Identification and Estimation of Production Function with Unobserved Heterogeneity

Hiroyuki Kasahara*
Vancouver School of Economics
University of British Columbia
hkasahar@mail.ubc.ca

Paul Schrimpf
Vancouver School of Economics
University of British Columbia
schrimpf@mail.ubc.ca

Michio Suzuki
Faculty of Economics
University of Tokyo
msuzuki@e.u-tokyo.ac.jp

December 23, 2015

Abstract

This paper examines non-parametric identifiability of production function when production functions are heterogenous across firms beyond Hicks-neutral technology terms. Using a finite mixture specification to capture unobserved heterogeneity in production technology, we shows that 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 publicly-traded manufacturing firms and compare it with the estimate of production function with fixed coefficients estimated by the method of Gandhi, Navarro, and Rivers (2013). 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 industry.

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

*Address for correspondence: Hiroyuki Kasahara, Vancouver School of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC, V6T 1Z1 Canada.

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 suffers from simultaneity bias because inputs are correlated with error term 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, fixed effects estimator may be used but such an assumption could be violated in practice, and, furthermore, 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, 2006, (ACF, hereafter); Wooldridge, 2009), which are now widely used in empirical applications.

Despite their popularity, however, potential identification issues of proxy variable approach have been pointed out in the literature. Bond and Sderbom (2005) and ACF discuss identification issue due to collinearity under two flexible inputs (i.e., material and labor) in Cobb-Douglas specification. Gandhi, Navarro, and Rivers (2013, 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 heterogenous 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.

Given that, except for the result of GNR, little formal identification result for production function estimation in the literature is available, our nonparametric identification result is an important contribution to the literature. Our identification result on production function with unobserved heterogeneity is also useful in practice as the random coefficient models for production function become increasingly popular in empirical analysis (e.g., Mairesse and Griliches, 1990; Van Biesebroeck, 2003; Doraszelski and Jaumandreu, 2014).

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 facilitates 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 less 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 provide empirical evidence that production functions are heterogeneous beyond Hicks-neutral technology term to motivate the necessity of considering production functions with unobserved heterogeneity in empirical applications. As analyzed by GNR, if Hicks-neutral technology term is the only source of permanent unobserved heterogeneity in production function and if intermediate input is a flexible input, then we expect that the ratio of intermediate input cost to output value after controlling for the difference in the input level of capital, labor, and intermediates should not exhibit any serial correlation. However, using the panel of Japanese manufacturing firms that belongs to machine industry, we find that the serial correlation of the ratio of intermediate input cost to output value is very high at 0.95 and that, even after controlling for the difference in the input level of capital, labor, and intermediates, the majority of variation in the ratio of intermediate input cost to output value can be explained by the firm-specific persistent component rather than the idiosyncratic component. These findings strongly suggest the presence of unobserved heterogeneity in production technology beyond Hicks-neutral term within the 3-digit industry classification.

We estimate a random coefficients production function using the panel data of Japanese publicly-traded manufacturing firms between 1980 and 2007 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. When we estimate production function without incorporating heterogeneity using the estimation procedure suggested by GNR, we found that the majority of variations in total factor productivity is coming from idiosyncratic ex-post shocks rather than serially correlated shocks. In contrast, when we estimate production function with random coefficients, the majority of variations in total factor productivity is explained by the variation in serially correlated shocks. Furthermore, the estimated serial correlation in ex-post shocks of random coefficients model is substantially lower than that of homogenous model. 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 The Model

Assume that we have panel data of firms $i = 1, ..., N$ over periods $t = 1, ..., T$ for output, labor, capital and intermediate inputs denoted by $(Y_{it}, L_{it}, K_{it}, M_{it})$, respectively. For brevity, let $X_{it} := (L_{it}, K_{it}, M_{it})'$ so that $(Y_{it}, L_{it}, K_{it}, M_{it}) = (Y_{it}, X_{it})$. Each firm's observation $\{Y_{it}, X_{it}\}_{t=1}^{T}$ is randomly sampled from a population distribution $P(\{Y_{it}, X_{it}\}_{t=1}^{T})$.

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. In the following, the superscript j indicates that functions are specific to type j. Denote the information available to a firm for making decisions on M_{it}, L_{it+1}, and K_{it+1} at period t by $\mathcal{I}_{it}$.

Assumption 1. *(a) Each firm belongs to one of J types, where the probability of belonging to type j is given by π^j, and J is known. (b) For the j-th type of production technology at period t, the output is related to inputs as*

$$Y_{it} = F_t^j(X_{it}) e^{\omega_{it} + \epsilon_{it}}, \tag{1}$$

where $F_t^j(X_{it}) = F_t^j(L_{it}, K_{it}, M_{it})$ is differentiable with respect to M_{it}.

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

$$\omega_{it} = h^j(\omega_{it-1}) + \eta_{it}, \tag{2}$$

where $\eta_{it}|\mathcal{I}_{it-1}$ is independently and identically distributed (i.i.d.) with the probability density function $g_\eta^j(\cdot)$, where $P_\omega^j(\omega_{it}|\mathcal{I}_{it-1}) = P_\omega^j(\omega_{it}|\omega_{it-1}) := g_\eta^j(\omega_{it} - h^j(\omega_{it-1}))$. (b) For the j-th type, $\epsilon_{it} \notin \mathcal{I}_{it}$ is an i.i.d. ex-post shock that is not known when M_{it} is chosen at period t. The probability density function of ϵ_{it} for type j is denoted by $g_\epsilon^j(\cdot)$, where $P_\epsilon^j(\epsilon_{it}|\mathcal{I}_{it}) = P_\epsilon^j(\epsilon_{it}) := g_\epsilon^j(\epsilon_{it})$.

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

Assumption 4. *(a) M_{it} is flexibly chosen at period t. (b) M_{it} is a type-specific deterministic function of L_{it}, K_{it}, and ω_{it} that can be written as $M_{it} = \mathbb{M}_t^j(L_{it}, K_{it}, \omega_{it})$, where $\mathbb{M}_t^j$ is strictly increasing in ω_{it} for any (L_{it}, K_{it}).*

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

Our Assumptions 1-4 correspond to Assumptions 1-4 of GNR, respectively, but our assumptions relax the assumptions of GNR in that we allow for permanent unobserved heterogeneity in production technology. In Assumption 1, 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. Our Assumption 2 is more stringent than that of GNR because we assume stochastic independence of η_{it} and that of ϵ_{it} while GNR only assumes mean independence. Assumption 3(a) is a simplifying assumption adopted by GNR which helps us to clarify the essence of our identification argument.[1] Assumption 3(b) can be justified by explicitly considering the dynamic model of investment and labor input choices. Assumption 4(b) implies that there is one-to-one mapping between (L_{it}, K_{it}, M_{it}) and $(L_{it}, K_{it}, \omega_{it})$ and, considering the inverse function of $\ln \mathbb{M}_t^j$ with respect to ω_{it}, we may write $\omega_{it} = \ln \mathbb{M}_t^{j^{-1}}(L_{it}, K_{it}, M_{it})$. 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). Under Assumptions 1-5, the information set $\mathcal{I}_{it}$ is given by $\mathcal{I}_{it} = \{\omega_{it}, L_{it}, K_{it}, P_{M,t}, P_{Y,t}, Z_{it-1}, Z_{it-2}, ...\}$, where $Z_{it} = \{\epsilon_{it}, \omega_{it}, Y_{it}, X_{it}, P_{M,t}, P_{Y,t}\}$.

Consider a firm i of which type is j. A firm chooses M_{it} by maximizing its expected profit given $\mathcal{I}_{it}$:

$$\mathbb{M}_t^j(L_{it}, K_{it}, \omega_{it}) = \underset{M_{it}}{\operatorname{argmax}} P_{Y,t} E_\epsilon[F_t^j(L_{it}, K_{it}, M_{it}) e^{\omega_{it}+\epsilon} | \mathcal{I}_{it}] - P_{M,t} M_{it}.$$

Under Assumptions 1, 2, 3(a), 4(a), and 5, the first order condition with respect to M_{it} gives

$$P_{Y,t} F_{M,t}^j(X_{it}) e^{\omega_{it}} \mathcal{E}^j = P_{M,t}, \tag{3}$$

where $\mathcal{E}^j = \int e^\epsilon g_\epsilon^j(\epsilon) d\epsilon$. Note that Assumption 4(b) holds when $F_{M,t}^j(L_{it}, K_{it}, M_{it}) := \frac{\partial F_t^j(L_{it}, K_{it}, M_{it})}{\partial M_{it}}$ is strictly decreasing in M_{it} given (L_{it}, K_{it}). Equations (1) and (3) give a system of equations

$$\begin{aligned} \ln Y_{it} &= \ln F_t^j(X_{it}) + \omega_{it} + \epsilon_{it}, \\ \ln S_{it} &= \ln G_{M,t}^j(X_{it}) + \ln \mathcal{E}^j - \epsilon_{it}, \end{aligned} \tag{4}$$

where

$$S_{it} := \frac{P_{M,t} M_{it}}{P_{Y,t} Y_{it}}$$

is the intermediate input share and $G_{M,t}^j(X_{it}; \xi_{it}) := \frac{F_{M,t}^j(X_{it}) M_{it}}{F_t^j(X_{it})}$.

In place of Assumption 5, we may alternatively consider the case where firms produce differentiated products and face a demand function with constant price elasticity as follows.

Assumption 6 (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 period* t*. (b) A firm is a price taker for intermediate input and the intermediate input price* $P_{M,t}$ *at period* t *is common across firms. (c)* $P_{M,t} \in \mathcal{I}_{it}$*. (d)* $P_{Y,it}$ *and*

[1] In particular, because we assume that L_{it} is predetermined, the source of non-identification of the coefficient of L_{it} discussed in Ackerberg, Caves, and Fraser (2006) is absent in this setup.

Y_{it} *are not separately observed in the data.*

Under Assumption 6, the "revenue" production function is given by $P_{Y,it}Y_{it} = \tilde{F}_t^j(X_{it})e^{\tilde{\omega}_{it}+\tilde{\epsilon}_{it}}$, where $\tilde{F}_t^j(X_{it}) := [F_t^j(X_{it})]^{\frac{\sigma_Y^j-1}{\sigma_Y^j}}$, $\tilde{\omega}_{it} := \frac{\sigma_Y^j-1}{\sigma_Y^j}\omega_{it}$, and $\tilde{\epsilon}_{it} := \epsilon_{it}^d + \frac{\sigma_Y^j-1}{\sigma_Y^j}\epsilon_{it}$. Then, in place of (4), we have

$$\begin{aligned} \ln P_{Y,it}Y_{it} &= \ln \tilde{F}_t^j(X_{it}) + \tilde{\omega}_{it} + \tilde{\epsilon}_{it}, \\ \ln S_{it} &= \ln \tilde{G}_{M,t}^j(X_{it}) + \ln \tilde{\mathcal{E}}^j - \tilde{\epsilon}_{it}, \end{aligned} \tag{5}$$

where $\tilde{G}_{M,t}^j(X_{it}) := \frac{\partial \tilde{F}_t^j(X_{it})/\partial M_{it}}{\tilde{F}_t^j(X_{it})}$. When $P_{Y,it}$ and Y_{it} are not separately observed in the data, the observable implication of (5) are the same as that of (4). 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.

3 Nonparametric identification of gross production functions with unobserved heterogeneity

In this section, we establish the non-parametric identification of production functions with unobserved heterogeneity using the second equation of (4) as an additional restriction. For notational brevity, we drop the subscript i in this section. The distribution of $\{Y_t, S_t, X_t\}_{t=1}^T$ follows an J-term mixture distribution

$$\begin{aligned} \mathrm{P}(\{Y_t, S_t, X_t\}_{t=1}^T) &= \sum_{j=1}^J \pi^j \mathrm{P}^j(\{Y_t, S_t, X_t\}_{t=1}^T) \\ &= \sum_{j=1}^J \pi^j \mathrm{P}_1^j(Y_1, S_1, X_1) \prod_{t=2}^T \mathrm{P}_t^j(Y_t, S_t, X_t | \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}). \end{aligned} \tag{6}$$

Proposition 1. *Suppose that Assumptions 1-5 hold. Then, the distribution of* $\{Y_t, S_t, X_t\}_{t=1}^T$ *defined in (6) can be written as*

$$P(\{Y_t, S_t, X_t\}_{t=1}^T) = \sum_{j=1}^J \pi^j P_1^j(Y_1, S_1, X_1) \prod_{t=2}^T P_t^j(Y_t, S_t, X_t | Y_{t-1}, S_{t-1}, X_{t-1}). \tag{7}$$

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

Assumption 7. *Let* $W_t := (Y_t, S_t, X_t)$ *and let* $\mathcal{W}_t$ *be the support of* W_t*. For every* $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$*, there exists* $(\bar{w}_2, \bar{w}_3) \in \mathcal{W}_2 \times \mathcal{W}_3$*,* $(a_1, ..., a_J) \in \mathcal{W}_1^J$ *and* $(b_1, ..., b_{J-1}) \in \mathcal{W}_4^{J-1}$ *such that (a)* L_{w_3}*,* $L_{\bar{w}_3}$*,* $\bar{L}_{w_2}$*, and* $\bar{L}_{\bar{w}_2}$ *defined in (33) are nonsingular, (b)* $P^j(W_3 = w_3|W_2 = \bar{w}_2) \neq 0$ *and* $P^j(W_3 = \bar{w}_3|W_2 = w_2) \neq 0$ *hold for* $j = 1, ..., J$*, and (c) all the diagonal elements of* $D_{w_2,\bar{w}_2,w_3,\bar{w}_3}$ *defined in (34) take distinct values.*

Proposition 2. *Suppose that Assumptions 1-5, and 7 hold and* $T \geq 4$*. Then,* $\{\pi^j, P_1^j(W_1), \{P_t^j(W_t|W_{t-1})\}_{t=2}^T\}_{j=1}^J$ *is uniquely determined from* $P(\{W_t\}_{t=1}^T)$*, where* $W_t := (Y_t, S_t, X_t)$*.*

Remark 1. *Under the additional assumption of the stationarity, i.e.,* $P_t^j(W_t|W_{t-1}) = P^j(W_t|W_{t-1})$ *for* $t = 2, ..., T$*, Kasahara and Shimotsu (2009) establishes the nonparametric identification of the model (7) when* $T = 6$ *while Hu and Shum (2013) shows that* $T = 4$ *suffices for identification.*

Remark 2. *Considering serially correlated continuous unobserved variables* $\{X_t^*\}$*, Hu and Shum (2013) analyze the nonparametric identification of the model*

$$P(\{W_t\}_{t=1}^T) = \int P_1(W_1, X_1^*) \prod_{t=2}^T P_t(W_t, X_t^*|W_{t-1}, X_{t-1}^*) d(\{X_t^*\}_{t=1}^T).$$

Given the panel data $\{W_t\}_{t=1}^T$ *with* $T = 5$*, Theorem 1 and Corollary 1 of Hu and Shum (2013) state that, under their Assumptions 1-4,* $P_3(W_3, X_3^*)$*,* $P_4(W_4, X_4^*|W_3, X_3^*)$*, and* $P_5(W_5, X_5^*|W_4, X_4^*)$ *are non-parametrically identified but the identification of* $P_1(W_1, X_1^*)$*,* $P_2(W_2, X_2^*|W_1, X_1^*)$*, and* $P_3(W_3, X_3^*|W_2, X_2^*)$ *remains unresolved. Our Proposition 2 shows that, for a model in which unobserved heterogeneity is discrete and finite, we can nonparametrically identify the type-specific distribution of* $\{W_t\}_{t=1}^T$ *including the first two periods of the data from* $T = 4$ *periods of panel data without imposing stationarity.*

Remark 3. *Assumption 7 assumes the rank condition of matrices* L_{w_3}*,* $L_{\bar{w}_3}$*,* $\bar{L}_{w_2}$*, and* $\bar{L}_{\bar{w}_2}$ *defined in (33), of which elements are constructed by evaluating* $P_4^j(W_4|W_3)$ *and* $\pi^j P_2^j(W_2|W_1) P_1^j(W_1)$ *at different points, where* $W_t := (Y_t, S_t, X_t)$*. 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* $(\bar{w}_2, \bar{w}_3) \in \mathcal{W}_2 \times \mathcal{W}_3$ *and one set of* $J - 1$ *and* J *points of* W_1 *and* W_4 *to construct nonsingular* L_{w_3}*,* $L_{\bar{w}_3}$*,* $\bar{L}_{w_2}$*, and* $\bar{L}_{\bar{w}_2}$ *for each* $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$ *and these rank conditions are not stringent when* W_t *has continuous support. The identification of* $P_4^j(W_4|W_3 = w_3)$ *and* $\pi^j P_2^j(W_2 = w_2|W_1) P_1^j(W_1)$ *at all other points of* W_4 *and* W_1*, respectively, follows without any further requirement on the rank condition.*

Once the type-specific distribution of $\{Y_t, S_t, X_t\}$ is identified, we may apply the argument of GNR to prove the nonparametric identification of $\theta = \{\{g_\epsilon^j(\cdot), g_\eta^j(\cdot), h^j(\cdot), \mathcal{E}^j, \pi^j, \{G_{M,t}^j(\cdot), F_t^j(\cdot)\}_{t=1}^T\}_{j=1}^J\}$.

Proposition 3. *Suppose that Assumptions 1-5, and 7 hold and* $T \geq 4$*. Then, (a)* $\theta_1 := \{\pi^j, g_\epsilon^j(\cdot), \mathcal{E}^j, \{G_{M,t}^j(\cdot)\}_{t=1}^T\}_{j=1}^J$ *is uniquely determined from* $P(\{S_t, X_t\}_{t=1}^T)$*. (b)* $\theta_2 := \{\{\{F_t^j(\cdot)\}_{t=1}^T\}_{j=1}^J, h^j(\cdot), g_\eta^j(\cdot)\}$ *is uniquely determined from* $P(\{Y_t, S_t, X_t\}_{t=1}^T)$ *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 (potentially time-varying) random coefficients:*

$$\ln F_t^j(X_t) = \alpha_t^j + \alpha_{\ell,t}^j \ell_t + \alpha_{k,t}^j k_t + \alpha_{m,t}^j m_t, \tag{8}$$

where the intermediate share equation is given by

$$\ln S_t = \ln(\alpha_{m,t}^j \mathcal{E}^j) - \epsilon_t. \tag{9}$$

Then, equation (7) holds with

$$P_1^j(S_1, X_1) = P_1^j(S_1) P_1^j(X_1) \quad \text{and} \quad P_t^j(S_t, X_t | X_{t-1}) = P_t^j(S_t) P_t^j(X_t | X_{t-1}),$$

where $P_t^j(S_t) = g_\epsilon^j(\ln(\alpha_{m,t}^j \mathcal{E}^j) - \ln S_t)$. Under Assumptions 1-5, and 7, we may nonparametrically identify $g_\epsilon^j(\cdot)$, $h^j(\cdot)$, $g_\eta^j(\cdot)$, $P_1^j(X_1)$, and $P_t^j(X_t|X_{t-1})$, and $\{\alpha_t^j, \alpha_{\ell,t}^j, \alpha_{k,t}^j, \alpha_{m,t}^j\}$ for $t = 1, ..., 4$ and $j = 1, ..., J$ from the panel data $\{Y_t, S_t, X_t\}_{t=1}^4$.

In the appendix, we discuss the conditions under which Assumption 7 holds when the production function is Cobb-Douglas.

The following corollary shows that type-specific distribution of the intermediate share, 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 7 hold and $T \geq 4$. Suppose that production function is Cobb-Douglas given by (8). Then, $\{\pi^j, \{P_1^j(S_t)\}_{t=1}^T\}_{j=1}^J$ is uniquely determined from $P(\{S_t\}_{t=1}^T)$.*

4 Estimation of production function with random coefficients

4.1 Cobb-Douglas production function with random coefficients

Denote the log values of $(Y_{it}, L_{it}, K_{it}, M_{it}, S_{it})$ by $(y_{it}, \ell_{it}, k_{it}, m_{it}, s_{it})$ and let $x_{it} := (\ell_{it}, k_{it}, m_{it})$. We assume that $f_t^j(x_{it}) = \alpha_t^j + \alpha_m^j m_{it} + \alpha_\ell^j \ell_{it} + \alpha_k^j k_{it}$ so that

$$y_{it} = \alpha_t^j + \alpha_m^j m_{it} + \alpha_\ell^j \ell_{it} + \alpha_k^j k_{it} + \omega_{it} + \epsilon_{it}, \tag{10}$$

for $j = 1, ..., J$. The first order condition with respect to M_{it} implies

$$s_{it} = \ln \mathcal{E}^j + \ln \alpha_m^j - \epsilon_{it}. \tag{11}$$

Therefore, the distribution of S_{it} does not depend on X_{it} under Cobb-Douglas assumption. We also assume that ω_{it} follows a first order autoregressive process

$$\omega_{it} = \rho^j \omega_{it-1} + \eta_{it}, \tag{12}$$

where $E[\eta_{it}|\omega_{it-1}] = 0$. Denote the true type of firm i by $j^*(i)$ and an index set of firms of type j by $\mathcal{I}^j := \{i : j^*(i) = j\}$.

Assumption 8. *(a) The production function of type j is given by (10). (b) The stochastic process of ω of type j is given by (12). (c) $\epsilon_{it}|x_{it}, i \in \mathcal{I}^j \overset{iid}{\sim} N(0, (\sigma_\epsilon^j)^2)$ for $j = 1, ..., J$. (d) J is known.*

The normality assumption in Assumption 8(c) can be 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} is non-parametrically specified although its asymptotic distribution is not known. Kasahara and Shimotsu (2015) develop a likelihood-based procedure for testing the number of components in normal mixture regression models.

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, where the likelihood function is a parametric version of (7). 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 2(b)(c). The EM algorithm can be used to facilitates 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 intermediate share equation (11) 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 by following the estimation procedure proposed by GNR.[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 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.

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

4.2 Maximum likelihood estimator

We make the parametric distributional assumptions and develop parametric maximum likelihood estimator. Let $z_{it} := (\ell_{it}, k_{it}, \omega_{it})'$ and $\tilde{z}_{it} := (1, z_{it}')'$.

Assumption 9. *(a) T is fixed at $T \geq 4$ and $N \to \infty$. (b) $z_{i1}|i \in \mathcal{I}^j \overset{iid}{\sim} N(\mu_1^j, \Sigma_1^j)$, where $\mu_1^j = (\mu_\ell^j, \mu_k^j, \mu_\omega^j)'$. (c) $(\ell_{it}, k_{it})'|z_{i,t-1}, i \in \mathcal{I}^j \overset{iid}{\sim} N((\rho_z^j)'\tilde{z}_{it-1}, \Sigma_z^j)$, where $\rho_z^j = (\rho_\ell^j, \rho_k^j)$ is an 4×2 matrix. (d) $\omega_{it} = \rho_\omega^j \omega_{it-1} + \eta_{it}$ with $\eta_{it}|k_{it}, \ell_{it}, i \in \mathcal{I}^j \overset{iid}{\sim} N(0, (\sigma_\eta^j)^2)$.*

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

$$\theta_1 = (\boldsymbol{\pi}', \theta_1^1, ..., \theta_1^J)' \quad \text{and} \quad \theta_2 = ((\theta_2^1)', ..., (\theta_2^J)')', \quad \text{where}$$

$$\theta_1^j = (\alpha_m^j, \sigma_\epsilon^j)' \text{ and}$$
$$\theta_2^j = (\alpha_2^j, ..., \alpha_T^j, \alpha_\ell^j, \alpha_k^j, (\mu_1^j)', \text{vech}(\Sigma_1^j)', \text{vech}(\rho_z^j)', \text{vech}(\Sigma_z^j)', \rho_\omega^j, \sigma_\eta^j)'.$$

Under Proposition 1 and Cobb-Douglas specification (10) and noting that $f_t^j(m_{it}|y_{it}, s_{it}) = 1$ under Assumption 5(c) because $m_{it} = s_{it} + y_{it} + p_t$ with $p_t := \ln P_{M,t}/P_{Y,t}$, we may write the probability density function of $\{y_{it}, s_{it}, m_{it}, k_{it}, \ell_{it}\}_{t=1}^T$ for type j as

$$\begin{aligned}\mathrm{f}_t^j(\{y_{it}, s_{it}, m_{it}, \ell_{it}, k_{it}\}_{t=1}^T) &= \underbrace{\prod_{t=1}^T \mathrm{f}_t^j(s_{it}; \theta_1^j)}_{=L_{1i}(\theta_1^j)} \\ &\times \underbrace{\mathrm{f}_1^j(y_{i1}, \ell_{it}, k_{it}|s_{i1}; \theta_1^j, \theta_2^j) \prod_{t=2}^T \mathrm{f}_t^j(y_{it}|s_{it}, \ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1}; \theta_1^j, \theta_2^j)\mathrm{f}_t(\ell_{it}, k_{it}|y_{it-1}, s_{it-1}, x_{it-1}; \theta_1^j, \theta_2^j)}_{=L_{2i}(\theta_2^j, \theta_1^j)},\end{aligned} \tag{13}$$

where the exact expressions for $L_{1i}(\theta_1^j)$ and $L_{2i}(\theta_2^j, \theta_1^j)$ are derived as follows.[3]

We consider a random sample of N independent observations $\{\{y_{it}, s_{it}, x_{it}\}_{t=1}^T\}_{i=1}^N$ from the true J-component mixture model $\sum_{j=1}^J \pi^j \mathrm{f}_t^j(\{y_{it}, s_{it}, x_{it}\}_{t=1}^T)$, where $\mathrm{f}_t^j(\{y_{it}, s_{it}, x_{it}\}_{t=1}^T)$ is given in (13). Given the decomposition (13), we estimate the model by two-stage maximum likelihood estimation procedure. In the first stage, we estimate $\boldsymbol{\pi}$ and θ_1 by maximizing $\sum_{i=1}^N \log(\sum_{j=1}^J \pi^j L_{1i}(\theta_1^j))$ over $\boldsymbol{\pi}$ and θ_1. In the second stage, we estimate θ_2 given the first stage estimate $\hat{\boldsymbol{\pi}}$ and $\hat{\theta}_1$ by maximizing $\sum_{i=1}^N \log(\sum_{j=1}^J \hat{\pi}^j L_{1i}(\hat{\theta}_1^j) L_{2i}(\theta_2^j, \hat{\theta}_1^j))$ over θ_2.

Given that $\ln \mathcal{E}^j = 0.5(\sigma_\epsilon^j)^2$ in equation (11), we can compute ϵ_{it} as

$$\epsilon^*(s_{it}; \theta_1^j) := -s_{it} + \ln \alpha_m^j + 0.5(\sigma_\epsilon^j)^2. \tag{14}$$

[3]Equation (13) holds because $\mathrm{f}_t^j(\{y_{it}, s_{it}, x_{it}\}_{t=1}^T) = \mathrm{f}_t^j(y_{i1}, s_{i1}, x_{i1}) \prod_{t=2}^T \mathrm{f}_t^j(y_{it}, s_{it}, x_{it}|y_{it-1}, s_{it-1}, x_{it-1}) = \mathrm{f}_t^j(m_{i1}|y_{i1}, k_{i1}, \ell_{i1}, s_{i1})\mathrm{f}_t^j(y_{i1}, \ell_{i1}, k_{i1}|s_{i1})\mathrm{f}_t^j(s_{i1}) \prod_{t=2}^T \mathrm{f}_t^j(m_{it}|y_{it}, s_{it}, \ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1})\mathrm{f}_t^j(y_{it}|s_{it}, \ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1}) \mathrm{f}_t^j(s_{it}|\ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1})\mathrm{f}_t^j(\ell_{it}, k_{it}|y_{it-1}, s_{it-1}, x_{it-1}) = \mathrm{f}_t^j(y_{i1}, \ell_{i1}, k_{i1}|s_{i1})\, f(s_{i1}) \prod_{t=2}^T \mathrm{f}_t^j(y_{it}|s_{it}, \ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1})\mathrm{f}_t^j(s_{it}) \mathrm{f}_t^j(\ell_{it}, k_{it}|y_{it-1}, s_{it-1}, x_{it-1})$ where the first equality follows from Proposition 1, the second equality always holds via repeated conditioning, and the third equality holds because $\mathrm{f}_t^j(m_{it}|y_{it}, s_{it}) = 1$ under Assumption 5(c) and $f(s_{it}|\ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1}) = f(s_{it})$ under Cobb-Douglas assumption.

Under Assumptions 8-9, in the first stage, we estimate θ_1 by the maximum likelihood estimator given by

$$\hat{\theta}_1 = \underset{\theta_1}{\operatorname{argmax}} \sum_{i=1}^{N} \ln \left(\sum_{j=1}^{J} \pi^j L_{1i}(\theta_1^j) \right) \quad \text{with} \quad L_{1i}(\theta_1^j) := \prod_{t=1}^{T} \frac{1}{\sigma_\epsilon^j} \phi \left(\frac{\epsilon^*(s_{it};\theta_1^j)}{\sigma_\epsilon^j} \right).$$

In the second stage, because $m_{it} = y_{it} + s_{it} + p_t$, we have

$$\omega_{it} = \omega^*(y_{it}, s_{it}, \ell_{it}, k_{it}; \theta_1^j, \theta_2^j) := (1-\alpha_m^j)y_{it} - \alpha_m^j s_{it} - \alpha_m^j p_t - \epsilon^*(s_{it};\theta_1^j) - \alpha_t^j - \alpha_\ell^j \ell_{it} - \alpha_k^j k_{it}. \tag{15}$$

By the change of variables in equation (15), we can relate the probability density function of y_{it} conditional on s_{it}, ℓ_{it}, and k_{it} to the probability of density function of ω_{it} as $\mathrm{f}_t^j(y_{it}|s_{it},\ell_{it},k_{it}) = (1-\alpha_m^j)\mathrm{f}_t^j(\omega_{it})$. Then, we have

$$\mathrm{f}_1^j(y_{i1}, \ell_{i1}, k_{i1}|s_{i1};\theta_1^j,\theta_2^j) = (1-\alpha_m)\mathrm{f}_1^j(\omega_{i1}, \ell_{i1}, k_{i1};\theta_1^j,\theta_2^j), \tag{16}$$

$$\begin{aligned}\mathrm{f}_t^j(y_{it}|s_{it}, \ell_{it}, k_{it}, y_{it-1}, s_{it-1}, x_{it-1};\theta_1^j,\theta_2^j) &= (1-\alpha_m)\mathrm{f}_t^j(\omega_{it}|y_{it-1}, \ell_{it}, k_{it}, s_{it-1}, x_{it-1};\theta_1^j,\theta_2^j) \\ &= (1-\alpha_m)\mathrm{f}_t^j(\omega_{it}|\omega_{it-1};\theta_1^j,\theta_2^j),\end{aligned} \tag{17}$$

$$\mathrm{f}_t^j(\ell_{it}, k_{it}|y_{it-1}, s_{it-1}, x_{it-1};\theta_1^j,\theta_2^j) = \mathrm{f}_t^j(\ell_{it}, k_{it}|\ell_{it-1}, k_{it-1}, \omega_{it-1};\theta_1^j,\theta_2^j), \tag{18}$$

where (16) and the first equality of (17) follow from the change of variables from y_{it} to ω_{it} in view of (15), the second equality of (17) holds because ω_{it} follows the exogenous Markov process under Assumption 2(a) where the value of ω_{it-1} can be recovered from $(y_{it-1}, s_{it-1}, \ell_{it-1}, k_{it-1})$ as in (15), and (18) follows from Assumption 3(b). Therefore, from (13), (16)-(18), we have

$$\begin{aligned}L_{2i}(\theta_2^j,\theta_1^j) &= (1-\alpha_m)\mathrm{f}_1(\omega_{i1}, \ell_{i1}, k_{i1};\theta_1^j,\theta_2^j) \\ &\quad \times \prod_{t=2}^{T}(1-\alpha_m)\mathrm{f}_t^j(\omega_{it}|\omega_{it-1};\theta_1^j,\theta_2^j)\mathrm{f}_t^j(\ell_{it}, k_{it}|\ell_{it-1}, k_{it-1}, \omega_{it-1};\theta_1^j,\theta_2^j).\end{aligned}$$

Recall $z_{it} := (\ell_{it}, k_{it}, \omega_{it})'$ and $\tilde{z}_{it} := (1, z_{it}')'$. As stated in Assumption 9, we approximate the distribution of the initial distribution of $(\ell_{it}, k_{it}, \omega_{it})$ by multivariate normal distribution while we approximate the distribution of (ℓ_{it}, k_{it}) conditional on $(\omega_{it-1}, \ell_{it-1}, k_{it-1})$ by a normal regression model so that, given θ_2^j, we have

$$\mathrm{f}_1(\omega_{i1}, \ell_{i1}, k_{i1};\theta_1^j,\theta_2^j) = (2\pi)^{-3/2}|\Sigma_1^j|^{-1/2}\exp\left(-\frac{1}{2}(z_{i1}-\mu_1^j)'(\Sigma_1^j)^{-1}(z_{i1}-\mu_1^j)\right)$$

and

$$
\begin{aligned}
&\mathrm{f}_t^j(\omega_{it}|\omega_{it-1};\theta_1^j,\theta_2^j) = \frac{1}{\sigma_\eta^j}\phi\left(\frac{\hat{\eta}^*_{it}(\theta_2^j;\hat{\theta}_1^j)}{\sigma_\eta^j}\right),\\
&\mathrm{f}_t^j(\ell_{it},k_{it}|\ell_{it-1},k_{it-1},\omega_{it-1};\theta_1^j,\theta_2^j) =\\
&\quad (2\pi)^{-1}|\Sigma_z^j|^{-1/2}\exp\left(-\frac{1}{2}((\ell_{it},k_{it})' - (\rho_z^j)'\tilde{z}_{it-1})'(\Sigma_z^j)^{-1}((\ell_{it},k_{it})' - (\rho_z^j)'\tilde{z}_{it-1})\right),
\end{aligned}
$$

where

$$
\hat{\eta}^*_{it}(\theta_2^j;\hat{\theta}_1^j) := \omega^*(y_{it},s_{it},\ell_{it},k_{it};\theta_1^j,\theta_2^j) - \rho_\omega^j\omega^*(y_{it-1},s_{it-1},\ell_{it-1},k_{it-1};\theta_1^j,\theta_2^j). \tag{19}
$$

Therefore,

$$
\begin{aligned}
&L_{2i}(\theta_2^j,\hat{\theta}_1^j) := (1-\alpha_m)^T(2\pi)^{-3/2}|\Sigma_1^j|^{-1/2}\exp\left(-\frac{1}{2}(z_{i1}-\mu_1^j)'(\Sigma_1^j)^{-1}(z_{i1}-\mu_1^j)\right)\\
&\times\prod_{t=2}^T\left\{\frac{1}{\sigma_\eta^j}\phi\left(\frac{\hat{\eta}^*_{it}(\theta_2^j;\hat{\theta}_1^j)}{\sigma_\eta^j}\right)(2\pi)^{-1}|\Sigma_z^j|^{-1/2}\exp\left(-\frac{1}{2}((\ell_{it},k_{it})' - (\rho_z^j)'\tilde{z}_{it-1})'(\Sigma_z^j)^{-1}((\ell_{it},k_{it})' - (\rho_z^j)'\tilde{z}_{it-1})\right)\right\}.
\end{aligned}
$$

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

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

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

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

Collect the model parameters into θ_1 and θ_2 as follows. Let $\theta_1 := (\boldsymbol{\pi}',\boldsymbol{\alpha_m}',\boldsymbol{\sigma_\epsilon}')'$ with $\boldsymbol{\pi}' = (\pi^1,...,\pi^J)$, $\boldsymbol{\alpha_m}' = (\alpha_m^1,...,\alpha_m^J)$, $\boldsymbol{\sigma_\epsilon}' = (\sigma_\epsilon^1,...,\sigma_\epsilon^J)$ and let $\theta_2 = ((\theta_2^1)',...,(\theta_2^J)')'$ with $\theta_2^j = (\{\alpha_t^j\}_{t=2}^T,\alpha_\ell^j,\alpha_k^j)'$ for $j=1,...,J$. We consider the following assumption.

Note that $P_t^j(S_{it}|X_{it}) = P_t^j(S_{it})$ under Cobb-Douglas assumption. Then, by integrating $\{Y_t,X_t\}_{t=1}^T$ in (7), we obtain the distribution of $\{S_{it}\}_{t=1}^T$ under Assumption 8 as

$$
\mathrm{P}(\{S_{it}\}_{t=1}^T) = \sum_{j=1}^J\pi^jL_{1i}(\theta_1^j;T), \tag{20}
$$

where

$$
L_{1i}(\theta_1^j;T) := \prod_{t=1}^T L_{1it}(\theta_1^j) \quad \text{with} \quad L_{1it}(\theta_1^j) := \frac{1}{\sigma_\epsilon^j}\phi\left(\frac{\epsilon^*(s_{it};\theta_1^j)}{\sigma_\epsilon^j}\right).
$$

The type-specific distribution of $\{S_{it}\}_{t=1}^T$ can be identified from the joint distribution of $\{S_{it}\}_{t=1}^T$ as in Corollary 1. Therefore, in the first stage, we estimate θ_1 by the maximum likelihood estimator as

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

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

$$\hat{\pi}_i^j = \frac{\hat{\pi}^j L_{1i}(\hat{\theta}_1^j; T)}{\sum_{k=1}^{J} \hat{\pi}^k L_{1i}(\hat{\theta}_1^k; T)} \quad \text{for } j = 1, ..., J. \tag{21}$$

We classify each firm into one of the J types by taking the type that gives the highest posterior probability as its type. Then, our estimator of $\mathcal{I}^j$ is given by

$$\hat{\mathcal{I}}^j := \{i : \hat{j}(i) = j\} \quad \text{with} \quad \hat{j}(i) = \underset{j=1,..,J}{\operatorname{argmax}} \{\hat{\pi}_i^j\}.$$

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

Proposition 4. *For each $i \in \mathcal{I}^j$, $\hat{\pi}_i^j - 1 = O_p(N^{-1/2})$ under Assumption 10.*

Proposition 4 implies that, when Assumption 10 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 a candidate value of θ_2^j given the first stage estimate $\hat{\theta}_1^j$ as

$$\tilde{\eta}_{it}^*(\theta_2^j; \hat{\theta}_1^j) := \omega^*(y_{it}, s_{it}, \ell_{it}, k_{it}; \hat{\theta}_1^j, \theta_2^j) - \hat{\rho}_\omega^j \omega^*(y_{it-1}, s_{it-1}, \ell_{it-1}, k_{it-1}; \hat{\theta}_1^j, \theta_2^j)$$

where $\hat{\rho}_\omega^j$ is the estimate of ρ_ω^j obtained by regressing $\omega^*(y_{it}, s_{it}, \ell_{it}, k_{it}; \hat{\theta}_1^j, \theta_2^j)$ on $\omega^*(y_{it-1}, s_{it-1}, \ell_{it-1}, k_{it-1}; \hat{\theta}_1^j, \theta_2^j)$ using the subsample of firms that belong to $\hat{\mathcal{I}}^j$. Then, stacking the moment conditions implied by $E[\hat{\eta}_{it}^*(\theta_2^j; \hat{\theta}_1^j) | k_{it}, \ell_{it}] = 0$ for $t = 2, ..., T$, we can use standard GMM procedure to estimate θ_2^j as

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

where $\#\hat{\mathcal{I}}^j$ is the number of elements of $\hat{\mathcal{I}}^j$ while $g_i(\theta_2) := (\tilde{\eta}_{i2}^*(\theta_2^j; \hat{\theta}_1^j) Z_{i2}', ..., \tilde{\eta}_{iT}^*(\theta_2^j; \hat{\theta}_1^j) Z_{iT}')'$ with $Z_{it} := (\ell_{it}, k_{it})'$.

5 Empirical applications

5.1 Data

We use Japanese publicly traded manufacturing firms, 1980-2007. The data set compiled by the Development Bank of Japan (DBJ) contains detailed corporate balance sheet/income statement data for the firms listed on the Tokyo Stock Exchange.[4] The initial value of capital (K) is defined as fixed asset less land from the firm's balance sheet and the subsequent values of capital are constructed by perpetual inventory method. The labor input (L) is the number of employees. The intermediate input (M) is defined as the sum of energy input, material input, transportation cost, outsourcing cost, and changes in input inventories. The output (Y) is defined as the value of total sales plus the changes in inventories of finished goods. The machine investment rate ($\frac{I_{m,it}}{K_{m,it}}$) is defined as the ratio of machine investment to machine capital stock. In this preliminary version, we focus on a sample from Machine industry. Table 1 presents summary statistics for the variables we use in our empirical analysis.

Table 1: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
$\ln Y_{it}$	5602	17.108	1.368	12.191	21.785
$\ln M_{it}$	5602	16.314	1.472	9.003	21.306
$\ln L_{it}$	5602	6.647	1.189	2.890	10.978
$\ln K_{it}$	5602	15.926	1.415	12.223	21.328
$\ln \frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$	5602	-0.777	0.510	-6.930	1.708
$\frac{I_{it}}{K_{it}}$	5602	0.100	0.151	-0.491	2.849

5.2 Evidence for unobserved heterogeneity

In the data, the material share is heterogenous across firms and persistent over time within firm. Figure 1 presents the histogram of $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ across all observations that belongs to Machine industry, which shows a large variation in material shares. In the model, the variation in material shares is coming from idiosyncratic ex-post shocks ϵ_{it}. We may eliminate most of idiosyncratic components by considering the firm-level average of material shares over 28 years; however, in Figure 2, the persistent component of the ratio of intermediate inputs to total sales substantially varies across firms. As shown in Figures 3 and 4, we also observe a large variation in the persistent component in the ratio of intermediate cost to the sum of intermediate cost and total wage bills, $\frac{P_{M,t}M_t}{P_{M,t}M_{it}+W_tL_{it}}$, of which variation is not likely to be driven by a variation in markups.

[4] Because firm's financial data do not necessarily refer to a calendar year, we assign year t to an observation if the given firm's closing date is between June of year t and May of year $t+1$. If firms change their closing dates, the data after the change may refer to less than 12 months. When it occurs, we multiply the data x_{it} by 12/m where m represents the number of months to which the data refer.

Figure 5 plots each firm's material share, output, and inputs from 1980 to 2007. The presence of heterogeneity across firms and the persistence within each firm in materials shares are apparent in the upper left panel of Figure 5. It also appears that labour and capital inputs are changing over time more smoothly than material input, suggesting that material input responds to idiosyncratic shocks more than labour and capital inputs do within a short period of time. As shown in Figure 6, the heterogeneity in material shares across firms do not disappear even when we examine firms within the subindustries of Machine industry, which roughly corresponds to 4-digit ISIC.

In view of the intermediate share equation in (4), a large cross-sectional variation in the persistent component of the ratio of intermediate inputs to total sales suggests either heterogeneity in production function or the persistence in inputs over time. To examine further, we regress $\ln S_{it}$ on the third order polynomials of (l_{it}, k_{it}, m_{it}) to get residuals, denoted by e_{it}, and decompose e_{it} into permanent components and idiosyncratic components as $\hat{\xi}_i := T^{-1}\sum_{t=1}^{T} e_{it}$ and $\hat{\zeta}_{it} := e_{it} - \xi_i$. Comparing the variance of $\hat{\xi}_i$ with that of $\hat{\zeta}_{it}$, we found that $\frac{\mathrm{Var}(\hat{\xi}_i)}{\mathrm{Var}(\hat{\xi}_i)+\mathrm{Var}(\hat{\zeta}_{it})} = 0.612$. Therefore, the majority of variation is coming from the permanent component even after controlling the observed input (l_{it}, k_{it}, m_{it}), suggesting that production function is heterogeneous beyond Hicks-neutral term.

We also estimated the value added Cobb-Douglas specification

$$y_{it}^{va} = \alpha_t^{va} + \alpha_\ell^{va}\ell_{it} + \alpha_k^{va}k_{it} + \omega_{it}^{va} + \epsilon_{it}^{va},$$

by the approach developed by Levinsohn and Petrin, where y_{it}^{va} is the logarithm of value added. When we compute the serial correlation of estimated values of ϵ_{it}^{va}, we found that the correlation coefficient of 0.85.[5] One possible reason for this high correlation of estimated values of ϵ_{it}^{va} is the presence of unobserved heterogeneity in $(\alpha_t^{va}, \alpha_\ell^{va}, \alpha_k^{va})$.

5.3 Estimation of production function

Given the relatively long length of our panel data, we apply our proposed estimation method based on classifying each firm into one of the J types. Table 2 presents the parameter estimates for the number of components equal to $J = 1, 3$, and 5. Setting $J = 1$ gives the homogenous production function specification considered by GNR.

The estimated coefficients across different types when $J = 3$ and 5 suggest that there are substantial differences in the output elasticities with respect to materials, labor, and capital across firms. For the model with $J = 3$, the material share is lowest for Type 1 and highest for Type 3 while Type 1 is more labor intensive than Type 2 or 3. For the model with $J = 5$, the material share of Type 1 is the highest while the material share of Type 2 is the lowest among three types. The degree of capital intensity is also different across five types, where Type 5 is the most capital intensive while Type 2 is the most labor intensive.

Figure 7 shows the distribution of posterior type probabilities, defined in (21), across firms for

[5] Using the OP approach with the value-added specification, Fox and Smeets (2007) also report the high serial correlation of estimated values of idiosyncratic shocks.

Table 2: Estimates of Production Function (10): Machine Industry in Japan, 1980-2008

	GNR	Random Coefficients Model							
	Estimation by Classification								
	$J=1$	$J=3$			$J=5$				
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 4	Type 5
β_m^j	0.340	0.623	0.184	0.438	0.112	0.642	0.387	0.267	0.509
β_ℓ^j	0.422	0.215	0.845	0.416	0.954	0.200	0.577	0.611	0.357
β_k^j	0.260	0.162	0.134	0.195	0.097	0.154	0.089	0.230	0.162
$\beta_m^j + \beta_\ell^j + \beta_k^j$	1.021	1.000	1.163	1.049	1.163	0.995	1.054	1.109	1.028
β_k^j / β_ℓ^j	0.617	0.753	0.159	0.470	0.102	0.772	0.155	0.376	0.455
π^j	1.000	0.467	0.200	0.333	0.082	0.373	0.155	0.113	0.277
No. of Obs.	5602								
No. of firms	240								

the model with $J = 3$ and 5, respectively. The posterior probabilities for each type are concentrated on around 0 or 1, which is consistent with the result of Proposition 4 where Assumption 10 could be roughly applied here given $T = 28$ in our data set. We assign one of the J types to each firm based on its posterior type probability that achieves the highest value across J types.

Figure 8 plots each firm's material share and the log of output from 1980 to 2007, where different colours represent different types for the model with $J = 3$. From the left panel of Figure 8, it is clear that each firm's type is identified with its average material share. On the other hand, it does not appear that there is any systematic difference across types in terms of the distribution of firm sizes measured by outputs.

Table 3 shows a fraction of firms belonging to each type of the J types within subindustries of Machine industry for the model with $J = 3$ and 5. See Figure 6 for the definition of subindustries. The distribution of types is quite different across subindustries. Figure 9 plots each firm's material share from 1980 to 2007 where different colours identify firm's type for the model with $J = 3$, suggesting that our framework flexibly captures the unobserved heterogeneity in production technology within more narrowly defined subindustries.

The first two rows of Table 4 present the standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_t^j$ and $\hat{\epsilon}_{it}$. To compute the standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_t^j$ and $\hat{\epsilon}_{it}$, we compute $\hat{\omega}_{it}$, $\hat{\xi}_i$, and $\hat{\epsilon}_{it}$ by assigning one of the $J = 3$ types to each firm based on its posterior type probabilities. As the number of components J increases from 1 to 3, and then to 5, the standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_t^j$ and $\hat{\epsilon}_{it}$ within each type decrease on average across types. This indicates a possibility of substantial upward bias in the estimated variation in ex-post shocks in homogenous model with $J = 1$ as a result of ignoring unobserved heterogeneity.

The third row of Table 4 reports the serial correlation in $\hat{\epsilon}_{it}$. The serial correlation in $\hat{\epsilon}_{it}$ is very high at 0.951 when $J = 1$. Given that the presence of high serial correlation indicates a possibility of misspecificaiton of the model, a smaller value of serial correlation is more desirable. When the

Table 3: A fraction of firms for each type by subindustry for the model with $J = 3$ and 5

	$J=3$			$J=5$				
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 4	Type 5
2511	0.890	0.000	0.110	0.000	0.890	0.110	0.000	0.000
2521	0.734	0.122	0.144	0.045	0.614	0.000	0.077	0.264
2522	0.599	0.288	0.114	0.220	0.379	0.000	0.068	0.333
2523	0.120	0.776	0.104	0.000	0.120	0.104	0.776	0.000
2529	0.000	0.417	0.583	0.000	0.000	0.583	0.417	0.000
2531	0.564	0.178	0.258	0.000	0.195	0.000	0.178	0.628
2532	0.231	0.391	0.378	0.338	0.102	0.120	0.053	0.387
2533	0.481	0.000	0.519	0.000	0.481	0.193	0.000	0.326
2534	0.762	0.140	0.098	0.140	0.738	0.000	0.000	0.122
2535	0.377	0.449	0.174	0.247	0.237	0.130	0.154	0.233
2536	0.506	0.090	0.404	0.008	0.364	0.138	0.082	0.408
2537	0.462	0.120	0.418	0.071	0.399	0.165	0.049	0.316
2541	0.154	0.215	0.631	0.046	0.133	0.471	0.170	0.181

Notes: Subindustries are 2511: Boiler prime mover, 2521: Metal machine tools, 2522: Metal working machinery, 2523: Machinery tool, 2531: Textile machinery, 2532: Agricultural machines, 2533: Construction and mining equipment, 2534: Chemical machinery, 2535: Office machinery, 2536: Special industrial machinery, 2537: General industrial machinery, 2541: General Mechanical Components.

number of components increases from $J = 1$ to $J = 3$, and then to $J = 5$, the average serial correlation in $\hat{\epsilon}_{it}$ across types decreases from 0.951 to 0.762, and then to 0.693, indicating that the very high serial correlation in $\hat{\epsilon}_{it}$ when $J = 1$ is partly due to ignoring unobserved heterogeneity in production function coefficients. On the other hand, the level of serial correlation in $\hat{\epsilon}_{it}$ is still high at 0.693 when $J = 5$. To examine this issue, we also consider a more flexible model in which the material share parameter is firm-specific. In this case, the material share equation is given by firm-fixed effects model: $s_{it} = \alpha_{m,i} - \epsilon_{it}$. The estimated serial correlation in ϵ_{it} for this firm-fixed effects model remains quite strong at 0.773, suggesting that the material share parameter could change over time persistently within the same firm.

Ignoring unobserved heterogeneity may lead to substantial biases in the measurement of productivity growth. To examine this issue, we take a specification with $J = 5$ 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 $\Delta\omega_{it} := \Delta y_{it} - (\hat{\alpha}_t^j + \hat{\alpha}_m^j \Delta m_{it} + \hat{\alpha}_\ell^j \Delta\ell_{it} + \hat{\alpha}_k^j \Delta k_{it} + \Delta\hat{\epsilon}_{it}^j)$ for $j = 1, 2, ..., 5$ be an estimated productivity growth when $J = 5$ and let $\Delta\tilde{\omega}_{it} := \Delta y_{it} - (\bar{\alpha}_t + \bar{\alpha}_m \Delta m_{it} + \bar{\alpha}_\ell \Delta\ell_{it} + \bar{\alpha}_k \Delta k_{it} + \Delta\bar{\epsilon}_{it})$ be an estimated productivity growth when $J = 1$, where $\{\hat{\alpha}_t^j, \hat{\alpha}_m^j, \hat{\alpha}_\ell^j, \hat{\alpha}_k^j\}_{j=1}^5$ and $\{\bar{\alpha}_t, \bar{\alpha}_m^j, \bar{\alpha}_\ell^j, \bar{\alpha}_k^j\}$ denote estimated coefficients when $J = 5$ and $J = 1$, respectively. Then, we compute the bias as

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

Table 5 reports the ratio of the average absolute value of bias to the average productivity growth within each of five subsamples, classified by types. The magnitude of the bias is high on average and substantially different across different types. In particular, the bias in the measured productivity growth when $J = 1$ is larger than 40 percent of the average productivity growth for Type 1 and 2, indicating that ignoring unobserved heterogeneity could result in serious bias in estimated productivity growth.

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

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

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

Table 6 reports the estimated coefficients of ω_{it} across different specifications and different types for $J = 1$, 3, and 5. The coefficient of ω_{it} is estimated significantly at 0.016 when $J = 1$. For the model with $J = 3$ and 5, the estimated coefficients of ω_{it} are substantially different across different types of firms. For $J = 3$, the coefficient of ω_{it} for Type 2 is insignificant at -0.007 while the coefficients of ω_{it} are estimated significantly at 0.090 and 0.068 for Type 1 and 3, respectively. As

Table 4: Random Coefficients Model (10): Machine Industry in Japan, 1980-2008

	Estimation by Classification										
	$J=1$	$J=3$				$J=5$					
		Type 1	Type 2	Type 3	Ave[†]	Type 1	Type 2	Type 3	Type 4	Type 5	Ave[†]
Std. Dev. $\hat{\omega}_{it} + \hat{\alpha}_t^j$	0.472	0.195	0.620	0.241	0.295	0.895	0.185	0.307	0.359	0.190	0.283
Std. Dev. $\hat{\epsilon}_{it}$	0.510	0.138	0.703	0.178	0.264	1.021	0.126	0.170	0.236	0.130	0.220
Corr$(\hat{\epsilon}_{it}, \hat{\epsilon}_{it-1})$	0.951	0.663	0.923	0.806	0.762	0.924	0.602	0.733	0.859	0.658	0.693

[†] The columns under "Ave." report the average across types using the estimated mixing proportions $\hat{\pi}^j$ as weights.

Table 5: Bias in the measurement of productivity: Machine Industry in Japan, 1980-2008

	Estimation by Classification				
	$J=5$				
	Type 1	Type 2	Type 3	Type 4	Type 5
$\frac{\text{Mean of } \lvert \text{Bias}_{it} \rvert}{\text{Mean of } \lvert \Delta\tilde{\omega}_{it} \rvert}$	0.403	0.453	0.170	0.123	0.257

Table 6: The Effect of ω_{it} on Investment

	by Classification								
	$J=1$	$J=3$			$J=5$				
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3	Type 4	Type 5
α_ω	0.016	0.090	-0.007	0.068	-0.016	0.094	0.054	0.010	0.083
	(0.004)	(0.018)	(0.006)	(0.014)	(0.007)	(0.022)	(0.014)	(0.013)	(0.020)
β_m^j	0.340	0.623	0.184	0.438	0.112	0.642	0.387	0.267	0.509
β_k^j/β_ℓ^j	0.617	0.753	0.159	0.470	0.102	0.772	0.155	0.376	0.455

Note: *p<0.1; **p<0.05; ***p<0.01

reported in the second and third rows of Table 6, the material shares of Type 1 and 3 are higher than that of Type 2 while Type 1 and 3 is more capital intensive than Type 2 for the model with $J=3$. For $J=5$, the coefficients of ω_{it} are high at 0.094 and 0.083 for Type 2 and 5, respectively, both of which have relatively higher material shares and more capital intensive technology than other three types. Therefore, we find that the correlation between productivity and investment is stronger among a type of firms with high material shares and high capital intensive production technology than other types of firms.

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., Caves, Kevin, and Garth Frazer (2006) "Structural Identification of Production Functions," working paper, UCLA.

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

[5] 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.

[6] 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.

[7] 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.

[8] Doraszelski, Ulrich, and Jordi Jaumandreu (2014) "Measuring the Bias of Technological Change," working paper, University of Pennsylvania.

[9] Gandhi, A., S. Navarro, and D. Rivers (2013) "On the Identification of Production Functions: How Heterogeneous id Productivity?," working paper, Western University.

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

[11] 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.

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

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

[14] 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.

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

[16] 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)*, forthcoming.

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

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

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

[20] 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,

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

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

[23] 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

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

$$
\begin{aligned}
&\mathrm{P}_t^j(\{Y_t, S_t, X_t\}_{t=1}^T) = \mathrm{P}_1^j(Y_1, S_1, X_1)\prod_{t=2}^T \mathrm{P}_t^j(Y_t, S_t, X_t|\{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_1^j(Y_1|S_1, X_1)\mathrm{P}_1^j(S_1, X_1) \\
&\quad \times \prod_{t=2}^T \mathrm{P}_t^j(Y_t|S_t, X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(S_t|X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(X_t|\{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}).
\end{aligned} \tag{22}
$$

For $t \geq 2$, in view of the second equation of (4), we have

$$
\mathrm{P}_t^j(S_t|X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(S_t|X_t) := g_\epsilon(\ln G_{M,t}^j(X_t) + \ln \mathcal{E} - \ln S_t). \tag{23}
$$

Furthermore,

$$
\begin{aligned}
&\mathrm{P}_t^j(X_t|\{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(L_t, K_t, \omega_t|\{Y_{t-s}, S_{t-s}, L_{t-s}, K_{t-s}, \omega_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_t^j(\omega_t|L_t, K_t, \{Y_{t-s}, S_{t-s}, L_{t-s}, K_{t-s}, \omega_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(L_t, K_t|\{Y_{t-s}, S_{t-s}, L_{t-s}, K_{t-s}, \omega_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_\omega^j(\omega_t|\omega_{t-1})\mathrm{P}_t^j(L_t, K_t|L_{t-1}, K_{t-1}, \omega_{t-1}) \\
&= \mathrm{P}_t^j(L_t, K_t, \omega_t|L_{t-1}, K_{t-1}, \omega_{t-1}) \\
&= \mathrm{P}_t^j(X_t|X_{t-1}),
\end{aligned} \tag{24}
$$

where the first and the last equality hold because there is one-to-one mapping between X_t and (L_t, K_t, ω) in view of Assumption 4(b) while the third equality follows from Assumptions 2(a) and 3(b).

The stated result follows from (22)-(24) if we show that $\mathrm{P}_t^j(Y_t|S_t, X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(Y_t|S_t, X_t, Y_{t-1}, S_{t-1}, X_{t-1})$. Given $\theta_1 := \{g_\epsilon^j(\cdot), \mathcal{E}^j, \{G_{M,t}^j(\cdot)\}_{t=1}^T\}_{j=1}^J$, define

$$
\begin{aligned}
\epsilon_t^j(S_t, X_t; \theta_1) &:= \ln(G_{M,t}^j(X_t)\mathcal{E}^j) - \ln S_t, \\
\mathcal{Y}_t^j(Y_t, S_t, X_t; \theta_1) &:= \ln Y_t - \int \frac{G_{M,t}^j(L_t, K_t, M_t)}{M_t} dM_t - \epsilon_t^j(S_t, X_t; \theta_1).
\end{aligned} \tag{25}
$$

Integrating both sides of $\frac{G_{M,t}^j(L_t,K_t,M_t)}{M_t} = (\partial/\partial M_t)\ln F_t^j(L_t, K_t, M_t)$ with respect to M_t gives

$$
\int \frac{G_{M,t}^j(L_t, K_t, M_t)}{M_t} dM_t = \ln F_t^j(L_t, K_t, M_t) + \mathcal{C}_t^j(L_t, K_t), \tag{26}
$$

where $\mathcal{C}_t^j(L_t, K_t)$ is a function that only depends on (L_t, K_t). Then, substituting (26) into the second equation of (25) and combining it with (4) gives

$$\omega_{it} = \mathcal{Y}_t^j(Y_t, S_t, X_t; \theta_1) + \mathcal{C}_t^j(K_t, L_t). \tag{27}$$

Then, it follows from (2), (4), (25), and (27) that

$$\mathcal{Y}_t^j(Y_t, S_t, X_t; \theta_1) = -\mathcal{C}_t^j(K_t, L_t) + \underbrace{h^j\left(\mathcal{Y}_{t-1}^j(Y_{t-1}, S_{t-1}, X_{t-1}; \theta_1) + \mathcal{C}_{t-1}^j(K_{t-1}, L_{t-1})\right)}_{=\tilde{h}^j(\mathcal{Y}_{t-1}^j(Y_{t-1}, S_{t-1}, X_{t-1}; \theta_1), K_{t-1}, L_{t-1})} + \eta_t. \tag{28}$$

Because η_t is i.i.d., equation (28) implies that

$$\mathrm{P}_t^j(Y_t | S_t, X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(Y_t | S_t, X_t, Y_{t-1}, S_{t-1}, X_{t-1}), \tag{29}$$

and the stated result follows. □

A.2 Proof of Proposition 2

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 (W_2, W_3) at (w_2, w_3) and choose $(\bar{w}_2, \bar{w}_3) \in \mathcal{W}_2 \times \mathcal{W}_3$, $(a_1, ..., a_J) \in \mathcal{W}_1^J$ and $(b_1, ..., b_{J-1}) \in \mathcal{W}_4^{J-1}$ that satisfy Assumption 7. Evaluating (7) at $(W_2, W_3) = (w_2, w_3)$ gives

$$\begin{aligned} P(\{W_t\}_{t=1}^4) &= \sum_{j=1}^J \pi^j P_4^j(W_4|w_3) P_3^j(w_3|w_2) P_2^j(w_2|W_1) P_1^j(W_1) \\ &= \sum_{j=1}^J \lambda_4^j(W_4|w_3) \lambda_3^j(w_3|w_2) \bar{\lambda}_2^j(W_1, w_2), \end{aligned} \tag{30}$$

where $\lambda_4^j(W_4|w_3) := P_4^j(W_4|W_3 = w_3)$, $\lambda_3^j(w_3|w_2) := P_3^j(W_3 = w_3|W_2 = w_2)$, and $\bar{\lambda}_2^j(W_1, w_2) := \pi^j P_2^j(W_2 = w_2|W_1) P_1^j(W_1)$. Integrating out W_4 from (30) gives

$$P(\{W_t\}_{t=1}^3) = \sum_{j=1}^J \lambda_3^j(w_3|w_2) \bar{\lambda}_2^j(W_1, w_2). \tag{31}$$

Let $f_{w_2,w_3}(a, b) := P((W_1, W_2, W_3, W_4) = (a, w_2, w_3, b))$ and $\bar{f}_{w_2,w_3}(a) := P((W_1, W_2, W_3) = (a, w_2, w_3))$. Evaluating (30) at $W_1 = a_1, ..., a_J$ and $W_4 = b_1, ..., b_{J-1}$ gives $M(M-1)$ equations while evaluating (31) at $W_1 = a_1, ..., a_J$ gives M equations. Collecting these $M(M-1) + M = M^2$ equations and denoting them using matrix notation, we have

$$P_{w_2,w_3} = L_{w_3}' D_{w_3|w_2} \bar{L}_{w_2}, \tag{32}$$

where

$$P_{w_2,w_3} := \begin{bmatrix} \bar{f}_{w_2,w_3}(a_1) & \bar{f}_{w_2,w_3}(a_2) & \cdots & \bar{f}_{w_2,w_3}(a_J) \\ f_{w_2,w_3}(a_1,b_1) & f_{w_2,w_3}(a_2,b_1) & \cdots & f_{w_2,w_3}(a_J,b_1) \\ \vdots & \vdots & \ldots & \vdots \\ f_{w_2,w_3}(a_1,b_{J-1}) & f_{w_2,w_3}(a_2,b_1) & \cdots & f_{w_2,w_3}(a_J,b_{J-1}) \end{bmatrix},$$

$$L_{w_3} := \begin{bmatrix} 1 & \lambda_4^1(b_1|w_3) & \cdots & \lambda_4^1(b_{J-1}|w_3) \\ \vdots & \vdots & \ldots & \vdots \\ 1 & \lambda_4^J(b_1|w_3) & \cdots & \lambda_4^J(b_{J-1}|w_3) \end{bmatrix}, \quad \bar{L}_{w_2} := \begin{bmatrix} \bar{\lambda}_2^1(a_1,w_2) & \cdots & \bar{\lambda}_2^1(a_J,w_2) \\ \vdots & \vdots & \ldots \\ \bar{\lambda}_2^J(a_1,w_2) & \cdots & \bar{\lambda}_2^J(a_J,w_2) \end{bmatrix}, \tag{33}$$

and $D_{w_3|w_2} := \text{diag}\left(\lambda_3^1(w_3|w_2), ..., \lambda_3^J(w_3|w_2)\right)$. Evaluating (32) at four different points, (w_2, w_3), $(\bar{w}_2, w_3)$, $(w_2, \bar{w}_3)$, and $(\bar{w}_2, \bar{w}_3)$ gives

$$\begin{aligned} P_{w_2,w_3} &= L'_{w_3} D_{w_3|w_2} \bar{L}_{w_2}, \quad P_{\bar{w}_2,w_3} = L'_{w_3} D_{w_3|\bar{w}_2} \bar{L}_{\bar{w}_2}, \\ P_{w_2,\bar{w}_3} &= L'_{\bar{w}_3} D_{\bar{w}_3|w_2} \bar{L}_{w_2}, \quad P_{\bar{w}_2,\bar{w}_3} = L'_{\bar{w}_3} D_{\bar{w}_3|\bar{w}_2} \bar{L}_{\bar{w}_2}. \end{aligned}$$

Then, under Assumption 7,

$$A := P_{w_2,w_3}(P_{w_2,\bar{w}_3})^{-1} P_{\bar{w}_2,\bar{w}_3}(P_{\bar{w}_2,w_3})^{-1} = L'_{w_3} D_{w_2,\bar{w}_2,w_3,\bar{w}_3} (L'_{w_3})^{-1},$$

where

$$D_{w_2,\bar{w}_2,w_3,\bar{w}_3} := D_{w_3|w_2}(D_{\bar{w}_3|w_2})^{-1} D_{\bar{w}_3|\bar{w}_2}(D_{w_3|\bar{w}_2})^{-1}. \tag{34}$$

Because $AL'_{w_3} = L'_{w_3} D_{w_2,\bar{w}_2,w_3,\bar{w}_3}$, the eignvalues of A determine the diagonal elements of $D_{w_2,\bar{w}_2,w_3,\bar{w}_3}$ while the right eigenvectors of A determine the columns of L'_{w_3} up to multiplicative constant. Denote the right eigenvectors of A by $L'_{w_3}C$, where C is some diagonal matrix. Now we can determine the diagonal matrix $D_{w_2,\bar{w}_2,w_3,\bar{w}_3}C$ from the first row of $AL'_{w_3}C = L'_{w_3} D_{w_2,\bar{w}_2,w_3,\bar{w}_3}C$ because the first row of L'_{w_3} is a vector of ones. Then, L'_{w_3} is determined uniquely from $AL'_{w_3}C$ and $D_{w_2,\bar{w}_2,w_3,\bar{w}_3}C$ as $L'_{w_3} = (AL'_{w_3}C)(D_{w_2,\bar{w}_2,w_3,\bar{w}_3}C)^{-1}$ in view of $AL'_{w_3} = L'_{w_3} D_{w_2,\bar{w}_2,w_3,\bar{w}_3}$. Therefore, L_{w_3} is identified. Repeating the above argument for all values of $w_3 \in \mathcal{W}_3$ identifies $\{P_4^j(W_4|W_3 = w_3)\}_{j=1}^J$ for each $w_3 \in \mathcal{W}_3$ for $W_4 = (b_1, ..., b_{J-1})$ that satisfies Assumption 7(a).

Evaluating $P(W_4, W_3|W_2)$ at $(W_2, W_3) = (w_2, w_3)$, we have

$$P(W_4, W_3 = w_3|W_2 = w_2) = \sum_{j=1}^J \tilde{\pi}_{w_2}^j P_4^j(W_4|w_3) P_3^j(w_3|w_2) = \sum_{j=1}^J \lambda_4^j(W_4|w_3) \tilde{\lambda}_3^j(w_3|w_2), \tag{35}$$

where $\tilde{\pi}_{w_2}^j := \frac{\pi^j P_2^j(W_2=w_2)}{P_2(W_2=w_2)}$ and $\tilde{\lambda}_3^j(w_3|w_2) := \tilde{\pi}_{w_2}^j P_3^j(W_3 = w_3|W_2 = w_2)$. Then, evaluating (35) at $W_4 = b_1, ..., b_{J-1}$ and collecting them into a vector together with $P(W_3 = w_3|W_2 = w_2) = \sum_{j=1}^J \tilde{\lambda}_3^j(w_3|w_2)$ gives

$$p_{w_3|w_2} = L'_{w_3} d_{w_3|w_2},$$

where $d_{w_3|w_2} = (\tilde{\lambda}_3^1(w_3|w_2), ..., \tilde{\lambda}_3^J(w_3|w_2))'$ and $p_{w_3|w_2} = (P(W_3 = w_3|W_2 = w_2), P((W_4, W_3) = (b_1, w_3)|W_2 = w_2), ..., P((W_4, W_3) = (b_{J-1}, w_3)|W_2 = w_2))'$. Therefore, we uniquely determine $\tilde{\pi}_{w_2}^j P_3^j(W_3 = w_3|W_2 = w_2)$ from $d_{w_3|w_2} = (L'_{w_3})^{-1} p_{w_3|w_2}$. Repeating the above argument across all possible values of $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$ determines the value of $\tilde{\pi}_{w_2}^j P_3^j(W_3 = w_3|W_2 = w_2)$ for every $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$. Then, $\tilde{\pi}_{w_2}^j$ and $P_3^j(W_3 = w_3|W_2 = w_2)$ are uniquely identified as $\tilde{\pi}_{w_2}^j = \int_{\mathcal{W}_3} \tilde{\pi}_{w_2}^j P_3^j(W_3|W_2 = w_2) dW_3$ and $P^j(W_3 = w_3|W_2 = w_2) = [\tilde{\pi}_{w_2}^j P_3^j(W_3 = w_3|W_2 = w_2)]/\tilde{\pi}_{w_2}^j$. Therefore, $\{P_3^j(W_3|W_2)\}_{j=1}^J$ is identified.

Evaluating $P_3^j(W_3|W_2)$ at $(W_2, W_3) = (w_3, w_2)$ for $j = 1, ..., J$ identifies $D_{w_3|w_2}$ and, from (32), $\bar{L}_{w_2}$ is identified as $\bar{L}_{w_2} = (D_{w_3|w_2})^{-1}(L'_{w_3})^{-1} P_{w_2,w_3}$. Once $D_{w_3|w_2}$ and $\bar{L}_{w_2}$ are identified, we can determine $\ell_{w_3}(\zeta) = (\lambda_4^1(\zeta|w_3), ..., \lambda_4^J(\zeta|w_3))'$ for any $\zeta \in \mathcal{W}_4$ by constructing $p_{w_2,w_3}(\zeta) = (f_{w_2,w_3}(a_1, \zeta), f_{w_2,w_3}(a_2, \zeta), ..., f_{w_2,w_3}(a_J, \zeta))$ from the observed data, and using the relationship $\ell_{w_3}(\zeta) = (D_{w_3|w_2})^{-1}(\bar{L}'_{w_2})^{-1} p_{w_2,w_3}(\zeta)'$. Similarly, we can determine $\bar{\ell}_{w_2}(\xi) = (\bar{\lambda}_2^1(\xi, w_2), ..., \bar{\lambda}_2^J(\xi, w_2))'$ for any $\xi \in \mathcal{W}_1$ by constructing $\bar{p}_{w_2,w_3}(\xi) = (\bar{f}_{w_2,w_3}(\xi), f_{w_2,w_3}(\xi, b_1), f_{w_2,w_3}(\xi, b_2), ..., f_{w_2,w_3}(\xi, b_{J-1}))'$ and using the relationship $\bar{\ell}_{w_2}(\xi) = (D_{w_3|w_2})^{-1}(L'_{w_3})^{-1}\bar{p}_{w_2,w_3}(\xi)$. Therefore, $\{P_4^j(W_4|W_3 = w_3), \pi^j P_2^j(W_2 = w_2|W_1)P_1^j(W_1)\}_{j=1}^J$ is identified. Repeating this argument for all possible values of $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$ identifies $\{P_4^j(W_4|W_3), \pi^j P_2^j(W_2|W_1)P_1^j(W_1)\}_{j=1}^J$. Finally, $\{\pi^j, P_2^j(W_2|W_1), P_1^j(W_1)\}_{j=1}^J$ is identified from $\{\pi^j P_2^j(W_2|W_1)P_1^j(W_1)\}_{j=1}^J$ as $\pi^j = \int_{\mathcal{W}_1}\int_{\mathcal{W}_2}[\pi^j P_2^j(W_2|W_1)P_1^j(W_1)]dW_2 dW_1$, $P_1^j(W_1) = [\int_{\mathcal{W}_2}[\pi^j P_2^j(W_2|W_1)P_1^j(W_1)]dW_2]/\pi^j$, and $P_2^j(W_2|W_1) = [\pi^j P_2^j(W_2|W_1)P_1^j(W_1)]/[\pi^j \times P_1^j(W_1)]$. This proves the stated result. □

A.3 Proof of Proposition 3

Given that $\mathrm{P}_t^j(Y_t, S_t, X_t)$ is identified from Proposition 2, we may apply the argument in section 3 of GNR to prove the stated result. For part (a), we may generate infinitely many observations of $\{S_t, X_t\}$ from $\mathrm{P}_t^j(S_t, X_t)$ and nonparametric regression of $\ln S_t$ on X_t identifies $G_{M,t}^j \mathcal{E}^j$ and $\epsilon_t = \ln(G_{M,t}^j(X_t)\mathcal{E}^j) - \ln S_t$. Consequently, we may identify g_ϵ^j from $\{\epsilon_t\}$, and $\mathcal{E}^j$ and $G_{M,t}^j$ are identified as $\mathcal{E}^j = E_\epsilon[e^\epsilon|\text{type} = j]$ and $G_{M,t}^j = (G_{M,t}^j \mathcal{E}^j)/\mathcal{E}^j$, respectively. Therefore, θ_1 is identified, proving part (a).

For part (b), we can identify $\mathcal{C}_t^j(K_t, L_t)$ from (28) given infinitely many observations randomly drawn from $\mathrm{P}^j(\{Y_t, S_t, X_t\}_{t=1}^T)$ via non-parametric regression of $\mathcal{Y}_t^j := \mathcal{Y}_t^j(Y_t, S_t, X_t; \theta_1)$ on $(K_t, L_t, \mathcal{Y}_{t-1}^j, K_{t-1}, L_{t-1})$ while imposing additive separability implied by (28). Given $\mathcal{C}_t^j(K_t, L_t)$, the identification of $F_t^j(\cdot)$ and $\{\omega_t\}_{t=1}^T$ follows from (26) and (27), respectively, and $h^j(\cdot)$ is identified from $\{\omega_t\}_{t=1}^T$. Finally, because η_t is identified from (28), $g_\eta^j(\cdot)$ is also identified. This proves part (b). □

A.4 Proof of Proposition 4

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,

$$\hat{\pi}_i^j - 1 = (\hat{\pi}_i^j - \pi_T^j) + (\pi_T^j - 1). \tag{36}$$

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 (36), 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

$$\pi_T^j = \frac{1}{1 + \sum_{k \neq j} (\pi^{k^*}/\pi^{j*}) \exp\left(-\sum_{t=1}^T \xi_{it}^{jk}\right)}. \tag{37}$$

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 $\sum_{k \neq j} (\pi^{k^*}/\pi^{j*}) \exp\left(-\sum_{t=1}^T \xi_{it}^{jk}\right)$ is $O_p\left(\exp(-a^j T)T^{1/2}\right)$, where $a^j := \min_{k \neq j} a^{jk}$. Therefore, in view of (37), the consistency of π_T^j as $T \to \infty$ and the mean value theorem give

$$\pi_T^j - 1 = O_p\left(\exp(-a^j T)T^{1/2}\right). \tag{38}$$

Then, the stated result follows from (36), (38), and $\hat{\pi}_i^j - \pi_T^j = O_p(N^{-1/2})$ because $O_p\left(\exp(-a^j T)T^{1/2}\right) = o_p(N^{-1/2})$ as $N, T \to \infty$ under Assumption 10. $\square$

A.5 Assumption 7 under Cobb-Douglas production function

In the following, we discuss the conditions under which Assumption 7 holds when the production function is Cobb-Douglas.

Example 1 (continued)**.** *For random coefficients model (8), we may write $L_{\bar{w}_3}$, $\bar{L}_{\bar{w}_2}$, and $D_{\bar{w}_3|\bar{w}_2}$ 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 W_1 and W_4. Denote $\bar{w}_3 = (y_3, \bar{s}_3, \bar{x}_3)$*

and $b_h = (y_3, b_h^s, \bar{x}_4)$ *for* $h = 1, ..., J-1$. *Then,*

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

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

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

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

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

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*

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

For Assumption 7(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* $L_{\bar{w}_3}$ *and* $\bar{L}_{\bar{w}_2}$ *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* $L_{\bar{w}_3}$ *and* $\bar{L}_{\bar{w}_2}$ *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 $\bar{w}_3 = (\bar{s}_3, \bar{x}_3)$ *and* $\bar{w}_2 = (\bar{s}_2, \bar{x}_2)$. *Then,*

$$\lambda^j(\bar{w}_3 | \bar{w}_2) = \pi^j g_\epsilon^j(\ln \bar{s}_3 - \ln(\alpha_{m,3}^j \mathcal{E}^j)) P_3^j(X_3 = \bar{x}_3 | X_2 = \bar{x}_2). \tag{39}$$

Pick $w_3 = (s_3, x_3)$ *and* $w_2 = (s_2, x_2)$. *Assumption 7(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*

$$D_{w_3|w_2}(D_{\bar{w}_3|w_2})^{-1} D_{\bar{w}_3|\bar{w}_2}(D_{w_3|\bar{w}_2})^{-1} = diag\left\{ \frac{P_3^1(x_3|x_2)}{P_3^1(\bar{x}_3|x_2)} \frac{P_3^1(\bar{x}_3|\bar{x}_2)}{P_3^1(x_3|\bar{x}_2)}, ..., \frac{P_3^J(x_3|x_2)}{P_3^J(\bar{x}_3|x_2)} \frac{P_3^J(\bar{x}_3|\bar{x}_2)}{P_3^J(x_3|\bar{x}_2)} \right\}.$$

Therefore, Assumption 7(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.

Figure 1: Histogram of $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$

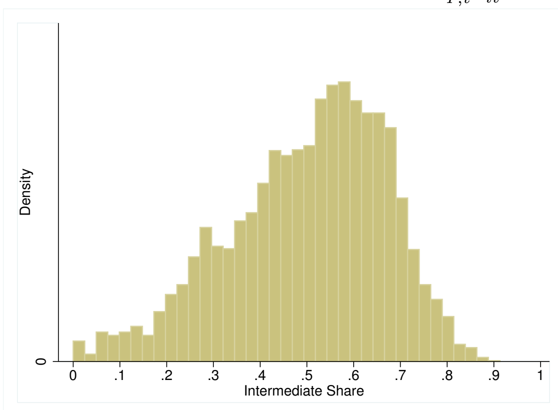


Figure 2: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)_i}$ over 28 years

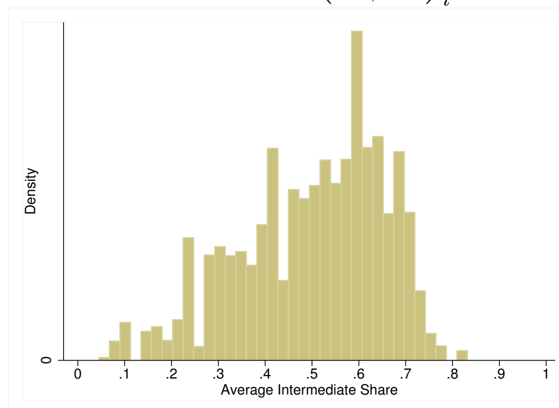


Figure 3: $\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}$

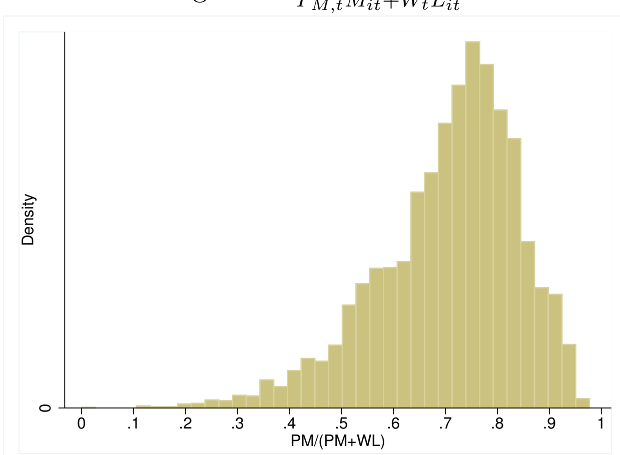


Figure 4: $\overline{\left(\frac{P_{M,t}M_{it}}{P_{M,t}M_{it}+W_tL_{it}}\right)_i}$ over 28 yrs

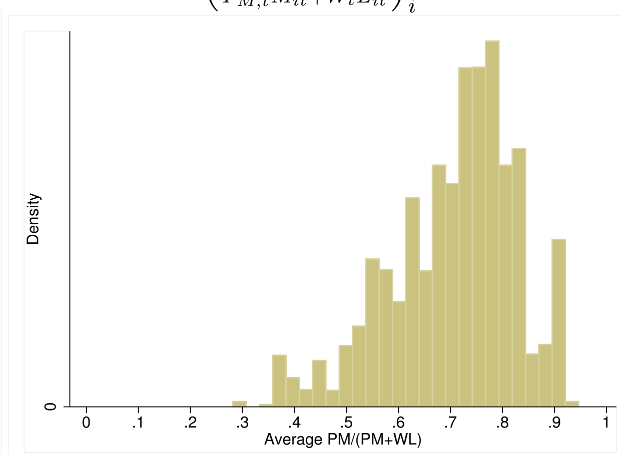
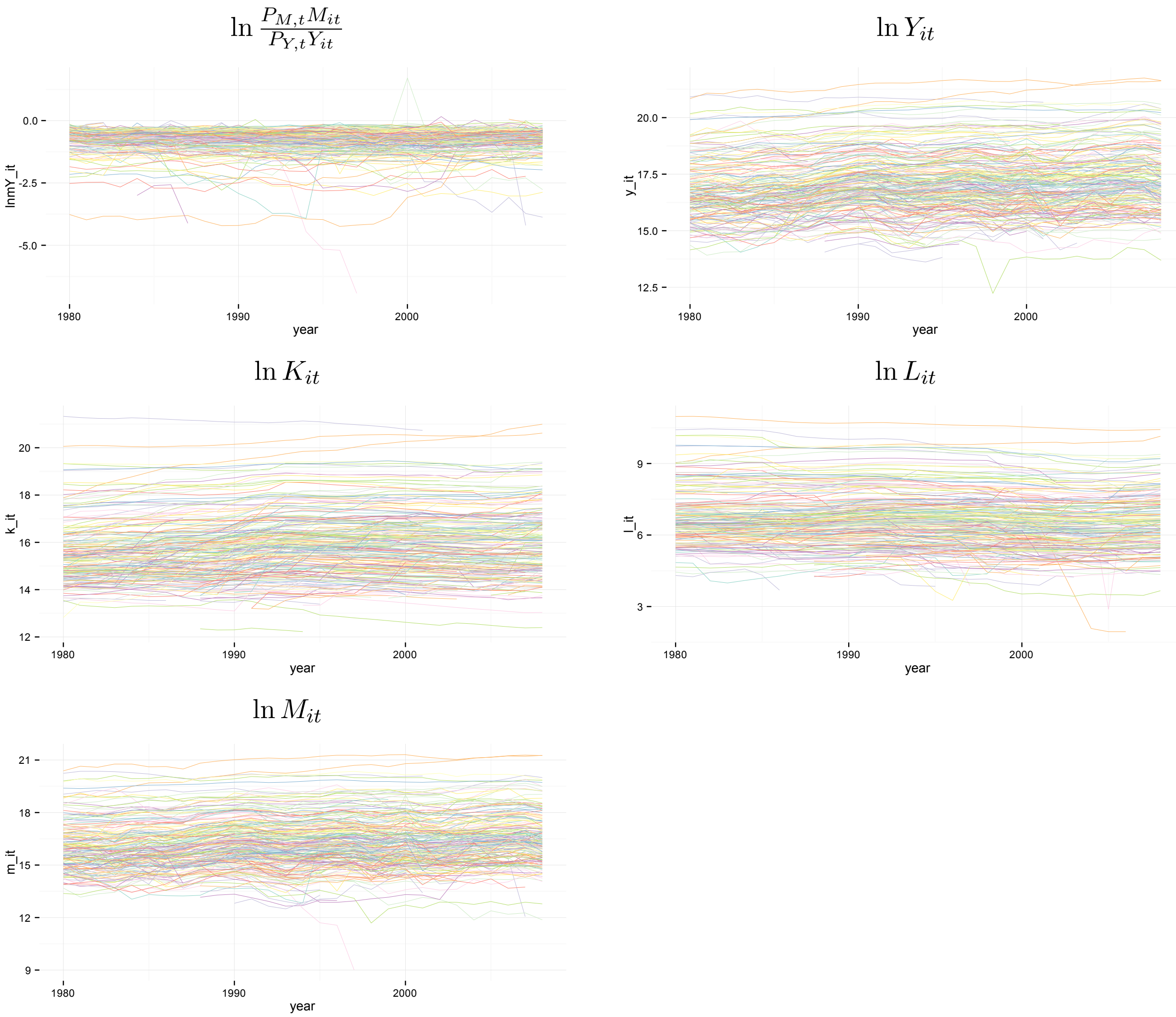
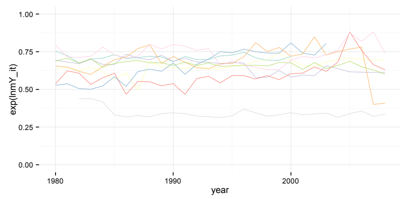


Figure 5: Trends in the log of $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$, output, and inputs in Machine industry

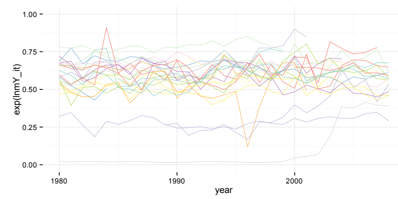


Notes: This figure shows each firms' inputs and outputs in each year. Each line represents a different firm.

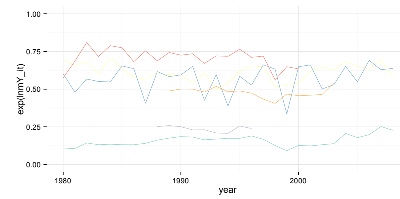
Figure 6: Trends in $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ for subindustries in Machine industry



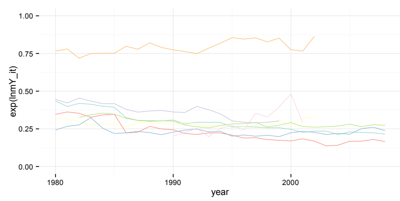
2511: Boiler prime mover



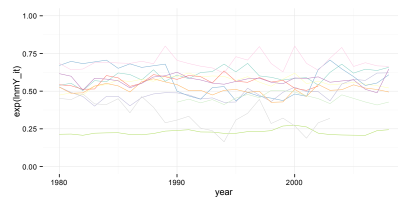
2521: Metal machine tools



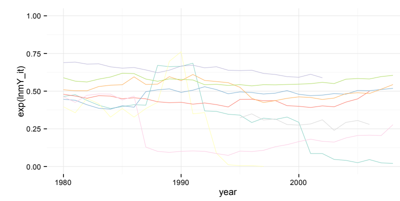
2522: Metal working machinery



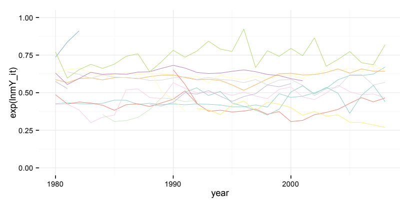
2523: Machinery tool



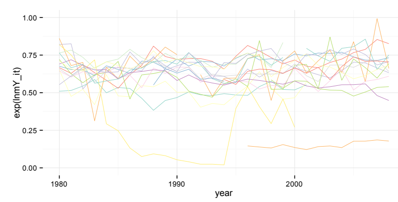
2531: Textile machinery



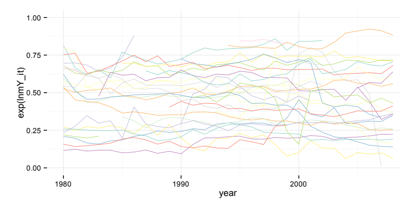
2532: Agricultural machines



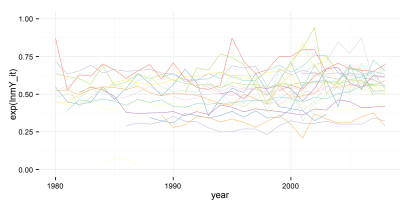
2533: Construction and mining equipment



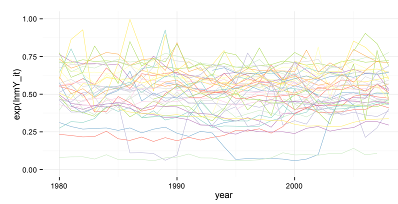
2534: Chemical machinery



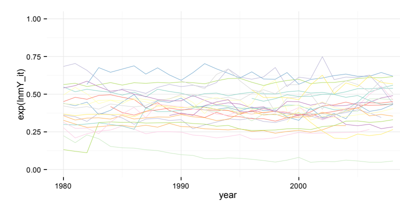
2535: Office machinery



2536: Special industrial machinery

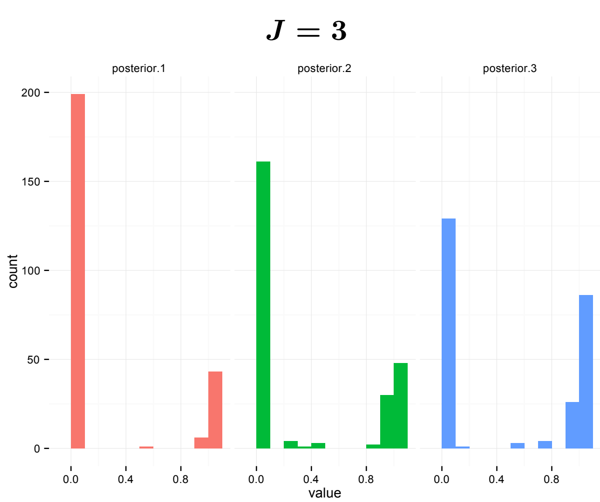


2537: General industrial machinery



2541: General Mechanical Components

Figure 7: Posterior probabilities for $J = 3$ and $J = 5$



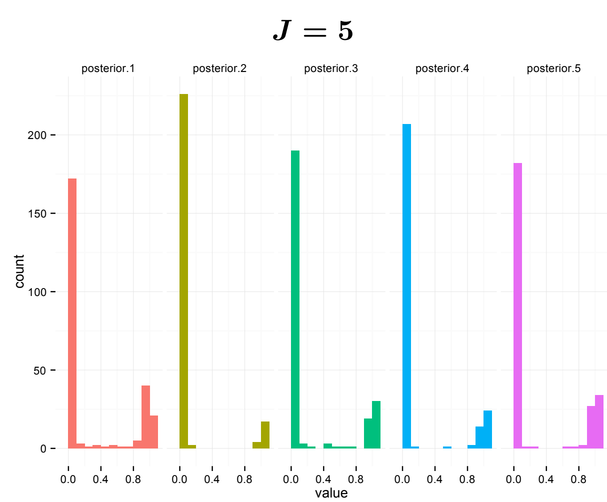


Figure 8: Trends in $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ and $\ln Y_{it}$ in Machine industry for $J = 3$

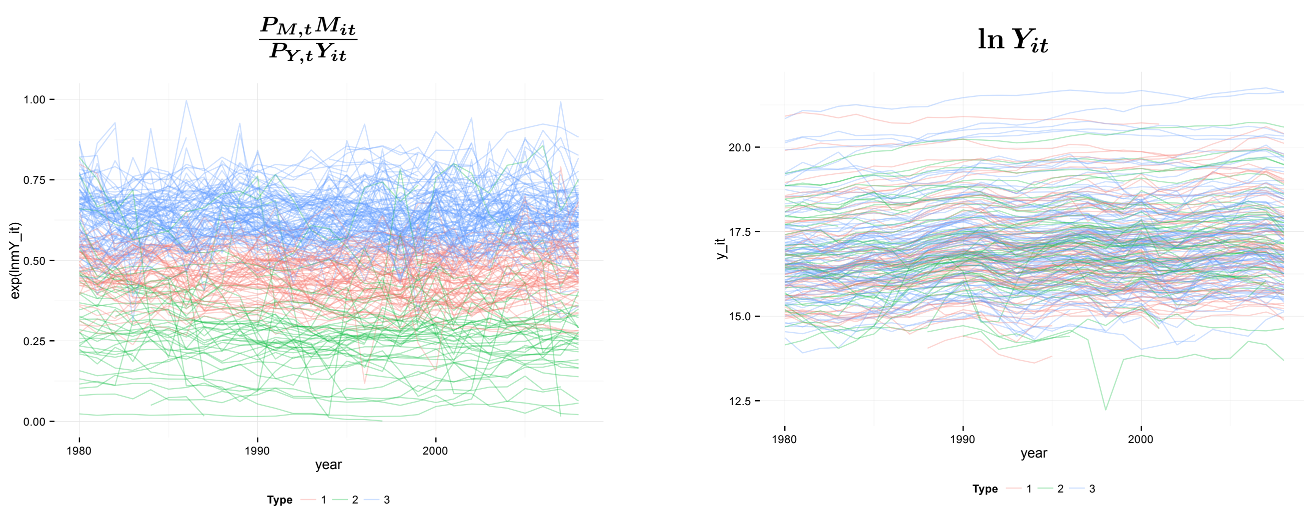


Figure 9: Trends in $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$ in subindustries when $J = 3$