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

January 14, 2015

Preliminary and Incomplete. Please Do Not Circulate.

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 permanent unobserved heterogeneity in production technology, we shows that production function for each unobserved type is non-parametrically identified under regularity conditions. We also propose an estimation procedure for production function with random coefficients based on EM algorithm. 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).

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).

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

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 methods (Arellano and Bond, 1991; Blundell and Bond, 1998; Blundell and Bond, 2000) or developing proxy variable methods (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves, and Frazer, 2006, (ACF, hereafter); Wooldridge, 2009), which are now popular in empirical applications. However, as shown in Gandhi, Navarro, and Rivers (2013, GNR hereafter), both dynamic panel methods and proxy variable methods suffer from a fundamental identification problem when flexible inputs is present. Specifically, GNR formally prove that, if intermediate inputs are chosen flexibly in each period with no adjustment cost, then production function is nonparametrically non-identifiable under a typical model of production used in the literature.[1] Then, GNR show that exploiting the first order condition with respect to flexible inputs under profit maximization makes it possible to establish the nonparametric identification of production function without making any functional form assumptions. Furthermore, 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 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 production technologies and each firm belongs to one of J types. 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 assumptions on production functions that are based on the assumptions in GNR. Based on the identification argument, we consider a random coefficient specification for production function and develop an estimation procedure in which EM algorithm is used to facilitate the computational complexity associated with estimation of a finite mixture model.

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 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).

We provide empirical evidence that production functions are heterogeneous beyond Hicks-neutral technology term to motivate the necessity of considering production functions with un-

[1] Bond and Sderbom (2005) and ACF also discuss the related identification issues.

observed 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, we find that the serial correlation of the ratio of intermediate input cost to output value is as high as 0.89 even after controlling for the difference in the input level of capital, labor, and intermediates, which strongly suggests the presence of unobserved heterogeneity beyond Hicks-neutral term.

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. 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.

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. (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. *(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) $(P_{M,t}, P_{Y,t}) \in \mathcal{I}_{it}$.*

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. 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 $\mathbb{M}_t^j$ with respect to ω_{it}, we may write $\omega_{it} = \mathbb{M}_t^{j^{-1}}(L_{it}, K_{it}, M_{it})$. 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 $G_{M,t}^j(X_{it};\xi_{it}) := \frac{F_{M,t}^j(X_{it})M_t}{F_t^j(X_{it})}$.

3 Nonparametric identification of π^j, $F_t^j(\cdot)$, $h^j(\cdot)$, $g_\eta^j(\cdot)$, and $g_\epsilon^j(\cdot)$

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

$$\begin{aligned} \mathrm{P}(\{Y_t, S_t, X_t\}_{t=1}^T) &= \sum_{j=1}^J \pi^j \mathrm{P}^j(\{Y_t, S_t, X_t\}_{t=1}^T) \\ &= \sum_{j=1}^J \pi^j \mathrm{P}_1^j(Y_1, S_1, X_1) \prod_{t=2}^T \mathrm{P}_t^j(Y_t, S_t, X_t | \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}). \end{aligned} \tag{5}$$

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

$$P(\{Y_t, S_t, X_t\}_{t=1}^T) = \sum_{j=1}^J \pi^j P_1^j(Y_1|S_1, X_1) P_1^j(S_1, X_1) \prod_{t=2}^T P_t^j(Y_t|S_t, X_t, Y_{t-1}, S_{t-1}, X_{t-1}) P_t^j(S_t, X_t|X_{t-1}). \tag{6}$$

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 use the argument in Kasahara and Shimotsu (2009) and Hu and Shum (2012) to establish the nonparametric identification of $\{\pi^j, \{\mathrm{P}_t^j(Y_t, S_t, X_t)\}_{t=1}^T\}_{j=1}^J$ as follows.

Assumption 6. *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 (30) 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 (31) take distinct values.*

Proposition 2. *Suppose that Assumptions 1-6 hold and* $T \geq 4$*. Then,* $\{\pi^j, P_1^j(Y_1, S_1, X_1), \{P_t^j(Y_t, S_t, X_t|Y_{t-1}, S_{t-1}, X_{t-1})\}_{t=2}^T\}_{j=1}^J$ *is uniquely determined from* $P(\{Y_t, S_t, X_t\}_{t=1}^T)$*.*

Remark 1. *Considering serially correlated continuous unobserved variables* $\{X_t^*\}$, *Hu and Shum (2013) analyze the nonparametric identification of the model*

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

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,* $P_3^j(W_3, X_3^*)$, $P_4^j(W_4, X_4^*|W_3, X_3^*)$, *and* $P_5^j(W_5, X_5^*|W_4, X_4^*)$ *are non-parametrically identified but the identification of* $P_1^j(W_1, X_1^*)$, $P_2^j(W_2, X_2^*|W_1, X_1^*)$, *and* $P_3^j(W_3, X_3^*|W_2, X_2^*)$ *remains unresolved. Our Proposition 2 shows that, for a model in which unobserved heterogeneity is permanent and discrete, 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 2. *Assumption 6 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 (30), 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* $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$. *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 1-6 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:*

$$\ln F_t^j(X_t) = \alpha_{\ell,t}^j \ell_t + \alpha_{k,t}^j k_t + \alpha_{m,t}^j m_t, \tag{7}$$

where the intermediate share equation is given by

$$\ln S_t = \ln(\alpha_{m,t}^j \mathcal{E}^j) - \epsilon_t. \tag{8}$$

Then, equation (6) holds with

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

where $P_t^j(S_t) = g_\epsilon^j(\ln(\alpha_{m,t}^j \mathcal{E}^j) - \ln S_t)$*. Under Assumptions 1-6, we may nonparametrically identify* $g_\epsilon^j(\cdot)$*,* $h(\cdot)$*,* $g_\eta(\cdot)$*,* $P_1^j(X_1)$*, and* $P_t^j(X_t|X_{t-1})$*, and* $\{\alpha_{\ell,t}^j, \alpha_{k,t}^j, \alpha_{m,t}^j\}$ *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 6 holds when the production function is Cobb-Douglas.

4 Estimation of production function with random coefficients

4.1 Cobb-Douglas production function with random coefficients

We consider a 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 $x_{it} := (\ell_{it}, k_{it}, m_{it})$. A scalar unobserved variable ξ_i summarizes the permanent unobserved hererogeneity for firm i in the coefficients of production function. Specifically, if firm i is of type j, then $\xi_i = \xi^j$ and we assume that $\ln F_t^j(X_{it}) = f_t(x_{it}; \xi^j)$, where

$$f_t(x_{it}; \xi^j) = \alpha_t + e^{\xi^j} m_{it} + \alpha_\ell e^{\delta_\ell \xi^j} \ell_{it} + \alpha_k e^{\delta_k \xi^j} k_{it} + \alpha_\xi \xi^j,$$

where α_t depends on t to reflect period t aggregate shock. Then, the production function of type j is given by

$$y_{it} = \alpha_t + e^{\xi^j} m_{it} + \alpha_\ell e^{\delta_\ell \xi^j} \ell_{it} + \alpha_k e^{\delta_k \xi^j} k_{it} + \omega_{it} + \alpha_\xi \xi^j + \epsilon_{it}. \tag{9}$$

The first order condition with respect to M_{it} implies

$$s_{it} = \ln \mathcal{E}^j + \xi^j - \epsilon_{it}. \tag{10}$$

Therefore, the distribution of S_{it} does not depend on X_{it} under Cobb-Douglas assumption.

The stochastic process of ω_{it} is assumed to follow

$$\omega_{it} = \rho \omega_{it-1} + \eta_{it}. \tag{11}$$

Let $x_{it} = (m_{it}, \ell_{it}, k_{it})'$. We consider the following parametric random effects model.

Assumption 7. *(a) The production function of type* j *is given by (9). (b)* $\xi_i \sim_{iid} \sum_{j=1}^J \pi^j 1(\xi = \xi^j)$*. (c)* $\eta_{it}|\xi_i, k_{it}, \ell_{it} \sim_{iid} N(0, \sigma_\eta^2)$*. (d)* $\epsilon_{it}|\xi_i = \xi^j, x_{it} \sim_{iid} N(0, (\sigma_\epsilon^j)^2)$*. (e)* $x_{i1}|\xi_i \sim_{iid} N(c_0 + c_1\xi_i, \Sigma_1)$*, where* $c_0 = (c_{\ell 0}, c_{k0}, c_{m0})'$ *and* $c_1 = (c_{\ell 1}, c_{k1}, c_{m1})'$*. (f)* $\omega_{i1}|x_{i1}, \xi_i \sim_{iid} N(\beta_0 + x_{i1}'\beta_x + \beta_\xi \xi_i, \sigma_\omega^2)$*, where* $\beta_x = (\beta_\ell, \beta_k, \beta_m)'$ *(g)* $x_{it}|x_{i,t-1}, \xi_i \sim_{iid} N(\rho_x' \tilde{x}_{it-1} + \rho_{x\xi}\xi_i, \Sigma_x)$*, where* $\tilde{x}_{it-1} = (1, x_{it-1}')'$*,* $\rho_x = (\rho_\ell, \rho_k, \rho_m)$ *is an* 4×3 *matrix, and* $\rho_{x\xi}$ *is a* 3×1 *vector.*

Collect the model parameters into θ_1, and θ_2 as follows. Let $\theta_1 := (\theta_{1,s}', \theta_{1,x}')'$, where $\theta_{1,s} := (\boldsymbol{\pi}', \boldsymbol{\xi}', \boldsymbol{\sigma_\epsilon}')'$ and $\theta_{1,x} := (c_0', c_1', (\mathrm{vech}(\Sigma_1))', \rho_{x0}', \rho_{x1}', \rho_{x\xi}', (\mathrm{vech}(\Sigma_x))')'$ with $\boldsymbol{\pi} := (\pi^1, ..., \pi^J)$, $\boldsymbol{\xi} := (\xi^1, ..., \xi^J)'$, and $\boldsymbol{\sigma_\epsilon}' := (\sigma_\epsilon^1, ..., \sigma_\epsilon^J)'$. Let $\theta_2 := (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\delta}', \rho, \sigma_\eta, \sigma_\omega)'$ with $\boldsymbol{\alpha} := (\alpha_1, ..., \alpha_T, \alpha_\ell, \alpha_k, \alpha_\xi)'$, $\boldsymbol{\beta} := (\beta_0, \beta_x', \beta_\xi)'$, and $\boldsymbol{\delta} := (\delta_\ell, \delta_k)'$.

As shown in Proposition 1, we may write the distribution of $\{Y_{it}, S_{it}, X_{it}\}_{t=1}^T$ for type j as

$$\begin{aligned}\mathrm{P}_t^j(\{Y_{it}, S_{it}, X_{it}\}_{t=1}^T) = & \underbrace{\mathrm{P}_1^j(S_{i1}|X_{i1})\mathrm{P}_1^j(X_{i1}) \prod_{t=2}^T \mathrm{P}_t^j(S_{it}|X_{it})\mathrm{P}_t^j(X_{it}|X_{it-1})}_{=L_{1i}(\xi^j;\theta_1)} \\ & \times \underbrace{\mathrm{P}_1^j(Y_{i1}|S_{i1}, X_{i1}) \prod_{t=2}^T \mathrm{P}_t^j(Y_{it}|S_{it}, X_{it}, Y_{it-1}, S_{it-1}, X_{it-1})}_{=L_{2i}(\xi^j;\theta_2,\theta_1)} .\end{aligned} \tag{12}$$

The exact expressions for $L_{1i}(\xi^j; \theta_1)$ and $L_{2i}(\xi^j; \theta_2, \theta_1)$ are derived below.

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 $\sum_{j=1}^J \pi^j \mathrm{P}_t^j(\{Y_{it}, S_{it}, X_{it}\}_{t=1}^T)$, where $\mathrm{P}_t^j(\{Y_{it}, S_{it}, X_{it}\}_{t=1}^T)$ is given in (12). Given the decomposition (12), we estimate the model by two-stage maximum likelihood estimators. In the first stage, we estimate $\boldsymbol{\pi}$ and θ_1 by maximizing $\sum_{i=1}^N \log(\sum_{j=1}^J \pi^j L_{1i}(\xi^j; \theta_1))$ 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 $\sum_{i=1}^N \log(\sum_{j=1}^J \hat{\pi}^j L_{1i}(\xi^j; \theta_2, \hat{\theta}_1))$ over θ_2.

4.2 The first stage estimation

Our estimation procedure follows the two-stage identification proof of Proposition 2(b)(c). Given that $\ln \mathcal{E}^j = 0.5(\sigma_\epsilon^j)^2$ in equation (10), we can compute ϵ_{it} as

$$\epsilon_{it}^*(\xi^j, \sigma_\epsilon^j) := -s_{it} + \xi^j + 0.5(\sigma_\epsilon^j)^2.$$

Because $P_t^j(S_{it}|X_{it}) = P_t^j(S_{it})$ under Cobb-Douglas assumption, the distribution of $\{S_{it}, X_{it}\}_{t=1}^T$ conditioning on being type j in equation (6) under Assumption 7 is written as

$$\mathrm{P}_t^j(\{S_{it}, X_{it}\}_{t=1}^T) = \left(\prod_{t=1}^T \mathrm{P}_t^j(S_{it})\right) \times \left(\mathrm{P}_1^j(X_{i1}) \prod_{t=2}^T \mathrm{P}_t^j(X_{it}|X_{it-1})\right) = L_{1i}^s(\xi^j, \sigma_\epsilon^j) L_{1i}^x(\xi^j; \theta_{1,x}), \tag{13}$$

where

$$L_{1i}^s(\xi^j, \sigma_\epsilon^j) := \prod_{t=1}^T \frac{1}{\sigma_\epsilon^j} \phi\left(\frac{\epsilon_{it}^*(\xi^j, \sigma_\epsilon^j)}{\sigma_\epsilon^j}\right) \quad \text{and} \quad L_{1i}^x(\xi^j; \theta_{1,x}) := P_1(x_{i1}; \xi^j; \theta_{1,x}) \prod_{t=2}^T P_t(x_{it}|x_{it-1}; \xi^j, \theta_{1,x}) \tag{14}$$

with

$$P_1(x_{i1};\xi,\theta_{1,x}) = (2\pi)^{-3/2}|\Sigma_1|^{-1/2}\exp\left(-\frac{1}{2}(x_{i1}-c_0-c_1\xi)'\Sigma_1^{-1}(x_{i1}-c_0-c_1\xi)\right) \quad \text{and}$$

$$P_t(x_{it}|x_{it-1};\xi,\theta_{1,x}) = (2\pi)^{-3/2}|\Sigma_x|^{-1/2}\exp\left(-\frac{1}{2}(x_{it}-\rho_{x0}-\rho_{x1}x_{it-1}-\rho_{x\xi}\xi)'\Sigma_x^{-1}(x_{it}-\rho_{x0}-\rho_{x1}x_{it-1}-\rho_{x\xi}\xi)\right). \tag{15}$$

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

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

where

$$L_{1i}(\xi^j;\theta_1) := L_{1i}^s(\xi^j,\sigma_\epsilon^j)L_{1i}^x(\xi^j;\theta_{1,x}). \tag{16}$$

Alternatively, we may first estimate $\theta_{1,s}$ by

$$\hat{\theta}_{1,s} = \underset{\theta_{1,s}}{\operatorname{argmax}} \sum_{i=1}^{N} \ln\left(\sum_{j=1}^{J} \pi^j L_{1i}^s(\xi^j,\sigma_\epsilon^j)\right),$$

and, then, estimate $\theta_{1,x}$ given $\hat{\theta}_{1,s}$ by

$$\hat{\theta}_{1,x} = \underset{\theta_{1,x}}{\operatorname{argmax}} \sum_{i=1}^{N} \ln\left(\sum_{j=1}^{J} \hat{\pi}^j L_{1i}^s(\hat{\xi}^j,\hat{\sigma}_\epsilon^j)L_{1i}^x(\hat{\xi}^j;\theta_{1,x})\right).$$

Given $\hat{\theta}_1$, define

$$\hat{\epsilon}_{it}^j := \epsilon_{it}^*(\hat{\xi}^j,\hat{\sigma}_\epsilon^j).$$

4.3 The second stage estimation

Given the estimates of $\hat{\theta}_1$ and $\{\hat{\epsilon}_{it}^j\}$, we estimate $\theta_2 = (\boldsymbol{\alpha}',\boldsymbol{\beta}',\boldsymbol{\delta}',\rho,\sigma_\eta,\sigma_\omega)'$ with $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_T,\alpha_\ell,\alpha_k,\alpha_\xi)'$, $\boldsymbol{\beta} = (\beta_0,\beta_x',\beta_\xi)'$, $\boldsymbol{\delta} = (\delta_\ell,\delta_k)'$ in the second stage as follows.

Suppose that we know the value of θ_1 and $\{\epsilon_{it}^j\}$. Then, given θ_2, we may express $\mathrm{P}_1^j(Y_{i1}|S_{i1},X_{i1})$ and $\mathrm{P}_t^j(Y_{it}|S_{it},X_{it},Y_{it-1},S_{it-1},X_{it-1})$ as

$$\mathrm{P}_1^j(Y_{i1}|S_{i1},X_{i1}) = \frac{1}{\sigma_\omega}\phi\left(\frac{\omega_{i1}^*(\theta_2,\xi^j,\theta_1,\epsilon_{i1}^j)-(\beta_0+x_{i1}'\beta_x+\beta_\xi\xi^j)}{\sigma_\omega}\right) \quad \text{and}$$

$$\mathrm{P}_t^j(Y_{it}|S_{it},X_{it},Y_{it-1},S_{it-1},X_{it-1}) = \frac{1}{\sigma_\eta}\phi\left(\frac{\eta_{it}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it}^j,\epsilon_{it-1}^j)}{\sigma_\eta}\right),$$

where

$$\omega_{it}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it}^j) := \tilde{y}_{it}(\xi^j,\theta_1,\epsilon_{it}^j)-\alpha_t-\alpha_\ell e^{\delta_\ell\xi^j}\ell_{it}-\alpha_k e^{\delta_k\xi^j}k_{it}-\alpha_\xi\xi^j,$$

with

$$\tilde{y}_{it}(\xi^j,\theta_1,\epsilon_{it}^j) := y_{it} - e^{\xi^j} m_{it} - \epsilon_{it}^j$$

and

$$\eta_{it}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it}^j,\epsilon_{it-1}^j) := \omega_{it}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it}^j) - \rho\omega_{it-1}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it-1}^j).$$

Therefore,

$$L_{2i}(\xi^j;\theta_2,\theta_1) := \frac{1}{\sigma_\omega}\phi\left(\frac{\omega_{i1}^*(\theta_2,\xi^j,\theta_1,\epsilon_{i1}^j) - (\beta_0 + x_{i1}'\beta_x + \beta_\xi\xi^j)}{\sigma_\omega}\right)\prod_{t=2}^{T}\frac{1}{\sigma_\eta}\phi\left(\frac{\eta_{it}^*(\theta_2,\xi^j,\theta_1,\epsilon_{it}^j,\epsilon_{it-1}^j)}{\sigma_\eta}\right).$$

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

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

4.4 More flexible production function with random coefficients

We also consider the following specification of production function which is more flexible than Cobb-Douglas specification, in which the cost share of intermediate goods depend on the values of ℓ_{it} and k_{it}. Specifically, if firm i is of type j, then $\xi_i = \xi^j$ and we assume that

$$f_t(x_{it};\xi^j) = e^{\gamma_0+\gamma_\ell\ell_{it}+\gamma_k k_{it}+\xi^j} m_{it} + \alpha_\ell e^{\delta_\ell\xi^j}\ell_{it} + \alpha_k e^{\delta_k\xi^j} k_{it} + \alpha_\xi\xi^j$$

so that the production function of type j is given by

$$y_{it} = e^{\gamma_0+\gamma_\ell\ell_{it}+\gamma_k k_{it}+\xi^j} m_{it} + \alpha_\ell e^{\delta_\ell\xi^j}\ell_{it} + \alpha_k e^{\delta_k\xi^j} k_{it} + \alpha_\xi\xi^j + \omega_{it} + \epsilon_{it}. \tag{17}$$

The first order condition with respect to M_{it} implies

$$s_{it} = \underbrace{(\gamma_0 + \ln E[e^{\epsilon_{it}}])}_{:=\tilde{\gamma}_0} + \gamma_\ell\ell_{it} + \gamma_k k_{it} + \xi_i - \epsilon_{it} = z_{1,it}'\gamma + \xi_i - \epsilon_{it}, \tag{18}$$

where $z_{1,it} := (1,\ell_{it},k_{it})'$ and $\gamma = (\tilde{\gamma}_0,\gamma_\ell,\gamma_k)'$ with $\tilde{\gamma}_0 = \gamma_0 + E[e^{\epsilon_{it}}]$. The stochastic process of ω_{it} is assumed to follow (11).

We consider the following parametric random effects model.

Assumption 8. *Assumption 7(c)-(g) holds and assume that The production function of type j is given by (17) and that $\xi_i \sim_{iid} \sum_{j=1}^{J}\pi^j 1(\xi = \xi^j)$ that approximates a normal distribution given by $N(0,\sigma_\xi^2)$. Specifically, we approximate normal distribution using a (known) set of Gauss-Hermit quadrature points and weights, and the values of $\{\pi^j,\xi^j\}_{j=1}^J$ are uniquely determined from σ_ξ^2. See the appendix for the detail.*

We estimate the production function (17) using the two-stage method as discussed in the previous sections.

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.[2] 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. In this preliminary version, we focus on a sample from Machine industry.

5.2 Evidence for unobserved heterogeneity: the persistence in the ratio of intermediate inputs to total sales

In the data, the ratio of intermediate inputs to total sales is heterogenous across firms and persistent over time within firm. Figure 1 presents the histogram of $\frac{P_{M,t}M_t}{P_{Y,t}Y_t}$ across all observations that belongs to Machine industry, which shows a large variation in the share of intermediate inputs. In the model, the variation in the share of intermediate inputs is coming from idiosyncratic ex-post shocks ϵ_{it}. We may eliminate most of idiosyncratic components by considering the firm-level average of the share of intermediate inputs over 28 years; as shown in Figure 2, the persistent component of the ratio of intermediate inputs to total sales substantially varies across firms.

In view of the intermediate share equation in (4), a large cross-section 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 $\hat{\epsilon}_{it} := e_{it} - \xi_i$. Comparing the variance of $\hat{\xi}_i$ with that of $\hat{\epsilon}_{it}$, we found that $\frac{\text{Var}(\hat{\xi}_i)}{\text{Var}(\hat{\xi}_i)+\text{Var}(\hat{\epsilon}_{it})} = 0.609$. Therefore, the majority of variation is coming from the permanent component, suggesting that production function is heterogeneous beyond Hicks-neutral term.

[2]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 1: Histogram of $\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$

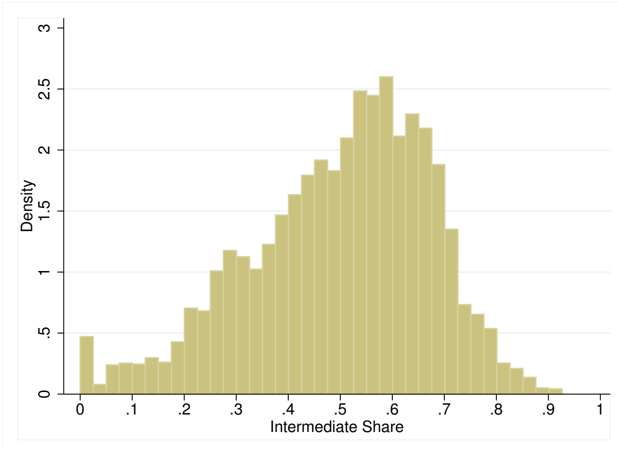
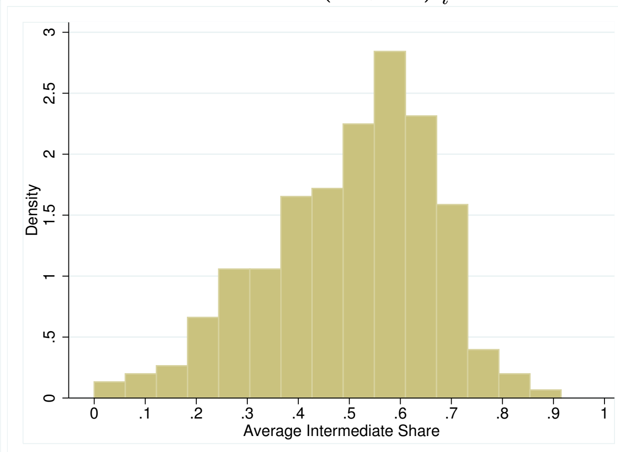


Figure 2: Histogram of $\overline{\left(\frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}\right)_i}$ over 28 years



5.3 Estimation of production function

Table 1 presents the estimated average elasticities of output with respect to labor, capital, and intermediate inputs as well as the parameter estimates for the number of components equal to $J = 1$, 3, and 5 under Assumption 8. Setting $J = 1$ gives the homogenous production function specification considered by GNR.

The estimated average elasticities of output with respect to intermediate goods in random coefficients model with $J = 3$ or 5 is larger than those in homogenous model with $J = 1$ while the estimated average elasticities with respect to labor and capital in random coefficients model than those in homogenous model. As the number of components increases, the estimated standard deviation of ex-post shock, ϵ, decreases, indicating a possibility of substantial downward bias in the estimated variation in ex-post shocks in homogenous model with $J = 1$ as a result of ignoring unobserved heterogeneity. In fact, consistent with our analysis in the previous section, the estimated standard deviations of permanent unobserved heterogeneity σ_ξ is larger than those of ex-post shocks σ_ϵ for the model with $J = 3$ and 5.

The first two rows of Table 2 present the standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_\xi\hat{\xi}_i$ and $\hat{\epsilon}_{it}$. To compute the standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_\xi\hat{\xi}_i$ and $\hat{\epsilon}_{it}$, we first compute $\hat{\omega}_{it}$, $\hat{\xi}_i$, and $\hat{\epsilon}_{it}$ for each i and t using the posterior probability of being j-th type for every firm as described in the appendix and then we compute the sample standard deviations of $\hat{\omega}_{it} + \hat{\alpha}_\xi\hat{\xi}_i$ and $\hat{\epsilon}_{it}$.[3]

As the number of components J increases, the standard deviation of $\hat{\omega}_{it} + \hat{\alpha}_\xi\hat{\xi}_i$ while that of $\hat{\epsilon}_{it}$ decreases. This suggests that the role of ex-post shocks to explain the residual in production function can be exaggerated if homogenous model rather than random coefficients model is

[3] We found that, for each firm, the posterior probability is concentrated in one of the types; the average highest posterior probability across firms for the model with $J = 5$ is given by 0.989 so that a typical firm puts more about 99 percent of posterior probability in one of the five types.

Table 1: Estimates of Production Function (17): Machine Industry in Japan, 1980-2007

	Original GNR	Random Coefficients Model	
	$J = 1$	$J = 3$	$J = 5$
Ave. Elast. for Labor	0.269	0.207	0.240
Ave. Elast. for Capital	0.302	0.150	0.124
Ave. Elast. for Intermed.	0.408	0.482	0.482
Sum	0.979	0.839	0.846
$\hat{\sigma}_\epsilon$	0.507	0.315	0.291
$\hat{\sigma}_\xi$	—	0.391	0.422
$(\hat{\alpha}_\xi, \hat{\delta}_\ell, \hat{\delta}_k)$	—	(-1.068,7.267,-0.038)	(-1.291,6.422,-0.068)
$(\hat{\rho}_1, \hat{\sigma}_\eta)$	(0.936,0.149)	(0.967,0.133)	(0.964,0.135)
$(\gamma_0, \gamma_\ell, \gamma_k)$	(-0.464,0.044,-0.046)	(-0.569,0.076,-0.075)	(-0.396,0.072,-0.075)
$(\hat{\alpha}_\ell, \hat{\alpha}_k)$	(-0.022,0.645)	(-0.004,0.740)	(-0.013, 0.721)
No. of Observations		5518	
No. of firms		247	

Table 2: GNR vs. Random Coefficients Model (17): Machine Industry in Japan, 1980-2007

	Original GNR	Random Coeff.	
	$J = 1$	$J = 3$	$J = 5$
Std. Dev. $\hat{\omega}_{it} + \hat{\alpha}_\xi \hat{\xi}_i$	0.337	0.788	0.826
Std. Dev. $\hat{\epsilon}_{it}$	0.506	0.331	0.306
$\text{Corr}(\hat{\epsilon}_{it}, \hat{\epsilon}_{it-1})$	0.887	0.867	0.710

Notes: p-values in square brackets and standard errors in round brackets.

estimated.

The serial correlation in $\hat{\epsilon}_{it}$ also decreases from 0.887 to 0.710 as the number of components increases from $J = 1$ to $J = 5$. 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. However, the level of serial correlation in $\hat{\epsilon}_{it}$ is still high at 0.71 for the model with $J = 5$. One possible reason for this high serial correlation is the presence of transitory serially correlated shock. Namely, the share parameter for intermediate goods could change over time within the same firm.

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

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

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

[10] 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.

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

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

[13] 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.

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

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

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

[17] 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,

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

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

[20] 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

$$\begin{aligned}
&\mathrm{P}_t^j(\{Y_t, S_t, X_t\}_{t=1}^T) = \mathrm{P}_1^j(Y_1, S_1, X_1) \prod_{t=2}^T \mathrm{P}_t^j(Y_t, S_t, X_t | \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) \\
&= \mathrm{P}_1^j(Y_1|S_1, X_1)\mathrm{P}_1^j(S_1, X_1) \\
&\quad \times \prod_{t=2}^T \mathrm{P}_t^j(Y_t|S_t, X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(S_t|X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(X_t|\{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}).
\end{aligned} \tag{19}$$

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

$$\mathrm{P}_t^j(S_t|X_t, \{Y_{t-s}, S_{t-s}, X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(S_t|X_t) := g_\epsilon(\ln G_{M,t}^j(X_t) + \ln \mathcal{E} - \ln S_t). \tag{20}$$

Furthermore,

$$
\begin{aligned}
&\mathrm{P}_t^j(X_t|\{Y_{t-s},S_{t-s},X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(L_t,K_t,\omega_t|\{Y_{t-s},S_{t-s},L_{t-s},K_{t-s},\omega_{t-s}\}_{s=1}^{t-1})\\
&= \mathrm{P}_t^j(\omega_t|L_t,K_t,\{Y_{t-s},S_{t-s},L_{t-s},K_{t-s},\omega_{t-s}\}_{s=1}^{t-1})\mathrm{P}_t^j(L_t,K_t|\{Y_{t-s},S_{t-s},L_{t-s},K_{t-s},\omega_{t-s}\}_{s=1}^{t-1})\\
&= \mathrm{P}_\omega^j(\omega_t|\omega_{t-1})\mathrm{P}_t^j(L_t,K_t|L_{t-1},K_{t-1},\omega_{t-1})\\
&= \mathrm{P}_t^j(L_t,K_t,\omega_t|L_{t-1},K_{t-1},\omega_{t-1})\\
&= \mathrm{P}_t^j(X_t|X_{t-1}),
\end{aligned}
\tag{21}
$$

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 (19)-(21) 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

$$
\begin{aligned}
\epsilon_t^j(S_t,X_t;\theta_1) &:= \ln(G_{M,t}^j(X_t)\mathcal{E}^j) - \ln S_t,\\
\mathcal{Y}_t^j(Y_t,S_t,X_t;\theta_1) &:= \ln Y_t - \int \frac{G_{M,t}^j(L_t,K_t,M_t)}{M_t}dM_t - \epsilon_t^j(S_t,X_t;\theta_1).
\end{aligned}
\tag{22}
$$

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

$$
\int \frac{G_{M,t}^j(L_t,K_t,M_t)}{M_t}dM_t = \ln F_t^j(L_t,K_t,M_t) + \mathcal{C}_t^j(L_t,K_t), \tag{23}
$$

where $\mathcal{C}_t^j(L_t,K_t)$ is a function that only depends on (L_t,K_t). Then, substituting (23) into the second equation of (22) and combining it with (4) gives

$$
\omega_{it} = \mathcal{Y}_t^j(Y_t,S_t,X_t;\theta_1) + \mathcal{C}_t^j(K_t,L_t). \tag{24}
$$

Then, it follows from (2), (4), (22), and (24) that

$$
\mathcal{Y}_t^j(Y_t,S_t,X_t;\theta_1) = -\mathcal{C}_t^j(K_t,L_t) + \underbrace{h^j\left(\mathcal{Y}_{t-1}^j(Y_{t-1},S_{t-1},X_{t-1};\theta_1) + \mathcal{C}_{t-1}^j(K_{t-1},L_{t-1})\right)}_{=\tilde{h}^j(\mathcal{Y}_{t-1}^j(Y_{t-1},S_{t-1},X_{t-1};\theta_1),K_{t-1},L_{t-1})} + \eta_t. \tag{25}
$$

Because η_t is i.i.d., equation (25) implies that

$$
\mathrm{P}_t^j(Y_t|S_t,X_t,\{Y_{t-s},S_{t-s},X_{t-s}\}_{s=1}^{t-1}) = \mathrm{P}_t^j(Y_t|S_t,X_t,Y_{t-1},S_{t-1},X_{t-1}), \tag{26}
$$

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 6. Evaluating (6) at $(W_2, W_3) = (w_2, w_3)$ gives

$$\begin{aligned} P(\{W_t\}_{t=1}^4) &= \sum_{j=1}^J \pi^j P_4^j(W_4|w_3) P_3^j(w_3|w_2) P_2^j(w_2|W_1) P_1^j(W_1) \\ &= \sum_{j=1}^J \lambda_4^j(W_4|w_3) \lambda_3^j(w_3|w_2) \bar{\lambda}_2^j(W_1, w_2), \end{aligned} \tag{27}$$

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 (27) gives

$$P(\{W_t\}_{t=1}^3) = \sum_{j=1}^J \lambda_3^j(w_3|w_2) \bar{\lambda}_2^j(W_1, w_2). \tag{28}$$

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 (27) at $W_1 = a_1, ..., a_J$ and $W_4 = b_1, ..., b_{J-1}$ gives $M(M-1)$ equations while evaluating (28) 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

$$P_{w_2,w_3} = L'_{w_3} D_{w_3|w_2} \bar{L}_{w_2}, \tag{29}$$

where

$$\begin{aligned} P_{w_2,w_3} &:= \begin{bmatrix} \bar{f}_{w_2,w_3}(a_1) & \bar{f}_{w_2,w_3}(a_2) & \cdots & \bar{f}_{w_2,w_3}(a_J) \\ f_{w_2,w_3}(a_1, b_1) & f_{w_2,w_3}(a_2, b_1) & \cdots & f_{w_2,w_3}(a_J, b_1) \\ \vdots & \vdots & \ldots & \vdots \\ f_{w_2,w_3}(a_1, b_{J-1}) & f_{w_2,w_3}(a_2, b_1) & \cdots & f_{w_2,w_3}(a_J, b_{J-1}) \end{bmatrix}, \\ L_{w_3} &:= \begin{bmatrix} 1 & \lambda_4^1(b_1|w_3) & \cdots & \lambda_4^1(b_{J-1}|w_3) \\ \vdots & \vdots & \ldots & \vdots \\ 1 & \lambda_4^J(b_1|w_3) & \cdots & \lambda_4^J(b_{J-1}|w_3) \end{bmatrix}, \quad \bar{L}_{w_2} := \begin{bmatrix} \bar{\lambda}_2^1(a_1, w_2) & \cdots & \bar{\lambda}_2^1(a_J, w_2) \\ \vdots & \vdots & \ldots \\ \bar{\lambda}_2^J(a_1, w_2) & \cdots & \bar{\lambda}_2^J(a_J, w_2) \end{bmatrix}, \end{aligned} \tag{30}$$

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 (29) 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 6,

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

where

$$D_{w_2,\bar{w}_2,w_3,\bar{w}_3} := D_{w_3|w_2}(D_{\bar{w}_3|w_2})^{-1}D_{\bar{w}_3|\bar{w}_2}(D_{w_3|\bar{w}_2})^{-1}. \tag{31}$$

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 6(a).

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

$$P(W_4, W_3 = w_3|W_2 = w_2) = \sum_{j=1}^J \tilde{\pi}^j_{w_2}P_4^j(W_4|w_3)P_3^j(w_3|w_2) = \sum_{j=1}^J \lambda_4^j(W_4|w_3)\tilde{\lambda}_3^j(w_3|w_2), \tag{32}$$

where $\tilde{\pi}^j_{w_2} := \frac{\pi^j P_2^j(W_2=w_2)}{P_2(W_2=w_2)}$ and $\tilde{\lambda}_3^j(w_3|w_2) := \tilde{\pi}^j_{w_2}P_3^j(W_3 = w_3|W_2 = w_2)$. Then, evaluating (32) 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 $\tilde{\pi}^j_{w_2}P_3^j(W_3 = w_3|W_2 = w_2)$ 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 $\tilde{\pi}^j_{w_2}P_3^j(W_3 = w_3|W_2 = w_2)$ for every $(w_2, w_3) \in \mathcal{W}_2 \times \mathcal{W}_3$. Then, $\tilde{\pi}^j_{w_2}$ and $P_3^j(W_3 = w_3|W_2 = w_2)$ are uniquely identified as $\tilde{\pi}^j_{w_2} = \int_{\mathcal{W}_3} \tilde{\pi}^j_{w_2}P_3^j(W_3|W_2 = w_2)dW_3$ and $P^j(W_3 = w_3|W_2 = w_2) = [\tilde{\pi}^j_{w_2}P_3^j(W_3 = w_3|W_2 = w_2)]/\tilde{\pi}^j_{w_2}$. 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 (29), $\bar{L}_{w_2}$ 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, $\{P_4^j(W_4|W_3 = w_3), \pi^jP_2^j(W_2 = w_2|W_1)P_1^j(W_1)\}_{j=1}^J$ 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. $\square$

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 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 (25) 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 (25). Given $\mathcal{C}_t^j(K_t, L_t)$, the identification of $F_t^j(\cdot)$ and $\{\omega_t\}_{t=1}^T$ follows from (23) and (24), respectively, and $h^j(\cdot)$ is identified from $\{\omega_t\}_{t=1}^T$. Finally, because η_t is identified from (25), $g_\eta^j(\cdot)$ is also identified. This proves part (b). $\square$

A.4 Assumption 6 under Cobb-Douglas production function

In the following, we discuss the conditions under which Assumption 6 holds when the production function is Cobb-Douglas.

Example 1 (continued)**.** *For random coefficients model (7), 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

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

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 6(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,*

$$\lambda^j(\bar{w}_3 | \bar{w}_2) = \pi^j g_\epsilon^j(\ln \bar{s}_3 - \ln(\alpha_{m,3}^j \mathcal{E}^j)) P_3^j(X_3 = \bar{x}_3 | X_2 = \bar{x}_2). \tag{33}$$

Pick $w_3 = (s_3, x_3)$ *and* $w_2 = (s_2, x_2)$. *Assumption 6(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 6(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.

B Guass-Hermit quadrature

Suppose that ξ is distributed as $N(0, \sigma_\xi^2)$ and we are interested in evaluating

$$\int L_{1i}(\xi; \theta_1) g_\xi(\xi) d\xi = \int L_{1i}(\xi; \theta_1) \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(- \left(\frac{\xi}{\sqrt{2}\sigma_\xi} \right)^2 \right) d\xi.$$

Let $v = \frac{\xi}{\sqrt{2}\sigma_\xi}$. Then,

$$\int L_{1i}(\xi; \theta_1) \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(- \left(\frac{\xi}{\sqrt{2}\sigma_\xi} \right)^2 \right) d\xi = \int L_{1i}(\sqrt{2}\sigma_\xi v, \theta_1) \frac{1}{\sqrt{\pi}} \exp\left(-v^2\right) dv.$$

The Gauss-Hermite quadrature approximates the integrals of the form $\int_{-\infty}^{\infty} f(v) \exp(-v^2) dv$ as $\sum_{j=1}^{K} w^j f(v^j)$, where $\{w^j, v^j\}_{j=1}^J$ is a set of quadrature points and weights such that the integration is exact when $f(v)$ is a polynomial of maximum order $2J-1$. Therefore, we may approximate

$$\int L_{1i}(\xi;\theta_1)\frac{1}{\sqrt{2\pi}\sigma_\xi}\exp\left(-\left(\frac{\xi}{\sqrt{2}\sigma_\xi}\right)^2\right)d\xi \approx \sum_{j=1}^{J}\frac{w^j}{\sqrt{\pi}}L_{1i}(\xi^j;\theta_1),$$

where $\xi^j = \sqrt{2}\sigma_\xi v^j$.