Cabinet Office

ESRI Discussion Paper Series No.368

Identification and Estimation of Production Function with Unobserved Heterogeneity

Hiroyuki Kasahara
Paul Schrimpf
Michio Suzuki

April 2022

Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

Identification and Estimation of Production Function with Unobserved Heterogeneity *

Hiroyuki Kasahara†
Vancouver School of Economics
University of British Columbia

Paul Schrimpf
Vancouver School of Economics
University of British Columbia

Michio Suzuki
Cabinet Office, Government of Japan
Tohoku University

March 25, 2022

Abstract

This paper examines the non-parametric identifiability of production function when production functions are heterogeneous across firms beyond Hicks-neutral technology terms. Using a finite mixture specification to capture unobserved heterogeneity in production technology, we show that the production function for each unobserved type is non-parametrically identified under regularity conditions. We estimate a random coefficients production function using the panel data of Japanese manufacturing plants and compare it with the estimate of the production function with fixed coefficients estimated by the method of Gandhi, Navarro, and Rivers (2020). Our estimates for random coefficients production function suggest that there exists substantial heterogeneity in production function coefficients beyond Hicks neutral term across firms within narrowly defined industries.

*The third author gratefully acknowledges financial support from JSPS Grant-in-Aid for Young Scientists (B) No.16K21004 and Grant-in-Aid for Scientic Research (C) No.19K01584. All remaining errors are our own. This paper represents the views of the authors and not necessarily those of the Cabinet Office, the Government of Japan or the institutions they represent.

†Address for correspondence: Hiroyuki Kasahara, Vancouver School of Economics, University of British Columbia, 6000 Iona Dr., Vancouver, BC, V6T 1L4 Canada.

1 Introduction

Estimation of production function is one of the most important topics in empirical economics. Understanding how the input is related to the output is a fundamental issue in empirical industrial organization (see, for example, Ackerberg, Benkard, Berry, and Pakes, 2007). In empirical trade and macroeconomics, researchers are often interested in estimating production function to obtain a measure of total factor productivity to examine the effect of trade policy on productivity and to analyze the role of resource allocation on aggregate productivity (e.g., Pavcnik, 2002; Kasahara and Rodrigue, 2008; Hsieh and Klenow, 2009).

As first discussed by Marschak and Andrews (1944), the ordinary least square estimates of production function suffer from simultaneity bias because inputs are correlated with error terms when a firm makes an input decision based on their productivity level (Griliches and Mairesse, 1998). Under the assumption that error terms could be decomposed into permanent and idiosyncratic components, a fixed-effects estimator may be used but such an assumption could be violated in practice, and the coefficient of inputs that are persistent over time could be severely biased downward due to measurement errors (Griliches and Hausman, 1986). More recent literature attempts to address the simultaneity issue by employing dynamic panel approach (Arellano and Bond, 1991; Blundell and Bond, 1998; Blundell and Bond, 2000) or developing proxy variable approach (Olley and Pakes, 1996 (OP, hereafter); Levinsohn and Petrin, 2003 (LP, hereafter); Ackerberg, Caves, and Frazer, 2015, (ACF, hereafter); Wooldridge, 2009), which are now widely used in empirical applications.

Despite their popularity, however, potential identification issues of the proxy variable approach have been pointed out in the literature. Bond and Sderbom (2005) and ACF discuss identification issues due to collinearity under two flexible inputs (i.e., material and labor) in Cobb-Douglas specification. Gandhi, Navarro, and Rivers (2020, GNR hereafter) argue that, if the firm's decision follows a Markovian strategy, then the conditional moment restriction implied by proxy variable approach may not provide enough restriction for non-parametrically identifying gross production function. GNR exploit the first order condition with respect to flexible input under profit maximization and establish the identification of production function without making any functional form assumption. Based on their identification strategy, GNR proposes an estimation procedure that does not suffer from simultaneity bias.

This paper extends the identification result of GNR based on the first-order condition to the case where production functions are heterogeneous across firms beyond Hicks-neutral technology terms. We consider a finite mixture specification in which there are J distinct time-varying production technologies and each firm belongs to one of J types. Econometricians do not observe the type of firms. Without making any functional form assumption on each type of production technology, we establish nonparametric identification of J distinct production functions and a population proportion of each type under the reasonable assumption.

Recent literature includes studies examining the identification of extended production functions (e.g. Li and Sasaki, 2017; Demirer, 2020; Chen, Igami, Sawada, and Xiao, 2021; Doty, 2022). However, since our extension differs from those existing models, our nonparametric identification result is an important contribution to this literature. Our identification result on production function with unobserved heterogeneity is also useful in practice as the random coefficient models of production functions are increasingly popular in empirical analysis (e.g., Mairesse and Griliches, 1990; Van Biesebroeck, 2003; Doraszelski and Jaumandreu, 2018; Balat, Brambilla, and Sasaki, 2019).

In estimation, we consider a random coefficient specification for production function and propose two different estimation procedures. The first procedure follows our two-stage identification proof and directly maximizes the log-likelihood function of a finite mixture model of production functions under parametric assumptions, where the EM algorithm can be used to facilitate the computational complication of maximizing the log-likelihood function of the finite mixture model. In the second procedure, we first estimate the partial likelihood function under the normality assumption and use the posterior distribution of type probabilities to classify each firm observation into one of the J types, generating J data sets; using each of J data sets, we estimate the rest of the type-specific parameters. The second procedure is computationally much simpler and requires fewer auxiliary parametric assumptions than the first one although the second procedure could lead to a biased estimator due to misclassification of types when T is small.

We estimate a random coefficients production function using the panel data of Japanese manufacturing plants between 1986 and 2010 and compare the results with those from the original GNR specification without unobserved heterogeneity. Our estimates suggest that there exists substantial heterogeneity in production function coefficients beyond Hicks neutral term. Ignoring unobserved heterogeneity may lead to substantial biases in the measurement of productivity growth. To examine this issue, we take a specification with unobserved heterogeneity as the true model and compute the bias in the measurement of productivity growth when we use a misspecified homogenous model. The results suggest that ignoring unobserved heterogeneity could result in serious bias in the estimated productivity growth and the bias is likely to have a systematic pattern depending on the heterogeneous parameter estimates. We also find that the correlation between estimated productivity and investment is different across different types of firms, where the correlation is stronger among a type of firms with capital-intensive production technology than other types of firms.

2 Evidence for unobserved heterogeneity

To motivate the necessity of considering production functions with unobserved heterogeneity, we first present stylized facts that production functions are heterogeneous beyond Hicks-neutral technology term using the panel data of Japanese manufacturing plants from 1986 to 2010. Section 6.1 discusses the details of our data.

To fix the idea, consider a plant with the Cobb-Douglas production technology:

$$\log Y_{it} = \beta_0 + \beta_m^i \log M_{it} + \beta_\ell^i \log L_{it} + \beta_k^i \log K_{it} + \omega_{it},$$

where Y_{it} is output, M_{it}, L_{it}, and K_{it} are intermediate input, labor, and capital, and ω_{it} is the total factor productivity (TFP) that follows the first order Markov process. We assume that firms take their output and input prices given and that intermediate input and labor are flexibly chosen after ω_{it} is fully observed. Then, a plant's profit maximization implies:

$$\beta_m^i = \frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}} \quad \text{and} \quad \frac{\beta_m^i}{\beta_m^i + \beta_\ell^i} = \frac{P_{M,t}M_{it}}{P_{M,t}M_{it} + W_t L_{it}}, \tag{1}$$

where $P_{Y,t}$, $P_{M,t}$, and W_t are the prices of output, intermediate input, and labor.

In most existing empirical work, production function is estimated under the assumption that the coefficients β_m^i, β_ℓ^i, and β_k^i do not vary across plants. This assumption can be tested in view of (1) by examining whether the intermediate input share, $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$, and the ratio of intermediate cost to the sum of intermediate and labor costs, $\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}$, are constant across plants.

Figures 1a and 1b present the histogram of the plant-level average of the intermediate input share, $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$, over the maximum 25 years across all plants that belong to the concrete products and electric audio equipment industries. Both figures show a large variation in intermediate shares. In particular, the dispersion in the electric audio equipment industry is very large with the 90th to the 10th difference given by 0.63, indicating that the elasticities of output with respect to intermediate input dramatically differ across firms.

Figures 2a and 2b show the histogram of the plant-level average of the material cost to variable costs, $\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}$, for the concrete products and electric audio equipment industries, respectively. The large variation in the cost shares suggests that heterogeneous markups are not the main reason for the variation in the intermediate input share shown in Figures 1a and 1b.

Comparing the degrees of dispersions in cost shares within 2-digit industry classification against that within 3 or 4-digit industry classification, we may examine how classifying industries at a finer level helps control for heterogeneity in production technology.

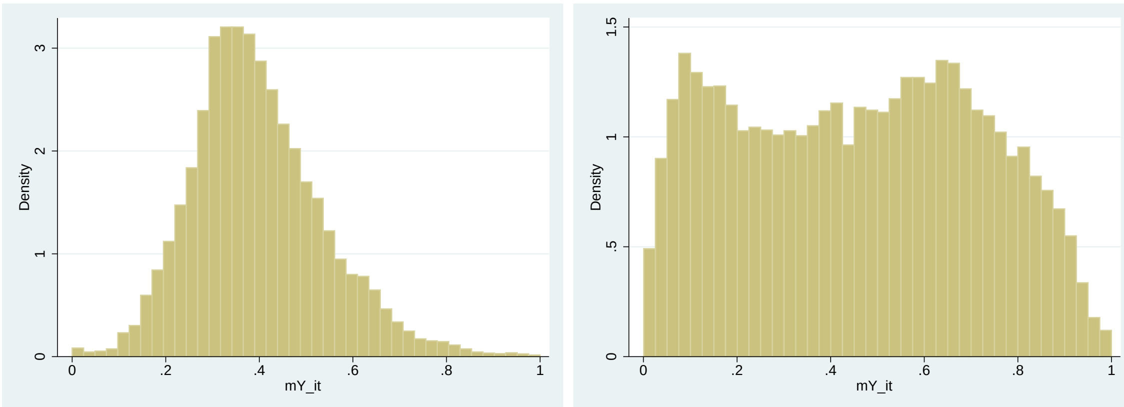
Figures 3a, 3b, and 3c compare the histogram of the plant-level average of the material cost to variable costs for ceramics and clay (2 digit), cement product (3 digit) and concrete products (4 digit). These figures indicate that the dispersion somewhat decreases from 2 digit to 4 digit, but the difference is not large. Figures 4a, 4b, and 4c similarly shows that the dispersion of material shares does not decline much by moving from electric parts, device, and circuit (2 digit) to electric device (3 digit) and then to electric audio equipment (4 digit),

Table 1 reports the difference between the 90th percentile and 10th percentile for the plant-level averages of $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ and $\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}$ for 2, 3, and 4 digit industry classifications. For concrete products, the 90th-10th percentile difference in intermediate input shares decreases from 0.38 (2 digit) to 0.28 (4 digit), while the 90th-10th percentile difference changes only slightly from 2 digit to 4 digit for the electric audio equipment industry, ranging from 0.61 to 0.62. The patterns are similar for the plant-level average of the cost shares. These patterns hold across all industries; Table 2 reports the averages of the 90th-10th differences for all industries at 2-digit, 3-digit, and 4-digit classifications and shows that the dispersion measure decreases only slightly by considering a finer industry classification from 2 digit to 4 digit. Overall, this suggests that using the data set at a finer industry classification may not be a solution to control for production technology heterogeneity; furthermore, information on industry classifications finer than 4-digit classification is often not available in the data set.

Implications in (1) only hold under the Cobb-Douglass production function. For a more general production function, the elasticities of output with respect to inputs depend on the level of material, labor, and capital inputs even in the absence of heterogeneity in production technology. Therefore, we also examine if the ratio of intermediate input cost to output value is similar across firms after controlling for differences in capital, labor, and intermediate inputs. Specifically, we first regress $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ on the second order polynomials of the log of materials, the number of workers, and capital to obtain the residuals, denoted by e_{it}. Then, we compute the plant-level average $\hat{\xi}_i := T^{-1}\sum_{t=1}^{T} e_{it}$ to measure production technology heterogeneity conditioning on inputs.

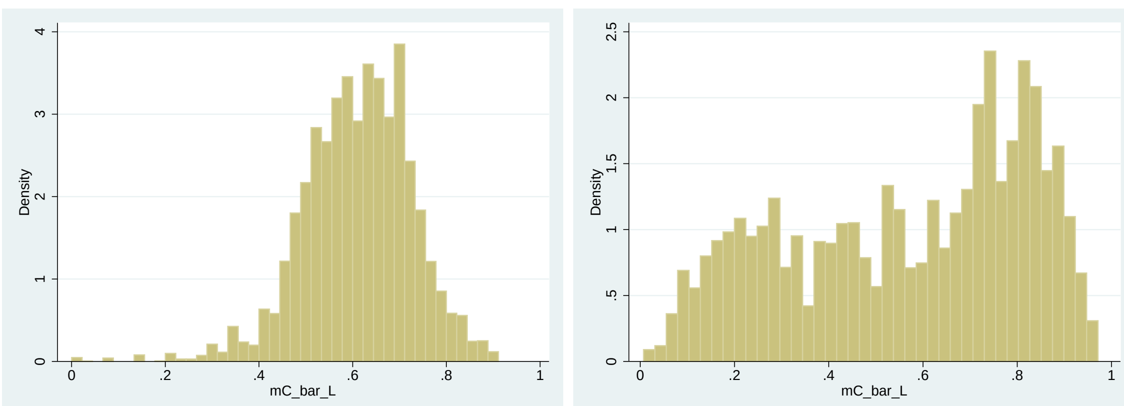
The 90th to 10th percentile differences in $\hat{\xi}_i$ for the concrete industry and electric audio equipment industry are 0.27 and 0.29, respectively, indicating a substantial variation in

Figure 1: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)}_i$



(a) Concrete Product (b) Electric Audio

Figure 2: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}\right)}_i$



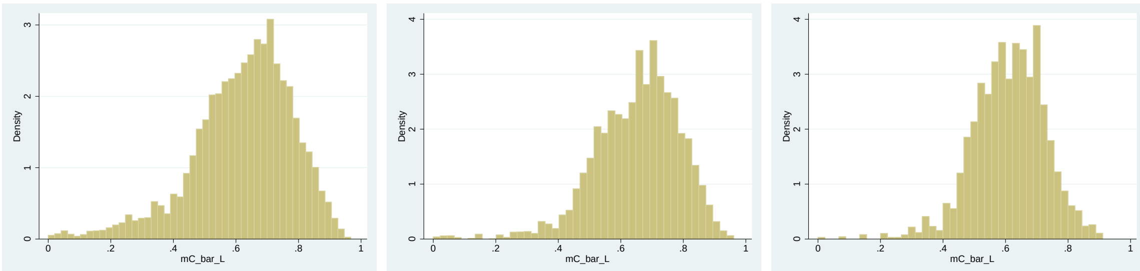
(a) Concrete Product (b) Electric Audio

production technology after controlling for observable inputs.

Table 1: The 90th-10th percentile ratio of Intermediate and Labor Cost Shares for Concrete Product and Electric Audio

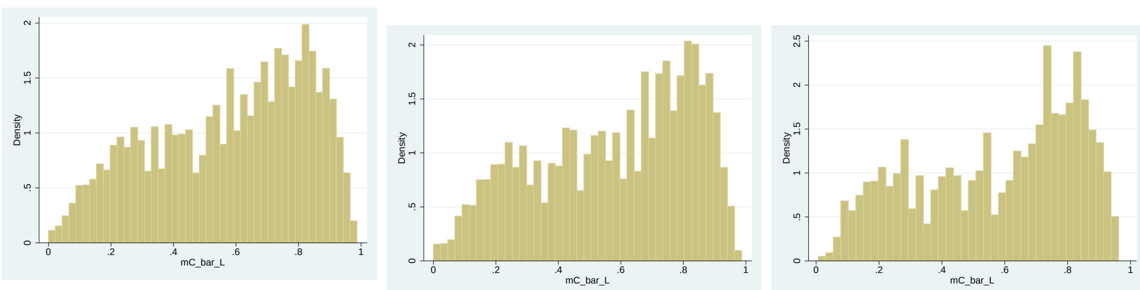
Industry Code : Name	No. of Obs.	90-10 diff in $\overline{\left(\frac{PM_{it}}{PY_{it}}\right)}_i$	90-10 diff in $\overline{\left(\frac{PM_{it}}{PM_{it}+WL_{it}}\right)}_i$
22: Ceramics and Clay	53,042	0.38	0.38
222: Cement Product	22,834	0.35	0.32
2223: Concrete Product	14,463	0.28	0.27
28: Electric Parts/Devise/Circuit	30,814	0.61	0.66
281: Electric Device	19,901	0.62	0.66
2814: Electric Audio	11,325	0.62	0.67

Figure 3: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)}_i$



(a) 2-digit Ceramics and Clay (b) 3-digit Cement Product (c) 4-digit Concrete Product

Figure 4: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)}_i$



(a) 2-digit Electric Parts, Device, Circuit (b) 3-digit Electric Device (c) 4-digit Electric Audio

Table 2: The 90th-10th percentile ratio of Intermediate and Labor Cost Shares

Industry Classifications	No. of Industries	Ave. 90-10 diff in $\overline{\left(\frac{PM_{it}}{PY_{it}}\right)}_i$	Ave. 90-10 diff in $\overline{\left(\frac{PM_{it}}{PM_{it}+WL_{it}}\right)}_i$	Ave. No. of Obs.
2-digit	24	0.46	0.44	49,512
3-digit	149	0.42	0.39	7,975
4-digit	279	0.38	0.35	2,481

3 The Model

Denote output, capital, intermediate inputs, labour input in effective unit of labour, and total wage bills, denoted by $(Y_{it}, K_{it}, M_{it}, L_{it}, B_{it}) \in \mathcal{Y}\times\mathcal{K}\times\mathcal{M}\times\mathcal{L}\times\mathcal{B}$, respectively, where $\mathcal{Y}$, $\mathcal{K}$, $\mathcal{M}$, $\mathcal{L}$, and $\mathcal{B}$ are the supports of corresponding variables. We collect the three inputs (capital, intermediate, and labour) into a vector as $X_{it} := (K_{it}, M_{it}, L_{it})' \in \mathcal{X} := \mathcal{K}\times\mathcal{M}\times\mathcal{L}$.

We consider a possibility that firms are different in production technology beyond Hick's neutral productivity shock. Specifically, we use a finite mixture specification to capture permanent unobserved heterogeneity in firm's production technology. Define the latent random variable $D_i \in \{1, 2, ..., J\}$ that represents the type of firm i so that $D_i = j$ when firm i has the j-th type of technology. In the following, the superscript j indicates that functions are specific to type j while the subscript t indicates that functions are specific to period t. In particular, for a random variable Z_{it}, we denote the probability distribution and the expectation conditional on $D_i = j$ as $P^j(Z_{it}) := P(Z_{it}|D_i = j)$ and $E^j[Z_{it}] := E[Z_{it}|D_i = j]$.

We assume that both M_{it} and L_{it} are flexibly chosen after observing serially correlated productivity shock ω_{it}. On the other hand, K_{it} is predetermined at the end of the last period. Denote the information available to a firm for making decisions on M_{it} and L_{it} by $\mathcal{I}_{it}$.

Assumption 1. *(a) Each firm belongs to one of the J types, where the probability of belonging to type j is given by $\pi^j = \mathrm{P}(D_i = j)$, and J is known. A firm knows its type, i.e., $D_i \in \mathcal{I}_{it}$. (b) For the j-th type of production technology at time t, the output is related to inputs as*

$$Y_{it} = e^{\omega_{it}+\epsilon_{it}} F_t^j(X_{it}). \tag{2}$$

(c) The total wage bills is related to the labour input in effective unit as

$$B_{it} = e^{v_{it}+\zeta_{it}} P_{L,t} L_{it}, \tag{3}$$

where $P_{L,t}$ is the market wage.

Assumption 2. *(a) $(v_{it}, \omega_{it}) \in \mathcal{I}_{it}$. For the j-th type, ω_{it} follows an exogenous first order stationary Markov process given by*

$$\omega_{it} = h^j(\omega_{it-1}) + \eta_{it} \tag{4}$$

where, conditional on $\mathcal{I}_{it-1}$, η_{it} and v_{it} are mean-zero i.i.d. random variables on $\mathbb{R}$ with the probability density functions $g_\eta^j(\cdot)$ and $g_v^j(\cdot)$, respectively. Furthermore, the unconditional expectation of ω_{it} is zero, i.e., $E^j[\omega_{it}] = 0$. (b) $(\epsilon_{it}, \zeta_{it}) \notin \mathcal{I}_{it}$ so that $(\epsilon_{it}, \zeta_{it})$ is not known when L_{it} and M_{it} are chosen. For the j-th type, conditional on $\mathcal{I}_{it}$, $(\epsilon_{it}, \zeta_{it})$ is a mean-zero i.i.d. random variable on $\mathbb{R}^2$ with the probability density function $g_{\epsilon\zeta,t}^j(\cdot)$.

Assumption 3. *(a) $K_{it} \in \mathcal{I}_{it}$ but $K_{it} \notin \mathcal{I}_{it-1}$. (b) the conditional distribution of K_{it} given $\mathcal{I}_{t-1}$ is type specific and only depends on K_{it-1} and ω_{it-1}, i.e., $P_t(K_{it}|\mathcal{I}_{t-1}, D_i = j) = P_t^j(K_{it}|K_{it-1}, \omega_{it-1})$.*

Assumption 4. *(a) M_{it} and L_{it} are chosen at time t by maximizing the expected profit* *conditional on* $\mathcal{I}_{it}$ *as*

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

where $(\mathbb{M}_t^j(K_{it}, \omega_{it}, v_{it}), \mathbb{L}_t^j(K_{it}, \omega_{it}, v_{it}))$ *is a type-specific deterministic function of* $(K_{it}, \omega_{it}, v_{it})$*. (b) For any given* $K_{it} \in \mathcal{K}$*,* $(\mathbb{M}_t^j(K_{it}, \omega_{it}, v_{it}), \mathbb{L}_t^j(K_{it}, \omega_{it}, v_{it}))$ *is invertible with respect to* (ω_{it}, v_{it}) *with probability one.*

Assumption 5. *(a) A firm is a price taker. (b) The intermediate input price* $P_{M,t}$*, the output price* $P_{Y,t}$*, and the market wage* $P_{L,t}$ *at time* t *are common across firms. (c)* $(P_{M,t}, P_{Y,t}, P_{L,t}) \in \mathcal{I}_{it}$ *and* $(P_{M,t}, P_{Y,t})$ *is known to an econometrician.*

Assumption 6. *(a) The labour input in effective unit of labour* L_{it} *is not directly observable. (b)* L_{it} *is related to the number of workers, denoted by* $\widetilde{L}_{it}$*, as*

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

with $\sum_{j=1}^J \pi^j e^{\psi_t^j} = 1$*.*

In Assumption 1(b), as indicated by the subscript t in $F_t^j(\cdot)$, type-specific production function could be different across periods because of type-specific aggregate shocks or type-specific biased technological changes. In Assumption 1(c), we relate the labor input in an effective unit of labor to the total wage bills.

Assumption 2 assume that (ω_{it}, v_{it}) is known when L_{it} and M_{it} are chosen while $(\epsilon_{it}, \zeta_{it})$ is not known when L_{it} and M_{it} are chosen. The presence of wage shock v_{it} provides an additional source of variation for L_{it} beyond ω_{it} and K_{it}; consequently, L_{it} and M_{it} are not collinear, preventing the identification problem discussed by Bond and Sderbom (2005) and ACF.

Assumption 3(a) assumes that K_{it} is determined at time $t-1$ so that $(\eta_{it}, \omega_{it}, v_{it})$ is not known when K_{it} is chosen. Assumption 3(b) can be justified by explicitly considering the dynamic model of investment decisions. Assumption 4(b) holds when there exists one-to-one relationship between (M_{it}, L_{it}) and (ω_{it}, v_{it}) conditional on the value of K_{it}, and is satisfied in the case of Cobb-Douglas function.

Under Assumption 5(b), the intermediate input price $P_{M,t}$ cannot be used for instrumenting M_{it}; when intermediate prices are exogenous and heterogenous across firms, production function could be identified using the intermediate input prices as instruments (see Doraszelski and Jaumandreu, 2014). In Assumption 5(c), we may alternatively assume that a firm is subject to idiosyncratic price shock ξ_{it} such that, for example, $P_{Y,it} = \exp(\xi_{it})P_{Y,t}$ with $\xi_{it} \notin \mathcal{I}_{it}$, then ξ_{it} plays the similar role to ϵ_{it}. We may assume that $(P_{M,t}, P_{Y,t})$ is not known to econometrician by treating $P_{M,t}/P_{Y,t}$ as parameters to be estimated; in such a case, we may identify the production function up to scale.

As stated in Assumption 6(a), we assume that L_{it} is not directly observable because of the firm-level differences in both labor quality and working hours. Assumption 6(b) imposes a specific structure on how the labor input is related to the observed number of workers, where ψ_t^j represents worker quality and working hours which are specific type j, and the variation in labor inputs within each type is fully captured by the number of workers $\widetilde{L}_{it}$. The assumption that $\sum_{j=1}^J \pi^j e^{\psi_t^j} = 1$ is normalization. The assumption of flexibly chosen labor input (i.e., Assumption 4) is more plausible when we don't impose Assumption 6(b) because working hours are more flexibly adjustable than the number of workers and Assumption 6(b) does not allow working hours to be different across firms within each type.[1] For this reason, we provide two different identification results: the one with Assumption 6(b) and the other without Assumption 6(b).

Under Assumptions 1-6, we have $\mathcal{I}_{it} = \{D_i, \omega_{it}, \upsilon_{it}, K_{it}, P_{M,t}, P_{Y,t}, P_{L,t}, V_{it-1}, V_{it-2}, ...\}$, where $V_{it} = \{\zeta_{it}, \epsilon_{it}, \omega_{it}, \upsilon_{it}, K_{it}, P_{M,t}, P_{Y,t}, P_{L,t}\}$.

Let $g_{\epsilon,t}(\epsilon) := \int g_{\epsilon\zeta,t}(\epsilon,\zeta)d\zeta$ and $g_{\zeta,t}(\zeta) := \int g_{\epsilon\zeta,t}(\epsilon,\zeta)d\epsilon$. Under Assumptions 1, 2, 3(a), 4(a), and 5, the first order conditions with respect to M_{it} and L_{it} give

$$P_{Y,t}F^j_{M,t}(X_{it})E^j_t(e^{\epsilon})e^{\omega_{it}} = P_{M,t}, \quad P_{Y,t}F^j_{L,t}(X_{it})E^j_t(e^{\epsilon})e^{\omega_{it}} = E^j_t(e^{\zeta})e^{\upsilon_{it}}P_{L,t}, \tag{5}$$

where $F^j_{M,t}(X) := \frac{\partial F^j_t(X)}{\partial M}$, $F^j_{L,t}(X) := \frac{\partial F^j_t(X)}{\partial L}$, $E^j_t[e^{\epsilon}] := \int e^{\epsilon} g^j_{\epsilon,t}(\epsilon)d\epsilon$, and $E^j_t[e^{\zeta}] := \int e^{\zeta} g^j_{\zeta,t}(\zeta)d\zeta$. Equations (2), (3), and (5) give a system of equations

$$\begin{gathered}\ln Y_{it} = \ln F^j_t(X_{it}) + \omega_{it} + \epsilon_{it}, \quad \ln S^m_{it} = \ln\left(G^j_{M,t}(X_{it})E^j_t[e^{\epsilon}]\right) - \epsilon_{it},\\ \ln S^{\ell}_{it} - \ln S^m_{it} = \ln\left(\frac{G^j_{L,t}(X_{it})}{G^j_{M,t}(X_{it})E^j_t[e^{\zeta}]}\right) + \zeta_{it}, \quad \ln P_{M,t}M_{it} = \ln\left(\frac{P_{L,t}L_{it}G^j_{M,t}(X_{it})E^j_t[e^{\zeta}]}{G^j_{L,t}(X_{it})}\right) + \upsilon_{it},\end{gathered} \tag{6}$$

where

$$S^m_{it} := \frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}, \ S^{\ell}_{it} := \frac{B_{it}}{P_{Y,t}Y_{it}}, \ G^j_{M,t}(X_{it}) := \frac{F^j_{M,t}(X_{it})M_{it}}{F^j_t(X_{it})}, \ \text{and } G^j_{L,t}(X_{it}) := \frac{F^j_{L,t}(X_{it})L_{it}}{F^j_t(X_{it})}.$$

In place of Assumption 5, we may alternatively consider the case where firms produce

[1] See, for example, Miyamoto et al. (2017) for the evidence that working hours are adjusted more flexibly than the number of workers over the business cycle in Japan.

differentiated products and face a demand function with constant price elasticity as follows.

Assumption 7 (Constant Demand Elasticity)**.** *(a) A firm faces an inverse demand function with constant elasticity given by* $P_{Y,it} = Y_{it}^{-1/\sigma_Y^j} e^{\epsilon_{d,it}^j}$*, where* $\epsilon_{d,it} \notin \mathcal{I}_{it}$ *is an i.i.d. ex-post shock that is not known when* M_{it} *is chosen at time* t*. (b) A firm is a price taker for intermediate and labour inputs and the intermediate price and the market wage at time* t*,* $P_{M,t}$ *and* $P_{L,t}$*, are common across firms. (c)* $P_{Y,it}$ *and* Y_{it} *are not separately observed in the data.*

Under Assumption 6, the "revenue" production function is given by $P_{Y,it}Y_{it} = \overline{F}_t^j(X_{it})e^{\overline{\omega}_{it}+\overline{\epsilon}_{it}}$, where $\overline{F}_t^j(X_{it}) := [F_t^j(X_{it})]^{\frac{\sigma_Y^j-1}{\sigma_Y^j}}$, $\overline{\omega}_{it} := \frac{\sigma_Y^j-1}{\sigma_Y^j}\omega_{it}$, $\overline{\zeta}_{it} := \frac{\sigma_Y^j-1}{\sigma_Y^j}\zeta_{it}$, and $\overline{\epsilon}_{it} := \epsilon_{it}^d + \frac{\sigma_Y^j-1}{\sigma_Y^j}\epsilon_{it}$. Then, in place of (6), we have

$$\ln P_{Y,it}Y_{it} = \ln \overline{F}_t^j(X_{it}) + \overline{\omega}_{it} + \overline{\epsilon}_{it}, \quad \ln S_{it}^m = \ln\left(\overline{G}_{M,t}^j(X_{it})\right) + \ln\left(E_t^j[e^{\overline{\epsilon}}]\right) - \overline{\epsilon}_{it},$$
$$\ln S_{it}^\ell - \ln S_{it}^m = \ln\left(\frac{\overline{G}_{L,t}^j(X_{it})}{\overline{G}_{M,t}^j(X_{it})E_t^j[e^{\overline{\zeta}}]}\right) + \overline{\zeta}_{it}, \quad \ln P_{M,t}M_{it} = \ln\left(\frac{P_{L,t}L_{it}\overline{G}_{M,t}^j(X_{it})E_t^j[e^{\overline{\zeta}}]}{\overline{G}_{L,t}^j(X_{it})}\right) + v_{it}, \tag{7}$$

where $\overline{G}_{M,t}^j(X_{it}) := \frac{\overline{F}_{M,t}^j(X_{it})M_{it}}{\overline{F}_t^j(X_{it})}$ and $\overline{G}_{L,t}^j(X_{it}) := \frac{\overline{F}_{L,t}^j(X_{it})L_{it}}{\overline{F}_t^j(X_{it})}$. When $P_{Y,it}$ and Y_{it} are not separately observed in the data, the observable implication of (7) are the same as that of (6). In particular, we cannot separately identify the parameter σ_Y^j and the production function F_t^j. Therefore, we focus on the identification analysis under Assumption 5 although we should be careful in interpreting the empirical result because the unobserved heterogeneity in revenue production function could partly reflect in difference in demand elasticity.

4 Nonparametric identification

Assume that we have panel data of firms $i = 1, ..., N$ over periods $t = 1, ..., T$ for output, capital, intermediate inputs, the number of workers, and the total wage bills, denoted by $(Y_{it}, B_{it}, K_{it}, M_{it}, \widetilde{L}_{it}) \in \mathcal{Y} \times \mathcal{B} \times \mathcal{K} \times \mathcal{M} \times \tilde{\mathcal{L}}$, respectively. For brevity, define $\widetilde{X} := (K_{it}, M_{it}, \widetilde{L}_{it}) \in \widetilde{\mathcal{X}} := \mathcal{K} \times \mathcal{M} \times \tilde{\mathcal{L}}$. Each firm's observation $\{Y_{it}, B_{it}, \widetilde{X}_{it}\}_{t=1}^T$ is randomly sampled from a population distribution $P(\{Y_{it}, B_{it}, \widetilde{X}_{it}\}_{t=1}^T)$.

We first establish the non-parametric identification of production functions with unobserved heterogeneity under Assumptions 1-6. For notational brevity, we drop the subscript i in this section and denote $S_t = (S_t^m, S_t^\ell)$. Note that, by definition of S_t^ℓ and S_t^m, we have $Y_t = \frac{P_{M,t}M_t}{S_t^m P_{Y,t}}$ and $B_t = \frac{S_t^\ell P_{M,t}M_t}{S_t^m}$ so that the value of Y_t and B_t is known given $(S_t, \widetilde{X}_t)$ under Assumption 5. Therefore, we consider $\{S_t, \widetilde{X}_t\}_{t=1}^T$ as our data. Let $Z_t := (S_t, \widetilde{X}_t) \in \mathcal{Z} := \mathcal{S} \times \widetilde{\mathcal{X}}$.

We first establish the nonparametric identification of model structures when $J = 1$ as follows.

Proposition 1. *Suppose that $J = 1$ and Assumption 1-6 holds with $T \geq 3$. Then, (a) $\theta_1 := \{g_v(\cdot), g_{\epsilon\zeta,t}(\cdot), G_{M,t}(\cdot), G_{L,t}(\cdot), P_{L,t}\}_{t=1}^{T}$ is uniquely determined from $P(\{Z_t\}_{t=1}^{T})$. (b) $\theta_2 := \{\{F_t(\cdot)\}_{t=2}^{T}, h(\cdot), g_\eta(\cdot)\}$ is uniquely determined from $P(\{Z_t\}_{t=1}^{T})$ and θ_1.*

Remark 1. *Proposition 1 extends the identification result of GNR to the setting where L_{it} is contemporaneously determined rather than predetermined.*

When $J \geq 2$, the distribution of $\{Z_t\}_{t=1}^{T}$ follows an J-term mixture distribution

$$\mathrm{P}(\{Z_t\}_{t=1}^{T}) = \sum_{j=1}^{J} \pi^j \mathrm{P}_1^j(Z_1) \prod_{t=2}^{T} \mathrm{P}_t^j(Z_t | \{Z_{t-s}\}_{s=1}^{t-1}). \tag{8}$$

Proposition 2. *Suppose that Assumptions 1-6 hold. Then, the distribution of $\{Z_t\}_{t=1}^{T}$ defined in (8) can be written as*

$$P(\{Z_t\}_{t=1}^{T}) = \sum_{j=1}^{J} \pi^j \left(P_1^j(S_1|\widetilde{X}_1) \prod_{t=2}^{T} P_t^j(S_t|\widetilde{X}_t) \right) \times \left(P_1^j(\widetilde{X}_1) \prod_{t=2}^{T} P_t^j(\widetilde{X}_t|\widetilde{X}_{t-1}) \right). \tag{9}$$

Therefore, $\{Z_t\}_{t=1}^{T}$ follows a first order Markov process within subpopulation specified by type. The result of Proposition 2 allows us to establish the nonparametric identification of $\{\pi^j, \{\mathrm{P}_t^j(Z_t)\}_{t=1}^{T}\}_{j=1}^{J}$ by extending the argument in Kasahara and Shimotsu (2009) and Hu and Shum (2012).

Assumption 8. *Let $\mathcal{W}_t$ be the support of W_t. For every $(z_2, z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$, there exists $(\bar{z}_2, \bar{z}_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$, $(a_1, ..., a_J) \in \mathcal{Z}_1^J$ and $(b_1, ..., b_{J-1}) \in \mathcal{Z}_4^{J-1}$ such that (a) L_{z_3}, $L_{\bar{z}_3}$, $\bar{L}_{z_2}$, and $\bar{L}_{\bar{z}_2}$ defined in (34) are nonsingular, (b) $P^j(Z_3 = z_3|Z_2 = \bar{z}_2) \neq 0$ and $P^j(Z_3 = \bar{z}_3|Z_2 = z_2) \neq 0$ hold for $j = 1, ..., J$, and (c) all the diagonal elements of $D_{z_2,\bar{z}_2,z_3,\bar{z}_3}$ defined in (35) take distinct values.*

Proposition 3. *Suppose that Assumptions 1-5, and 8 hold and $T \geq 4$. Then, $\{\pi^j, P_1^j(Z_1), \{P_t^j(Z_t|Z_{t-1})\}_{t=2}^{T}\}_{j=1}^{J}$ is uniquely determined from $P(\{Z_t\}_{t=1}^{T})$.*

Remark 2. *Under the additional assumption of the stationarity, i.e., $P_t^j(Z_t|Z_{t-1}) = P^j(Z_t|Z_{t-1})$ for $t = 2, ..., T$, Kasahara and Shimotsu (2009) establishes the nonparametric identification of the model (9) when $T = 6$ while Hu and Shum (2013) shows that $T = 4$ suffices for identification.*

Remark 3. *Considering serially correlated continuous unobserved variables* $\{X_t^*\}$*, Hu and Shum (2013) analyze the nonparametric identification of the model*

$$P(\{Z_t\}_{t=1}^T) = \int P_1(Z_1, X_1^*) \prod_{t=2}^T P_t(Z_t, X_t^*|Z_{t-1}, X_{t-1}^*) d(\{X_t^*\}_{t=1}^T).$$

Given the panel data $\{Z_t\}_{t=1}^T$ *with* $T = 5$*, Theorem 1 and Corollary 1 of Hu and Shum (2013) state that, under their Assumptions 1-4,* $P_3(Z_3, X_3^*)$*,* $P_4(Z_4, X_4^*|Z_3, X_3^*)$*, and* $P_5(Z_5, X_5^*|Z_4, X_4^*)$ *are non-parametrically identified but the identification of* $P_1(Z_1, X_1^*)$*,* $P_2(Z_2, X_2^*|Z_1, X_1^*)$*, and* $P_3(Z_3, X_3^*|Z_2, X_2^*)$ *remains unresolved. Our Proposition 3 shows that, for a model in which unobserved heterogeneity is discrete and finite, we can nonparametrically identify the type-specific distribution of* $\{Z_t\}_{t=1}^T$ *including the first two periods of the data from* $T = 4$ *periods of panel data without imposing stationarity.*

Remark 4. *Assumption 8 assumes the rank condition of matrices* L_{z_3}*,* $L_{\bar{z}_3}$*,* $\bar{L}_{z_2}$*, and* $\bar{L}_{\bar{z}_2}$ *defined in (34), of which elements are constructed by evaluating* $P_4^j(Z_4|Z_3)$ *and* $\pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)$ *at different points. These conditions are similar to the assumption stated in Proposition 1 of Kasahara and Shimotsu (2009). Please refer to Remark 2 of Kasahara and Shimotsu (2009) for their interpretations. One needs to find only one pair of values* $(\bar{Z}_2, \bar{Z}_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$ *and one set of* $J-1$ *and* J *points of* Z_1 *and* Z_4 *to construct nonsingular* L_{z_3}*,* $L_{\bar{z}_3}$*,* $\bar{L}_{z_2}$*, and* $\bar{L}_{\bar{z}_2}$ *for each* $(Z_2, Z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$ *and these rank conditions are not stringent when* W_t *has continuous support. The identification of* $P_4^j(Z_4|Z_3 = Z_3)$ *and* $\pi^j P_2^j(Z_2 = Z_2|Z_1)P_1^j(Z_1)$ *at all other points of* Z_4 *and* Z_1*, respectively, follows without any further requirement on the rank condition.*

Once the type-specific distribution of $\{Z_t\}$ is identified, we may apply the argument in the proof of Proposition 1 to prove the nonparametric identification for each type's model structure.

Proposition 4. *Suppose that Assumptions 1-6, and 8 hold and* $T \geq 4$*. Then, (a)* $\theta_1 := \{\pi^j, g_v^j(\cdot), \{g_{\epsilon\zeta,t}^j(\cdot),, G_{M,t}^j(\cdot), G_{L,t}^j(\cdot), P_{L,t}, \psi_t^j\}_{t=1}^T\}_{j=1}^J$ *is uniquely determined from* $P(\{Z_t\}_{t=1}^T)$*. (b)* $\theta_2 := \{\{\{F_t^j(\cdot)\}_{t=1}^T\}_{j=2}^J, h^j(\cdot), g_\eta^j(\cdot)\}$ *is uniquely determined from* $P(\{Z_t\}_{t=1}^T)$ *and* θ_1*.*

Therefore, type-specific production functions as well as the distribution of unobserved variables can be non-parametrically identified. In estimation, we focus our attention to the case where type-specific function is given by Cobb-Douglas production function with random coefficients.

Example 1 (Random Coefficients Model)**.** *Consider a Cobb-Douglas production function with time-varying random coefficients:*

$$\tilde{f}_t^j(\tilde{x}_t) = \beta_{0,t}^j + \beta_{k,t}^j k_t + \beta_{\ell,t}^j(\psi_t^j + \tilde{\ell}_t) + \beta_{m,t}^j m_t, \tag{10}$$

where $\tilde{f}_t^j(\tilde{x}_t) := \ln F_t^j(K_t, e^{\psi_t^j}\widetilde{L}_t, M_t)$ *while the intermediate and labor cost share equations are given by*

$$s_t^m = \ln(\beta_{m,t}^j) + \ln E_t^j[e^{\epsilon}] - \epsilon_t, \quad s_t^\ell - s_t^m = \ln(\beta_{\ell,t}^j/\beta_{m,t}^j) - \ln\left(E_t^j[e^{\zeta}]\right) + \zeta_t,$$
$$m_{it} - \tilde{\ell}_{it} = \ln(P_{L,t}/P_{M,t}) + \ln\left(\beta_{m,t}^j/\beta_{\ell,t}^j\right) + \ln\left(E_t^j[e^{\zeta}]\right) + \psi_t^j + v_{it},$$

Under Assumptions 1-6, and 8, $\{\pi^j, h^j(\cdot), g_\eta^j(\cdot), g_v(\cdot), \{\beta_{\ell,t}^j, \beta_{m,t}^j, g_{\epsilon,t}^j(\cdot), g_{\zeta,t}^j(\cdot), \psi_t^j\}_{t=1}^4, \{\beta_{0,t}^j, \beta_{k,t}^j\}_{t=2}^4\}$ *for* $j = 1, ..., J$ *is nonparametrically identified from the panel data* $\{S_t, X_t\}_{t=1}^4$.

In the appendix, we discuss the conditions under which Assumption 8 holds when the production function is Cobb-Douglas. The following corollary shows that type-specific distribution of S_t can be identified from the joint distribution of $\{S_t\}_{t=1}^T$ for Cobb-Douglas specification.

Corollary 1. *Suppose that Assumptions 1-5, and 8 hold and* $T \geq 4$*. Suppose that production function is Cobb-Douglas given by (10). Then,* $\{\pi^j, \{P_t^j(S_t)\}_{t=1}^T\}_{j=1}^J$ *is uniquely determined from* $P(\{S_t\}_{t=1}^T)$.

5 Estimation of production function with random coefficients

We consider a random sample of N independent observations $\{\{Y_{it}, B_{it}, \widetilde{X}_{it}\}_{t=1}^T\}_{i=1}^N$ from the J-component mixture model $\sum_{j=1}^J \pi^j \mathrm{P}_t^j(\{Y_{it}, B_{it}, \widetilde{X}_{it}\}_{t=1}^T) = \sum_{j=1}^J \pi^j \mathrm{P}_t^j(\{S_{it}, \widetilde{X}_{it}\}_{t=1}^T)$.

We impose the following parametric assumptions for estimation.

Assumption 9. *(a) Assumption 1 holds with*

$$Y_{it} = F_t^j(K_{it}, e^{\psi_t^j}\widetilde{L}_{it}, M_{it})e^{\omega_{it}+\epsilon_{it}} \quad \text{with} \quad F_t^j(K_{it}, e^{\psi_t^j}\widetilde{L}_{it}, M_{it}) = \exp((\beta_{0,t}^j + \beta_\ell^j\psi_t^j) + \beta_k^j k_{it} + \beta_\ell^j \tilde{\ell}_{it} + \beta_m^j m_{it}). \tag{11}$$

(b) Assumption 2 holds with $\begin{pmatrix}\epsilon_{it}\\ \zeta_{it}\end{pmatrix}\Bigg| D_i = j \overset{d}{\sim} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}(\sigma_\epsilon^j)^2 & \rho_{\epsilon\zeta}^j\sigma_\epsilon^j\sigma_\zeta^j\\ \rho_{\epsilon\zeta}^j\sigma_\epsilon^j\sigma_\zeta^j & (\sigma_\zeta^j)^2\end{pmatrix}\right)$.

In (11), because $\log L_{it} = \psi_t^j + \tilde{\ell}_{it}$, the intercept term contains both $\beta_{0,t}^j$ and $\beta_\ell^j \psi_t^j$, where the latter captures the difference in the quality of workers across types. The normality assumption in Assumption 9(b) can be potentially relaxed, for example, using the maximum smoothed likelihood estimator of finite mixture models of Levine et al. (2011) in which the type-specific distribution of ϵ_{it} and ζ_{it} is non-parametrically specified. Kasahara and Shimotsu (2015) develop a likelihood-based procedure for testing the number of components in normal mixture regression models.

Denote the log values of $(Y_{it}, \tilde{L}_{it}, K_{it}, M_{it}, S_{it}^m, S_{it}^\ell, W_{it})$ by $(y_{it}, \tilde{\ell}_{it}, k_{it}, m_{it}, s_{it}^m, s_{it}^\ell, w_{it})$ and let $s_{it} := (s_{it}^\ell, s_{it}^m)$ and $\tilde{x}_{it} := (\tilde{\ell}_{it}, k_{it}, m_{it})$. Under Assumptions 3-6, 9, the first order conditions for the expected profit maximization imply that

$$s_{it}^m = \ln \beta_m^j + 0.5(\sigma_\epsilon^j)^2 - \epsilon_{it}, \quad s_{it}^\ell = \ln \beta_\ell^j + 0.5\left\{(\sigma_\epsilon^j)^2 - (\sigma_\zeta^j)^2\right\} - \epsilon_{it} + \zeta_{it}, \tag{12}$$

$$m_{it} - \tilde{\ell}_{it} = \alpha_t + \ln(\beta_m^j/\beta_\ell^j) + 0.5(\sigma_\zeta^j)^2 + \psi^j + v_{it}, \tag{13}$$

where $\alpha_t := \ln(P_{L,t}/P_{M,t})$ and (13) follows from (12), $v_{it} = b_{it} - (\psi^j + \tilde{\ell}_t + \ln P_{L,t} + \zeta_{it})$, and $s_{it}^\ell - s_{it}^m = b_{it} - (\ln P_{M,t} + m_{it})$.

We propose two different estimation procedures. The first procedure directly maximizes the log-likelihood function of a finite mixture model of production functions under additional parametric assumptions on the law of motion for k_{it} and the initial distribution of (k_{it}, ω_{it}), where the likelihood function is a parametric version of (9). Because the maximum likelihood estimator utilizes the distributional information, it is consistent even when T is small as long as $T \geq 4$. Our estimation procedure follows the two-stage identification proof of Proposition 3. The EM algorithm can be used to facilitate the computational complication of maximizing the log-likelihood function of the finite mixture model.

In the second procedure, we first estimate the partial likelihood function of the input share equations (12) under the normality assumption and use the posterior distribution of type probabilities to classify each firm observation into one of the J types under the assumption that $T \to \infty$. This generates J data sets, where a firm's production technology becomes increasingly homogenous within each of the J data sets as $T \to \infty$. In the second stage, we estimate the rest of the type-specific parameters by using each of J data sets.[2]

The first procedure can consistently estimate the parameter even when T is small as long as $T \geq 4$ and $N \to \infty$ but it is computationally more complicated and requires more auxiliary parametric assumptions than the second one. We introduce the second procedure

[2]Note that the identification of production function immediately follows from $T \to \infty$ without appealing to Proposition 3 because, in principle, each firm's production function can be identified from the time-series data of each firm.

because it is computationally much simpler than the first one although when T is small, the second procedure leads to a biased estimator due to misclassification of types.

5.1 Maximum likelihood estimator

We make the parametric distributional assumptions and develop parametric maximum likelihood estimator.

Assumption 10. *(a) T is fixed at $T \geq 4$ and $N \to \infty$. (b) Assumption 2 holds with $h^j(\omega_{it}) = \rho_\omega^j \omega_{it}$ so that*

$$\omega_{it} = \rho_\omega^j \omega_{it-1} + \eta_{it}, \tag{14}$$

$g_\eta^j(\eta) = \phi(\eta/\sigma_\eta^j)/\sigma_\eta^j$, and $g_v^j(v) = \phi(v/\sigma_v^j)/\sigma_v^j$. (c) Assumption 3 holds with the additional assumption that, conditional on being type j, k_{it} given $(k_{it-1}, \omega_{it-1})$ is normally distributed with mean $\rho_{k0}^j + \rho_{kk}^j k_{it-1} + \rho_{k\omega}^j \omega_{it-1}$ and variance $(\sigma_k^j)^2$ while the distribution of (k_{i1}, ω_{i1}) follows a bivariate normal distribution with mean μ_1^j and variance Σ_1^j.

Collect the model parameters into θ_1, and θ_2 as follows. Let

$$\theta_1 = (\boldsymbol{\pi}', \boldsymbol{\alpha}', \theta_1^1, ..., \theta_1^J)',\ \theta_2 = ((\theta_2^1)', ..., (\theta_2^J)')',\ \text{and}\ \theta^j = ((\theta_1^j)', (\theta_2^j)')',\ \text{where}\ \boldsymbol{\alpha} = (\alpha_1, ..., \alpha_T)',$$

$$\theta_1^j = (\beta_m^j, \beta_\ell^j, \psi^j, (\sigma_\epsilon^j)^2, (\sigma_\zeta^j)^2, (\sigma_v^j)^2)',\ \text{and}\ \theta_2^j = (\beta_2^j, ..., \beta_T^j, \beta_k^j, (\mu_1^j)', \mathrm{vech}(\Sigma_1^j)', \rho_{k0}^j, \rho_{kk}^j, \rho_{k\omega}^j, \sigma_k^2, \rho_\omega^j, \sigma_\eta^j)'.$$

We may write the probability density function of $\{s_{it}, \tilde{x}_{it}\}_{t=1}^T$ for type j as

$$\mathrm{f}_t^j(\{s_{it}, \tilde{x}_{it}\}_{t=1}^T) = \underbrace{\prod_{t=1}^T \mathrm{f}_t^j(s_{it}, \tilde{\ell}_{it} - m_{it}; \theta_1^j, \alpha_t)}_{=L_{1i}(\theta_1^j, \boldsymbol{\alpha})} \times \underbrace{\mathrm{f}_1^j(\tilde{x}_{it}|\tilde{\ell}_{i1} - m_{i1}; \theta^j) \prod_{t=2}^T \mathrm{f}_t^j(\tilde{x}_{it}|\tilde{\ell}_{it} - m_{it}, \tilde{x}_{it-1}; \theta^j)}_{=L_{2i}(\theta_1^j, \theta_2^j)}, \tag{15}$$

where the exact expression for $L_{1i}(\theta_1^j, \boldsymbol{\alpha})$ and $L_{2i}(\theta_2^j, \theta_1^j)$ is derived below.

Given the decomposition (15), we estimate the model by two-stage maximum likelihood estimation procedure. In the first stage, we estimate $\boldsymbol{\pi}$, $\boldsymbol{\alpha}$, and θ_1 by maximizing $\sum_{i=1}^N \log(\sum_{j=1}^J \pi^j L_{1i}(\theta_1^j, \boldsymbol{\alpha}))$ over $\boldsymbol{\pi}$ and θ_1. In the second stage, we estimate $\boldsymbol{\pi}$ and θ_2 given the first stage estimate $\hat{\boldsymbol{\alpha}}$ and $\hat{\theta}_1$ by maximizing $\sum_{i=1}^N \log(\sum_{j=1}^J \pi^j L_{1i}(\hat{\theta}_1^j, \hat{\boldsymbol{\alpha}}) L_{2i}(\hat{\theta}_1^j, \theta_2^j))$ over $\boldsymbol{\pi}$ and θ_2.

From equations (12)-(13), we can compute ϵ_{it}, ζ_{it}, and v_{it} as

$$\epsilon^*(s_{it}; \theta_1^j) := -s_{it}^m + \ln \beta_m^j + 0.5(\sigma_\epsilon^j)^2, \quad \zeta^*(s_{it}; \theta_1^j) := s_{it}^\ell - s_{it}^m - \ln(\beta_\ell^j/\beta_m^j) + 0.5(\sigma_\zeta^j)^2, \tag{16}$$

$$v^*(\tilde{\ell}_{it} - m_{it}; \theta_1^j, \alpha_t) := -\left(\tilde{\ell}_{it} - m_{it} + \alpha_t + \ln(\beta_m^j/\beta_\ell^j) + 0.5(\sigma_\zeta^j)^2 + \psi^j\right). \tag{17}$$

In the first stage, we estimate θ_1 by the maximum likelihood estimator given by

$$\hat{\theta}_1 = \underset{\theta_1}{\operatorname{argmax}} \sum_{i=1}^{N} \ln\left(\sum_{j=1}^{J} \pi^j L_{1i}(\theta_1^j)\right) \quad \text{with}$$

$$L_{1i}(\theta_1^j, \boldsymbol{\alpha}) := \prod_{t=1}^{T} \frac{1}{\sqrt{1-(\rho_{\epsilon\zeta}^j)^2}\sigma_\epsilon^j\sigma_\zeta^j}\phi\left(\frac{\epsilon^*(s_{it};\theta_1^j)}{\sigma_\epsilon^j}\right)\phi\left(\frac{\zeta^*(s_{it};\theta_1^j)-\rho_{\epsilon\zeta}^j(\sigma_\zeta^j/\sigma_\epsilon^j)\epsilon^*(s_{it};\theta_1^j)}{\sqrt{1-(\rho_{\epsilon\zeta}^j)^2}\sigma_\zeta^j}\right)$$
$$\times \frac{1}{\sigma_v^j}\phi\left(\frac{v^*(m_{it}-\tilde{\ell}_{it};\theta_1^j,\alpha_t)}{\sigma_v^j}\right).$$

In the second stage, from (11), $\epsilon_t = E[s_t^m|x_t] - s^m$, and $y_t + s_t^m = m_t + \ln(P_{M,t}/P_{Y,t})$, we have

$$\omega_{it} = \omega_t^*(m_{it}, \tilde{\ell}_{it}-m_{it}, k_{it};\theta^j) := (1-\beta_m^j-\beta_\ell^j)m_{it} - \beta_\ell^j\psi^j - \beta_t^j - \beta_\ell^j(\tilde{\ell}_{it}-m_{it}) - \alpha_k^j k_{it}, \quad (18)$$

where $\beta_t^j = \beta_{0,t}^j + \ln(P_{M,t}/P_{Y,t})$.

By the change of variables in equation (18), we can relate the density function of m_{it} conditional on $\tilde{\ell}_{it}-m_{it}$ and k_{it} to the density function of ω_{it}, denoted by $g_{\omega,t}$, as $\mathrm{f}_t^j(m_{it}|\ell_{it}-m_{it},k_{it}) = (1-\beta_m^j-\beta_\ell^j)g_{\omega,t}^j(\omega_t^*(m_{it},\tilde{\ell}_{it}-m_{it},k_{it};\theta^j))$. Then, from (18)-(17) and Assumptions 2-3, we have

$$\mathrm{f}_1^j(m_{i1}|\ell_{i1}-m_{i1},k_{i1};\theta^j) = (1-\beta_m^j-\beta_\ell^j)g_{\omega|k,1}^j(\omega_{i1}^*(\theta^j)|k_{i1}), \quad (19)$$
$$\mathrm{f}_t^j(m_{it}|\ell_{it}-m_{it},k_{it},x_{it-1};\theta^j) = (1-\beta_m^j-\beta_\ell^j)g_\eta^j(\eta_{it}^*(\theta^j)) \quad \text{for } t\geq 2, \quad (20)$$
$$\mathrm{f}_t^j(k_{it}|x_{it-1};\theta^j) = g_{k,t}^j(k_{it}|k_{it-1},\omega_{i,t-1}^*(\theta^j)) \quad \text{for } t\geq 2, \quad (21)$$

where $g_{\omega|k,1}^j(\omega_{i1}|k_{i1})$ is the density function of ω_{i1} conditional on k_{i1}, $g_{k,t}^j(k_{it}|k_{it-1},\omega_{it-1})$ is the density function of k_{it} given (k_{it-1},ω_{it-1}), $\omega_{it}^*(\theta^j) := \omega_t^*(m_{it},\ell_{it}-m_{it},k_{it};\theta^j)$, and

$$\eta_{it}^*(\theta^j) := \omega_{it}^*(\theta^j) - \rho_\omega^j\omega_{i,t-1}^*(\theta^j). \quad (22)$$

Therefore, under Assumption 10, it follows from (15) and (19)-(21) that

$$L_{2i}(\theta^j) = \mathrm{f}_1^j(m_{i1}|\ell_{i1}-m_{i1},k_{i1};\theta^j)\mathrm{f}_1^j(k_{i1};\theta^j) \times \prod_{t=2}^{T}\mathrm{f}_t^j(m_{it}|\ell_{it}-m_{it},k_{it},x_{it-1};\theta^j)\mathrm{f}_t^j(k_{it}|x_{it-1};\theta^j)$$
$$= (1-\beta_m^j-\beta_\ell^j)^T g_{\omega k,1}^j(\omega_{i1}^*(\theta^j),k_{i1})\prod_{t=2}^{T}g_\eta^j(\eta_{it}^*(\theta^j))g_{k,t}^j(k_{it}|k_{it-1},\omega_{i,t-1}^*(\theta^j))$$

where

$$g_\eta^j(\eta_{it}^*(\theta^j)) = \frac{1}{\sigma_\eta^j}\phi\left(\frac{\eta_{it}^*(\theta^j)}{\sigma_\eta^j}\right),$$

$$g_{\omega k,1}^j(\omega_{i1}^*(\theta^j), k_{i1}) = (2\pi)^{-3/2}|\Sigma_1^j|^{-1/2}\exp\left(-\frac{1}{2}\left(\begin{pmatrix} k_{i1} \\ \omega_{i1}^*(\theta^j)\end{pmatrix} - \mu_1^j\right)'(\Sigma_1^j)^{-1}\left(\begin{pmatrix} k_{i1} \\ \omega_{i1}^*(\theta^j)\end{pmatrix} - \mu_1^j\right)\right),$$

$$g_{k,t}^j(k_{it}|k_{it-1}, \omega_{i,t-1}^*(\theta^j)) = \frac{1}{\sigma_k^j}\phi\left(\frac{k_{it} - (\rho_{k0}^j + \rho_{kk}^j k_{it-1} + \rho_{k\omega}^j\omega_{it-1})}{\sigma_k^j}\right).$$

Given the first stage estimate $\hat{\theta}_1$ and $\hat{\boldsymbol{\alpha}}$, the parameter $\boldsymbol{\pi}$ and θ_2 can be estimated by maximizing the log-likelihood function as

$$(\hat{\boldsymbol{\pi}}, \hat{\theta}_2) = \underset{\boldsymbol{\pi},\theta_2}{\operatorname{argmax}} \sum_{i=1}^N \log\left(\sum_{j=1}^J \pi^j L_{1i}(\hat{\theta}_1^j, \hat{\boldsymbol{\alpha}}) L_{2i}(\hat{\theta}_1^j, \theta_2^j)\right).$$

In practice, we use EM algorithm to estimate θ_1, θ_2, and $\boldsymbol{\pi}$ as discussed in the Appendix.

5.2 Estimation by classifying each observation into one of the J types

Given the first stage estimate $\hat{\theta}_1$, define the posterior probability of being type j for each firm i by

$$\hat{\pi}_i^j = \frac{\hat{\pi}^j L_{1i}(\hat{\theta}_1^j; T)}{\sum_{k=1}^J \hat{\pi}^k L_{1i}(\hat{\theta}_1^k; T)} \quad \text{for } j = 1, ..., J, \tag{23}$$

where we explicitly write the dependence of the likelihood on the length of panel data T in $L_{1i}(\hat{\theta}_1^k; T)$. We classify each firm into one of the J types by taking the type that gives the highest posterior probability as its type. Then, for each i, our estimator of D_i is given by

$$\hat{D}_i = \underset{j=1,..,J}{\operatorname{argmax}}\{\hat{\pi}_i^j\}.$$

Denote the true value of θ_1^j by θ_1^{j*}. We assume that $T \to \infty$ but require that T goes to ∞ at much slower rate than N.

Assumption 11. *$N, T \to \infty$ and $\frac{\sqrt{N}}{exp(a^j T)/\sqrt{T}} \to 0$ for $j = 1, ..., J$, where $a^j = \min_{k \neq j} E[\ln L_{1it}(\theta_1^{j*}) - \ln L_{1it}(\theta_1^{k*}) | i \in \mathcal{I}^j] > 0$.*

Proposition 5. *For each $i \in \mathcal{I}^j$, $\hat{\pi}_i^j - 1 = o_p(N^{-1/2})$ under Assumption 11.*

Proposition 5 implies that, when Assumption 11 holds, the possible classification error across types does not affect our inference.

In the second stage, we compute the estimate of η_{it}^j for $t = 2, ..., T$ for each candidate value of θ_2^j given the first stage estimate $\hat{\theta}_1^j$ as in (22) using the subsample of firms for which $\hat{D}_i = j$. Then, stacking the moment conditions implied by $E[\hat{\eta}_{it}^*(\hat{\theta}_1^j, \theta_2^j)|k_{it}, x_{it-1}] = 0$ for $t = 2, ..., T$, we can use standard GMM procedure to estimate θ_2^j as

$$\hat{\theta}_2^j = \underset{\theta_2}{\operatorname{argmin}} \left(\frac{1}{\#\{i : \hat{D}_i = j\}} \sum_{i \in \{i:\hat{D}_i = j\}} g_i(\theta_2) \right) \left(\frac{1}{\#\{i : \hat{D}_i = j\}} \sum_{i \in \{i:\hat{D}_i = j\}} g_i(\theta_2) \right)' \quad \text{for } j = 1, ..., J,$$

where $\#\{i : \hat{D}_i = j\}$ is the number of firms with $\hat{D}_i = j$ while $g_i(\theta_2) := (\eta_{i2}^*(\hat{\theta}_1^j, \theta_2^j) Z_{i2}(\hat{\theta}_1^j, \theta_2^j)', ..., \eta_{iT}^*(\hat{\theta}_1^j, \theta_2^j) Z_{iT}(\hat{\theta}_1^j, \theta_2^j)')'$ with $Z_{it}(\hat{\theta}_1^j, \theta_2^j) := (1, k_{it}, \omega_{it-1}^*(\hat{\theta}_1^j, \theta_2^j))'$.

6 Empirical Application

6.1 Data

We use plant-level panel data from the Census of Manufacture of Japan for 1986-2010. The data set contains production information for the manufacturing industry in Japan. We focus on plants with 30 or more employees because detailed data are consistently available for these plants only.[3] At the 4-digit level of industry classification available in the Census of Manufacture, we have 276 industries in total. In the empirical application, we mainly focus on concrete products and electric audio equipment because 1. both industries have a large number of observations; 2. the former features a relatively small variation in the intermediate input share, while the latter features a large variation as we show in Section 2. Therefore, it is useful to discuss these two industries to examine the importance of unobserved heterogeneity.

The output (Y) is defined as the sum of shipments, revenues from repairing and fixing services, and revenue from performing subcontracted work. The initial value of capital (K) is defined as fixed asset less land and the subsequent values of capital are constructed by perpetual inventory method. The observed labor input ($\tilde{L}$) is the number of employees. The intermediate input (M) is defined as the sum of material input, energy input, subcontracting

[3]The survey uses different questionnaires depending on the plant size: 1. Plants with 30 or more employees; 2. 4-29 employees; 3. 1-3 employees. The questionnaire for the plants with 30 or more employees is more detailed. For example, from 2000, the census collects fixed asset data every 5 years only for plants with less than 30 employees.

Table 3: Summary statistics

	Concrete Products					Electric Audio Equipment				
	Obs	Mean	Std. Dev	Min	Max	Obs	Mean	Std. Dev	Min	Max
y_{it}	13892	11.37	0.68	7.6	14.39	10955	11.24	1.79	5.51	17.81
$\tilde{\ell}_{it}$	13892	3.97	0.41	3.4	6.81	10955	4.52	0.9	3.4	8.53
k_{it}	13892	10.91	0.86	5.22	14.06	10955	10.04	1.99	1.67	16.22
m_{it}	13892	10.34	0.83	-0.13	13.89	10955	10.37	2.31	3.27	16.89
s_{it}^m	13892	-0.98	0.35	-1.83	-0.33	10955	-0.99	0.79	-3.13	-0.08
s_{it}^ℓ	13892	-1.47	0.43	-2.39	-0.5	10955	-1.39	0.78	-3.11	-0.16
I_{it}/K_{it}	13892	0.1	1.06	-1.04	115.02	10955	0.33	11.46	-1.28	1030.83

expenses for consigned production. Flow data such as shipments and various production costs refer to the calendar year. The number of employees refers to the value at the end of the year, while the stock of fixed assets refers to the beginning of the period. Table 3 presents summary statistics for the variables we use in our empirical analysis.

6.2 Estimation of production function

This section presents estimation results for a random-coefficient Cobb-Douglas production function with 3 technology types and 2 unobserved labor type within each technology type. Tables 4 and 5 report parameter estimates with unobserved heterogeneity ($J = 3\times 2 = 6$) and with homogeneous production function ($J = 1$) for the concrete products and electric audio equipment industries. For both industries, the estimated coefficients suggest substantial differences in the output elasticities with respect to labor, capital, and intermediate inputs across different types of firms.

Comparing the two industries, the variation in $\hat{\beta}_m^j$ across types is larger for electric audio equipment than concrete products, which is expected from the dispersion of the intermediate input share shown in Section 2. Because $\hat{\beta}_\ell^j$ and $\hat{\beta}_k^j$ also varies across types, we have variation in the capital-labor ratio $\hat{\beta}_k^j/\hat{\beta}_\ell^j$ across types as well. For electric audio equipment, it ranges from 0.24 (Type 1 and 2) to 1.53 (Type 5 and 6), while it ranges from 0.52 to 0.79 for concrete products. As we show below, $\hat{\beta}_k^j/\hat{\beta}_\ell^j$ is an important determinant of the response of capital investment to productivity $\hat{\omega}_{it}$. The returns to scale $(\hat{\beta}_m^j + \hat{\beta}_\ell^j + \hat{\beta}_k^j)$ is around 0.7 for concrete products, while it ranges from 0.77 to 0.93 for electric audio equipment. The estimates of $\hat{\psi}^j$ suggest that there is substantial unobserved heterogeneity in labor inputs across types.

Figures 5 and 6 show the distribution of posterior type probabilities, defined by $\hat{\pi}_i^j = \frac{\hat{\pi}^j L_i(\hat{\theta}^j)}{\sum_{k=1}^J \hat{\pi}^k L_i(\hat{\theta}^k)}$ for $j = 1, ..., J$, across plants for the model with $J = 6$. The posterior probabil-

Table 4: Estimates of Production Function (Concrete Products)

	J = 1	J = 6					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
β_m^j	0.332	0.273		0.344		0.448	
	(0.004)	(0.005)		(0.006)		(0.009)	
β_ℓ^j	0.224	0.283		0.212		0.161	
	(0.003)	(0.006)		(0.003)		(0.004)	
β_k^j	0.075	0.146		0.168		0.102	
	(0.015)	(0.021)		(0.007)		(0.012)	
$\beta_m^j + \beta_\ell^j + \beta_k^j$	0.632	0.701		0.723		0.711	
β_k^j/β_ℓ^j	0.336	0.517		0.791		0.637	
ψ^j	0.000	-0.654	-0.041	-0.235	0.182	0.002	0.746
	(NA)	(0.178)	(0.050)	(0.040)	(0.041)	(0.044)	(-)
π	1.000	0.046	0.268	0.230	0.224	0.175	0.058
	(NA)	(0.009)	(0.019)	(0.016)	(0.019)	(0.022)	(0.012)
Obs	13892						
No. Plants	914						

ities for each type are concentrated on around 0 or 1. In the subsequent analysis, we assign one of the J types to each plant based on its posterior type probability that achieves the highest value across J types.

Ignoring unobserved heterogeneity may lead to substantial biases in the measurement of productivity growth. To examine this issue, we take a specification with $J = 6$ as the true model and compute the bias in the measurement of productivity growth when we use a misspecified model with $J = 1$. Specifically, let $\Delta\omega_{it} := \Delta y_{it} - (\hat{\beta}_t^j + \hat{\beta}_m^j \Delta m_{it} + \hat{\beta}_\ell^j \Delta\tilde{\ell}_{it} + \hat{\beta}_k^j \Delta k_{it} + \Delta\hat{\epsilon}_{it}^j)$ for $j = 1, 2, ..., 6$ be an estimated productivity growth when $J = 6$ and let $\Delta\tilde{\omega}_{it} := \Delta y_{it} - (\bar{\beta}_t + \bar{\beta}_m \Delta m_{it} + \bar{\beta}_\ell \Delta\tilde{\ell}_{it} + \bar{\beta}_k \Delta k_{it} + \Delta\bar{\epsilon}_{it})$ be an estimated productivity growth when $J = 1$, where $\{\hat{\beta}_t^j, \hat{\beta}_m^j, \hat{\beta}_\ell^j, \hat{\beta}_k^j\}_{j=1}^6$ and $\{\bar{\beta}_t, \bar{\beta}_m^j, \bar{\beta}_\ell^j, \bar{\beta}_k^j\}$ denote estimated coefficients when $J = 6$ and $J = 1$, respectively. Then, we compute the bias as

$$\Delta\tilde{\omega}_{it} = \Delta\omega_{it} + \underbrace{(\bar{\beta}_m - \hat{\beta}_m^j)\Delta m_{it} + (\bar{\beta}_\ell - \hat{\beta}_\ell^j)\Delta\ell_{it} + (\bar{\beta}_k^j - \hat{\beta}_k^j)\Delta k_{it} + (\Delta\bar{\epsilon}_{it} - \Delta\hat{\epsilon}_{it}^j)}_{:=\text{Bias}_{it}}.$$

The first row of Table 6, designated by $\frac{\text{Mean of } |\text{Bias}_{it}|}{\text{Mean of } |\Delta\tilde{\omega}_{it}|}$, reports the ratio of the average absolute value of bias to the average productivity growth within each of three subsamples, classified by technology types. The magnitude of the bias ranges from 0.07 to 0.18 for concrete products, while it ranges from 0.14 to 0.44 for electric audio equipment. The second row of Table 6, designated by $\left.\frac{\text{Mean of Bias}_{it}}{\text{Mean of } |\Delta\tilde{\omega}_{it}|}\right|_{\Delta\omega_{it}>0}$, reports the ratio of the average value of bias to the average

Table 5: Estimates of Production Function (Electric Audio Equipment)

	J = 1	J = 6					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
β_m^j	0.281	0.136		0.364		0.597	
	(0.008)	(0.007)		(0.022)		(0.01)	
β_ℓ^j	0.296	0.507		0.228		0.132	
	(0.006)	(0.012)		(0.012)		(0.006)	
β_k^j	0.076	0.122		0.185		0.202	
	(0.016)	(0.01)		(0.021)		(0.017)	
$\beta_m^j + \beta_\ell^j + \beta_k^j$	0.652	0.765		0.778		0.932	
β_k^j/β_ℓ^j	0.256	0.242		0.813		1.526	
ψ^j	0.000	-0.999	0.474	-0.258	0.201	-0.215	0.798
	(NA)	(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
	(NA)	(0.023)	(0.017)	(0.163)	(0.149)	(0.015)	(0.014)
Obs	10913						
No. Plants	907						

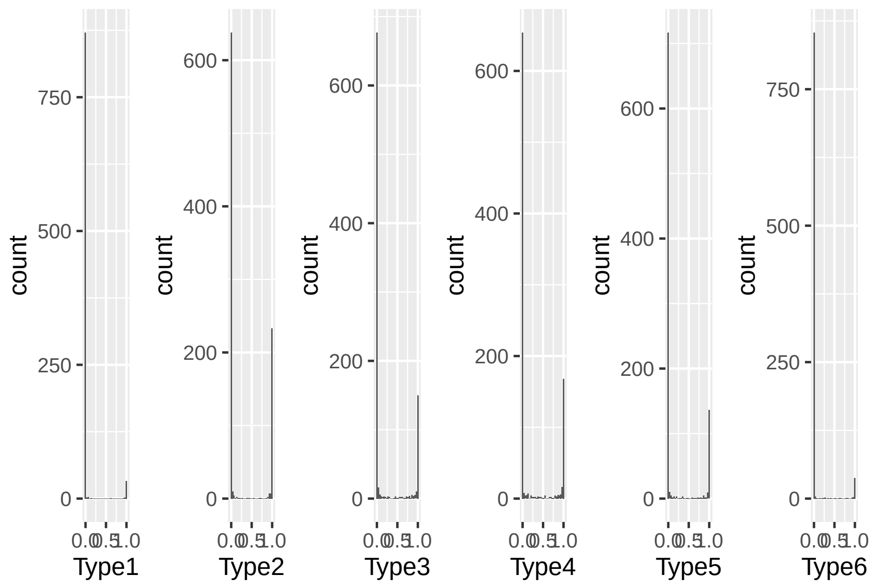
Table 6: Bias in $\Delta\tilde{\omega}$

	Concrete Products			Electric Audio Equipment		
	J = 6			J = 6		
	Type 1 - 2	Type 3 - 4	Type 5 - 6	Type 1 - 2	Type 3 - 4	Type 5 - 6
$\frac{\text{Mean of } \lvert\text{Bias}_{it}\rvert}{\text{Mean of } \lvert\Delta\tilde{\omega}_{it}\rvert}$	0.126	0.074	0.181	0.226	0.140	0.440
$\left.\frac{\text{Mean of Bias}_{it}}{\text{Mean of } \lvert\Delta\tilde{\omega}_{it}\rvert}\right\rvert_{\Delta\omega_{it}>0}$	-0.098	0.002	0.167	-0.200	0.084	0.247
β_m^j	0.273	0.344	0.448	0.136	0.364	0.597

productivity growth conditional on positive productivity growth measured by the model with $J = 6$. Note that $\text{Bias}_{it} \approx (\hat{\beta}_m^j - \bar{\beta}_m)\Delta m_{it}$ and $Corr(\Delta\omega_{it}, \Delta m_{it}) > 0$. Thus, the average bias conditional on $\Delta\omega_{it} > 0$ tends to be positive when $\hat{\beta}_m^j > \bar{\beta}_m^j$. The empirical results confirm this pattern: for both concrete products and electric audio equipment, Types 3-4 and 5-6 have $\hat{\beta}_m^j$ higher than $\bar{\beta}_m$ and thus the estimated bias $\left.\frac{\text{Mean of Bias}_{it}}{\text{Mean of } \lvert\Delta\tilde{\omega}_{it}\rvert}\right\rvert_{\Delta\omega_{it}>0}$ are positive for these types. These results suggest that ignoring unobserved heterogeneity could result in serious bias in estimated productivity growth and the bias is likely to have a systematic pattern depending on the values of $\hat{\beta}_m^j$.

As an example of using the estimated productivity growth in empirical analysis, we now examine whether unobserved heterogeneity captured by type-specific production function parameters is important for an investment decision. Specifically, for each subsample classified

Figure 5: Posterior Probabilities (Concrete Products)



by type, we estimate the following linear investment model

$$\frac{I_{it}}{K_{it}} = \alpha_0 + \alpha_\omega^j \hat{\omega}_{it} + \text{quadratic of } k_{it} + \zeta_{it},$$

where I_{it}/K_{it} denotes the ratio of investment to capital stock.

Table 7 reports the OLS estimates of α_ω^j in the first row as well as the quantile regression estimates of α_ω^j at the 10th, 25th, 50th, 75th, and 90th percentiles across different types for $J = 1$ and 6 for concrete products. Table 8 reports the same estimates for the electric audio equipment industry. When $J = 1$, the OLS coefficient of ω_{it} is estimated significantly at 0.06 for concrete products and at 0.5 for electric audio equipment.

For the model with $J = 6$, the estimated coefficients of ω_{it} substantially differ across different types of plants, suggesting that the investment response to a productivity shock differs across plants. In both OLS and quantile regression results, the estimated coefficients tend to be higher for the types with higher $\hat{\beta}_m^j$ and $\hat{\beta}_k^j/\hat{\beta}_\ell^j$, suggesting that firms invest more given a positive productivity shock if they have production technology with high material shares and high capital-labor ratios. In the case of quantile regressions, this pattern is particularly pronounced for firms with high investment ratios.

Figure 6: Posterior Probabilities (Electric Audio Equipment)

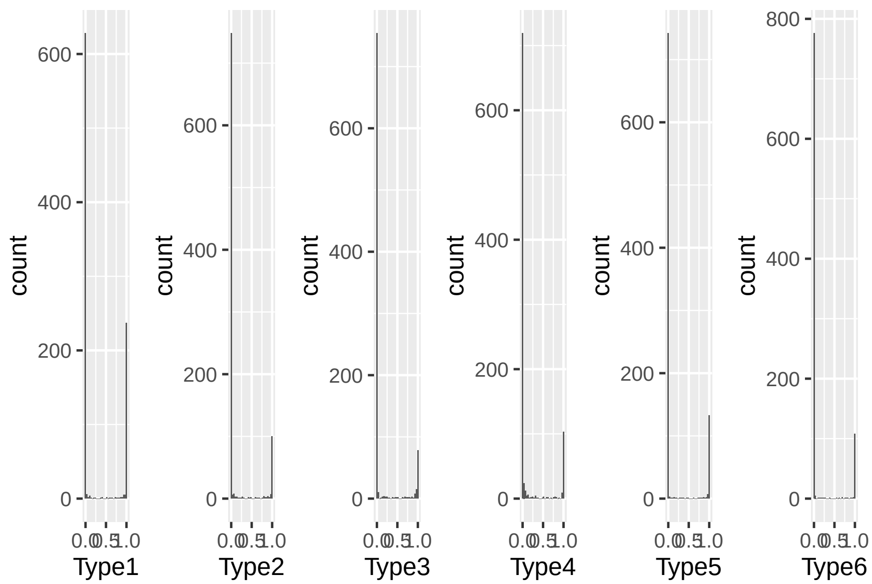


Table 7: The Effect of ω_{it} on Investment (Concrete Products)

	J = 1	J=6					
		Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
α_ω^j	0.06	0.06	−0.02	0.08	0.06	0.01	0.12
	(0.02)	(0.04)	(0.10)	(0.01)	(0.01)	(0.06)	(0.03)
$\alpha_\omega^j(0.10)$	0.00	−0.00	0.00	0.00	0.01	0.02	0.00
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
$\alpha_\omega^j(0.25)$	0.01	0.00	0.00	0.03	0.02	0.03	0.02
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
$\alpha_\omega^j(0.50)$	0.02	0.02	0.02	0.04	0.04	0.04	0.04
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
$\alpha_\omega^j(0.75)$	0.04	0.05	0.03	0.08	0.07	0.07	0.08
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
$\alpha_\omega^j(0.90)$	0.07	0.06	0.07	0.13	0.10	0.09	0.14
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)
β_m^j	0.33	0.27	0.27	0.34	0.34	0.45	0.45
β_k^j/β_ℓ^j	0.34	0.52	0.52	0.79	0.79	0.64	0.64
$\overline{(I_{it}/K_{it})}$	0.10	0.13	0.14	0.08	0.07	0.12	0.09

Table 8: The Effect of ω_{it} on Investment (Electric Audio Equipment)

	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)
$\alpha_\omega^j(0.10)$	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)
$\alpha_\omega^j(0.25)$	0.00	0.00	0.00	0.01	0.00	0.01	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(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)
$\alpha_\omega^j(0.75)$	0.03	0.02	0.04	0.08	0.03	0.06	0.07
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
$\alpha_\omega^j(0.90)$	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	0.14	0.36	0.36	0.60	0.60
β_k^j/β_ℓ^j	0.26	0.24	0.24	0.81	0.81	1.53	1.53
$\overline{(I_{it}/K_{it})}$	0.34	0.34	0.97	0.15	0.47	0.08	0.11

References

[1] Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes (2007) "Econometric Tools For Analyzing Market Outcomes." In *Handbook of Econometrics*, vol. 6, edited by James J. Heckman and Edward E. Leamer. Amsterdam: Elsevier, 4171-4276.

[2] Ackerberg, Daniel A., Kevin Caves, and Garth Frazer (2015)"Identification Properties of Recent Production Function Estimators," *Econometrica*, 83(6): 2411-2451.

[3] Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* 58: 277- 297.

[4] Balat, Jorge, Irene Brambilla, and Yuya Sasaki (2019) "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports," mimeo.

[5] Blundell, R. and Bond, S. (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87: 115-143.

[6] Blundell, R.W. and S.R. Bond (2000) "GMM Estimation with Persistent Panel Data: an Application to Production Functions," *Econometric Reviews*, 19(3): 321-340.

[7] Bond, Stephen and Mans Sderbom (2005) "Adjustment Costs and the Identification of Cobb Douglas Production Functions," The Institute for Fiscal Studies, Working Paper Series No. 05/4.

[8] Basu, Susanto, and John G. Fernald (1997) "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105(2): 249-283.

[9] Chen, Yuyu, Mitsuru Igami, Masayuki Sawada, and Mo Xiao (2021) "Privatization and productivity in China," *RAND Journal of Economics*, 52(4): 884-916.

[10] Demirer, Mert (2020) "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups," mimeo.

[11] Doraszelski, Ulrich, and Jordi Jaumandreu (2018) "Measuring the Bias of Technological Change," *Journal of Political Economy*, 126(3): 1027-1084.

[12] Doty, Justin (2022) "A Dynamic Framework for Identification and Estimation of Nonseparable Production Functions," mimeo.

[13] Gandhi, A., S. Navarro, and D. Rivers (2020) "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 128(8): 2973-3016.

[14] Griliches, Z. and J. Hausman (1986) "Errors in Variables in Panel Data," *Journal of Econometrics*, 31(1): 93-118.

[15] Griliches, Zvi and Jacques Mairesse (1998) "Production Functions: The Search for Identification." In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium.* New York: Cambridge University Press.

[16] Hsieh, Chang-Tai and Peter J. Klenow (2009) "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124 (4):1403-1448.

[17] Hu, Yingyao and Matthew Shum (2012) "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171(1): 32-44.

[18] Kasahara, Hiroyuki and Joel Rodrigue (2008) "Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence." *Journal of Development Economics* 87 (1):106-118.

[19] Kasahara, Hiroyuki and Katsumi Shimotsu(2009) "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77: 135–175.

[20] Kasahara, Hiroyuki and Katsumi Shimotsu (2015) "Testing the Number of Components in Normal Mixture Regression Models," *Journal of the American Statistical Association (Theory and Methods)*, 110(512): 1632-1645.

[21] Levine, M., Hunter, D. R., and Chauveau, D. (2011) "Maximum smoothed likelihood for multivariate mixtures," *Biometrika*, 98, 403-416.

[22] Levinsohn, James and Amil Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70: 317-341.

[23] Li, Tong and Yuya Sasaki (2017) "Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function," mimeo.

[24] Marschak, J. and Andrews, W.H. (1944) "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12(3,4): 143-205.

[25] Mairesse, Jacques, and Zvi Griliches (1990) "Heterogeneity in Panel Data: Are There Stable Production Functions?" In *Essays in Honor of Edmond Malinvaud, Volume 3:*

Empirical Economics, edited by Paul Champsaur et al., pp. 192-231, Cambridge, MA: MIT Press.

[26] Olley, S. and Pakes, A. (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 65(1): 292-332.

[27] Pavcnik, Nina (2002) "Trade Liberalization, Exit, and Productivity Improvements: Evidence From Chilean Plants," *Review of Economic Studies*, 69(1): 245-276.

[28] Van Biesebroeck, Johannes (2003) "Productivity Dynamics with Technology Choice: An Application to Automobile Assembly," *Review of Economic Studies* 70: 167-198.

A Appendix

A.1 Proof of Proposition 1

We drop the superscript j and we have $L_t = \widetilde{L}_t$ and $X_t = \widetilde{X}_t$ because $J = 1$. Let $(s_t^\ell, s_t^m) = (\ln S_t^\ell, \ln S_t^m)$ and let $\Delta s_t := s_t^\ell - s_t^m$. Denote the density function of $(s_t^m, \Delta s_t, X_t)$ by $p_t(s_t^m, \Delta s_t, X_t)$, which can be identified from $P_t(S_t, X_t)$, and let $g_{\epsilon,t}(\cdot)$ and $g_{\zeta,t}(\cdot)$ be the marginal densities of ϵ_t and ζ_t, respectively. Because $E[s_t^m|X_t] = \ln(G_{M,t}(X_t)E_t[e^\epsilon])$ and $E[\Delta s_t|X_t] = \ln\left(G_{L,t}(X_t)/G_{M,t}E_t[e^\zeta]\right)$, we have $s_t^m = E[s_t^m|X_t] - \epsilon_t$ and $\Delta s_t = E[\Delta s_t|X_t] - \zeta_t$ from the second and third equations in (6). Then, we may identify $g_{\epsilon\zeta,t}(\cdot)$ as $g_{\epsilon\zeta,t}(\epsilon,\zeta) = \int p_t(E[s_t^m|X_t = x] - \epsilon, E[\Delta s_t|X_t = x] - \zeta, x)dx$. Furthermore, from $E[s_t^m|X_t] = \ln G_{M,t}(X_t) + \ln \int e^\epsilon g_{\epsilon,t}(\epsilon)d\epsilon$, we may identify $G_{M,t}(X_t)$ as $G_{M,t}(X_t) = \exp\left(E[s_t^m|X_t] - \ln \int e^\epsilon g_{\epsilon,t}(\epsilon)d\epsilon\right)$ and, similarly, $G_{L,t}(X_t) = \exp\left(E[\Delta s_t|X_t] + \ln \int e^\zeta g_{\zeta,t}(\zeta)d\zeta\right)$. Given the identification of $G_{M,t}(X)$, $G_{L,t}(X)$, and $g_{\zeta,t}(\zeta)$, we may identify $g_{v,t}(v)$ from the density of X_t because $v_t = \ln P_{M,t}M_t - \ln\left(L_{it}G_{L,t}(X_t)\int e^\zeta g_{\zeta,t}(\zeta)d\zeta/G_{M,t}(X_t)\right)$. $P_{L,t}$ is identified as $\ln P_{L,t} = E_t[\ln\left(S_t^\ell P_{M,t}M_t/S_t^m L_t\right)]$ given that v_t and ζ_t are mean zero random variables. This proves part (a).

We proceed to prove part (b). Fix $(L_0, M_0) \in \mathcal{L} \times \mathcal{M}$ such that $L_0 < L_t$ and $M_0 < M_t$. Because $\frac{G_{L,t}(X_t)}{L_t} = \frac{\partial \ln F_t(X_t)}{\partial L_t}$ and $\frac{G_{M,t}(X_t)}{M_t} = \frac{\partial \ln F_t(X_t)}{\partial M_t}$, we have

$$\ln F_t(K_t, L_t, M_t) = \int_{L_0}^{L_t} \frac{G_{L,t}(K_t, L, M_t)}{L} dL + \int_{M_0}^{M_t} \frac{G_{M,t}(K_t, L_0, M)}{M} dM + \ln F_t(K_t, L_0, M_0). \tag{24}$$

It follows from (2), (24), $\epsilon_t = E[s_t^m|X_t] - s^m$, and $\ln Y_t + s_t^m = \ln M_t + \ln(P_{M,t}/P_{Y,t})$. that

$$\omega_t = \tilde{y}_t(X_t;\theta_1) - \ln F_t(K_t, L_0, M_0), \quad \text{where} \tag{25}$$
$$\tilde{y}_t(X_t;\theta_1) := \ln M_t + \ln(P_{M,t}/P_{Y,t}) - \left\{ \int_{L_0}^{L_t} \frac{G_{L,t}(K_t, L, M_t)}{L} dL + \int_{M_0}^{M_t} \frac{G_{M,t}(K_t, L_0, M)}{M} dM - E[s_t^m|X_t] \right\}.$$

Substituting the right-hand side of (25) to $\omega_t = h(\omega_{t-1}) + \eta_t$ and rearranging terms give

$$\tilde{y}_t(X_t;\theta_1) = \ln F_t(K_t, L_0, M_0) + h\left(\tilde{y}_{t-1}(X_{t-1};\theta_1) - \ln F_{t-1}(K_{t-1}, L_0, M_0)\right) + \eta_t, \tag{26}$$

where the second term on the right hand side only depends on X_{t-1}. Fix $K_0 \in \mathcal{K}$ and let $C_t := \ln F_t(K_0, L_0, M_0)$. Then, from (26) and $E[\eta_t|\mathcal{I}_{t-1}] = 0$, $\ln F_t(K_t, L_0, M_0)$ is identified up to constant C_t as

$$\ln F_t(K_t, L_0, M_0) = C_t + E[\tilde{y}_t(X_t;\theta_1)|K_t, X_{t-1}] - E[\tilde{y}_t(X_t;\theta_1)|K_t = K_0, X_{t-1}]. \tag{27}$$

It follows from the moment restriction $E[\omega_t] = 0$ with (25) and (27) that we may identify C_t as

$$C_t = E\left\{\tilde{y}_t(X_t;\theta_1) - E[\tilde{y}_t(X_t;\theta_1)|K_t, X_{t-1}] + E[\tilde{y}_t(X_t;\theta_1)|K_t = K_0, X_{t-1}]\right\}.$$

Therefore, $\ln F_t(K_t, L_0, M_0)$ is identified from (27), and the identification of $\ln F_t(L_t, K_t, M_t)$ for $t \geq 2$ follows from (24) given that the first two terms on the right hand side of (24) is identified from and $G_{L,t}(X_t)$ and $G_{M,t}(X_t)$.

Finally, we prove the identification of $g_\eta(\cdot)$ and $h(\cdot)$. Because $\omega_t = \tilde{y}_t(X_t;\theta_1) - \ln F_t(K_t, L_0, M_0)$, we may identify the joint density function of ω_t and ω_{t-1}, denoted by $p_\omega(\omega_t, \omega_{t-1})$, from the joint distribution of (X_t, X_{t-1}) for $t \geq 3$. Then, $h(\omega_{t-1})$ is identified as $h(\omega_{t-1}) = E_t[\omega_t|\omega_{t-1}] = \int \omega_t p_\omega(\omega_t|\omega_{t-1}) d\omega_t$, where $p_\omega(\omega_t|\omega_{t-1}) = p_\omega(\omega_t, \omega_{t-1}) / \int p_\omega(\omega_t, \omega_{t-1}) d\omega_t$, while the density function of η_t is identified as $g_\eta(\eta) = \int p_\omega(h(\omega_{t-1}) + \eta, \omega_{t-1}) d\omega_{t-1}$. This proves part (b). $\square$

A.2 Proof of Proposition 2

The distribution of $\{S_t, \widetilde{X}_t\}_{t=1}^T$ for type j is given by

$$\begin{aligned}
\mathrm{P}_t^j(\{S_t, \widetilde{X}_t\}_{t=1}^T) &= \mathrm{P}_1^j(S_1, \widetilde{X}_1)\prod_{t=2}^T \mathrm{P}_t^j(S_t, \widetilde{X}_t|\{S_{t-s}, \widetilde{X}_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_1^j(S_1|\widetilde{X}_1)\mathrm{P}_1^j(\widetilde{X}_1)\prod_{t=2}^T \mathrm{P}_t^j(S_t|\widetilde{X}_t, \{S_{t-s}, \widetilde{X}_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(\widetilde{X}_t|\{S_{t-s}, \widetilde{X}_{t-s}\}_{s=1}^{t-1}).
\end{aligned} \tag{28}$$

In view of the second and the third equations of (6), we have

$$\mathrm{P}_t^j(S_t|\widetilde{X}_t, \{S_{t-s}, \widetilde{X}_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(S_t|\widetilde{X}_t). \tag{29}$$

Furthermore,

$$\begin{aligned}
&\mathrm{P}_t^j(\widetilde{X}_t|\{S_{t-s}, \widetilde{X}_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(K_t, \omega_t, v_t|\{S_{t-s}, K_{t-s}, \omega_{t-s}, v_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_t^j(\omega_t, v_t|K_t, \{S_{t-s}, K_{t-s}, \omega_{t-s}, v_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(K_t|\{S_{t-s}, K_{t-s}, \omega_{t-s}, v_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_\omega^j(\omega_t|\omega_{t-1})\mathrm{P}_v^j(v_t)\mathrm{P}_t^j(K_t|K_{t-1}, \omega_{t-1}) \\
&= \mathrm{P}_t^j(K_t, \omega_t, v_t|K_{t-1}, \omega_{t-1}) \\
&= \mathrm{P}_t^j(K_t, \omega_t, v_t|K_{t-1}, \omega_{t-1}, v_{t-1}) \\
&= \mathrm{P}_t^j(\widetilde{X}_t|\widetilde{X}_{t-1}),
\end{aligned} \tag{30}$$

where the first equality and the last equality hold because there is a one-to-one mapping between $\widetilde{X}_t$ and (K_t, ω_t, v_t) in view of Assumption 4(b), the third equality follows from Assumptions 2 and 3, the fifth equality holds because v_t is i.i.d.. Therefore, the stated result follows from (28)-(30). □

A.3 Proof of Proposition 3

We apply the argument of Kasahara and Shimotsu (2009) and Hu and Shum (2012) under the assumption that unobserved heterogeneity is permanent and discrete. The proof is constructive.

Consider the case that $T = 4$. Fix (Z_2, Z_3) at (Z_2, Z_3) and choose $(\bar{Z}_2, \bar{Z}_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$, $(a_1, ..., a_J) \in \mathcal{Z}_1^J$ and $(b_1, ..., b_{J-1}) \in \mathcal{Z}_4^{J-1}$ that satisfy Assumption 8. Evaluating (9) at

$(Z_2, Z_3) = (Z_2, Z_3)$ gives

$$
\begin{aligned}
P(\{Z_t\}_{t=1}^4) &= \sum_{j=1}^J \pi^j P_4^j(Z_4|Z_3) P_3^j(Z_3|Z_2) P_2^j(Z_2|Z_1) P_1^j(Z_1) \\
&= \sum_{j=1}^J \lambda_4^j(Z_4|Z_3) \lambda_3^j(Z_3|Z_2) \bar{\lambda}_2^j(Z_1, Z_2),
\end{aligned}
\tag{31}
$$

where $\lambda_4^j(Z_4|Z_3) := P_4^j(Z_4|Z_3 = Z_3)$, $\lambda_3^j(Z_3|Z_2) := P_3^j(Z_3 = Z_3|Z_2 = Z_2)$, and $\bar{\lambda}_2^j(Z_1, Z_2) := \pi^j P_2^j(Z_2 = Z_2|Z_1) P_1^j(Z_1)$. Integrating out Z_4 from (31) gives

$$
P(\{Z_t\}_{t=1}^3) = \sum_{j=1}^J \lambda_3^j(Z_3|Z_2) \bar{\lambda}_2^j(Z_1, Z_2).
\tag{32}
$$

Let $f_{Z_2,Z_3}(a, b) := P((Z_1, Z_2, Z_3, Z_4) = (a, Z_2, Z_3, b))$ and $\bar{f}_{Z_2,Z_3}(a) := P((Z_1, Z_2, Z_3) = (a, Z_2, Z_3))$. Evaluating (31) at $Z_1 = a_1, ..., a_J$ and $Z_4 = b_1, ..., b_{J-1}$ gives $M(M-1)$ equations while evaluating (32) at $Z_1 = a_1, ..., a_J$ gives M equations. Collecting these $M(M-1) + M = M^2$ equations and denoting them using matrix notation, we have

$$
P_{Z_2,Z_3} = L'_{z_3} D_{Z_3|Z_2} \bar{L}_{z_2},
\tag{33}
$$

where

$$
P_{Z_2,Z_3} := \begin{bmatrix}
\bar{f}_{Z_2,Z_3}(a_1) & \bar{f}_{Z_2,Z_3}(a_2) & \cdots & \bar{f}_{Z_2,Z_3}(a_J) \\
f_{Z_2,Z_3}(a_1, b_1) & f_{Z_2,Z_3}(a_2, b_1) & \cdots & f_{Z_2,Z_3}(a_J, b_1) \\
\vdots & \vdots & \ldots & \vdots \\
f_{Z_2,Z_3}(a_1, b_{J-1}) & f_{Z_2,Z_3}(a_2, b_1) & \cdots & f_{Z_2,Z_3}(a_J, b_{J-1})
\end{bmatrix},
$$

$$
L_{z_3} := \begin{bmatrix}
1 & \lambda_4^1(b_1|Z_3) & \cdots & \lambda_4^1(b_{J-1}|Z_3) \\
\vdots & \vdots & \ldots & \vdots \\
1 & \lambda_4^J(b_1|Z_3) & \cdots & \lambda_4^J(b_{J-1}|Z_3)
\end{bmatrix}, \quad
\bar{L}_{z_2} := \begin{bmatrix}
\bar{\lambda}_2^1(a_1, Z_2) & \cdots & \bar{\lambda}_2^1(a_J, Z_2) \\
\vdots & \vdots & \ldots \\
\bar{\lambda}_2^J(a_1, Z_2) & \cdots & \bar{\lambda}_2^J(a_J, Z_2)
\end{bmatrix},
\tag{34}
$$

and $D_{Z_3|Z_2} := \mathrm{diag}\left(\lambda_3^1(Z_3|Z_2), ..., \lambda_3^J(Z_3|Z_2)\right)$. Evaluating (33) at four different points, (Z_2, Z_3), $(\bar{Z}_2, Z_3)$, $(Z_2, \bar{Z}_3)$, and $(\bar{Z}_2, \bar{Z}_3)$ gives

$$
\begin{aligned}
P_{Z_2,Z_3} &= L'_{z_3} D_{Z_3|Z_2} \bar{L}_{z_2}, \quad & P_{\bar{Z}_2,Z_3} &= L'_{z_3} D_{Z_3|\bar{Z}_2} \bar{L}_{\bar{z}_2}, \\
P_{Z_2,\bar{Z}_3} &= L'_{\bar{z}_3} D_{\bar{Z}_3|Z_2} \bar{L}_{z_2}, \quad & P_{\bar{Z}_2,\bar{Z}_3} &= L'_{\bar{z}_3} D_{\bar{Z}_3|\bar{Z}_2} \bar{L}_{\bar{z}_2}.
\end{aligned}
$$

Then, under Assumption 8,

$$A := P_{Z_2,Z_3}(P_{Z_2,\bar{Z}_3})^{-1}P_{\bar{Z}_2,\bar{Z}_3}(P_{\bar{Z}_2,Z_3})^{-1} = L'_{z_3}D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}(L'_{z_3})^{-1},$$

where

$$D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3} := D_{Z_3|Z_2}(D_{\bar{Z}_3|Z_2})^{-1}D_{\bar{Z}_3|\bar{Z}_2}(D_{Z_3|\bar{Z}_2})^{-1}. \tag{35}$$

Because $AL'_{z_3} = L'_{z_3}D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}$, the eignvalues of A determine the diagonal elements of $D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}$ while the right eigenvectors of A determine the columns of L'_{z_3} up to multiplicative constant. Denote the right eigenvectors of A by $L'_{z_3}C$, where C is some diagonal matrix. Now we can determine the diagonal matrix $D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}C$ from the first row of $AL'_{z_3}C = L'_{z_3}D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}C$ because the first row of L'_{z_3} is a vector of ones. Then, L'_{z_3} is determined uniquely from $AL'_{z_3}C$ and $D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}C$ as $L'_{z_3} = (AL'_{z_3}C)(D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}C)^{-1}$ in view of $AL'_{z_3} = L'_{z_3}D_{Z_2,\bar{Z}_2,Z_3,\bar{Z}_3}$. Therefore, L_{z_3} is identified. Repeating the above argument for all values of $Z_3 \in \mathcal{Z}_3$ identifies $\{P_4^j(Z_4|Z_3 = Z_3)\}_{j=1}^J$ for each $Z_3 \in \mathcal{Z}_3$ for $Z_4 = (b_1, ..., b_{J-1})$ that satisfies Assumption 8(a).

Evaluating $P(Z_4, Z_3|Z_2)$ at $(Z_2, Z_3) = (Z_2, Z_3)$, we have

$$P(Z_4, Z_3 = Z_3|Z_2 = Z_2) = \sum_{j=1}^{J} \tilde{\pi}^j_{Z_2} P_4^j(Z_4|Z_3)P_3^j(Z_3|Z_2) = \sum_{j=1}^{J} \lambda_4^j(Z_4|Z_3)\tilde{\lambda}_3^j(Z_3|Z_2), \tag{36}$$

where $\tilde{\pi}^j_{Z_2} := \frac{\pi^j P_2^j(Z_2=Z_2)}{P_2(Z_2=Z_2)}$ and $\tilde{\lambda}_3^j(Z_3|Z_2) := \tilde{\pi}^j_{Z_2}P_3^j(Z_3 = Z_3|Z_2 = Z_2)$. Then, evaluating (36) at $Z_4 = b_1, ..., b_{J-1}$ and collecting them into a vector together with $P(Z_3 = Z_3|Z_2 = Z_2) = \sum_{j=1}^J \tilde{\lambda}_3^j(Z_3|Z_2)$ gives

$$p_{Z_3|Z_2} = L'_{z_3}d_{Z_3|Z_2},$$

where $d_{Z_3|Z_2} = (\tilde{\lambda}_3^1(Z_3|Z_2), ..., \tilde{\lambda}_3^J(Z_3|Z_2))'$ and $p_{Z_3|Z_2} = (P(Z_3 = Z_3|Z_2 = Z_2), P((Z_4, Z_3) = (b_1, Z_3)|Z_2 = Z_2), ..., P((Z_4, Z_3) = (b_{J-1}, Z_3)|Z_2 = Z_2))'$. Therefore, we uniquely determine $\tilde{\pi}^j_{Z_2}P_3^j(Z_3 = Z_3|Z_2 = Z_2)$ from $d_{Z_3|Z_2} = (L'_{z_3})^{-1}p_{Z_3|Z_2}$. Repeating the above argument across all possible values of $(Z_2, Z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$ determines the value of $\tilde{\pi}^j_{Z_2}P_3^j(Z_3 = Z_3|Z_2 = Z_2)$ for every $(Z_2, Z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$. Then, $\tilde{\pi}^j_{Z_2}$ and $P_3^j(Z_3 = Z_3|Z_2 = Z_2)$ are uniquely identified as $\tilde{\pi}^j_{Z_2} = \int_{\mathcal{Z}_3} \tilde{\pi}^j_{Z_2}P_3^j(Z_3|Z_2 = Z_2)dZ_3$ and $P^j(Z_3 = Z_3|Z_2 = Z_2) = [\tilde{\pi}^j_{Z_2}P_3^j(Z_3 = Z_3|Z_2 = Z_2)]/\tilde{\pi}^j_{Z_2}$. Therefore, $\{P_3^j(Z_3|Z_2)\}_{j=1}^J$ is identified.

Evaluating $P_3^j(Z_3|Z_2)$ at $(Z_2, Z_3) = (Z_3, Z_2)$ for $j = 1, ..., J$ identifies $D_{Z_3|Z_2}$ and, from (33), $\bar{L}_{z_2}$ is identified as $\bar{L}_{z_2} = (D_{Z_3|Z_2})^{-1}(L'_{z_3})^{-1}P_{Z_2,Z_3}$. Once $D_{Z_3|Z_2}$ and $\bar{L}_{z_2}$ are identified, we can determine $L_{z_3}(\zeta) = (\lambda_4^1(\zeta|Z_3), ..., \lambda_4^J(\zeta|Z_3))'$ for any $\zeta \in \mathcal{Z}_4$ by constructing $p_{Z_2,Z_3}(\zeta) = (f_{Z_2,Z_3}(a_1, \zeta), f_{Z_2,Z_3}(a_2, \zeta), ..., f_{Z_2,Z_3}(a_J, \zeta))$ from the observed data, and using the relationship $L_{z_3}(\zeta) = (D_{Z_3|Z_2})^{-1}(\bar{L}'_{z_2})^{-1}p_{Z_2,Z_3}(\zeta)'$. Similarly, we can determine $\bar{L}_{z_2}(\xi) = (\bar{\lambda}_2^1(\xi, Z_2), ..., \bar{\lambda}_2^J(\xi, Z_2))'$ for any $\xi \in \mathcal{Z}_1$ by constructing $\bar{p}_{Z_2,Z_3}(\xi) = (\bar{f}_{Z_2,Z_3}(\xi), f_{Z_2,Z_3}(\xi, b_1), f_{Z_2,Z_3}(\xi, b_2), ...,$ and using the relationship $\bar{L}_{z_2}(\xi) = (D_{Z_3|Z_2})^{-1}(L'_{z_3})^{-1}\bar{p}_{Z_2,Z_3}(\xi)$. Therefore, $\{P_4^j(Z_4|Z_3 = Z_3), \pi^j P_2^j(Z_2 = Z_2|Z_1)P_1^j(Z_1)\}_{j=1}^J$ is identified. Repeating this argument for all possible values of $(Z_2, Z_3) \in \mathcal{Z}_2 \times \mathcal{Z}_3$ identifies $\{P_4^j(Z_4|Z_3), \pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)\}_{j=1}^J$. Finally, $\{\pi^j, P_2^j(Z_2|Z_1), P_1^j(Z_1)\}_{j=1}^J$ is identified from $\{\pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)\}_{j=1}^J$ as $\pi^j = \int_{\mathcal{Z}_1}\int_{\mathcal{Z}_2}[\pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)]$ $P_1^j(Z_1) = [\int_{\mathcal{Z}_2}[\pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)]dZ_2]/\pi^j$, and $P_2^j(Z_2|Z_1) = [\pi^j P_2^j(Z_2|Z_1)P_1^j(Z_1)]/[\pi^j \times P_1^j(Z_1)]$. This proves the stated result. □

A.4 Proof of Proposition 4

We first show that $P_{L,t}$ and $\{\psi_t^j\}_{j=1}^J$ are identified from $\{\pi^j, P_t^j(B_t)\}_{j=1}^J$. Because $E_t^j[\ln B_t] = \ln(P_{L,t}e^{\psi_t^j})$, we may have $\psi_t^j = E_t^j[\ln B_t] - P_{L,t}$ for $j = 1, ..., J$, where $E_t^j[\ln B_t]$ is identified from $P_t^j(W_t)$. Then, $P_{L,t}$ is identified from $\sum_{j=1}^J \pi^j e^{E_t^j[\ln B_t] - \ln P_{L,t}} = 1$ as $\ln P_{L,t} = \ln\left(\sum_{j=1}^J \pi^j e^{E_t^j[\ln B_t]}\right)$. Once $P_{L,t}$ and $\{\psi_t^j\}_{j=1}^J$ are identified, then repeating the argument in the proof of Proposition 1 for each type proves the stated result. □

A.5 Proof of Proposition 5

Consider $i \in \mathcal{I}^j$ so that $j = j^*(i)$. For each T, let $\pi_T^j := \frac{\pi^{j*}L_{1i}(\alpha_m^{j*}, \sigma_\epsilon^{j*}; T)}{\sum_{k=1}^J \pi^{k*}L_{1i}(\alpha_m^{k*}, \sigma_\epsilon^{k*}; T)}$, where $(\pi^{j*}, \alpha_m^{j*}, \sigma_\epsilon^{j*})$ is the true value of $(\pi^j, \alpha_m^j, \sigma_\epsilon^j)$. Then,

$$\hat{\pi}_i^j - 1 = (\hat{\pi}_i^j - \pi_T^j) + (\pi_T^j - 1). \tag{37}$$

For the first term, $\hat{\pi}_i^j - \pi_T^j = O_p(N^{-1/2})$ as $N \to \infty$ because the maximum likelihood estimator $(\hat{\pi}^j, \hat{\alpha}_m^j, \hat{\sigma}_\epsilon^j)$ is a root-N consistent estimator of $(\pi^{j*}, \alpha_m^{j*}, \sigma_\epsilon^{j*})$ when the number of components J is correctly specified.

For the second term of (37), define $\xi_{it}^{jk} := \ln L_{1it}(\alpha_m^{j*}, \sigma_\epsilon^{j*}) - \ln L_{1it}(\alpha_m^{k*}, \sigma_\epsilon^{k*})$ and $a^{jk} := E[\xi_{it}^{jk}|i \in \mathcal{I}^j] > 0$, and we have

$$\pi_T^j = \frac{1}{1 + \sum_{k \neq j}(\pi^{k*}/\pi^{j*})\exp\left(-\sum_{t=1}^T \xi_{it}^{jk}\right)}. \tag{38}$$

For $i \in \mathcal{I}^j$, $k \neq j$,

$$\begin{aligned}
\exp\left(-\sum_{t=1}^{T} \xi_{it}^{jk}\right) &= \left\{\exp\left(-\sum_{t=1}^{T} \xi_{it}^{jk}\right) - \exp(-a^{jk}T)\right\} + \exp(-a^{jk}T) \\
&= \exp(-a^{jk}T) \underbrace{\left\{\exp\left(-\sum_{t=1}^{T} (\xi_{it}^{jk} - a^{jk})\right) - 1\right\}}_{O_p(T^{1/2})} + \exp(-a^{jk}T) \\
&= O_p\left(\exp(-a^{jk}T)T^{1/2}\right)
\end{aligned}$$

as $T \to \infty$. It follows that $\sum_{k \neq j} \left(\pi^{k^*}/\pi^{j*}\right) \exp\left(-\sum_{t=1}^{T} \xi_{it}^{jk}\right)$ is $O_p\left(\exp(-a^jT)T^{1/2}\right)$, where $a^j := \min_{k \neq j} a^{jk}$. Therefore, in view of (38), the consistency of π_T^j as $T \to \infty$ and the mean value theorem give

$$\pi_T^j - 1 = O_p\left(\exp(-a^jT)T^{1/2}\right). \tag{39}$$

Then, the stated result follows from (37), (39), and $\hat{\pi}_i^j - \pi_T^j = o_p(N^{-1/2})$ because $O_p\left(\exp(-a^jT)T^{1/2}\right) = o_p(N^{-1/2})$ as $N, T \to \infty$ under Assumption 11. $\square$

A.6 Assumption 8 under Cobb-Douglas production function

In the following, we discuss the conditions under which Assumption 8 holds when the production function is Cobb-Douglas.

Example 1 (continued)**.** *For random coefficients model (10), we may write $L_{\bar{z}_3}$, $\bar{L}_{\bar{z}_2}$, and $D_{\bar{Z}_3|\bar{Z}_2}$ as follows. Throughout the analysis, we fix the value of $\{Y_t\}_{t=1}^T$ at, say, $\{y_t\}_{t=1}^T$ so that the variation in the values of a_j's and b_j's are due the variation in the values of Z_1 and Z_4. Denote $\bar{Z}_3 = (y_3, \bar{s}_3, \bar{x}_3)$ and $b_h = (y_3, b_h^s, \bar{x}_4)$ for $h = 1, ..., J-1$. Then,*

$$\lambda_{\bar{Z}_3}^j(b_h) = P_4^j(S_4 = b_h^s | X_4 = \bar{x}_4) P_4^j(X_4 = \bar{x}_4 | X_3 = \bar{x}_3) = c_4^j g_\epsilon^j(\ln(\alpha_{m,4}^j \mathcal{E}^j) - \ln b_h^s),$$

where $c_4^j = P_4^j(X_4 = \bar{x}_4 | X_3 = \bar{x}_3)$. Therefore, we have

$$L_{\bar{z}_3} = diag\{c_4^1, ..., c_4^J\} \begin{bmatrix} 1 & g_\epsilon^1(\ln(\alpha_{m,4}^1 \mathcal{E}^1) - \ln b_1^s) & \cdots & g_\epsilon^1(\ln(\alpha_{m,4}^1 \mathcal{E}^1) - \ln b_{J-1}^s) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & g_\epsilon^J(\ln(\alpha_{m,4}^J \mathcal{E}^J) - \ln b_1^s) & \cdots & g_\epsilon^J(\ln(\alpha_{m,4}^J \mathcal{E}^J) - \ln b_{J-1}^s) \end{bmatrix}.$$

Similarly, denote $\bar{Z}_2 = (\bar{s}_2, \bar{x}_2)$ *and* $a_h = (a_h^s, \bar{x}_1)$ *for* $h = 1, ..., J$. *Then,*

$$\lambda^j_{\bar{Z}_2}(a_h) = P_2^j(S_2 = \bar{s}_2|X_2 = \bar{x}_2)P_1^j(S_1 = a_h^s|X_1 = \bar{x}_1)P_1^j(X_1 = \bar{x}_1) = c_2^j g_\epsilon^j(\ln a_h^s - \ln(\alpha_m^j \mathcal{E})),$$

where $c_2^j = P_2^j(S_2 = \bar{s}_2|X_2 = \bar{x}_2)P_1^j(X_1 = \bar{x}_1)$. *Then, we have*

$$\bar{L}_{\bar{z}_2} = diag\{c_2^1, ..., c_2^J\} \begin{bmatrix} g_\epsilon^1(\ln(\alpha_{m,1}^1 \mathcal{E}^1) - \ln a_1^s) & \cdots & g_\epsilon^1(\ln(\alpha_{m,1}^1 \mathcal{E}^1) - \ln a_J^s) \\ \vdots & \ldots & \vdots \\ g_\epsilon^J(\ln(\alpha_{m,1}^J \mathcal{E}^J) - \ln a_1^s) & \cdots & g_\epsilon^J(\ln(\alpha_{m,1}^J \mathcal{E}^J) - \ln a_J^s) \end{bmatrix}.$$

For Assumption 8(a), we choose $\bar{x}_4$, $\bar{x}_3$, $\bar{x}_2$, $\bar{x}_1$, *and* $\bar{s}_2$ *so that* $c_2^j \neq 0$ *and* $c_3^j \neq 0$ *for any* j *and find* $(a_1^s, ..., a_J^s)$ *and* $(b_1^s, ..., b_{J-1}^s)$ *such that* $L_{\bar{z}_3}$ *and* $\bar{L}_{\bar{z}_2}$ *are nonsingular. Because each point of* $(a_1^s, ..., a_J^s)$ *and* $(b_1^s, ..., b_{J-1}^s)$ *refers to a value of* $\ln S_1$ *and* $\ln S_4$, *the full rank condition of* $L_{\bar{z}_3}$ *and* $\bar{L}_{\bar{z}_2}$ *holds if the value of probability density function of* $\ln S_1$ *and* $\ln S_4$ *changes heterogeneously across types when we change the value of* $\ln S_1$ *and* $\ln S_4$.

Let $\bar{Z}_3 = (\bar{s}_3, \bar{x}_3)$ *and* $\bar{Z}_2 = (\bar{s}_2, \bar{x}_2)$. *Then,*

$$\lambda^j(\bar{Z}_3|\bar{Z}_2) = \pi^j g_\epsilon^j(\ln \bar{s}_3 - \ln(\alpha_{m,3}^j \mathcal{E}^j))P_3^j(X_3 = \bar{x}_3|X_2 = \bar{x}_2). \tag{40}$$

Pick $Z_3 = (s_3, x_3)$ *and* $Z_2 = (s_2, x_2)$. *Assumption 8(b) holds if* $P_3^j(\bar{x}_3|x_2) \neq 0$ *and* $P_3^j(x_3|\bar{x}_2) \neq 0$ *for all* j. *Then, we have*

$$D_{Z_3|Z_2}(D_{\bar{Z}_3|Z_2})^{-1}D_{\bar{Z}_3|\bar{Z}_2}(D_{Z_3|\bar{Z}_2})^{-1} = diag\left\{\frac{P_3^1(x_3|x_2)}{P_3^1(\bar{x}_3|x_2)}\frac{P_3^1(\bar{x}_3|\bar{x}_2)}{P_3^1(x_3|\bar{x}_2)}, ..., \frac{P_3^J(x_3|x_2)}{P_3^J(\bar{x}_3|x_2)}\frac{P_3^J(\bar{x}_3|\bar{x}_2)}{P_3^J(x_3|\bar{x}_2)}\right\}.$$

Therefore, Assumption 8(c) requires that $\frac{P_3^j(x_3|x_2)}{P_3^j(\bar{x}_3|x_2)}\frac{P_3^j(\bar{x}_3|\bar{x}_2)}{P_3^j(x_3|\bar{x}_2)}$ *takes different values across different j's, namely, the transition probability of* X_3 *given* X_2 *changes heterogeneously across types when we change the value of* X_2.