.67
2814: Electric Audio	12,441	0.63	0.67

The average difference between the 90th and the 10th percentiles at 2-, 3-, and 4-digit industry classifications

Industry Classifications	No. of Industries	Ave. 90-10 diff in $\left(\frac{PM_{it}}{PQ_{it}}\right)_j$	Ave. 90-10 diff in $\left(\frac{PM_{it}}{PM_{it}+WL_{it}}\right)_j$	Ave. No. of Obs.
2-digit	24	0.46	0.44	53,205
3-digit	149	0.42	0.41	8,570
4-digit	279	0.38	0.37	2,666

More general than Cobb-Douglas? (Electric Audio)

$$\frac{P_{M,t} M_{it}}{P_{Y,t} Y_{it}} = \text{2nd order polynomials of } (l_{it}, k_{it}, m_{it}) + e_{it}$$

$$\hat{\xi}_i := T^{-1} \sum_{t=1}^{T} \hat{e}_{it}$$

- 90-10 diff of $\hat{\xi}_i = 0.52$
- 90-10 diff of $\left(\frac{PM_{it}}{PQ_{it}}\right)_i = 0.63$
- Evidence for **persistent heterogeneity** under more general functions than Cobb-Douglas.

Issues

- Can we identify production function in the presence of unobserved heterogeneity?
- Can we estimate production function with random coefficients using a typical firm/plant-level panel data?

Issues in Estimation of Firm-Level Production Function

Simultaneity Bias

- Corr. between productivity & input factors (Marschak and Andrews, 1944)
- Control function/"Proxy variable" approach: OP (Olley and Pakes, 1996), LP (Levinsohn and Petrin, 2003), Wooldrige (2009), ACF (Ackerberg, Caves and Frazer, 2015)
- Dynamic Panel: Arellano and Bond (1991), Blundell and Bond (1998)
- "Exogenous" Input Price as IV: Doraszelski and Jaumandreu (2018)

→ OP/LP is widely used in empirical analysis.

Identification of Firm-Level Production Function

Nonparametric Identification Problem

- Bond and Söderbom (2005) and ACF (2015): Colineality of material and labor.
- GNR (Gandhi, Navarro, and Rivers, 2020): No exogenous variation to identify flexible input's elasticity after conditioning on quasi-fixed input factors.
- GNR exploit the first order condition w.r.t. flexible inputs for identification.

Production Function with Random Coefficients

- Mairesse and Griliches (1990), Van Biesebroeck (2003), Doraszelski and Jaumandreu (2018)
- Identification without input price instruments?
  $\rightarrow$ Unresolved issue.

Identification/Estimation of Panel Data Models

- Nonparametric Identification **with** T **fixed**: Kasahara and Shimotsu (2009), Carroll, Chen, and Hu (2010), Hu and Shum (2012), Higgins and Jochmans (2021).
- Group Heterogeneity **with** $T \to \infty$: Bonhomme and Manresa (2015), Cheng, Schorfheide, and Shao (2021)

Our Paper's contribution

- Non-parametric identification of finite mixture dynamic panel data models with $T = 4$ under non-stationarity
- Non-parametric identification of production function with unobserved heterogeneity
- Estimation of production function with random coefficients
- Empirical application: Japanese Census of Manufacture

Nonparametric Identification

Setup

- Panel Data: $\{\{Y_{it}, \tilde{L}_{it}, K_{it}, M_{it}, B_{it}\}_{t=1}^{T}\}_{i=1}^{n}$
  - Y_{it} Output; $\tilde{L}_{it}$ # of workers; K_{it} Capital stock; M_{it} Intermediate input; B_{it} Total wage bills
- K_{it}: Predetermined
- $\tilde{L}_{it}$, M_{it}: Flexibly chosen
- $\mathcal{I}_{it}$: Information available for choosing L_{it} & M_{it}

Production function and Capital

- J unobserved types with $\pi^j := \Pr(\text{type}=j)$

$$Y_{it} = e^{\omega_{it} + \epsilon_{it}} F_t^j(L_{it}, K_{it}, M_{it}), \quad \epsilon_{it}|\mathcal{I}_{it} \sim_{iid} g_{\epsilon,t}^j(\epsilon_{it})$$

$$\omega_{it} = h^j(\omega_{it-1}) + \eta_{it}, \quad \eta_{it}|\mathcal{I}_{it-1} \sim_{iid} g_\eta^j(\eta_{it})$$

- J is assumed to be known.
- K_{it} predetermined with the conditional probability density function:

$$g_t(K_{it}|\mathcal{I}_{t-1}, D_i = j) = g_t^j(K_{it}|K_{it-1}, \omega_{it-1})$$

Labor Input and Wage Bills

- L_{it} not directly observed but related to # of workers $\tilde{L}_{it}$ as

$$L_{it} = e^{\psi_t^j} \tilde{L}_{it}$$

- Total Wage Bills:

$$B_{it} = e^{v_{it} + \zeta_{it}} P_{L,t} L_{it}, \quad v_{it} | \mathcal{I}_{it-1} \sim_{iid} g_v^j(v_{it})$$

Average wage $B_{it} / \tilde{L}_{it}$

$\Rightarrow$ Identification of unobserved quality/hours ψ_t^j

L and M are flexible inputs

- L_{it} & M_{it} are chosen flexibly before observing ϵ_{it} & ζ_{it}:

$$
\begin{aligned}
(M_{it}, L_{it}) &= (\mathbb{M}_t^j(K_{it}, \omega_{it}, v_{it}), \mathbb{L}_t^j(K_{it}, \omega_{it}, v_{it})) \\
:= \underset{(M,L)\in\mathcal{M}\times\mathcal{L}}{\text{argmax}} &\ P_{Y,t} E^j[e^{\epsilon_{it}}|\mathcal{I}_{it}] e^{\omega_{it}} F_t^j(K_{it}, L, M) - P_{M,t} M \\
&- E^j[e^{\zeta_{it}}|\mathcal{I}_{it}] e^{v_{it}} P_{L,t} L,
\end{aligned}
$$

- $(\mathbb{M}_t^j(K_{it}, \omega_{it}, v_{it}), \mathbb{L}_t^j(K_{it}, \omega_{it}, v_{it}))$ is invertible w.r.t. (ω_{it}, v_{it}) with probability one

- Correlation between (L_{it}, M_{it}) and ω_{it} $\Rightarrow$ Endogeneity

First Order Conditions as Identifying Restrictions

- Material & Labor Share Equations:

$$\underbrace{\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}}_{:=S^m_{it}} = \underbrace{\frac{\partial F^j_t(L_{it}, K_{it}, M_{it})/\partial M}{F^j_t(L_{it}, K_{it}, M_{it})/M_{it}}}_{:=G^j_{M,t}(L_{it},K_{it},M_{it})} E[e^{\epsilon}|\text{type=j}]e^{-\epsilon_{it}}$$

$$\underbrace{\frac{B_{it}}{P_{Y,t}Y_{it}}}_{:=S^\ell_{it}} = \underbrace{\frac{\partial F^j_t(L_{it}, K_{it}, M_{it})/\partial L}{F^j_t(L_{it}, K_{it}, M_{it})/L_{it}}}_{:=G^j_{L,t}(L_{it},K_{it},M_{it})} \frac{E[e^{\epsilon}|\text{type=j}]}{E[e^{\zeta}|\text{type=j}]} e^{\zeta_{it}-\epsilon_{it}}$$

- Material/labor share equation $\Rightarrow$ Source of Identificaiton

Main Nonparametric Identification Results

Under regularity conditions, the model structure

$$\boldsymbol{\theta} = \{g^j_{\epsilon\zeta,t}(\cdot), g^j_v(\cdot), g^j_\eta(\cdot), h^j(\cdot), \pi^j, \{G^j_{M,t}(\cdot), G^j_{L,t}(\cdot), F^j_t(\cdot)\}^T_{t=1}, P_{L,t}, \psi^j_t\}^J_{j=1},$$

is nonparametrically identified from the population probability density function of $\underline{T = 4}$ periods of the panel data,

$$\mathrm{g}(\{Y_t, B_t, M_t, \tilde{L}_t, K_t\}^4_{t=1}) = \mathrm{g}(\{\underbrace{S^m_t, S^\ell_t, M_t, \tilde{L}_t, K_t}_{\boldsymbol{z}_t}\}^4_{t=1}).$$

Proposition 1: Identification of $\boldsymbol{\theta}$ when $J = 1$

Suppose $J = 1$ and Assumptions 1-6 holds with $T \geq 3$. Then, $\boldsymbol{\theta}$ is uniquely determined from the population density function $g_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_T}(\{\boldsymbol{z}_t\}_{t=1}^T)$.

Proposition 1 extends Gandhi, Navarro, and Rivers (2020) to variable labor input case

Proposition 2: a First-order Markov process

Under Assumptions 1-6, $\tilde{X}_t := (M_t, \tilde{L}_t, K_t)$, $S_t := (S_t^m, S_t^\ell)$, $Z_t := (S_t, \tilde{X}_t) = (S_t^m, S_t^\ell, M_t, \tilde{L}_t, K_t)$,

$$\underbrace{g_{\mathbf{Z}_1,\ldots,\mathbf{Z}_T}(\{\mathbf{z}_t\}_{t=1}^T)}_{\text{observed}} = \sum_{j=1}^J \pi^j g^j_{\mathbf{Z}_1}(\mathbf{z}_1) \prod_{t=2}^T g^j_{\mathbf{Z}_t|\mathbf{Z}_{t-1},..,\mathbf{Z}_1}(\mathbf{z}_t|\mathbf{z}_{t-1},..,\mathbf{z}_1)$$

$$= \sum_{j=1}^J \pi^j g^j_{\mathbf{Z}_1}(\mathbf{z}_1) \prod_{t=2}^T g^j_{\mathbf{Z}_t|\mathbf{Z}_{t-1}}(\mathbf{z}_t|\mathbf{z}_{t-1}).$$

$\Rightarrow$ Conditioning on latent type, Z_t follows a 1st-order Markov process under the model assumption.

Markov process is key for identification

Consider discrete variable case with $Z_t \in \{1, 2, ..., |\mathcal{Z}|\}$. Then,

- The number of restrictions $\approx |\mathcal{Z}|^T$.
- The number of unknowns $\approx J \times (1 + |\mathcal{Z}| + (T-1)|\mathcal{Z}|^2)$.

Proposition 3: Identification of Type-specific Distribution

Suppose that $T \geq 4$. *Under regularity conditions,*

$$\{\pi^j, g^j_{\boldsymbol{z}_1}(z_1), g^j_{\boldsymbol{z}_2|\boldsymbol{z}_1}(z_2|z_1), \ldots, g^j_{\boldsymbol{z}_T|\boldsymbol{z}_{T-1}}(z_T|z_{T-1})\}_{j=1}^J$$

is identified from the joint probability density function

$$g_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_T}(\{z_t\}_{t=1}^T)$$

up to a common permutation across the latent types.

Proposition 3 and the result of Kasahara and Shimotsu (2009) and Hu and Shum (2012)

1. Kasahara and Shimotsu (2009): identification under stationarity with $T = 6$.

   $\Rightarrow$ Here, identification under non-stationarity with $T = 4$.

2. Hu and Shum (2012): with continuous time-varying latent variables, under non-stationarity with $T = 5$, identification of $g^j_{\boldsymbol{z}_3}(z_3)$, $g^j_{\boldsymbol{z}_4|\boldsymbol{z}_3}(z_4|z_3)$, and $g^j_{\boldsymbol{z}_5|\boldsymbol{z}_4}(z_5|z_3)$ but the identification of $g^j_{\boldsymbol{z}_1}(z_1)$, $g^j_{\boldsymbol{z}_2|\boldsymbol{z}_1}(z_2|z_1)$, and $g^j_{\boldsymbol{z}_3|\boldsymbol{z}_2}(z_3|z_2)$ remains unresolved.

   $\Rightarrow$ We identify $g^j_{\boldsymbol{z}_1}(z_1)$, $g^j_{\boldsymbol{z}_2|\boldsymbol{z}_1}(z_2|z_1)$, $g^j_{\boldsymbol{z}_3|\boldsymbol{z}_2}(z_3|z_2)$ with $T = 4$ when unobserved latent variables are discrete.

Sketch of Proof

$$\underbrace{g_{\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4}(\boldsymbol{a},z_2,z_3,\boldsymbol{b})}_{:=q_{z_2,z_3}(\boldsymbol{a},\boldsymbol{b})} = \sum_{j=1}^{J} \underbrace{\pi^j g^j_{\boldsymbol{Z}_1}(\boldsymbol{a}) g^j_{\boldsymbol{Z}_2|\boldsymbol{Z}_1}(z_2|\boldsymbol{a})}_{:=\bar{\lambda}^j_2(\boldsymbol{a},z_2)} \underbrace{g^j_{\boldsymbol{Z}_3|\boldsymbol{Z}_2}(z_3|z_2)}_{:=\lambda^j_3(z_3,z_2)} \underbrace{g^j_{\boldsymbol{Z}_4|\boldsymbol{Z}_3}(\boldsymbol{b}|z_3)}_{:=\lambda^j_4(\boldsymbol{b},z_3)}, \quad (1)$$

$$\underbrace{g_{\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a},z_2,z_3)}_{:=\bar{q}_{z_2,z_3}(\boldsymbol{a})} = \sum_{j=1}^{J} \bar{\lambda}^j_2(\boldsymbol{a},z_2)\lambda^j_3(z_3|z_2). \quad (2)$$

Evaluating (1)-(2) at $\boldsymbol{a} = \boldsymbol{a}_1, ..., \boldsymbol{a}_J$ and $\boldsymbol{b} = \boldsymbol{b}_1, ..., \boldsymbol{b}_{J-1}$ gives $J(J-1) + J = J^2$ restrictions.

Sketch of Proof

Collect J^2 equations into matrix as:

$$\underbrace{\boldsymbol{Q}_{\boldsymbol{z}_2,\boldsymbol{z}_3}}_{\text{data}} = \boldsymbol{L}_{\boldsymbol{z}_3}\boldsymbol{D}_{\boldsymbol{z}_3|\boldsymbol{z}_2}\bar{\boldsymbol{L}}_{\boldsymbol{z}_2}^{\top}, \tag{3}$$

$$\boldsymbol{Q}_{\boldsymbol{z}_2,\boldsymbol{z}_3} := \begin{bmatrix} \bar{q}_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_1) & \bar{q}_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_2) & \cdots & \bar{q}_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_J) \\ q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_1,\boldsymbol{b}_1) & q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_2,\boldsymbol{b}_1) & \cdots & q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_J,\boldsymbol{b}_1) \\ \vdots & \vdots & \ddots & \vdots \\ q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_1,\boldsymbol{b}_{J-1}) & q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_2,\boldsymbol{b}_1) & \cdots & q_{\boldsymbol{z}_2,\boldsymbol{z}_3}(\boldsymbol{a}_J,\boldsymbol{b}_{J-1}) \end{bmatrix},$$

$$\boldsymbol{L}_{\boldsymbol{z}_3} := \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_4^1(\boldsymbol{b}_1|\boldsymbol{z}_3) & \cdots & \lambda_4^J(\boldsymbol{b}_1|\boldsymbol{z}_3) \\ \vdots & \ddots & \vdots \\ \lambda_4^1(\boldsymbol{b}_{J-1}|\boldsymbol{z}_3) & \cdots & \lambda_4^J(\boldsymbol{b}_{J-1}|\boldsymbol{z}_3) \end{bmatrix}, \quad \bar{\boldsymbol{L}}_{\boldsymbol{z}_2} := \begin{bmatrix} \bar{\lambda}_2^1(\boldsymbol{a}_1,\boldsymbol{z}_2) & \cdots & \bar{\lambda}_2^J(\boldsymbol{a}_1,\boldsymbol{z}_2) \\ \vdots & \ddots & \ldots \\ \bar{\lambda}_2^1(\boldsymbol{a}_J,\boldsymbol{z}_2) & \cdots & \bar{\lambda}_2^J(\boldsymbol{a}_J,\boldsymbol{z}_2) \end{bmatrix},$$

$$\boldsymbol{D}_{\boldsymbol{z}_3|\boldsymbol{z}_2} := \operatorname{diag}\left(\lambda_3^1(\boldsymbol{z}_3|\boldsymbol{z}_2), \ldots, \lambda_3^J(\boldsymbol{z}_3|\boldsymbol{z}_2)\right).$$

Sketch of Proof

Evaluating (3) at $(\check{z}_2, z_3)$, $(\bar{z}_2, z_3)$, $(\check{z}_2, \bar{z}_3)$, and $(\bar{z}_2, \bar{z}_3)$ gives

$$Q_{\check{z}_2,z_3} = L_{z_3} D_{z_3|\check{z}_2} \bar{L}_{\check{z}_2}^\top, \quad Q_{\bar{z}_2,\bar{z}_3} = L_{\bar{z}_3} D_{\bar{z}_3|\bar{z}_2} \bar{L}_{\bar{z}_2}^\top,$$
$$Q_{\check{z}_2,\bar{z}_3} = L_{\bar{z}_3} D_{\bar{z}_3|\check{z}_2} \bar{L}_{\check{z}_2}^\top, \quad Q_{\bar{z}_2,z_3} = L_{z_3} D_{z_3|\bar{z}_2} \bar{L}_{\bar{z}_2}^\top.$$

$\Rightarrow$

$$\underbrace{A}_{\text{observed}} := Q_{\check{z}_2,z_3} Q_{\check{z}_2,\bar{z}_3}^{-1} Q_{\bar{z}_2,\bar{z}_3} Q_{\bar{z}_2,z_3}^{-1} = L_{z_3} D L_{z_3}^{-1},$$

with

$$D := D_{z_3|\check{z}_2} D_{\bar{z}_3|\check{z}_2}^{-1} D_{\bar{z}_3|\bar{z}_2} D_{z_3|\bar{z}_2}^{-1}.$$

Sketch of Proof

$$\boldsymbol{A} = \boldsymbol{L}_{z_3} \boldsymbol{D} \boldsymbol{L}_{z_3}^{-1} \quad \Rightarrow \quad \boldsymbol{A}\boldsymbol{L}_{z_3} = \boldsymbol{L}_{z_3} \boldsymbol{D}$$

1. The right eigenvectors of $\boldsymbol{A}$ determines $\boldsymbol{L}_{z_3}$ up to multiplicative constant and the ordering of its columns as

   $$\boldsymbol{B} := \boldsymbol{L}_{z_3} \Delta_{z_3} \boldsymbol{C}, \text{ where } \boldsymbol{AB} = \boldsymbol{BD},$$

   Δ_{z_3} is a permutation matrix, and $\boldsymbol{C}$ is a diagonal matrix.
2. $\boldsymbol{CD}$ is identified from the first row of $\boldsymbol{AB}$ because

   $$\boldsymbol{AB} = \boldsymbol{L}_{z_3} \Delta_{z_3} \boldsymbol{CD}.$$

3. $\boldsymbol{L}_{z_3}$ is identified up to a permutation of columns as

   $$\boldsymbol{L}_{z_3} \Delta_{z_3} = \boldsymbol{AB}(\boldsymbol{CD})^{-1}.$$

Sketch of Proof

- We identify $\Delta_{\boldsymbol{z}_3}$ in a common order of latent types across different values of $\boldsymbol{z}_3$ using the identification argument in Higgins and Jochmans (2021).
- Once $\boldsymbol{L}_{\boldsymbol{z}_3}$ is identified, we may identify $\boldsymbol{D}_{\boldsymbol{z}_3|\boldsymbol{z}_2}$, $(\bar{\lambda}_2^1(\boldsymbol{z}_1, \boldsymbol{z}_2), \ldots, \bar{\lambda}_2^J(\boldsymbol{z}_1, \boldsymbol{z}_2))^\top$, and $(\lambda_4^1(\boldsymbol{z}_4|\boldsymbol{z}_3), \ldots, \lambda_4^J(\boldsymbol{z}_4|\boldsymbol{z}_3))^\top$ across different values of $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and $\boldsymbol{z}_4$.

Proposition 4: Identification of $\boldsymbol{\theta}$ when $J \geq 2$.

Suppose that $T \geq 4$. Under regularity conditions,

$$\boldsymbol{\theta} = \{g^j_{\epsilon\zeta,t}(\cdot), g^j_v(\cdot), g^j_\eta(\cdot), h^j(\cdot), \pi^j, \{G^j_{M,t}(\cdot), G^j_{L,t}(\cdot), F^j_t(\cdot)\}_{t=1}^T, P_{L,t}, \psi^j_t\}_{j=1}^J,$$

is nonparametrically identified.

Once the type-specific distribution is identified, the identification follows from Proposition 1.

Cobb-Douglas Case

$$
\begin{aligned}
y_t &= \beta^j_{0,t} + \beta^j_{m,t} m_t + \beta^j_{\ell,t}(\psi^j_t + \tilde{\ell}_t) + \beta^j_{k,t} k_t + \omega_t + \epsilon_t \\
s^m_t &= \ln \beta^j_{m,t} + \ln E^j_t[e^{\epsilon}] - \epsilon_t \\
s^\ell_t - s^m_t &= \ln(\beta^j_{\ell,t}/\beta^j_{m,t}) - \ln\left(E^j_t[e^{\zeta}]\right) + \zeta_t \\
m_t - \tilde{\ell}_t &= \underbrace{\ln(P_{L,t}/P_{M,t})}_{:=\alpha_t} + \ln\left(\beta^j_{m,t}/\beta^j_{\ell,t}\right) + \ln\left(E^j_t[e^{\zeta}]\right) + \psi^j_t + v_t
\end{aligned}
$$

We may identify:

- $\{\beta^j_{0,t}, \beta^j_{m,t}, \beta^j_{\ell,t}, \beta^j_{k,t}, \psi^j_t, g^j_{\epsilon\zeta,t}(\cdot)\}_{j,t}$
- $\{\pi^j, h^j(\cdot), g^j_\eta(\cdot), g^j_v(\cdot)\}_j,\ \{\alpha_t\}^T_{t=1}$

from the distribution of $\{S_t, \tilde{X}_t\}^T_{t=1}$ for $T \geq 4$.

Estimation

Penalized MLE

- J unobserved types
- Cobb-Douglas Production Function
- Assume normal distribution:

$$(\epsilon_{it}, \zeta_{it}, v_{it})' \overset{d}{\sim} N\left(0, \Sigma^j\right)$$
$$\omega_{it} = \rho^j_\omega \omega_{it-1} + \eta_{it}, \quad \eta_i \overset{d}{\sim} N(0, \sigma^j_\eta)$$
$$k_{it} | (k_{it-1}, \omega_{it-1}) \overset{d}{\sim} N(\rho^j_{k0} + \rho^j_{kk} k_{it-1} + \rho^j_{k\omega} \omega_{it-1}, (\sigma^j_k)^2)$$
$$(k_{i1}, \omega_{i1}) \overset{d}{\sim} N(\mu^j_1, \Sigma^j_1).$$

- Penalty term to deal with unbounded likelihood.

Penalized MLE

Probability Density Function of $(s_{it}, \tilde{x}_{it})_{t=1}^{T}$ for type j

$$g_t^j(\{s_{it}, \tilde{x}_{it}\}_{t=1}^{T}) = \underbrace{\prod_{t=1}^{T} g_t(s_{it}, \tilde{\ell}_{it} - m_{it}; \boldsymbol{\theta}_1^j, \alpha_t)}_{=L_{1i}(\boldsymbol{\theta}_1^j, \boldsymbol{\alpha})}$$

$$\times \quad \underbrace{g_1(\tilde{x}_{i1} | \tilde{\ell}_{i1} - m_{i1}; \boldsymbol{\theta}^j) \prod_{t=2}^{T} g_t(\tilde{x}_{it} | \tilde{\ell}_{it} - m_{it}, \tilde{x}_{it-1}; \boldsymbol{\theta}^j)}_{=L_{2i}(\boldsymbol{\theta}_1^j, \boldsymbol{\theta}_2^j)},$$

Penalized MLE

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{N} \log \left(\sum_{j=1}^{J} \pi^j L_{1i}(\boldsymbol{\theta}_1^j, \boldsymbol{\alpha}) L_{2i}(\boldsymbol{\theta}_1^j, \boldsymbol{\theta}_2^j) \right)$$
$$+ \sum_{j=1}^{J} \left\{ \sum_{s \in \{k,\eta,v\}} p_n((\sigma_s^j)^2; \hat{\sigma}_{s,0}^2) + p_n(\Sigma_{\epsilon\zeta}^j; \hat{\Sigma}_{\epsilon\zeta,0}) + p_n(\Sigma_1^j; \hat{\Sigma}_{1,0}) \right\}$$

- Three-stage procedure:
  1. max $\log \left(\sum_{j=1}^{J} \pi^j L_{1i}(\boldsymbol{\theta}_1^j, \boldsymbol{\alpha}) \right)$ over $\boldsymbol{\pi}$, $\boldsymbol{\theta}_1$, and $\alpha_t \Rightarrow \tilde{\boldsymbol{\pi}}$, $\tilde{\boldsymbol{\theta}}_1$, $\tilde{\alpha}_t$,
  2. max the full log-likelihood over $\boldsymbol{\theta}_2$ given $\tilde{\boldsymbol{\pi}}$, $\tilde{\boldsymbol{\theta}}_1$, $\tilde{\alpha}_t$,
  3. max the full log-likelihood over $\boldsymbol{\theta}$.
- EM algorithm

Empical Application: Japanese Census of Manufacute, 1986-2010

Data set

- Unbalanced panel data, Plants with 30+ workers
- Variable Construction
  - K_{it}: Fixed asset less land. Perpetual inventory method.
  - L_{it}: Number of employees
  - M_{it}: Material + Energy + Subcontracting expenses for consigned production
  - Y_{it}: Sales
- Electric Audio Equipment: 907 plants $\times$ 4$\sim$25 years

Specification

- Cobb-Douglas (3 Types):

$$Y_i = \beta_0^j + \beta_m^j m_{it} + \beta_k^j k_{it} + \beta_\ell^j \ell_{it} + \omega_{it} + \epsilon_{it}, \quad \text{for } j = 1, 2, 3$$

- Unobserved hours or labor quality ($2 \times 3 = 6$ Types):

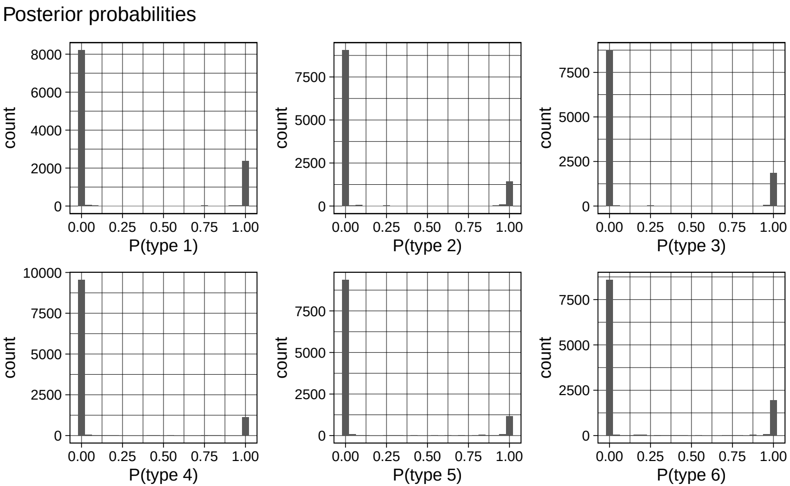
$$L_{it} = \psi^j \tilde{L}_{it} \quad \text{for } j = 1, 2, ..., 6$$

Estimates of Production Function (Electric Audio)

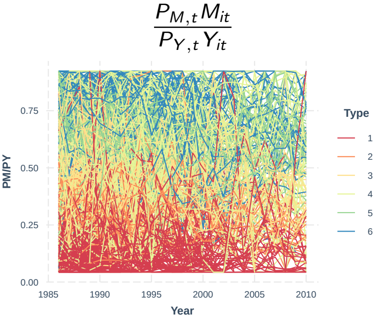
	J = 1	J = 6					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
β^j_m	0.281	0.136		0.364		0.597	
	(0.008)	(0.007)		(0.022)		(0.010)	
β^j_ℓ	0.296	0.507		0.228		0.132	
	(0.006)	(0.012)		(0.012)		(0.006)	
β^j_k	0.076	0.122		0.185		0.202	
	(0.016)	(0.010)		(0.021)		(0.017)	
$\beta^j_m + \beta^j_\ell + \beta^j_k$	0.652	0.765		0.778		0.932	
β^j_k / β^j_ℓ	0.256	0.242		0.813		1.526	
ψ^j	0.000	-0.999	0.474	-0.258	0.201	-0.215	0.798
		(0.135)	(0.139)	(0.389)	(0.142)	(0.115)	(—)
π	1.000	0.278	0.137	0.134	0.158	0.164	0.128
		(0.023)	(0.017)	(0.163)	(0.149)	(0.015)	(0.014)

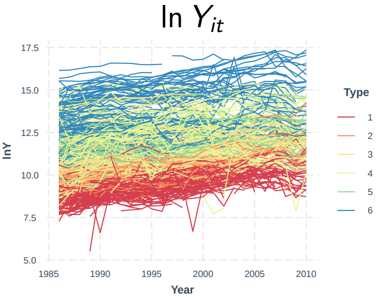
Posterior Probabilities for $J = 6$ Types (Electric Audio)

$$\hat{\pi}_i^j = \frac{\hat{\pi}^j L_i(\hat{\theta}^j)}{\sum_{k=1}^J \hat{\pi}^k L_i(\hat{\theta}^k)} \quad \text{for } j = 1, ..., J.$$



Trends in $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ and $\ln Y_{it}$ across types (Electric Audio)





Standard deviations of ω and ϵ (Electric Audio)

	J = 1	J = 6					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
S. D. $\hat{\omega}_{it} + \beta^j_{0t}$	1.39	0.83	0.62	0.47	0.91	0.29	0.36
S. D. $\hat{\epsilon}_{it}$	0.79	0.70	0.44	0.39	0.67	0.24	0.25
$Corr(\hat{\epsilon}_{it}, \hat{\epsilon}_{it-1})$	0.88	0.74	0.67	0.72	0.66	0.71	0.74

Bias in $\Delta\omega_{it}$ due to ignoring unobserved heterogeneity

Take $\underline{J=6}$ as a true model:

$$\Delta\omega_{it} := \Delta y_{it} - (\hat{\beta}^j_m \Delta m_{it} + \hat{\beta}^j_\ell \Delta \ell_{it} + \hat{\beta}^j_k \Delta k_{it} + \Delta \hat{\epsilon}^j_{it}).$$

Take $\underline{J=1}$ as a mis-specified model:

$$\Delta\tilde{\omega}_{it} := \Delta y_{it} - (\bar{\beta}_m \Delta m_{it} + \bar{\beta}_\ell \Delta \ell_{it} + \bar{\beta}_k \Delta k_{it} + \Delta \bar{\epsilon}_{it}).$$

Then,

$$\text{Bias}_{it} = (\hat{\beta}^j_m - \bar{\beta}_m)\Delta m_{it} + (\hat{\beta}^j_\ell - \bar{\beta}_\ell)\Delta \ell_{it} + (\hat{\beta}^j_k - \bar{\beta}^j_k)\Delta k_{it} + (\Delta \hat{\epsilon}^j_{it} - \Delta \bar{\epsilon}_{it}).$$

Substantial Bias in $\Delta\tilde{\omega}_{it}$ for $J = 1$ (Electric Audio)

$$\frac{\textbf{Mean of } |\textbf{Bias}_{it}|}{\textbf{Mean of } |\Delta\tilde{\omega}_{it}|}$$

	J = 6		
	Type 1 - 2	Type 3 - 4	Type 5 - 6
$\frac{\text{Mean of } \lvert\text{Bias}_{it}\rvert}{\text{Mean of } \lvert\Delta\tilde{\omega}_{it}\rvert}$	0.226	0.140	0.440
β_m^j	0.136	0.364	0.597

14-44 percent of Bias in $\Delta\tilde{\omega}_{it}$ relative to the mean absolute value of $\Delta\tilde{\omega}_{it}$.

Systematic Bias in $\Delta\tilde{\omega}_{it}$ for $J = 1$

$\frac{\textbf{Mean of Bias}_{it}}{\textbf{Mean of } |\Delta\tilde{\omega}_{it}|}$ conditional on $\Delta\omega_{it} > 0$

	J = 6		
	Type 1 - 2	Type 3 - 4	Type 5 - 6
$\left.\frac{\text{Mean of Bias}_{it}}{\text{Mean of } \lvert\Delta\tilde{\omega}_{it}\rvert}\right\rvert_{\Delta\omega_{it}>0}$	-0.200	0.084	0.247
β_m^j	0.136	0.364	0.597

$\text{Bias}_{it} \approx (\hat{\beta}_m^j - \bar{\beta}_m)\Delta m_{it}$ and $Corr(\Delta\omega_{it}, \Delta m_{it}) > 0$

$\Rightarrow$ Mean of Bias_{it} conditional on $\Delta\omega_{it} > 0$ is positive when $\hat{\beta}_m^j > \bar{\beta}_m$

The Heterogeneous Effect of ω_{it} on Investment (Electric Audio)

$$I_{it}/K_{it} = \alpha_0 + \alpha_\omega^j \omega_{it} + \text{quadratic of } k_{it}$$

	$J=1$	$J=6$					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
α_ω^j	0.50	-0.19	-0.47	0.29	0.18	0.11	0.10
	(0.11)	(0.13)	(1.26)	(0.07)	(0.34)	(0.02)	(0.03)
β_m^j	0.28	0.14		0.36		0.60	
β_k^j/β_ℓ^j	0.26	0.24		0.81		1.53	

The effect of ω_{it} on I_{it}/K_{it} is larger among firms with

- high material share
- high capital-labor ratio

Quantile Regression (Electric Audio)

$$I_{it}/K_{it} = \alpha_0 + \alpha_\omega^j \omega_{it} + \text{quadratic of } k_{it}$$

	$J=1$	$J=6$					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
$\alpha_\omega^j(0.10)$	0.00	0.00	0.00	0.01	0.00	-0.00	0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
$\alpha_\omega^j(0.50)$	0.01	0.00	0.02	0.04	0.01	0.04	0.04
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
$\alpha_\omega^j(0.90)$	0.06	0.01	0.01	0.09	0.09	0.18	0.10
	(0.00)	(0.01)	(0.02)	(0.03)	(0.02)	(0.05)	(0.04)
β_m^j	0.28	0.14		0.36		0.60	
β_k^j/β_ℓ^j	0.26	0.24		0.81		1.53	

Conclusions

- Evidence for heterogeneity beyond Hick's neutral term
- Non-parametric identification
- Maximum likelihood estimation
- Empirical applications to Japanese Census Data
- Heterogeneous investment elasticities across unobserved types