Testing the Number of Components in Finite Mixture Normal Regression Model with Panel Data

Yu Hao*
Vancouver School of Economics
University of British Columbia
yu.hao@alumni.ubc.ca

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

October 22, 2020

Abstract

This paper develops the likelihood-ratio based test of the null hypothesis of a m_0-component model against an alternative of $(m_0 + 1)$-component model in the normal mixture panel regression. I show that the normal mixture panel regression does not suffer from the Fisher Information matrix degeneracy under the reparameterization proposed in Kasahara and Shimotsu (2012). As a result, the likelihood ratio test statistic can be approximated by a local quadratic expansion of squares and products of the reparameterized parameters. Moreover, I obtain the data-driven penalty function via computational experiments to attend to unbounded likelihood ratio. In addition, I apply the test to random coefficient Cobb-Douglas production function estimation following the framework of Gandhi et al. (2013) and Kasahara and Shimotsu (2015). The empirical findings suggest evidence of heterogeneous production technology beyond Hicks-neutral technology factor.

Keywords: likelihood ratio test; panel regression; production function

*Address for correspondence: Yu(Jasmine) Hao, Vancouver School of Economics, University of British Columbia, 6000 Iona Dr., Vancouver, BC V6T 1L4, Canada. We are very grateful for comments from Vadim Marmer, Kevin Song and all my colleagues. We also thank to the IAAE grant at the 2019 IAAE Conference. This research is support by the Natural Science and Engineering Research Council of Canada.

1 Introduction

Finite mixture models provide a natural representation of heterogeneity in a finite number of classes. It implements a flexible way to represent a non-standard distribution by a mixture of other distributions. Finite mixture models have been applied in diverse fields of economics, for example, modeling unobserved heterogeneous ability in labor economic topics. A finite mixture model was proposed by Pearson (1894) firstly, describing a two-component normal mixture. Since then, the model has been applied in different areas to play a fundamental role in cluster analysis.

For example, finite mixtures are often used to model unobserved individual-specific effect in labor economics. Heckman and Singer (1984) use the finite mixture model to provide an alternative method to account for the unobserved heterogeneity in the analysis of single-spell duration times of unemployed workers. Keane and Wolpin (1997) and Cameron and Heckman (1998) analyze a dynamic model of schooling and occupational choices with unobserved heterogeneous human capital. Likewise, finite mixture models have been applied in health economics. Deb and Trivedi (1997) develop a finite mixture negative binomial count model that accounts for unobserved dispersion of elderly medical care utilization. In industrial organization, modeling consumer segmentation in marketing such as Kamakura and Russell (1989) and Andrews and Currim (2003) is a venue of application. There have been several papers discussing comprehensive theoretical accounts with stylized examples, including Brännäs and Rosenqvist (1994), Lindsay and Lesperance (1995), Titterington et al. (1985), and McLachlan and Peel (2004).

The number of components is an important parameter in the finite mixture models. In economics applications, the number of components often represents the numbers of unobservable types or abilities. With arbitrarily chosen number of parameters, the level of heterogeneity can be over-estimated or under-estimated. With too few components, overlooking the heterogeneity can result in biased estimation. On the other hand, with too many components, estimation is costly and ill-behaved due to potential identification problems. Therefore, it is important to develop a statistical procedure to determine the number of components.

I account for the likelihood ratio test of the null hypothesis of M_0 components against the alternative of $(M_0 + 1)$ components in the finite normal mixture panel regression model. The likelihood ratio test is based on the theoretical framework proposed by Kasahara and Shimotsu (2012). One difficulty of the testing procedure on the number of components in the finite mixture models is the singular Fisher Information matrix caused by collinearity of the score functions. Kasahara and Shimotsu (2012) propose a reparameterization orthogonal to the direction in which the Fisher information matrix is singular. The likelihood ratio of $(M_0 + 1)$-component against the M_0-component model is approximated with local quadratic expansion with squares and cross-products of the reparameterized parameters. This leads to a comparatively simple characterization of the asymptotic distribution of the likelihood ratio test (LRT) statistic. See section 3 for detailed discussions.

Another difficulty in the tests lies in that the alternative model parameters can be described by various elements in the null parameter space. I use the modified EM test as proposed in Kasahara and Shimotsu (2012) to partition the null hypothesis into several sub-hypotheses, each of which corresponds to one of the elements in the alternative parameter space that gives rise to the null model. The modified EM test statistic has the same asymptotic distribution as the likelihood ratio test statistic. See section 4 for detailed discussions on the partitions.

As a motivating example, I apply the testing procedure to the random coefficient Cobb-Douglas production function estimation to determine the number of types of intermediate good elasticity coefficients. In the past literature, the unobserved heterogeneity in production functions is not sufficiently discussed. The finite mixture models can be used to describe the unobserved heterogeneity in production functions.

This paper makes several contributions. Under the reparameterization by Kasahara and Shimotsu (2012), the log-likelihood function can be locally approximated by a quadratic expansion of squares and cross-products of the reparameterized parameters only if the Fisher Information matrix is non-singular. As shown in Kasahara and Shimotsu (2015), the finite normal mixture regression model with cross-sectional data has Fisher information matrix that does not satisfy this assumption. This is because the score functions are collinear under reparameterization as shown in section 3.3. In this paper, I show that with panel data, finite normal mixture panel regression model has a non-singular Fisher information matrix under this reparameterization.

The second contribution is that I use computational experiments to determine the data-driven penalty function in the modified EM test proposed by Kasahara and Shimotsu (2015). When testing the null M_0-component model against an alternative (M_0+1)-component model with the normal density, the empirical likelihood ratio tends to suffer from unboundedness. Suppose econometricians observe data from N firms in T periods. Each of the firms belong to one type, and the type is time invariant. If there exists a type j such that only one firm belongs to the type. The estimated mixing probability of type j is $\frac{1}{N}$, and the sample variance of type j is very small. By the property of normal density function, the likelihood is unbounded. Increasing the panel length T will increase the sample variance. As a result, the unbounded likelihood is a less severe problem. However, the unbounded likelihood ratio still causes over-rejection when the panel length T is small. The computational experiments to obtain the penalty function are similar to those in Chen et al. (2008), Chen and Li (2009) and Kasahara et al. (2015). The penalty function is a function of number of firms N, the panel length T and the misclassification probability as defined in Melnykov and Maitra (2010). The simulations of the modified EM algorithm with the penalty function exhibit correct finite sample Type I errors and good power properties. The computation is empowered by R Core Team (2013). I develop an R package Hao (2017). The package consists of the data generating module, the estimation module using EM algorithm and the likelihood ratio asymptotic distribution simulation module. The data generating module takes in parameters of mixture panel regression model and generates dependent and independent variables. The estimation module uses the penalized EM algorithm to obtain the empirical likelihood ratio as introduced in section 5. As shown in section 4, the asymptotic distribution is a non-standard distribution, and therefore needs to be simulated. The package contains an asymptotic distribution simulation module.

Lastly, I apply the likelihood ratio test to Japanese and Chilean plant-level producer data. The model follows the random coefficient production function as proposed by Gandhi et al. (2013). The model uses the firms first order conditions for profit maximization as constraints to estimate the input elasticity of the production functions non-parametrically. In particular, assuming Cobb-Douglas production functions, the revenue shares of flexible inputs(such as labor and intermediate good) can be used to identify the elasticity coefficients. Given the above identification, I estimate the input elasticity of production functions across firms in the same industry using the revenue share of intermediate good. I apply the likelihood ratio test on the input elasticity to determine the number of types of production functions. The empirical results show that if the panel length increases, the observations can be categorized into more types.

The rest of the paper is organized as follows. In Section 2, I define the finite normal mixture panel regression model. In Section 3, I demonstrate the likelihood ratio test(LRT) of homogeneity of normal mixture panel regression against two-component model as a precursor to general M_0 components test. Section 4 generalizes the test result to testing M_0 components against $M_0 + 1$ components. Section 5 shows the details of the modified EM test. Section 6 introduces the penalty function and reports the simulated results of the tests. Section 7 reports the likelihood ratio test result with empirical data.

2 Heteroskedastic finite mixture panel normal regression model

We consider finite mixture normal regression models with panel data, where the panel length T is fixed and the number of cross-sectional observations N goes to infinity. Define $\boldsymbol{w} := \{y_t, \boldsymbol{x}_t, \boldsymbol{z}_t\}_{t=1}^{T}$ with $y_t \in \mathbb{R}, \boldsymbol{x}_t \in \mathbb{R}^q, \boldsymbol{x}_t \in \mathbb{R}^p$. Given $M \geq 2$, denote the density of a $M-$component finite mixture panel normal regression model as

$$f_M(\boldsymbol{w}; \boldsymbol{\vartheta}_M) = \sum_{j=1}^{M} \alpha_j f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}_j), \tag{1}$$

where $\boldsymbol{\vartheta}_M = (\boldsymbol{\alpha}^\top, \boldsymbol{\theta}_1^\top, ..., \boldsymbol{\theta}_M^\top, \gamma^\top)^\top$ and

$$f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}_j) = \prod_{t=1}^{T} \frac{1}{\sigma_j} \phi\left(\frac{y_t - \mu_j - \boldsymbol{x}_t^\top \boldsymbol{\beta}_j - \boldsymbol{z}_t^\top \boldsymbol{\gamma}}{\sigma_j}\right) \tag{2}$$

is the j-th component density function with $\boldsymbol{\theta}_j = (\mu_j, \boldsymbol{\beta}_j^\top, \sigma_j^2)^\top \in \Theta_\theta$, $\mu_j \in \Theta_\mu \subset \mathbb{R}, \beta_j \in \Theta_\beta \subset \mathbb{R}^q$, $\sigma_j \in \Theta_\sigma \subset \mathbb{R}_{++}$ and $\gamma \in \Theta_\gamma \subset \mathbb{R}^p$. $\phi(t) = (2\pi)^{-1/2} \exp(-\frac{t^2}{2})$ is the standard normal p.d.f.

The number of components, denoted by M_0, is defined as the smallest integer M such that the data density of $\boldsymbol{w}$ admits the representation (1). Our goal is to test

$$H_0 : M = M_0 \text{ against } H_A : M = M_0 + 1. \tag{3}$$

2.1 Likelihood ratio test of $H_0 : M = 1$ against $H_A : M = 2$

We first develop the likelihood ratio test for testing $H_0 : M = 1$ against $H_1 : M = 2$. Consider a random sample of N with panel length of T independent observations $\{\boldsymbol{W}_i\}_{i=1}^N$ where $\boldsymbol{W}_i = \{(Y_{it}, \boldsymbol{X}_{it}^\top, \boldsymbol{Z}_{it}^\top)^\top\}_{t=1}^T$ from a true one-component density $f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$ defined in equation (2). The superscript $*$ signifies the true parameter value. Now consider a two-component mixture density function

$$f_2(\boldsymbol{w}; \boldsymbol{\vartheta}_2) = \alpha f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}^1) + (1-\alpha) f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}^2), \tag{4}$$

where $\boldsymbol{\vartheta}_2 = (\alpha, \boldsymbol{\gamma}, \boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top) \in \Theta_{\vartheta_2}$ and α is the mixing probability of the first component. The two-component model can generate the true one-component density in two cases:(1) $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$; (2) $\alpha = 0$ or 1. The null hypothesis $H_0 : M_0 = 1$ can be partitioned into two sub-hypotheses: H_{01} : $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ and $H_{02} : \alpha(1-\alpha) = 0$. This paper focuses on the first case. Define $\Upsilon_1^* : \{(\alpha, \boldsymbol{\gamma}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta_{\vartheta_2} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^* \text{ and } \boldsymbol{\gamma} = \boldsymbol{\gamma}^*\}$ the subspace of Θ_{ϑ_2} that corresponds to H_{01}.

For normal mixture models without panel dimension (i.e., $T = 1$), the log-likelihood function is known to be unbounded for any sample size N if σ_j is not restricted to be bounded from 0 (Hartigan, 1985). The following proposition shows that, for normal mixture models with panel data, the likelihood ratio test statistic becomes unbounded as the sample size N goes to ∞.

define the likelihood ratio statistics with respect to the true parameter under H_0 as:

$$LR^*(\hat{\boldsymbol{\vartheta}}_2) := 2\left\{\sum_{i=1} \log f_2(\boldsymbol{W}_i; \hat{\boldsymbol{\vartheta}}_2) - \sum_{i=1} \log f_1(\boldsymbol{W}_i; \boldsymbol{\vartheta}_1^*)\right\}$$

where f_M is the density of M-component finite mixture distribution; $\boldsymbol{\vartheta}_1^* = ((\gamma^*)^\top, (\boldsymbol{\theta}^*)^\top)^\top$ is the true parameter value under H_0.

Therefore, we consider a maximum penalized likelihood estimator (PMLE) introduced by Chen and Li (2009). Similar to Chen and Li (2009), we use the following penalty function with

$M = 2$:

$$p_n(\boldsymbol{\vartheta}_M) := -a_n \sum_{j=1}^{M} \{(\hat{\sigma}_{0,j})^2/(\sigma_j)^2 + \log((\sigma_j)^2/(\hat{\sigma}_{0,j})^2) - 1\}, \tag{5}$$

where $\hat{\sigma}_{0,j}$ is the estimator of σ_0 from one-component model and $\hat{\sigma}_j^2$ is the parameter under the 2-component model. $a_n = o_p(n^{1/4})$ is the data-driven penalty term following the discussion of Section 4.1. Let $\hat{\vartheta}_2 = \arg\max_{\vartheta_2 \in \Theta_{\vartheta_2}} \sum_{i=1}^{N} \log f_2(\boldsymbol{W}_i; \vartheta_2)$ denote the MLE under the 2-component model.

Assumption 1. *$\boldsymbol{X}$ has finite second moment, and Pr$(\boldsymbol{X}^\top \boldsymbol{\beta} \neq \boldsymbol{X}^\top \boldsymbol{\beta}^*) > 0$ for $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$.*

Proposition 1. *Suppose that Assumption 1 holds. Then under the null hypothesis $H_0 : M_0 = 1$, $\inf_{\boldsymbol{\vartheta}_2 \in \Theta_2^*} |\hat{\boldsymbol{\vartheta}}_2 - \boldsymbol{\vartheta}_2| \to 0$.*

In testing $H_{01} : \boldsymbol{\theta}_1 = \boldsymbol{\theta_2} = \boldsymbol{\theta}^*$, the standard LRTS analysis breaks down. The first order derivative at the true value $\vartheta_2^* = (\theta^*, \theta^*, \alpha, \gamma^*)$ are linear dependent as

$$\nabla_{\boldsymbol{\theta}_1} \log f_2(\boldsymbol{w}; \vartheta_2^*) = \frac{\alpha}{1-\alpha} \nabla_{\boldsymbol{\theta}_2} \log f_2(\boldsymbol{w}; \vartheta_2^*).$$

The linear dependency leads to loss of strong identifiability and difficulty in analysis. We analyze the LRTS by developing a higher order approximation for the log-likelihood function.

2.2 Reparameterization

To extract the direction of Fisher Information matrix singularity, I adapt the reparameterization approach by Kasahara and Shimotsu (2012), following the result of Rotnitzky et al. (2000). Now consider the one-to-one reparameterization of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ given α:

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\nu} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \\ \alpha\boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2 \end{pmatrix} \text{ so that } \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\nu} + (1-\alpha)\boldsymbol{\lambda} \\ \boldsymbol{\nu} - \alpha\boldsymbol{\lambda} \end{pmatrix}, \tag{6}$$

where ν and λ are both $(q+2) \times 1$ reparameterized parameter vectors. Define the space for reparameterized parameters as

$$\boldsymbol{\psi}_\alpha := (\boldsymbol{\gamma}^\top, \boldsymbol{\nu}^\top, \boldsymbol{\lambda}^\top) \in \Theta_{\psi_\alpha}, \tag{7}$$

where $\Theta_{\boldsymbol{\psi}_\alpha} = \{\boldsymbol{\psi}_\alpha : \boldsymbol{\gamma} \in \Theta_\gamma, \boldsymbol{\nu} + (1-\alpha)\boldsymbol{\lambda} \in \Theta_\theta, \boldsymbol{\nu} - \alpha\boldsymbol{\lambda} \in \Theta_\theta\}$. Under the null hypothesis $H_{01} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$, we have $\boldsymbol{\lambda} = (0, \ldots, 0)^\top$ and $\boldsymbol{\nu} = \boldsymbol{\theta}^*$. I rewrite the reparameterized parameters under null hypothesis to be $(\boldsymbol{\psi}_\alpha^*)^\top = ((\boldsymbol{\gamma}^*)^\top, (\boldsymbol{\theta}^*)^\top, 0, \ldots, 0)^\top$. Under the reparameterized parameter space, the density function and its logarithm are expressed as

$$g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha) = \alpha f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}+(1-\alpha)\boldsymbol{\lambda}) + (1-\alpha) f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda}); \tag{8}$$

$$l(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha) = \log g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha), \tag{9}$$

Partition the parameter $\boldsymbol{\psi}_\alpha = (\boldsymbol{\eta}^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\eta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\nu}^\top)^\top$, where $\boldsymbol{\eta}^* = ((\boldsymbol{\gamma}^*)^\top, (\boldsymbol{\nu}^*)^\top)^\top$, $\boldsymbol{\lambda}^* = \mathbf{0}$. Recall that $\theta = (\mu, \beta^\top, \sigma^2)$, and $\beta \in \mathbb{R}^q$, then $\lambda \in \mathbb{R}^{q+2}$ has $q+2$ components. $\lambda = (\lambda_\mu, (\lambda_\beta)^\top, \lambda_\sigma)$, where $\lambda_\mu = \mu^1 - \mu^2$, $\lambda_\beta = \beta^1 - \beta^2 \in \mathbb{R}^q$, $\lambda_\sigma = (\sigma^1)^2 - (\sigma^2)^2$.

Under this reparameterization, the first order derivatives of the reparameterized log likelihood function with respect to the reparameterized parameters $\boldsymbol{\eta}$ is identical to those under the one-component model, and the first order derivative w.r.t $\boldsymbol{\lambda}$ is a zero vector.

$$\nabla_{\boldsymbol{\eta}} g(\boldsymbol{w};\boldsymbol{\psi}^*_\alpha,\alpha) = \nabla_{(\boldsymbol{\eta}^\top,\boldsymbol{\lambda}^\top)^\top} f(\boldsymbol{w};\gamma,\theta); \nabla_{\boldsymbol{\lambda}} g(\boldsymbol{w};\boldsymbol{\psi}^*_\alpha,\alpha) = 0. \tag{10}$$

With $\nabla_\lambda l(\boldsymbol{w};\boldsymbol{\psi}^*_\alpha,\alpha) = 0$, the information is singular under the usual quadratic expansion of the likelihood ratio test statistics, and the usual quadratic approximation fails. Consequently, the information on $\boldsymbol{\lambda}$ is provided by the second order derivative of $l(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha)$ w.r.t $\boldsymbol{\lambda}$. Instead of the first order condition with respect to $\boldsymbol{\lambda}$, I use second order derivative with respect to $\boldsymbol{\lambda}$ in place in the score functions as proposed by KS12:

$$\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} l(\boldsymbol{w};\boldsymbol{\psi}^*_\alpha,\alpha) = \alpha(1-\alpha)\frac{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}^\top} f(\boldsymbol{w};\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*)}{f(\boldsymbol{w};\boldsymbol{\gamma}^*,\theta^*)}. \tag{11}$$

We derive the asymptotic distribution of the LRTS. Let f^* and ∇f^* denote $f(\boldsymbol{w};\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*)$ and $\nabla f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\theta})$. Define the score vector $\boldsymbol{s}(\boldsymbol{w})$ as

$$\boldsymbol{s}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{s}_{\boldsymbol{\eta}} \\ \boldsymbol{s}_{\boldsymbol{\lambda}} \end{pmatrix} \text{ with } \underset{(p+q+2)\times 1}{\boldsymbol{s}_{\boldsymbol{\eta}}} = \frac{\nabla_{(\boldsymbol{\eta}^\top,\boldsymbol{\lambda}^\top)^\top} f^*}{f^*} \quad \text{and} \quad \underset{(q+2)(q+1)/2}{\boldsymbol{s}_{\boldsymbol{\lambda}}} = \frac{1}{2!}\frac{\nabla_{\boldsymbol{\theta}^{\otimes 2}} f^*}{f^*}. \tag{12}$$

Collect the relevant reparameterized parameters and define $\boldsymbol{t}(\boldsymbol{\psi}_\alpha,\alpha)$ as

$$\boldsymbol{t}(\boldsymbol{\psi}_\alpha,\alpha) = \begin{pmatrix} \boldsymbol{t}_{\boldsymbol{\eta}} \\ \boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda},\alpha) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\eta}-\boldsymbol{\eta}^* \\ \alpha(1-\alpha)\boldsymbol{v}(\boldsymbol{\lambda}) \end{pmatrix}, \tag{13}$$

where the vector of outer product of λ with itself as

$$v(\boldsymbol{\lambda}) = \boldsymbol{\lambda}\otimes\boldsymbol{\lambda}. \tag{14}$$

Let $L_n(\boldsymbol{\psi}_\alpha,\alpha) := \sum_{i=1}^N \log g(\boldsymbol{W}_i;\boldsymbol{\psi}^*_\alpha,\alpha)$ the re-parameterized log-likelihood function. Let

$\hat{\boldsymbol{\psi}}_{\boldsymbol{\psi}\in\Theta_\psi} = \arg\max PL_n(\boldsymbol{\psi}, \alpha)$ denote the PMLE of the $\boldsymbol{\psi}$, where Θ_ψ is defined as the space of $\boldsymbol{\psi}$ so that the $\boldsymbol{\vartheta}_2$ implied is in $\Theta_{\boldsymbol{\vartheta}}$. Let $(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{\theta}}_0)$ be the one-component MLE that maximizes the one-component likelihood function $L_{0,n}(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \sum_{i=1}^N f(\boldsymbol{W}_i; \boldsymbol{\gamma}, \boldsymbol{\theta})$. Define the LRTS of testing H_{01}, with $\epsilon \in (0,1)$ as

$$LR_n(\epsilon_1) := \max_{\alpha\in[\epsilon_1, 1-\epsilon_1]} 2\{L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0)\}. \tag{15}$$

We can use the penalized log-likelihood function and define the penalized likelihood ratio as $PLR_n(\epsilon_1)$, since the penalty term is negligible under our assumption. $PLR_n(\epsilon_1)$ has the same asymptotic distribution as $LR_n(\epsilon_1)$.

With $\boldsymbol{s}(\boldsymbol{w})$ defined in (12), define

$$\boldsymbol{\mathcal{I}} = \begin{pmatrix} \boldsymbol{\mathcal{I}}_\eta & \boldsymbol{\mathcal{I}}_{\eta\lambda} \\ \boldsymbol{\mathcal{I}}_{\lambda\eta} & \boldsymbol{\mathcal{I}}_\lambda \end{pmatrix}, \ \boldsymbol{\mathcal{I}}_\eta = E(s_\eta s_\eta^\top), \ \boldsymbol{\mathcal{I}}_{\lambda\eta} = E[s_\lambda s_\eta], \boldsymbol{\mathcal{I}}_{\eta\lambda} = \boldsymbol{\mathcal{I}}_{\lambda\eta}^\top, \ \boldsymbol{\mathcal{I}}_\lambda = E[s_\lambda s_\lambda^\top], \tag{16}$$

$$\boldsymbol{\mathcal{I}}_{\lambda,\eta} = \boldsymbol{\mathcal{I}}_{\lambda\lambda} - \boldsymbol{\mathcal{I}}_{\lambda\eta}\boldsymbol{\mathcal{I}}_\eta^{-1}\boldsymbol{\mathcal{I}}_{\eta\lambda}, \ \boldsymbol{Z}_{\boldsymbol{\lambda}} := (\boldsymbol{\mathcal{I}}_{\lambda,\eta})^{-1}\boldsymbol{G}_{\lambda,\eta}, \tag{17}$$

where $\boldsymbol{G}_{\lambda,\eta} \sim N(0, \boldsymbol{\mathcal{I}}_{\lambda,\eta})$. Define the set that characterize the admissible values of $\sqrt{n}\boldsymbol{t}(\boldsymbol{\lambda}, \alpha)$ when $n \to \infty$ the cone

$$\Lambda_\lambda = \left\{\sqrt{n}\alpha(1-\alpha)v(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Theta_\lambda\right\}, \tag{18}$$

where $v(\cdot)$ is the outer product defined in equation (14).

Define $\hat{t}_{\boldsymbol{\lambda}}$ by the conditional distribution of $\boldsymbol{\lambda}$ given $\boldsymbol{\eta}$. Then $\boldsymbol{t}_{\boldsymbol{\lambda}}$ is projecting the normal random vector W_λ onto the cone:

$$r_\lambda(\hat{t}_\lambda) = \inf_{\boldsymbol{t}_{\boldsymbol{\lambda}}\in\Lambda_\lambda} r_\lambda(\boldsymbol{t}_{\boldsymbol{\lambda}}), \quad r_\lambda(\boldsymbol{t}_{\boldsymbol{\lambda}}) := (\boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{Z}_{\boldsymbol{\lambda}})^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}(\boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{Z}_{\boldsymbol{\lambda}}). \tag{19}$$

The following proposition establishes the asymptotic null distribution of LRTS under the null hypothesis.

Assumption 2. *$\boldsymbol{X}$ has finite tenth moment.*

Assumption 3. *$\boldsymbol{\mathcal{I}}$ is finite and positive definite.*

Proposition 2. *Suppose that assumption 1, 2 and 3 hold, a_n in (5) satisfies $a_n = O(1)$. Under the null hypothesis $M = 1$, the likelihood ratio test statistic $LR_n(\epsilon_1) \to_d (\hat{\boldsymbol{t}}_\lambda)^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}\hat{\boldsymbol{t}}_\lambda$, where $\hat{\boldsymbol{t}}$ and $LR_n(\epsilon_1)$ are defined in (15) and (13).*

It is obvious that finite normal mixture panel regression model satisfy assumption 1 and 2, and I show that the model satisfies assumption 3.

2.3 Likelihood ratio test of $H_0 : M = M_0$ against $H_A : M = M_0 + 1$

Ihis section establishes the asymptotic distribution of the LRTS for testing the null hypothesis of M_0 components against alternative $(M_0 + 1)$ components for $M_0 \geq 1$. I follow Kasahara and Shimotsu (2015) and develop a partition of the $(M_0 + 1)$-component parameter space into M_0 sub-spaces, of which each sub-space corresponds to a specific way of generating the true model. Then I derive the asymptotic distribution of LRT statistic for each subset and characterize the asymptotic distribution of the LRT statistic by the maximum across the M_0 partitions.

Consider a random sample of N with panel length of T independent observations $\{\boldsymbol{W}_i\}_{i=1}^N$ where $\boldsymbol{W}_i = \{(y_{it}, \boldsymbol{X}_{it}^\top, \boldsymbol{Z}_{it}^\top)^\top\}_{t=1}^T$ from a M_0-component density $f_{M_0}(\boldsymbol{w}; \vartheta_{M_0})$ defined in equation (20):

$$f_{M_0}(\boldsymbol{w}; \boldsymbol{\vartheta}_{M_0}^*) = \sum_{j=1}^{M_0} \alpha_j^* f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_j^*), \tag{20}$$

where $\boldsymbol{\vartheta}_{M_0}^* = (\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_{M_0}^*, \alpha_1^*, \ldots, \alpha_{M_0-1}^*, \boldsymbol{\gamma}^*) \in \Theta_{\vartheta_{M_0}}$. Let the density of the $(M_0 + 1)$-component model be defined by:

$$f_{M_0+1}(\boldsymbol{w}; \boldsymbol{\vartheta}_{M_0+1}) = \sum_{j=1}^{M_0+1} \alpha_j f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}_j). \tag{21}$$

where $\boldsymbol{\vartheta}_{M_0+1} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{M_0+1}, \alpha_1, \ldots, \alpha_{M_0}, \boldsymbol{\gamma}) \in \Theta_{\vartheta_{M_0+1}}$ as defined in (20). Without generality, assume $\mu_0^{1*} < \mu_0^{2*}, \ldots, < \mu_0^{M_0 *}$ in the true parameters. The $(M_0 + 1)$-component model (21) gives rise to the true density (20) in two different cases: (1) two components have the same mixing parameter so that $\boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1} = \boldsymbol{\theta}_h^*$ for some $h = 1 \ldots, M_0$; and (2) one component has zero mixing proportion so that $\alpha_h = 0$ for some $h = 1, \ldots, M_0 + 1$. This paper focuses on case (1). Define the subsets of the parameter space $\Theta_{\vartheta_{M_0+1}}$ corresponding to the the null hypothesis of $H_{0,1h} : \boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1}$ as:

$$\begin{aligned} \Upsilon_{1h}^* := \Big\{ \boldsymbol{\vartheta}_{M_0+1} : \alpha_h + \alpha_{h+1} = \alpha_h^* \text{ and } \boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1} = \boldsymbol{\theta}_h^*; \boldsymbol{\gamma} = \boldsymbol{\gamma}^*; \\ \alpha_j = \alpha_j^* \text{ and } \boldsymbol{\theta}_j = \boldsymbol{\theta}_j^* \text{ for } 1 \leq j < h; \\ \alpha_j = \alpha_{j-1}^* \text{ and } \boldsymbol{\theta}_j = \boldsymbol{\theta}_{j-1}^* \text{ for } h+1 \leq j \leq M_0 + 1 \Big\}, \end{aligned} \tag{22}$$

for $h = 1, \ldots, M_0$. Note the null hypothesis under case 1 is $H_{01} = \cup_{h=1}^{M_0} H_{0,1h}$. Define $\Upsilon_1^* = \cup_{h=1}^{M_0} \Upsilon_{1h}^*$ the subspace of $\Theta_{\boldsymbol{\vartheta}_{M_0+1}}$ that corresponds to the null hypotehsis $H_{0,1}$. $H_{0,1h} : \boldsymbol{\vartheta}_{M_0+1} \in \Upsilon_{1h}^*$ for $h = 1, \ldots, M_0$.

Similar to the case of testing $M_0 = 1$, I approximate the log-likelihood function by expanding w.r.t the reparameterized parameters around the true parameter value. However, the difficulty rises because the true density function can be described by many different elements of the parameter space of the (M_0+1)-component model.

Consider a sufficiently small neighborhood of $\Upsilon^*_{1h} \subset \Theta_{\vartheta_{M_0+1}}$ such that $\mu_1 < \ldots < \mu_{h-1}, \mu_{h+1} < \ldots < \mu_{M_0+1}$ holds, and introduce the following one-to-one reparameterization from the (M_0+1)-component model parameter $\boldsymbol{\vartheta}_{M_0+1} = (\alpha_1, \ldots, \alpha_{M_0}, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_h^\top, \boldsymbol{\theta}_{h+1}^\top, \ldots, \boldsymbol{\theta}_{M_0+1}^\top, \boldsymbol{\gamma}^\top)^\top$ to $\boldsymbol{\psi}_{h,\tau}$ and τ in the following pattern:

$$\begin{aligned}
&\boldsymbol{\psi}_{h,\tau} = (\pi_1, \ldots, \pi_{M_0-1}, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_{h-1}^\top, \boldsymbol{\nu}_h^\top, \boldsymbol{\theta}_{h+2}^\top, \ldots, \boldsymbol{\theta}_{M_0+1}^\top, \boldsymbol{\gamma}, \boldsymbol{\lambda}_h^\top)^\top \\
&\quad \pi_h = \alpha_h + \alpha_{h+1},\ \tau := \frac{\alpha_h}{\alpha_h + \alpha_{h+1}},\ \boldsymbol{\lambda}_h = \boldsymbol{\theta}_h - \boldsymbol{\theta}_{h+1},\ \boldsymbol{\nu}_h = \tau\boldsymbol{\theta}_h + (1-\tau)\boldsymbol{\theta}_{h+1}, \\
&\quad \boldsymbol{\pi} = (\pi_1, \ldots, \pi_{h-1}, \pi_h, \pi_{h+1}, \ldots, \pi_{M_0-1})^\top = (\alpha_1, \ldots, \alpha_{h-1}, (\alpha_h + \alpha_{h+1}), \alpha_{h+2}, \ldots, \alpha_{M_0})^\top,
\end{aligned} \tag{23}$$

with $\pi_{M_0} = 1 - \sum_{j=1}^{M_0-1} \pi_j$. so that $\boldsymbol{\theta}_h = \boldsymbol{\nu}_h + (1-\tau)\boldsymbol{\lambda}_h$ and $\boldsymbol{\theta}_{h+1} = \boldsymbol{\nu}_h - \tau\boldsymbol{\lambda}_h$. In the reparameterized model, the null restriction $\boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1}$ implied by $H_{0,1h}$ holds if and only if $\boldsymbol{\lambda}_h = 0$. For $h \leq M_0$, collect the reparameterized model parameters other than τ and $\boldsymbol{\lambda}_h$ into

$$\begin{aligned}
\boldsymbol{\eta}_h &= (\pi_1, \ldots, \pi_{M_0-1}, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_{h-1}^\top, \boldsymbol{\nu}_h^\top, \boldsymbol{\theta}_{h+2}^\top, \ldots, \boldsymbol{\theta}_{M_0+1}^\top, \boldsymbol{\gamma}^\top)^\top, \\
\boldsymbol{\psi}_{h,\tau} &= (\boldsymbol{\eta}_h^\top, \boldsymbol{\lambda}_h^\top)^\top.
\end{aligned} \tag{24}$$

Define the sum of reparameterized log-likelihood functions by

$$L_n^h(\psi_\tau^h, \tau) := \sum_{i=1}^N l^h(\boldsymbol{w}; \boldsymbol{\psi}_{h,\tau}, \tau). \tag{25}$$

Note that under the null hypothesis, $\boldsymbol{\eta}_h^* = (\alpha_1^*, \ldots, \alpha_{M_0-1}^*, (\boldsymbol{\theta}_1^*)^\top, \ldots, (\boldsymbol{\theta}_{M_0-1}^*)^\top, (\gamma^*)\top)^\top$.

Let $\Theta_{\boldsymbol{\vartheta}_{M_0+1}}(\epsilon)$ be a subset of $\Theta_{\boldsymbol{\vartheta}_{M_0+1}}$ such that $\alpha_j \in [\epsilon_j, 1-\epsilon_j]$ for $j = 1, \ldots, M_0$. Define the PMLE by

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}_{M_0+1}(\epsilon_j) &= \arg \max_{\boldsymbol{\vartheta}_{M_0+1}(\epsilon_j) \in \Theta_{\boldsymbol{\vartheta}_{M_0+1}}} PL_n(\boldsymbol{\vartheta}_{M_0+1}), \\
\hat{\boldsymbol{\vartheta}}_{M_0} &= \arg \max_{\boldsymbol{\vartheta}_{M_0} \in \Theta_{\boldsymbol{\vartheta}_{M_0}}} PL_{0,n}(\boldsymbol{\vartheta}_{M_0}),
\end{aligned} \tag{26}$$

where $PL_n(\boldsymbol{\vartheta}_{M_0+1}) := L_n(\boldsymbol{\vartheta}_{M_0+1}) + p_n(\boldsymbol{\vartheta}_{M_0+1})$ and $PL_{0,n}(\boldsymbol{\vartheta}_{M_0}) := L_{0,n}(\boldsymbol{\vartheta}_{M_0}) + p_n(\boldsymbol{\vartheta}_{M_0})$ with $L_n(\boldsymbol{\vartheta}_{M_0+1}) = \sum_{i=1}^N \log f_{M_0+1}(\boldsymbol{W}_i; \boldsymbol{\gamma}, \boldsymbol{\vartheta}_{M_0+1})$ and $L_{0,n}(\boldsymbol{\vartheta}_{M_0}) = \sum_{i=1}^N \log f_{M_0+1}(\boldsymbol{W}_i, \boldsymbol{\gamma}, \boldsymbol{\vartheta}_{M_0})$.

Collect the score vector for testing $H_{0,1h}$ for $h = 1, \ldots, M_0$ into one vector as

$$\tilde{\boldsymbol{s}}(\boldsymbol{w}) = \begin{pmatrix} \tilde{\boldsymbol{s}}_{\boldsymbol{\eta}}(\boldsymbol{w}) \\ \tilde{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{w}) \end{pmatrix},\ \text{where} \underset{(M_0+p+q+1)\times 1}{\tilde{\boldsymbol{s}}_{\boldsymbol{\eta}}}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{s}_{\boldsymbol{\alpha}} \\ \boldsymbol{s}_{(\boldsymbol{\gamma},\boldsymbol{\nu})} \end{pmatrix} \text{ and } \tilde{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{s}_{\boldsymbol{\lambda}}^1(\boldsymbol{w}) \\ \vdots \\ \boldsymbol{s}_{\boldsymbol{\lambda}}^{M_0}(\boldsymbol{w}) \end{pmatrix}. \tag{27}$$

Let $f_0^* := f_{M_0+1}(\boldsymbol{w}; \boldsymbol{\vartheta}_{M_0+1})$. Then for $m = 1, \ldots, M_0$, the score functions are

$$\boldsymbol{s}_{\boldsymbol{\alpha}} = \begin{pmatrix} f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_0^*) - f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_{M_0}^*) \\ \vdots \\ f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_{M_0-1}^*) - f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_{M_0}^*) \end{pmatrix} \Bigg/ f_0^*,$$

$$\boldsymbol{s}_{(\boldsymbol{\gamma},\boldsymbol{\nu})} = \sum_{m=1}^{M_0} \alpha_m^* \nabla_{(\boldsymbol{\gamma},\boldsymbol{\nu})} f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_m^*)/f_0^*, \quad \boldsymbol{s}_{\boldsymbol{\lambda}}^m(\boldsymbol{w}) = \alpha_m^* \nabla_{\boldsymbol{\theta}^{\otimes 2}} f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}_m^*)/(2! f_0^*). \tag{28}$$

Define

$$\begin{aligned} &\tilde{\boldsymbol{\mathcal{I}}} := \mathbb{E}[\tilde{\boldsymbol{s}}(\boldsymbol{W})\tilde{\boldsymbol{s}}(\boldsymbol{W})^\top], \quad \tilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta}} = \mathbb{E}[\tilde{\boldsymbol{s}}_{\boldsymbol{\eta}}\tilde{\boldsymbol{s}}_{\boldsymbol{\eta}}^\top], \quad \tilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\lambda\eta}} = \mathbb{E}[\tilde{\boldsymbol{s}}_{\boldsymbol{\lambda}}\tilde{\boldsymbol{s}}_{\boldsymbol{\eta}}^\top], \\ &\tilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\eta\lambda}} = \tilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\lambda\eta}}^\top, \quad \tilde{\boldsymbol{\mathcal{I}}}_{\boldsymbol{\lambda}} = \mathbb{E}[\tilde{\boldsymbol{s}}_{\boldsymbol{\lambda}}\tilde{\boldsymbol{s}}_{\boldsymbol{\lambda}}^\top], \quad \tilde{\boldsymbol{\mathcal{I}}}_{\lambda,\eta} = \tilde{\boldsymbol{\mathcal{I}}}_{\lambda\lambda} - \tilde{\boldsymbol{\mathcal{I}}}_{\lambda\eta}\tilde{\boldsymbol{\mathcal{I}}}_{\eta}^{-1}\tilde{\boldsymbol{\mathcal{I}}}_{\eta\lambda}. \end{aligned} \tag{29}$$

Collect the relevant parameters as $\boldsymbol{t}_n^h(\boldsymbol{\psi}_{h,\tau}, \tau)$ and define the normalized score functions $\tilde{\boldsymbol{s}}$ in the similar pattern as (13):

$$\boldsymbol{t}_n^h(\boldsymbol{\psi}_{h,\tau}, \tau) := \begin{pmatrix} n^{1/2}(\boldsymbol{\eta}_h - \boldsymbol{\eta}_h^*) \\ n^{1/2}\tau(1-\tau)v(\boldsymbol{\lambda}_h). \end{pmatrix}, \tilde{S} := n^{-1/2}\sum_{i=1}^{N} \tilde{s}_i, \tag{30}$$

where $v(\boldsymbol{\lambda}_h) = \boldsymbol{\lambda}_h \otimes \boldsymbol{\lambda}_h$, where $\boldsymbol{\lambda}_h = (\lambda_{h,\mu}, \boldsymbol{\lambda}_{h,\beta}^\top, \lambda_{h,\sigma})^\top$, $\lambda_{h,\mu} = \mu_h - \mu_{h+1} \in \Theta_\mu$, $\boldsymbol{\lambda}_{h,\beta} = \boldsymbol{\beta}_h - \boldsymbol{\beta}_{h+1} \in \Theta_\beta$, $\lambda_{h,\sigma} = (\sigma_h)^2 - (\sigma_{h+1})^2 \in \mathbb{R}$. Let $\tilde{\boldsymbol{G}}_{\lambda,\eta} = (\tilde{\boldsymbol{G}}_{\lambda,\eta}^1, \ldots, \tilde{\boldsymbol{G}}_{\lambda,\eta}^{M_0})^\top \sim N(0, \tilde{\boldsymbol{\mathcal{I}}}_{\lambda,\eta})$ be a $\mathbb{R}^{M_0(q+2)(q+1)/2}$-valued random vector, and let $\tilde{\boldsymbol{\mathcal{I}}}_{\lambda,\eta}^m = \mathbb{E}[\tilde{\boldsymbol{G}}_{\lambda,\eta}\tilde{\boldsymbol{G}}_{\lambda,\eta}^\top]$ and $\tilde{\boldsymbol{Z}}_{\lambda,\eta} = (\tilde{\boldsymbol{\mathcal{I}}}_{\lambda,\eta}^m)^{-1}\tilde{\boldsymbol{G}}_{\lambda,\eta}$.

$L_n^h(\boldsymbol{\psi}_{h,\tau}, \tau) - L_n^h(\boldsymbol{\psi}_{h,\tau}^*, \tau)$ admits a similar quadratic expansion as derived in (15). Define the local LRT statistic for testing $H_{0,1h}$ as $LR_{n,1h}^\tau := 2\{L_n^h(\boldsymbol{\psi}_{h,\tau}, \tau) - L_{0,n}(\hat{\vartheta}_{M_0})\}$. Let $\Theta_\alpha(\epsilon) := \{\alpha \in \Theta_\alpha : \alpha^1, \ldots, \alpha^{M_0} \in [\epsilon, 1-\epsilon]\}$, and define the LRT statistic for testing H_{01} subject to $\alpha \in \Theta_\alpha(\epsilon)$ as $LR_{n,1}^{M_0}(\epsilon) := \max_{\psi \in \Theta_\psi, \alpha \in \Theta_\alpha(\epsilon)} 2\{L_n(\boldsymbol{\psi}_{h,\tau}, \tau) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\}$. In addition, define $\hat{t}_\lambda^h$ similar to $\hat{t}_\lambda$, so that $\hat{t}_\lambda^h$ is defined by:

$$r_\lambda^h(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h) = \inf_{\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h} r^h(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h); \quad r_\lambda^h(\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h) = (\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h - \tilde{\boldsymbol{Z}}_{\lambda,\eta})^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}^h (\hat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h - \tilde{\boldsymbol{Z}}_{\lambda,\eta}). \tag{31}$$

Assumption 4. *(a) $\alpha_j^* \in [\epsilon_1, 1-\epsilon_1]$ for $j = 1, \ldots, M_0$. (b) $\tilde{\boldsymbol{\mathcal{I}}}$ defined in (29) is non-singular.*

Proposition 3. *Assume that Assumption 2, 3 and 4 hold, and a_n satisfies (5) astisfies $a_n = o_n(1)$. Then under the null hypothesis $H_0 : M = M_0$, $LR_n^{M_0}(\epsilon_1) \to_d \max\{(\hat{\boldsymbol{t}}_\lambda^1)^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^1 (\hat{\boldsymbol{t}}_\lambda^1), \ldots, (\hat{\boldsymbol{t}}_\lambda^{M_0})^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^{M_0} (\hat{\boldsymbol{t}}_\lambda^{M_0})\}$.*

3 Modified EM test

EM algorithm is widely used in maximum likelihood estimation of finite mixture models, especially under normal density assumption. I extend the modified EM test of Kasahara and Shimotsu (2012) to test $H_0 : m = M_0$ against $H_A : m = M_0 + 1$ to the normal finite mixture panel regression model. The proposed modified EM statistic has the same asymptotic distribution as the LRT statistic for testing $H_{0,1h} : \boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1}$. Assume the null hypothesis is true, and the true density from the M_0-component model is $f_{M_0}(\boldsymbol{w}; \boldsymbol{\vartheta}^*_{M_0})$, where $\boldsymbol{\vartheta}^*_{M_0} = (\boldsymbol{\theta}^*_1, \boldsymbol{\theta}^*_2, \ldots, \boldsymbol{\theta}^*_{M_0}, \alpha^*_1, \ldots, \alpha^*_{M_0-1}, \boldsymbol{\gamma}^*) \in \Theta_{\vartheta_{M_0}}$. Because any parameters in $\Upsilon^*_1 = \cup_{h=1}^{M_0} \Upsilon^*_{1h}$ can generate the true density $f_{M_0}(\boldsymbol{w}; \boldsymbol{\vartheta}^*_{M_0}) = \sum_{j=1}^{M_0} \alpha_0^{j*} f(\boldsymbol{w}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}^*_j)$, I need to restrict the estimators under the (M_0+1)-component model to be in a neighborhood of Υ^*_{1h} in order to test $H_{0,1h}$. First suppose in $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{M_0}$, $\mu^*_1 < \mu^*_2 \ldots < \mu^*_{M_0}$. Let $\underline{\Theta}_\mu$ and $\overline{\Theta}_\mu$ denote the lower bound and upper bound of Θ_μ. Define $D^*_1 = [\underline{\Theta}_\mu, \frac{\mu^*_1 + \mu^*_2}{2}] \times \Theta_\beta \times \Theta_{\sigma^2}$, $D^*_h = [\frac{\mu^*_{h-1} + \mu^*_h}{2}, \frac{\mu^*_h + \mu^*_{h+1}}{2}] \times \Theta_\beta \times \Theta_{\sigma^2}$ for $h = 2, \ldots, M_0 - 1$, $D^*_{M_0} = [\frac{\mu^*_{M_0-1} + \mu^*_{M_0}}{2}, \overline{\Theta}_\mu] \times \Theta_\beta \times \Theta_{\sigma^2}$. Then $D^*_h \subset \Theta_\theta$ is a neighborhood that is close to θ_0^{h*} but not close to θ_0^{j*} for $j \neq h$. For $h = 1, \ldots M_0$, define a restricted parameter space $\boldsymbol{W}^*_h \subset \Theta_{\vartheta_{M_0+1}}$ as

$$\boldsymbol{W}^*_h = \left\{ \begin{aligned} \alpha_1, \ldots, \alpha_{M_0+1} > 0; \sum_{j=1}^{M_0+1} \alpha_j = 1; \gamma \in \Theta_\gamma; \theta \in \Theta_\theta : \boldsymbol{\theta}_j \in D^*_j \text{ for } j = 1, \ldots, h-1; \\ \boldsymbol{\theta}_h, \boldsymbol{\theta}_{h+1} \in D^*_h; \boldsymbol{\theta}_j \in D^*_{j-1} \text{ for } j = h+2, \ldots, M_0+1. \end{aligned} \right\} \tag{32}$$

Note that $\boldsymbol{W}^*_h \cap \Upsilon^*_{1h} \neq \emptyset$, and $\boldsymbol{W}^*_h \cap \Upsilon^*_{1l} = \emptyset$ if $h \neq l$.

To test $H_{0,1h}$, estimate the parameters $\boldsymbol{\vartheta}_{M_0+1}$ of (M_0+1)-component model under restriction of $\boldsymbol{\vartheta}_{M_0+1} \in \hat{\boldsymbol{W}}^*_h$. Under the null hypothesis, $\hat{\boldsymbol{\vartheta}}_{M_0}$ is a consistent estimator of $\boldsymbol{\vartheta}^*_{M_0}$ in the M_0-component model, we have $\hat{\boldsymbol{W}}^*_h$ and $\hat{D}^*_h$ are consistent estimators of $\boldsymbol{W}^*_h$ and D^*_h. Therefore, $\Pr(\hat{\boldsymbol{W}}^*_h \cap \Upsilon^*_{1h} \neq \emptyset) \to_p 1, \Pr(\hat{\boldsymbol{W}}^*_h \cap \Upsilon^*_{1l} = \emptyset) \to_p 1$. The resulting estimator $\hat{\boldsymbol{\vartheta}}_{M_0+1}$ approaches a neighborhood of Υ^*_{1h} under the null hypothesis.

To implement a modified EM test of $H_{0,1h}$, consider the reparameterization defined in (23). Then we have:

$$\begin{aligned} \boldsymbol{\pi} &= (\pi_1, \ldots, \pi_{h-1}, \pi_h, \pi_{h+1}, \ldots, \pi_{M_0-1})^\top, \\ \boldsymbol{\psi}_{h,\tau} &= (\pi_1, \ldots, \pi_{M_0-1}, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_{h-1}^\top, \boldsymbol{\nu}_h^\top, \boldsymbol{\theta}_{h+2}^\top, \ldots, \boldsymbol{\theta}_{M_0+1}^\top, \boldsymbol{\gamma}, \boldsymbol{\lambda}_h^\top)^\top, \\ \boldsymbol{\psi}^*_{h,\tau} &= (\alpha_1, \ldots, \alpha_h + \alpha_{h+1}, \ldots, \alpha_{M_0}, \boldsymbol{\theta}_1^{*\top}, \ldots, \boldsymbol{\theta}_{h-1}^{*\top}, \boldsymbol{\theta}_h^{*\top}, \boldsymbol{\theta}_{h+1}^{*\top}, \ldots, \boldsymbol{\theta}_{M_0}^{*\top}, \boldsymbol{\gamma}^*, \boldsymbol{0}^\top)^\top. \end{aligned} \tag{33}$$

Recall the reparameterized likelihood function is defined as (**??**), and define the penalized log-likelihood function for the (M_0+1)-component model as

$$PL_n^h(\boldsymbol{\psi}_{h,\tau},\tau) = L_n^h(\boldsymbol{\psi}_{h,\tau},\tau) + \textstyle\sum_{j=1}^{M_0+1} p_n((\sigma^j)^2), \tag{34}$$

$$p_n((\sigma^j)^2) = -a_n\{(\hat{\sigma}_{0,j})^2/(\sigma_j)^2 + \log((\sigma_j)^2/(\hat{\sigma}_{0,j})^2) - 1\}, \tag{35}$$

where $L_n^h(\cdot)$ is defined as (25) and the penalty function $p_n((\sigma^j)^2)$ need to satisfy Assumption **??** adopted from Kasahara and Shimotsu (2015). This assumption is adopted from Chen et al. (2008) and Chen and Li (2009).

Let $\mathcal{T}$ be a finite set of numbers in $(0, 0.5]$. For each $\tau_0 \in \mathcal{T}$, let $\tau^{(1)}(\tau_0) = \tau_0$, define the restricted penalized MLE $\boldsymbol{\psi}_h^{(1)}(\tau_0)$ by

$$\boldsymbol{\psi}_h^{(1)}(\tau_0) := \arg\max_{\boldsymbol{\psi}_{h,\tau_0}} PL_n^h(\boldsymbol{\psi}_{h,\tau_0},\tau_0), \tag{36}$$

Collect the mixing parameters of M_0+1-component model into one vector $\boldsymbol{\xi} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_{M_0+1}^\top)^\top \in \Theta_{\boldsymbol{\xi}}$. For $m = 1, \ldots, M_0$, define a restricted parameter space of $\boldsymbol{\xi}$ by $\Xi_m^* := \{\boldsymbol{\xi} \in \Theta_{\boldsymbol{\xi}} : \boldsymbol{\theta}_j \in D_j^* : j = 1, \ldots, m-1; \boldsymbol{\theta}_m, \boldsymbol{\theta}_{m+1} \in D_m^*; \}$

Define $\boldsymbol{\vartheta}_{M_0+1}^{h(1)}(\tau_0)$ to be the parameters that correspond to the reparameterized parameters $\boldsymbol{\psi}_h^{(1)}(\tau_0)$ and τ_0. Because the reparameterization is one-to-one from $\boldsymbol{\vartheta}_{M_0+1}$ to $\boldsymbol{\psi}_h^{(1)}(\tau_0)$, the problem defined by (36) is the same as

$$\boldsymbol{\vartheta}_{M_0+1}^{h(1)}(\tau_0) = \arg\max_{\boldsymbol{\vartheta}_{M_0+1} \in \Theta^m_{\boldsymbol{\vartheta}_{M_0+1}}(\tau)} \sum_{i=1}^{N} \log f_{M_0+1}(\boldsymbol{w}; \boldsymbol{\vartheta}_{M_0+1}), \tag{37}$$

where $\Theta^m_{\boldsymbol{\vartheta}_{M_0+1}}(\tau) = \{\theta \in \hat{\boldsymbol{W}}_h : \alpha^h/(\alpha^h + \alpha^{h+1}) = \tau_0\}$.

Starting from $(\boldsymbol{\vartheta}_{M_0+1}^{h(1)}(\tau_0), \tau^{(1)}(\tau_0))$, update $\boldsymbol{\vartheta}_{M_0+1}^{h(k)}(\tau_0)$ and $\tau^{h(k)}(\tau_0)$ by the following generalized EM algorithm in a similar way to the EM algorithm in Kasahara et al. (2015). Define $\vartheta_{M_0+1}^{h(k)}, \tau^{(k)}$ as the estimator and penalty term after k-th round of EM algorithm iteration. The details of the generalized EM algorithm are as follow. In the E-step, for $i = 1, \ldots, N$ and $j = 1, \ldots, M_0+1$, compute the weight for observation i and type j as:

$$\begin{aligned}
w_{ij}^{(k)} &= \begin{cases} \pi_j^{(k)} f(\boldsymbol{W}_i; \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}_j^{(k)}) / f_{M_0+1}(\boldsymbol{W}_i; \boldsymbol{\vartheta}_{M_0+1}^{h(k)}(\tau_0)), j = 1, \ldots, h-1, \\ \pi_{j-1}^{(k)} f(\boldsymbol{W}_i; \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}_j^{(k)}) / f_{M_0+1}(\boldsymbol{W}_i; \boldsymbol{\vartheta}_{M_0+1}^{h(k)}(\tau_0)), j = h+2, \ldots, M_0+1, \end{cases} \\
w_{ih}^{(k)} &= \tau^k \pi_h^{(k)} f(\boldsymbol{W}_i; \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}_h^{(k)}) / f_{M_0+1}(\boldsymbol{W}_i; \boldsymbol{\vartheta}_{M_0+1}^{h(k)}(\tau_0)), \\
w_{i,h+1}^{(k)} &= (1-\tau^k) \pi_h^{(k)} f(\boldsymbol{W}_i; \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}_{h+1}^{(k)}) / f_{M_0+1}(\boldsymbol{W}_i; \boldsymbol{\vartheta}_{M_0+1}^{h(k)}(\tau_0)).
\end{aligned} \tag{38}$$

In the M-step, update τ and $\boldsymbol{\alpha}$ in the following way:

$$\alpha_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} w_{ij}^{(k)}, \text{for } j = 1, \ldots, M_0+1,$$
$$\tau^{(k+1)} = \alpha_h^{(k+1)}/(\alpha_h^{(k+1)} + \alpha_{h+1}^{(k+1)}).$$

Update the estimation of parameters $\boldsymbol{\vartheta}_{M_0+1}^{h(k+1)}(\tau_0)$ in the following way:

$$(\sigma_j^{(k+1)})^2 = \arg\min_{(\sigma_j)^2} \left\{ \sum_{i=1}^{N} w_i^{j(k)} \sum_{t=1}^{T} (y_{it} - \mu^{j(k+1)} - \boldsymbol{z}_{it}^\top \boldsymbol{\gamma}^{(k+1)} - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_j^{(k+1)})^2 + p_n((\sigma^j)^2) \right\};$$
$$\boldsymbol{\gamma}^{(k+1)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{z}_{it}\boldsymbol{z}_{it}^\top \right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{z}_{it} \left(y_{it} - \sum_{j=1}^{M_0+1} w_{ij}^{(k)} \tilde{\boldsymbol{x}}_{it}^\top \begin{pmatrix} \mu_j^{(k)}) \\ \boldsymbol{\beta}_j^{(k)}) \end{pmatrix} \right) \right);$$
$$\begin{pmatrix} \mu_j^{(k+1)}) \\ \boldsymbol{\beta}_j^{(k+1)}) \end{pmatrix} = \left(\sum_{i=1}^{N} w_{ij}^{(k)} \sum_{t=1}^{T} \tilde{\boldsymbol{x}}_{it}\tilde{\boldsymbol{x}}_{it}^\top \right)^{-1} \left(\sum_{i=1}^{N} w_{ij}^{(k)} \sum_{t=1}^{T} \tilde{\boldsymbol{x}}_{it}(y_{it} - \boldsymbol{z}_{it}^\top \boldsymbol{\gamma}^{(k+1)}) \right),$$

where $\tilde{\boldsymbol{x}}_{it} = (1, \boldsymbol{x}_{it}^\top)^\top$. Note that in the updating procedure, $\boldsymbol{\vartheta}_{M_0+1}^{h(k+1)}(\tau_0)$ is not restricted to be in $\hat{\boldsymbol{W}}_h^*$.

For each of $\tau_0 \in \mathcal{T}$ and each step k, define

$$M_n^{h(k)}(\tau_0) := 2\{L_n^h(\boldsymbol{\psi}_h^{(k)}(\tau_0), \tau^{(k)}(\tau_0)) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} \tag{39}$$

With a pre-determined maximum iteration K, define the modified EM test statistic by taking maximum of

$$EM_n^{h(K)} := \max\{M_n^{h(K)}(\tau_0) : \tau_0 \in \mathcal{T}\}. \tag{40}$$

If $H_0 : m = M_0$ holds true, each of $EM_n^{h(K)}$ will have the same asymptotic size. On the other hand, different $EM_n^{h(K)}$ will have different powers under alternative hypothesis depending on the true parameter values. To obtain ideal power, take the maximum of M_0 modified local EM test statistics:

$$EM_n^{(K)} := \max\{EM_n^{1(K)}, \ldots, EM_n^{M_0(K)}\}. \tag{41}$$

Proposition 4. *Under Assumption 1 and 4, a_n in equation (35) satisfies $a_n = O_p(1)$ and $\{0.5\}$ is in $\mathcal{T}$. Then under the null hypothesis $H_0 : M = M_0$, for any fixed finite K, as $n \to \infty$, $\{EM_n^{h(K)}\}_{h=1}^{M_0} \to_d \{(\hat{t}_\lambda^h)^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}^h \hat{t}_\lambda^h\}_{h=1}^{M_0}$, and $EM_n^{(K)} \to_d \max_h\{(\hat{t}_\lambda^h)^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}^h \hat{t}_\lambda^h\}$, where $\hat{t}_\lambda^h$ is defined as in equation (31) for $h = 1, \ldots, M_0$.*

The asymptotic distribution, $\left\{\max\{(\hat{t}_\lambda^1)^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^1 (\hat{t}_\lambda^1), \ldots, (\hat{t}_\lambda^{M_0})^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^{M_0} (\hat{t}_\lambda^{M_0})\}\right\}$, is not a standard distribution; therefore the distribution needs to be simulated. First I draw multivariate random vector $\tilde{S}_{\lambda,\eta} \sim N(0, \tilde{\mathcal{I}}_{\lambda,\eta})$. Then for each draw, I calculate $\hat{t}^h_\lambda$ defined by equation (31) for $h = 1, \ldots, M_0$. Recall $\hat{t}^h_\lambda$ is locally approximated by a cone given in (18). Then I take maximum across $\{(\hat{t}^1_\lambda)' \mathcal{I}^1_{\eta,\lambda}(\hat{t}^1_\lambda), \ldots, (\hat{t}^{M_0}_\lambda)' \mathcal{I}^{M_0}_{\eta,\lambda}(\hat{t}^{M_0}_\lambda)\}$ for each draw.

4 Simulation

In this section, I examine the performance of the modified EM test for model without conditioning variables based on finite sample. I use Monte Carlo simulation to test $H_0 : m = M_0$ against $H_1 : m = M_0 + 1$ for a finite mixture model of normal distribution for $m = 2$. The critical values for likelihood ratio test statistics are obtained by simulation.

4.1 Choice of penalty function

To apply the modified EM test on panel regression, I need to specify $\mathcal{T}$ and the penalty function. Following Chen and Li (2009) and KS12, I use $\mathcal{T} = \{0.5\}$, and set the penalty function to be in the form of

$$p_n((\sigma_j)^2; (\hat{\sigma}^j_0)^2) := -a_n\{(\hat{\sigma}^j_0)^2/(\sigma_j)^2 + \log((\sigma_j)^2/(\hat{\sigma}^j_0)^2) - 1\}, \tag{42}$$

where $(\hat{\sigma}^j_0)^2$ is the estimator of $(\sigma^j_0)^2$ from M_0-component model, $a_n = o_p(n^{1/4})$. $\hat{\sigma}^2_j$ is the parameter under the $(M_0 + 1)$-component model.

For regression model with no conditioning variables $\{x_{it}\}, i = 1, \ldots, N, t = 1, \ldots, T$,

$$a_n = \begin{cases} 0.25, \text{ if } M_0 = 1; \\ a_n(N, T, \omega(\boldsymbol{\vartheta}_{M_0}; M_0); M_0) \text{ if } M_0 = 2, 3, 4; \end{cases} .$$

The empirical a_n-function is obtained by running empirical regression based on simulated data. In order to obtain the empirical a_n function, I collected data from different sets of N,T and $\boldsymbol{\vartheta}_{M_0}$. The regression I used to obtain empirical a_n-function is defined by

$$\log(\frac{\hat{s}}{0.1 - \hat{s}}) = \varrho_1(M_0) + \varrho_2(M_0)\frac{1}{T} + \varrho_3(M_0)\frac{1}{N} + \varrho_4(M_0)\log(\frac{\tilde{a_n}}{1 - \tilde{a_n}}) + \varrho_5 \log(\frac{\omega(\boldsymbol{\vartheta}_{M_0}; M_0)}{1 - \omega(\boldsymbol{\vartheta}_{M_0}; M_0)}). \tag{43}$$

For each test, define N as the number of firms, T as the number of time periods, M_0 as the number of components in the null model, $\boldsymbol{\vartheta}_{M_0}$ as the true density parameters under the null hypothesis, and $\tilde{a_n}$ the a_n-value I arbitrarily assign to test the rejection probability. $\omega(\boldsymbol{\vartheta}_{M_0}; M_0)$ is the misclassification probability as defined in section 7 in Kasahara and Shimotsu (2015). Define $\hat{s}$ to be estimated nominal sizes at 5% significance level, which are the estimated rejection probabilities given the simulated 5% critical values. $\hat{\varrho}(M_0)$'s are estimated values of $\varrho(M_0)$'s based on the simulation for data with null hypothesis of $M_0 = 2, 3, 4$. To obtain the estimated values of $\varrho(M_0)$'s, I use different sets of parameters to obtain the nominal size $\hat{s}$ at 5%-significance level. The parameter sets under which I obtain the nominal size at 5%-significance level $\hat{s}$ is given by table 9. For example, when tesitng $H_0 : M_0 = 2$, I set the null model parameters $(N, T, \alpha, \mu, \sigma) \in \{100, 500\} \times \{2, 5, 10\} \times \{(0.5, 0.5), (0.2, 0.8\} \times \{(-1, 1), (-0.5, 0.5)\} \times \{(1, 1), (1.5, 0.75)\}$, and obtain the nominal size $\hat{s}$ at 5%-significance level using different a_n-values $\tilde{a_n} \in \{0.05, 0.1, 0.15, 0.2, 0.3, 0.4\}$ respectively. Therefore, I have $2 * 3 * 2 * 2 * 2 * 6 = 288$ observations of $\{\hat{s}, N, T, \omega(\boldsymbol{\vartheta}_2; 2), \tilde{a_n}\}$. Then I run the regression defined in (43) to obtain the estimated values $\hat{\rho}(2)$'s.

Then I run the regression specified as equation (43) and obtain the estimated $\hat{\varrho}(M_0)$'s for the coefficients $\varrho(M_0)$'s for $M_0 = 2, 3, 4$. From the estimated coefficient, set the desirable nominal size to be 5%. The data-driven a_n as a function of N,T and misclassification probability $\omega(\boldsymbol{\vartheta}_{M_0}; M_0)$ is defined as:

$$
\begin{aligned}
&a_n(N, T, \omega(\boldsymbol{\vartheta}_{M_0}; M_0); M_0) \\
&= 0.5 * \left(1 + \exp\left\{\frac{\hat{\rho}_1(M_0)}{\hat{\rho}_4(M_0)} + \frac{\hat{\rho}_2(M_0)}{\hat{\rho}_4(M_0)}\frac{1}{T} + \frac{\hat{\rho}_3(M_0)}{\hat{\rho}_4(M_0)}\frac{1}{N} + \frac{\hat{\rho}_5(M_0)}{\hat{\rho}_4(M_0)}\log(\frac{\omega(\boldsymbol{\vartheta}_{M_0}; M_0)}{1 - \omega(\boldsymbol{\vartheta}_{M_0}; M_0)}\right\}\right)^{-1}.
\end{aligned} \tag{44}
$$

For regression model with conditioning variables $\{x_{i,t}\}$ for $i = 1, \ldots, N, t = 1, \ldots, T$, $a_n = \begin{cases} 0.25, & \text{if } M_0 = 1; \\ 0.1245674, & \text{if } M_0 = 2; \\ 0.07366668, & \text{if } M_0 = 3; \\ 0.05529925, & \text{if } M_0 = 4; \\ 0.5, & \text{otherwise.} \end{cases}$

4.2 Simulation result

In this section, I examine the type I and type II errors of the test of 2 components against 3 components. Table 1 reports the type I errors of the modified EM test using normal distribution using the empirical penalty term. The data are generated under 2-component models as specified in the footnotes. In general, The modified EM test has correct sizes and good powers. The size of the test is closer to 5% when μ is $\mu = (-1, 1)$ compared with $\mu = (-0.5, 0.5)$. A larger distance between two distributions reduces the mis-classification probability. Another observation is that when comparing different mixing proportions, the modified EM test has a better size when the mixing proportions are equal across components at $\boldsymbol{\alpha} = (0.5, 0.5)$ than when they are unequal $\boldsymbol{\alpha} = (0.2, 0.8)$.

Table 2 shows the powers of likelihood ratio test with null hypothesis $H_0 : M_0 = 2$. The data are generated from a 3-type mixture model of the normal distribution as indicated in the table footnote. Similar to the performance of type I error, the power of the test is better when μ the distance between μ_j's are larger: comparing with $\boldsymbol{\mu} = (-1, 0, 1)$, the test has better power in the model $\boldsymbol{\mu} = (-1.5, 0, 1.5)$. The tests perform better when the distance between μ_1 and μ_2 is equal to that between μ_2 and μ_3. The tests have higher power when $\boldsymbol{\mu} = (-1.5, 0, 1.5)$ and $(-1, 0, 1)$ comparing to that when $\boldsymbol{\mu} = (-1, 0, 2)$ and $(-0.5, 0, 1.5)$. As for mixing probability, the test has better power when the mixture probability is equal, comparing the powers when $\boldsymbol{\alpha} = (1/3, 1/3, 1/3)$ with those when $\boldsymbol{\alpha} = (1/4, 1/2, 1/4)$.

From the simulation results, as T gets larger, type I and type II error are both lower. Similarly, as N gets larger, we can observe a slight improve type I and type II error as well, but not as substantial as that when T increases. Intuitively, when the mis-classification probability of the null model is low, the test has better power.

In the empirical application, the test usually has the null hypothesis with M_0 greater than 3. With more types in the mixture model, the data-driven penalty term is less precise because the mis-classification probability consists of more terms and therefore harder to calculate. There is potential risk of penalty term being too high or too low. However, for tests with null hypothesis $H_0 : M_0$ where $M_0 \geq 5$, it is hard to calculate the critical values by simulation. Instead, I use the bootstrap method to obtain the critical values. To show that the bootstrapped critical values are more robust when changing the value of the penalty term, I test the nominal size using both simulated critical values and bootstrapped critical values using large and small penalty terms. Table 3 shows the nominal sizes when using penalty terms are 10 times of the empirical penalty function and 4 shows the nominal sizes using penalty terms are 0.1 times of the empirical penalty function. Both the tables indicate that when the sample size is large, the mis-specified penalty term will be less relevant. When using the bootstrapped asymptotic distribution, the size of the test is less affected by the penalty term compared with the simulated asymptotic distribution. When testing $H_0 : M_0$ where M_0 is large, I use the bootstrapped critical values instead of simulated critical values.

Table 1: Sizes(in %) in modified EM size test of $H_0 : M_0 = 2$ against $H_A : M_0 = 3$ at 5% level

	N = 100			N = 500		
Model	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10
(A, C)	5.25	5.4	4.2	4.55	4.5	4.0
(A, D)	2.6	3.55	3.4	3.4	3.15	4.3
(B, C)	4.05	5.75	3.5	4.35	4.85	4.1
(B, D)	2.45	3.25	3.4	4.5	3.95	5.25

Use A, B to denote $(\alpha_1, \alpha_2) = (0.5, 0.5)$ and $(0.2, 0.8)$; use C, D to denote $(\mu_1, \mu_2) = (-1, 1)$ and $(-0.5, 0.5)$; and set the variance $(\sigma_1, \sigma_2) = (0.8, 1.2)$.

Table 2: Powers (in %) of modified EM test of $H_0 : M_0 = 2$ against $H_A : M_0 = 3$ at 5% level

α	A				B			
	N = 100		N = 500		N = 100		N = 500	
(μ, σ)	T=2	T=5	T=2	T=5	T=2	T=5	T=2	T=5
(C, G)	8.8	64.8	18.2	100	8.2	74.0	26.6	100
(C, H)	56.4	100	100	100	40.6	99.8	99.8	100
(C, I)	73.2	100	100	100	84.6	100	100	100
(D, G)	34.8	100	98.0	100	40.0	100	99.8	100
(D, H)	93.4	100	100	100	84.0	100	100	100
(D, I)	99.8	100	100	100	100	100	100	100
(E, G)	21.8	97.8	70.6	100	25.2	97.4	70.6	100
(E, H)	83.0	100	100	100	68.0	100	100	100
(E, I)	87.6	100	100	100	91.4	100	100	100
(F, G)	4.4	15.0	6.6	59.8	5.6	13.6	9.6	63.4
(F, H)	45.8	99.6	100	100	32.4	97.8	98.6	100
(F, I)	13.2	81.6	47.8	100	12.6	81.4	48.4	100

A and B refers to $(\alpha_1, \alpha_2, \alpha_3) = (1/3, 1/3, 1/3)$ and $(1/4, 1/2, 1/4)$, respectively; C, D, E, F refers to $(\mu_1, \mu_2, \mu_3) = (-1, 0, 1), (-1.5, 0, 1.5), (-1, 0, 2), (-0.5, 0, 1.5)$; G, H, I refers to $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1), (0.6, 1.2, 0.6), (0.6, 0.6, 1.2)$.

Table 3: Sensitivity of modified EM size test of $H_0 : M_0 = 2$ against $H_A : M_0 = 3$ (10 x penalty)

	Asymptotic				Bootstrap			
	N = 100		N = 500		N = 100		N = 500	
	T = 2	T = 5	T = 2	T = 5	T = 2	T = 5	T = 2	T = 5
(A, C)	1.2	3.0	3.2	2.2	2.35	4.80	4.85	4.90
(A, D)	1.0	2.6	3.0	2.6	1.35	2.70	1.30	2.35
(B, C)	0.6	1.5	1.4	2.8	1.85	2.95	1.75	3.85
(B, D)	1.0	1.0	1.4	2.4	1.15	3.2	3.05	4.6

Use A, B to denote $(\alpha_1, \alpha_2) = (0.5, 0.5)$ and $(0.2, 0.8)$; use C, D to denote $(\mu_1, \mu_2) = (-1, 1)$ and $(-0.5, 0.5)$; and set the variance $(\sigma_1, \sigma_2) = (0.8, 1.2)$.

Table 4: Sensitivity of modified EM size test of $H_0 : M_0 = 2$ against $H_A : M_0 = 3$ (0.1 x penalty)

	Asymptotic				Bootstrap			
	N = 100		N = 500		N = 100		N = 500	
	T = 2	T = 5	T = 2	T = 5	T = 2	T = 5	T = 2	T = 5
(A, C)	8.2	6.2	6.8	3.2	10.5	8.45	8.05	5.75
(A, D)	6.0	6.0	5.6	4.2	7.1	5.4	3.30	3.35
(B, C)	5.6	3.8	4.4	3.2	8.8	5.7	5.6	4.65
(B, D)	5.6	3.8	4.4	3.2	8.6	7.6	7.9	6.2

Use A, B to denote $(\alpha_1, \alpha_2) = (0.5, 0.5)$ and $(0.2, 0.8)$; use C, D to denote $(\mu_1, \mu_2) = (-1, 1)$ and $(-0.5, 0.5)$; and set the variance $(\sigma_1, \sigma_2) = (0.8, 1.2)$.

5 Applications

The test of the number of components in finite mixture models can be applied to numerous potential fields, such as estimating demands in Dubé et al. (2010) and estimating production functions in Kasahara et al. (2015). When estimating production functions, economists face the problem of endogeneity. This is because that the input decisions are often related to the unobserved shocks. Past literature addresses the problem using two major methods. One stream of literature focus on dynamic panel approach (Chamberlain (1984); Arellano and Bover (1995); Blundell and Bond (1998); Blundell and Bond (2000)). The other stream of literature uses structural models to identify production function, such as Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg et al. (2006). One of the key assumptions in structural models is that the input decisions are optimal given the idiosyncratic shocks. Ackerberg et al. (2006) discuss the potential collinearity problem from structural identification strategies.Gandhi et al. (2013)(GNR thereafter) extend the method by identifying the production function elasticity of the flexible inputs using transformed first order condition. GNR's model allows for firms to be heterogeneous in terms of input-specific elasticity. The GNR paper finds evidence that there exists heterogeneity beyond Hicks-neutral production factor. Kasahara et al. (2015) extend the paper by classifying firms into finite classes. They found empirical evidence that production technologies are heterogeneous in terms of their input elasticity. The paper states that as the type of firms increases, the estimated coefficients are substantially different across different types of firms.

Except for the result of Gandhi et al. (2013) and Kasahara et al. (2015), few formal identification results for production function estimation in the past literature is available. This test of the component number in finite mixture normal panel regression is an important contribution to the literature. As random coefficient models for production function become increasingly popular in empirical analysis (e.g.,Mairesse and Griliches (1988);Van Biesebroeck (2003); Doraszelski and Jaumandreu (2014)), the identification result and test of a number of components on production function with unobserved heterogeneity become more important. The test of the component number can be used as a tool to determine the number of production technologies.

I extend the result of Kasahara et al. (2015) by testing the number of types of input elasticity. I apply the likelihood ratio test of number of components to the data from machine industry among Japanese publicly traded firms and Chilean producer data. I find that for panel data with longer more time periods, the observations are categorized into more types, providing a concrete evidence that the production functions are heterogeneous between firms.

5.1 Production Function and First Order Condition

Assume that the panel data consist the input and output data of firms $i = 1, \ldots, N$ over periods $t = 1, \ldots, T$. $(Y_{it}, L_{it}, K_{it}, M_{it})$ denote for output, labor, capital and intermediate good respectively. $(m_{it}, k_{it}, l_{it}, y_{it})$ are the logarithms of intermediate good, capital, labor and output. Econometricians observe $\{Y_{it}, M_{it}, L_{it}, K_{it}\}_{t=1}^{T}$ for each firm.

Now consider the case that firms are different in production technology. I use a finite mixture specification to capture permanent unobserved heterogeneity in firm's production technology. Define the latent random variable $D_i \in \{1, 2, \ldots, m\}$ to represent the type of firm i. If $D_i = j$, then firm i is type j. Assume there are m discrete types, each occurring with probability of α_j. The production function for type j is Cobb-Douglas with type specific coefficients $\{\beta_k^j, \beta_m^j, \beta_l^j, \sigma^j\}$. Define the Cobb-Douglas production function for type j as:

$$F^j(M, K, L) = M^{\beta_m^j} K^{\beta_k^j} L^{\beta_l^j},$$

where $\beta_m^j, \beta_k^j, \beta_l^j$ denote the production function's parameter with respect to intermediate good, capital and labor for type j.

In order to identify the intermediate good elasticity of the production function, I introduce the following modified assumptions on production function 5,6 and 7 proposed by Kasahara and Shimotsu (2015).

Assumption 5. *(a) Each firm belongs to one of m types, and the probability of being type j, given by $\alpha_j = P(D_i = j)$ is known and $\sum_{j=1}^{m} \alpha_j = 1$. (b) For the j^{th} type of production technology at time t, the output expressed in terms of input is $Y_{it} = \exp\{\sigma^{D_i} \epsilon_{it}\} F_t^{D_i}(K_{it}, L_{it}, M_{it})$, where $\epsilon_{it} \sim N(0, 1)$ are i.i.d across i's and t's. σ^{D_i} represents the variance of type-specific shock.*

Assumption 6. *M_{it}'s are chosen at time t by maximizing the expected profit conditional on information at* *time t. In mathematical expression,*

$$M_{it} = \arg\max_{M} P_{Y,t} E[\exp\{\sigma^{D_i}\epsilon_{it}\}] F_t^{D_i}(M, K_{it}, L) - P_{M,t} M. \tag{45}$$

Assumption 7. *(a) A firm is a price taker for intermediate good inputs, $P_{M,t}$ are common across firms. (b) $(P_{M,t}, P_{Y,t})$ are observed by firms at the beginning of the time period.*

In Assumption 5, as indicated by the subscript t in $F_t^j(\cdot)$, each firm's production function belongs to one of the m types. The output vary across time periods due to type-specific aggregate shocks $\sigma^{D_i}\epsilon_{it}$. The restriction $\sum_{j=1}^{m} \alpha_j = 1$ is necessary for identification. Assumption 6 assumes that M_{it} are chosen to maximize the current period profit. Assumption 7 states that the firms observe the input and output prices when making decision on M_{it}.

Given the above assumptions 5,6 and 7, the firm choose intermediate good M_{it} flexibly each time period without observing ϵ_{it}. The first order condition with respect to M_{it} gives $P_{Y,t}E(\exp\{\sigma^{D_i}\epsilon_{it}\})\frac{\partial F_i^{D_i}}{\partial M}(K_{it}, L_{it}, M_{it}) = P_{M,t}$. With Cobb-Douglas production function, the first order condition can be rewritten as $P_{Y,t}\beta_m^{D_i}\frac{F_i^{D_i}(X_{it})}{M}E[\exp\{\sigma^{D_i}\epsilon_{it}\}] = P_{M,t}$. Rewrite the first order condition as $\beta_m^{D_i}E[\exp\{\sigma^{D_i}\epsilon_{it}\}]) = \frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$. Define the revenue share of the intermediate good of firm i at time t as $S_{it} = \frac{P_{M,t}M_{it}}{P_{Y,t}Y_{it}}$, thus $S_{it} = \beta_m^{D_i}E[\exp\{\sigma^{D_i}\epsilon_{it}\}]$. Define s_{it} to be the logarithm of S_{it}. By the first order condition, s_{it} can be written as

$$s_{it} = \log \beta_m^{D_i} + \frac{1}{2}(\sigma^{D_i})^2 - \sigma^{D_i}\epsilon_{it}. \tag{46}$$

Each of the firm can be viewed as a random sample from the m types, and the likelihood of $s_{i1}, ..., s_{it}$ is written as $P(\{s_{it}\}_{t=1}^{T}) = \sum_{j=1}^{m} \alpha_j \prod_{t=1}^{T} \frac{1}{\sigma^j}\phi(\frac{\log \beta_m^j + \frac{1}{2}(\sigma^j)^2 - s_{it}}{\sigma^j})$, where $\phi(\cdot)$ is the standard normal probability density. Recall that $\epsilon_{it} \sim N(0,1)$ i.i.d across i and t. Let $\mu^j = \log \beta_m^j + \frac{1}{2}(\sigma^j)^2$. Collect the observed data as $\boldsymbol{\omega}_i = \{s_{it}\}_{t=1}^{T}$. Then rewrite the likelihood density as a finite normal mixture panel regression model density similar to equation (1):

$$f_m(\{s_{it}\}_{t=1}^{T}; \boldsymbol{\vartheta}_m) = \sum_{j=1}^{m} \alpha_j \prod_{t=1}^{T} \frac{1}{\sigma^j}\phi\left(\frac{s_{it} - \mu^j}{\sigma^j}\right). \tag{47}$$

Define a type-specific parameter to be $\theta^j = (\mu^j, \sigma^j)$. $\boldsymbol{\vartheta}_m = (\theta^1, \ldots, \theta^m, \alpha_1, \ldots, \alpha_{m-1})$. With the above parametric assumption, I identify type-specific intermediate good parameters β_m^j and mixing probability α_j given the number of types. Collect the parameters of each type and the mixing probability, $\theta = (\theta^1, \ldots, \theta^m)$ and $\alpha = (\alpha_1, \ldots, \alpha_m)$. The maximum likelihood estimator is defined as

$$\hat{\boldsymbol{\vartheta}}_m = \arg\max_{\boldsymbol{\vartheta}} \sum_{i=1}^{N} \log f_m(\{s_{it}\}_{t=1}^{T}; \boldsymbol{\vartheta}_m). \tag{48}$$

In practice, I use the modified EM algorithm to test the null hypothesis of $H_0 : M_0$ for $M_0 = 1, \ldots, 5$ as introduced in section 6.

5.2 Empirical result

The finite mixture panel regression model provides an approach to identify the underlying heterogeneity in production functions. With the Cobb-Douglas assumption on the production functions, the revenue share of intermediate material can be used to estimate the elasticity of the intermediate good. I apply the modified EM test on two producer data sets to determine the number of types of intermediate good elasticity in a certain industry. The first data set consists of production data from Japanese publicly traded manufacturing firms from 1980 to 2007, and the second data set is production data from Chilean producers from 1979 to 1996.

I apply the tests to two similar industries from the two data sets respectively, the machine industries from Japanese data set and Chilean data set. Table 5 presents the summary statistics in Japanese machine industry and 6 presents the summary statistics for Chilean machinery industry. The revenue share of intermediate material is close in these two countries.

Table 5: Descriptive data for Japan producer: Machine industry

Statistic	N	Mean	St. Dev.	Min	Max
$\log \frac{P_{M,t} M_{it}}{P_{Y,t} Y_{it}}$	5,446	−0.778	0.491	−4.452	−0.091
$\log Y_{it}$	5,446	17.092	1.280	13.688	21.319
$\log L_{it}$	5,446	6.645	1.116	2.890	10.767
$\log K_{it}$	5,446	15.916	1.315	12.382	20.340
$\log M_{it}$	5,446	16.314	1.372	11.860	20.899

The data of Japanese producer is compiled by the Development Bank of Japan (DBJ). This dataset contains detailed corporate balance sheet/income statement data from 1980 to 2008 for the firms listed on the Tokyo Stock Exchange.

The descriptive statistics of the intermediate good shares, the log output $\log Y_{it}$ and the log inputs $\log L_{it}, \log K_{it}$ and $\log M_{it}$ from Japanese machine industry are reported in table 5, and the descriptive statistics of Chilean machine industry are reported in table 6. There exist large variations in the revenue shares of intermediate good value in both industries.

To determine the number of components, I test the null hypothesis $H_0 : M_0$ v.s. $H_1 = M_0 + 1$ sequentially for $M_0 = 1, \ldots, 5$. If I fail to reject the null hypothesis at certain $M_0 = m$, then I can conclude that there are m types of intermediate good elasticity. I do the test using cross-sectional data, and panel data with panel length $T = 2, \ldots, 5$. I examine the test results for different panel length to show that when increasing the panel length, I cannot reject the null hypothesis $H_0 : M_0$

Table 6: Descriptive data for Chilean producer : Machinery industry, except electrical

Statistic	N	Mean	St. Dev.	Min	Max
$\log \frac{P_{M,t} M_{it}}{P_{Y,t} Y_{it}}$	2,410	−0.778	0.438	−3.931	1.241
$\log Y_{it}$	2,410	9.565	1.798	4.399	15.663
$\log L_{it}$	2,410	3.601	0.957	1.060	7.874
$\log K_{it}$	2,410	7.762	1.608	1.715	13.310
$\log M_{it}$	2,410	8.787	1.877	3.775	15.233

This dataset contains detailed corporate balance sheet/income statement data from 1979 to 1996 for the Chilean producer.

for a larger M_0. This can be explained by that when observing data from more time periods, the firm-specific input elasticity patterns are easier to identify, and the production functions can be categorized into more types.

The empirical likelihood ratio test results from Japanese machine industry are reported in table 7. For cross-sectional model, when $T = 1$, the test of one-component model against two-component model is rejected at 1% significance level. the test of $H_0 : m = 3$ against $H_1 : m = 4$ is rejected at 10% significance level. I cannot reject the null hypothesis of $m = 4$ against the alternative of $m = 5$. This result shows that the observations can be classified into four types when using cross-sectional data. For panel length $T \geq 2$, $H_0 : m = 5$ against $H_1 : m = 6$ is rejected at 1% significance level. The simulated critical values on $10\%, 5\%, 1\%$ significance level for Japanese machinery industry corresponding to the results in table 7 are reported in table 10 in the appendix.

The test results of Chilean machine industry are reported in table 8, with the critical values reported in table 11. When $T = 1$, I cannot reject the null hypothesis of $H_0 : m = 2$. When $T = 2$, I cannot reject null hypothesis of $H_0 : m = 5$. For $T \geq 3$, I reject the null hypothesis of $H_0 : m = 5$. The results are similar to those of Japanese machine industry.

The empirical results show that there exists heterogeneity across firm's production technology. The unobserved heterogeneity need to be explained by more than 5 types in the finite mixture model when econometricians observe data from more than 3 time periods.

Table 7: Estimated Likelihood Ratio for Japanese producer in Machine industry

Time	$M_0 = 1$	$M_0 = 2$	$M_0 = 3$	$M_0 = 4$	$M_0 = 5$
1	93.309***	19.851***	6.849*	0.163	-
2	297.487***	111.438***	90.417***	37.875***	32.410***
3	524.024***	205.256***	121.135***	78.498***	59.610***
4	732.667***	320.623***	159.470***	124.478***	89.004***
5	934.599***	419.310***	214.137***	156.153***	126.073***

The estimation is based on Machinery industry, with null model of $M_0 = 1, 2, 3, 4, 5$, respectively. For $T = 1$, I use the data from the latest year 2008. For $T = 2$, I use the data from 2007-2008. For $T = 3$, I use the data from 2006-2008. For $T = 4$, I use the data from 2005-2008. For $T = 5$, I use the data from 2004-2008. * indicates the result is significant at 10% level. ** indicates the result is significant at 5% level *** indicates the result is significant at 1% level

Table 8: Estimated Likelihood ratio for Chilean producer in Machinery industry, except electrical

Time	$M_0 = 1$	$M_0 = 2$	$M_0 = 3$	$M_0 = 4$	$M_0 = 5$
1	20.692***	1.155	-	-	-
2	83.806***	45.149***	18.519***	7.565*	2.932
3	122.470***	54.249***	22.266***	14.706***	15.632***
4	175.399***	109.734***	20.438***	10.923***	9.902***
5	230.308***	128.093***	23.414***	16.599***	14.234***

The estimation is based on Machinery industry, with null model of $M_0 = 1, 2, 3, 4, 5$, respectively. For $T = 1$, I use the data from the latest year 2008. For $T = 2$, I use the data from 2007-2008. For $T = 3$, I use the data from 2006-2008. For $T = 4$, I use the data from 2005-2008. For $T = 5$, I use the data from 2004-2008. * indicates the result is significant at 10% level. ** indicates the result is significant at 5% level *** indicates the result is significant at 1% level.

However, this can be a special case in machinery industries since machinery industries contain many products that are heterogeneous. To extend the result further, I apply the test to industries with sufficient amount of observations from Japan and Chile respectively. The results for Japanese industries using cross-sectional data are reported in table 14, and the results using panel data of length $T = 2$ and $T = 3$ are reported in table 15 and 16. For Japanese industries, when using cross-sectional data, many industries are concluded to have homogeneous production functions. When using the $T = 2$ panel data, most industries except for plastic industry and paper industry are concluded to have more than 4 types. When using the $T = 3$ panel data, some industries are concluded to have more than 5 types. The results show all industries have more than 3 types of intermediate good elasticity when using panel data of $T = 3$.

The results for Chilean industries using cross-sectional data are reported in table 17, and the results using panel data of length $T = 2$ and $T = 3$ are reported in table 18 and 19. The results show that for panel length $T = 2$, the null model of $m = 3$ is rejected for most of the industries, and for panel length $T = 3$, the null model of 5 components is rejected for most industries. Except for some industries like the paper products with homogeneous products, the production technology can be classified into more than 5 types.

The result suggests that with Cobb-Douglas production function, the intermediate good elasticity is heterogeneous. When observing panel data with more than 3 time periods, a 5-component mixture model is not sufficient to capture the unobserved heterogeneity. The result shows that past papers have overlooked the unobserved firm specific production technologies by assuming homogeneous production functions. As the number of observations increases, the estimated number of types increases. The too many components can be a result of the mis-specification of mixture model or mis-specification of the production functions. The unexplained heterogeneity under finite mixture model may be explained by infinite mixture models. It can also be the case that Cobb-Douglas is a mis-specified production function model. GNR has proposed a method to identify the input elasticity of general production functions non-parametrically by regressing the revenue shares of intermediate good against inputs. To extend the result, I plan to apply the test of components of the finite normal mixture panel regression model to the identification of input elasticity of general production functions.

6 Conclusion

This paper focuses on testing the number of components in the finite normal mixture panel regression model. The theoretical analysis extends the work of Kasahara and Shimotsu (2012). I show that unlike the finite mixture normal regression model with cross-sectional data, the finite mixture normal panel regression model has a positive definite Fisher Information matrix under the reparameterization. I can approximate the likelihood ratio using a quadratic expansion of squares and cross-products of the reparameterized parameters.

I obtained the data-driven penalty formula via computational experiments. To show that the penalty formula gives the modified EM test correct Type I errors and small Type II error, I run simulations of the modified EM test with the data-driven penalty formula. The results of the simulations are reported in section 6. I use R as the computational tool(R Core Team (2013)) and develop an R package normalRegPanelMix(Hao (2017)) that contains the modified EM test module and asymptotic distribution simulation module as in section 5, the experiments of simulations as in section 6 and the empirical experiments as in section 7.

As an empirical application, the likelihood ratio test of number of components can be used to determine the number of classes of unobserved heterogeneous productivity shocks. I applied the test of components of the finite normal mixture panel regression to plant level production data from Japan and Chile in various industries, and find strong evidence of heterogeneous production functions beyond the Hick-neutral factors under the Cobb-Douglas assumption. As an extension of the result, I plan to test the type of flexible input elasticity on the general form production functions as discussed in Gandhi et al. (2013). The extension will enrich the results on heterogeneous production technologies.

References

[1] Ackerberg, D., Caves, K., and Frazer, G. (2006). Structural identification of production functions.

[2] Alexandrovich, G. (2014). A note on the article 'inference for multivariate normal mixtures' by j. chen and x. tan. *Journal of Multivariate Analysis*, 129:245–248.

[3] Andrews, D. (1999). Estimation When a Parameter is on a Boundary. *Econometrica*, 67(6):1341–1383.

[4] Andrews, D. W. (1994). Empirical process methods in econometrics. In *Handbook of Econometrics*, volume 4, pages 2247–2294. Elsevier.

[5] Andrews, R. L. and Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, 20(4):315–321.

[6] Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51.

[7] Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

[8] Blundell, R. and Bond, S. (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19:321–340.

[9] Brännäs, K. and Rosenqvist, G. (1994). Semiparametric estimation of heterogeneous count data models. *European Journal of Operational Research*, 76(2):247–258.

[10] Cameron, S. V. and Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males. *Journal of Political Economy*, 106(2):262–333.

[11] Chamberlain, G. (1984). Panel data. In *Handbook of econometrics*, volume 2, pages 1247–1318.

[12] Chen, J. and Li, P. (2009). Hypothesis test for normal mixture models: The EM approach. *Annals of Statistics*, 37(5A):2523–2542.

[13] Chen, J., Tan, X., and Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statistica Sinica*.

[14] Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics*, 12(3):313–336.

[15] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

[16] Doraszelski, U. and Jaumandreu, J. (2014). Measuring the bias of technological change.

[17] Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, 41(3):417–445.

[18] Gandhi, A., Navarro, S., and Rivers, D. (2013). On the identification of production functions: How heterogeneous is productivity?

[19] Hao, J. (2017). *normalRegPanelMix: Finite Mixture Model with Normal Panel Data*. R package version 1.0.

[20] Hartigan, J. (1985). Failure of log-likelihood ratio test. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 807–810. University of California Press 2. Berkeley.

[21] Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320.

[22] Kamakura, W. and Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, 26(4):379–390.

[23] Kasahara, H., Schrimpf, P., and Suzuki, M. (2015). Identification and estimation of production function with unobserved heterogeneity.

[24] Kasahara, H. and Shimotsu, K. (2012). Testing the number of components in finite mixture models.

[25] Kasahara, H. and Shimotsu, K. (2015). Testing the number of components in normal mixture regression models. *Journal of the American Statistical Association*, 110(512):1632–1645.

[26] Kasahara, H. and Shimotsu, K. (2019). Testing the Order of Multivariate Normal Mixture Models.

[27] Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):473–522.

[28] Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(70):317–341.

[29] Lindsay, B. G. and Lesperance, M. L. (1995). A review of semiparametric mixture models. *Journal of Statistical Planning and Inference*, 47(1-2):29–39.

[30] Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics*, 31(3):807–832.

[31] Mairesse, J. and Griliches, Z. (1988). Heterogeneity in panel data: are there stable production functions?

[32] McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*.

[33] Melnykov, V. and Maitra, R. (2010). Finite mixture models and model-based clustering. *Statistics Survey*, 4:80–116.

[34] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier.

[35] Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263.

[36] Pollard (1990). Empirical processes: theory and applications.

[37] R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[38] Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). Likelihood-Based Inference with Singular Information Matrix. *Bernoulli*, 6(2):243.

[39] Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley.

[40] Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *The Review of Economic Studies*.

A Proofs of propositions

Proof of Proposition 1. As shown by Alexandrovich (2014)(p.248), $p_n(\boldsymbol{\vartheta}_M)$ satisfies Assumptions C1 - C3 of Chen and Li (2009) under the stated condition on a_n. Therefore, the stated result follows from Theorem 1 of Chen and Li (2009) and Corollary 3 of Alexandrovich (2014). □

Proof of Proposition 2. The proof is similar to that of Proposition 4 in Kasahara et al. (2015). Let $\boldsymbol{t}(\boldsymbol{\eta}) := \boldsymbol{\eta} - \boldsymbol{\eta}^*$ and Recall that $\boldsymbol{t}(\boldsymbol{\lambda}, \alpha) = (\boldsymbol{t}(\boldsymbol{\eta})^\top, \boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha)^\top)^\top$ as in (19). Define

$$\boldsymbol{G}_n := \begin{pmatrix} \boldsymbol{G}_{\boldsymbol{\eta}n} \\ \boldsymbol{G}_{\boldsymbol{\lambda}n} \end{pmatrix}, \quad \begin{matrix} \boldsymbol{G}_{\boldsymbol{\eta},\boldsymbol{\lambda}n} = \boldsymbol{G}_{\boldsymbol{\lambda}n} - \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\eta}} \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}}^{-1} \boldsymbol{G}_{\boldsymbol{\eta}n}, \quad \boldsymbol{Z}_{\boldsymbol{\eta},\boldsymbol{\lambda}n} = \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}}^{-1} \boldsymbol{G}_{\boldsymbol{\eta},\boldsymbol{\lambda}n}, \\ \boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}} = \boldsymbol{t}_{\boldsymbol{\eta}n} - \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}} \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}\boldsymbol{\lambda}}^{-1} \boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha). \end{matrix}$$

Write

$$\begin{aligned} LR_n(\epsilon_1) &= \max_{\alpha \in [\epsilon_1, 1-\epsilon_1]} 2\{(L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_n(\hat{\boldsymbol{\psi}}^*, \alpha)) - (L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0) - L_n(\hat{\boldsymbol{\psi}}^*, \alpha))\}, \\ &= \max_{\alpha \in [\epsilon_1, 1-\epsilon_1]} 2\{(L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_n(\hat{\boldsymbol{\psi}}^*, \alpha)) - (L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0) - L_{0,n}(\boldsymbol{\gamma}_0^*, \boldsymbol{\theta}_0^*))\}. \end{aligned}$$

Apply Lemma 1 and 3 to the terms and

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{A_{n\epsilon}}} |2(L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_n(\hat{\boldsymbol{\psi}}^*, \alpha)) - B_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) - C_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}})| = o_{p\epsilon}(1), \tag{49}$$

where

$$\begin{aligned} B_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) &= 2\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}^\top \boldsymbol{G}_{\boldsymbol{\eta}n} - \boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}} \\ C_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) &= 2\boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha)^\top \boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\eta}n} - \boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}} \boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha) \\ &= \boldsymbol{Z}_{\boldsymbol{\eta},\boldsymbol{\lambda}n}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}} \boldsymbol{Z}_{\boldsymbol{\eta},\boldsymbol{\lambda}n} - (\boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha) - \boldsymbol{Z}_{\boldsymbol{\eta},\boldsymbol{\lambda}n})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\eta}} (\boldsymbol{t}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \alpha) - \boldsymbol{Z}_{\boldsymbol{\eta},\boldsymbol{\lambda}n}). \end{aligned} \tag{50}$$

Observe that $L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0) - L_{0,n}(\boldsymbol{\gamma}_0^*, \boldsymbol{\theta}_0^*) = \max_{\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}} B_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) + o_p(1)$ from applying Lemma 3 to $L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0)$. Note that the possible values of both $\boldsymbol{t}_{\boldsymbol{\eta}}$ and $\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}$ approaches $\mathbb{R}^{q+2}$. Therefore, with $p_n(\boldsymbol{\vartheta}_{\boldsymbol{2}}) = o_p(1)$, we can write equation (49), we obtain

$$2\{L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_{0,n}(\hat{\gamma}_0, \hat{\boldsymbol{\theta}}_0)\} = C_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) + o_p(1). \tag{51}$$

We construct a parameter space $\tilde{\Lambda}_{\boldsymbol{\lambda}}$ that is locally equal to $\Lambda_{\boldsymbol{\lambda}}$ defined in (18). First, Assumption 2 of Andrews (1999) holds trivially for $C_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}})$. Second, Assumption 3 of Andrews (1999) holds with $B_T = n^{1/2}$ because $\boldsymbol{G}_{\boldsymbol{\eta},\boldsymbol{\lambda}n} \to_d \boldsymbol{G}_{\boldsymbol{\eta},\boldsymbol{\lambda}} \sim N(0, \boldsymbol{\mathcal{I}}_{\boldsymbol{\eta},\boldsymbol{\lambda}})$ and $\boldsymbol{\mathcal{I}}$ is nonsingular. Assumption 4 of Andrews (1999) holds from the same argument. Assumption 5 of Andrews (1999) follows from Assumption 5 of Andrews (1999) because $\Lambda_{\boldsymbol{\lambda}}$ locally equal to the cone $\tilde{\Lambda}_{\boldsymbol{\lambda}}$. Therefore, it follows from Theorem 3(c) of Andrews (1999) that $C_n(\sqrt{n}\boldsymbol{t}_{\boldsymbol{\eta},\boldsymbol{\lambda}}) \to_d (\hat{\boldsymbol{t}}_\lambda)^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta} \hat{\boldsymbol{t}}_\lambda$. □

Proof of Proposition 3. For $m = 1, \dots, M_0$, let $\mathcal{N}$, let $\mathcal{N}_m^* \subset \Theta_{\boldsymbol{\vartheta}_{M_0+1}}(\epsilon_1)$ be a sufficiently small closed neighbourhood of Υ_{1m}^*, such that $\alpha_m, \alpha_{m+1} > 0$ hold and $\Upsilon_{1k}^* \not\subset \mathcal{N}_m^*$ if $k \neq m$. Consider the one-to-one reparamterization as (23) for testing the null hypothesis $H_{0,1m}$:

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\nu} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\theta}_m - \boldsymbol{\theta}_{m+1} \\ \tau\boldsymbol{\theta}_m + (1-\tau)\boldsymbol{\theta}_{m+1} \end{pmatrix} \text{ so that } \begin{pmatrix} \boldsymbol{\theta}_m \\ \boldsymbol{\theta}_{m+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\nu} + (1-\tau)\boldsymbol{\lambda} \\ \boldsymbol{\nu} - \tau\boldsymbol{\lambda} \end{pmatrix}, \tag{52}$$

where $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$ are both $(q+2) \times 1$ reparameterized parameter vectors.

Collect the reparameterized parameters except τ into $\boldsymbol{\psi}_{m,\tau}$ as defined in (24), where $\boldsymbol{\psi}_{m,\tau}^*$ represent the true value of $\boldsymbol{\psi}_{m,\tau}^*$. Define the log-likelihood under the reparameterized parameters as

$$f_{M_0+1}^m(\boldsymbol{w}; \boldsymbol{\psi}^m, \tau) = \pi_m g^m(\boldsymbol{w}, \boldsymbol{\psi}_{m,\tau}, \tau) + \sum_{j=1}^{m-1} \pi_j f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}_j) + \sum_{j=m}^{M_0} \pi_j f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\theta}_{j+1}), \tag{53}$$

where $g^m(\boldsymbol{w}, \boldsymbol{\psi}_{m,\tau}, \tau)$ is defined similar to 8:

$$g^m(\boldsymbol{w}, \boldsymbol{\psi}_{m,\tau}, \tau) = \tau f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\nu} + (1-\tau)\boldsymbol{\lambda}) + (1-\tau) f(\boldsymbol{w}; \boldsymbol{\gamma}, \boldsymbol{\nu} - \tau\boldsymbol{\lambda}). \tag{54}$$

We apply Lemma 4 to (54) with replacing α with τ. Define $L_n^m(\boldsymbol{\psi}_{m,\tau}, \tau) := \sum_{i=1}^N \log g^m(\boldsymbol{W}_i; , \boldsymbol{\psi}_{m,\tau}, \tau)$. Then $L_n^M(\boldsymbol{\psi}_{m,\tau}, \tau) - L_n^M(\boldsymbol{\psi}_{m,\tau}^*, \tau)$ admits the same expansion as $L_n(\hat{\boldsymbol{\psi}}_\alpha, \alpha) - L_n(\hat{\boldsymbol{\psi}}_\alpha^*, \alpha))$ in (49) by replacing $(\boldsymbol{t}(\boldsymbol{\psi}_\alpha, \alpha), \boldsymbol{s}(\boldsymbol{W}), \boldsymbol{\mathcal{I}})$ with $(\boldsymbol{t}_m(\boldsymbol{\psi}_\alpha, \alpha), \boldsymbol{s}^m(\boldsymbol{W}), \boldsymbol{\mathcal{I}}^m)$. $(\boldsymbol{s}^m(\boldsymbol{w}), \boldsymbol{\mathcal{I}}^m)$ is defined in a same way to $(\boldsymbol{s}(\boldsymbol{w}), \boldsymbol{\mathcal{I}})$ by replacing $(\boldsymbol{s}_\eta, \boldsymbol{s}_\lambda)$ with $(\tilde{\boldsymbol{s}}_\eta, \boldsymbol{s}_\lambda^m)$.

Define the local penalized MLE of $\boldsymbol{\psi}^m$ by:

$$\hat{\boldsymbol{\psi}}_m := \arg\max_{\boldsymbol{\psi}_m \in \mathcal{N}_m^*} PL_{m,n}(\boldsymbol{\psi}_m, \tau); \quad \text{where} PL_{m,n}(\boldsymbol{\psi}^m, \tau) := L_n^M(\boldsymbol{\psi}_{m,\tau}, \tau) + p_n(\boldsymbol{\psi}^m). \tag{55}$$

Because $\boldsymbol{\psi}_m^*$ is the only parameter value in $\mathcal{N}_m^*$ that generates the true density, $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_m^* = o_p(1)$ follows Proposition 2. For $\epsilon_\tau \in (0, 1/2)$, define the LRTS for testing $H_{0,1m}$ as $LR_{n,1m} = \max_{\tau \in [\epsilon_\tau, 1-\epsilon_\tau]} 2(L_n^m(\boldsymbol{\psi}_{m,\tau}, \tau) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0}))$. Observe that $p_n(\hat{\boldsymbol{\psi}}_m) = o_p(1)$ because $a_n = o_p(1)$. By applying the proof of Proposition 2, we have that

$$(LR_{n,11}, \dots, LR_{n,1M_0})^\top \to_d (\hat{\boldsymbol{t}}_\lambda^1)^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^1 (\hat{\boldsymbol{t}}_\lambda^1), \dots, (\hat{\boldsymbol{t}}_\lambda^{M_0})^\top \boldsymbol{\mathcal{I}}_{\eta,\lambda}^{M_0} (\hat{\boldsymbol{t}}_\lambda^{M_0}))^\top. \tag{56}$$

□

Proof of Proposition 4. Let $\omega_{n,m}$ denote the sample counterpart of $(\hat{t}_\lambda^m)^\top \boldsymbol{\mathcal{I}}_{\lambda,\eta}^m \hat{t}_\lambda^m$ in Proposition 3 such that the LRTS satisfies $2\{L_n^m(\boldsymbol{\psi}_{m,\tau}, \tau) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} = \omega_{n,m} + o_p(1)$, where $\boldsymbol{\psi}_{m,\tau}$ is the local penalized MLE as defined (55).

First we show that $M_h^h(1) = \omega_{n,m} + o_p(1)$. For $\tau \in (0,1)$, define $\boldsymbol{\vartheta}_{M_0+1}^{m*}(\tau) \in \Sigma_{1m}^* : \alpha_m/(\alpha_m +$

$\alpha_{m+1} = \tau$), which gives the true density. $\boldsymbol{\vartheta}^{m*}_{M_0+1}(\tau_0)$ is the only value in $\Theta_{\boldsymbol{\vartheta}_{M_0+1}}$ that yields the true density, if $\boldsymbol{\xi}, \in \Xi^*_m$ and $\alpha_m/(\alpha_m + \alpha_{m+1}) = \tau_0$. Therefore $\boldsymbol{\vartheta}^{m(1)}_{M_0+1}(\tau_0)$ equals a reparameterzied local MLE in the neighbourhood of $\boldsymbol{\vartheta}^{m*}_{M_0+1}(\tau)$. Therefore,$2\{PL^m_n(\boldsymbol{\vartheta}^{m(1)}_{M_0+1}(\tau_0)) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} = \omega_{n,m} + o_p(1)$, follows from repeating the proof of Proposition 3.

We proceed to show that $M^{m(K)}_n(\tau_0) = \omega_{n,m} + o_p(1)$ for any finite K. Because a generalized EM step never decrease likelihood(Dempster et al. (1977)), we have $PL^m_n(\boldsymbol{\vartheta}^{m(K)}_{M_0+1}(\tau_0)) + p(\tau^{(K)}) > PL^m_n(\boldsymbol{\vartheta}^{m(1)}_{M_0+1}(\tau_0)) + p(\tau^{(1)})$. Therefore, it follows from Theorem 1 of Chen and Li (2009), Lemma 10 in Appendix D of Kasahara and Shimotsu (2019), and induction that $\boldsymbol{\vartheta}^{m(K)}_{M_0+1}(\tau_0) - \boldsymbol{\vartheta}^{m*}_{M_0+1} = o_p(1)$ for any finite K. Let $\tilde{\boldsymbol{\vartheta}}^m_{M_0+1}$ be the maximizer of $PL^m_{M_0+1}(\boldsymbol{\vartheta}_{M_0+1})$ under the constraint of $\alpha_m/(\alpha_m + \alpha_{m+1} = \tau$ in an arbitrary small neighbourhood of $\boldsymbol{\vartheta}^{m*}_{M_0+1}(\tau^{(K)})$. Then we have $PL_{M_0+1}(\tilde{\boldsymbol{\vartheta}}^m_{M_0+1}) \geq PL^m_{M_0+1}(\boldsymbol{\vartheta}^{m(K)}_{M_0+1}(\tau_0)) + o_p(1)$ from the consistency of $\tilde{\boldsymbol{\vartheta}}^m_{M_0+1}$. $2\{PL^m_n(\tilde{\boldsymbol{\vartheta}}^m_{M_0+1}) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} = \omega_{n,m} + o_p(1)$ holds from the definition of $\tilde{\boldsymbol{\vartheta}}^m_{M_0+1}$. Furthermore, note that $PL^m_{M_0+1}(\boldsymbol{\vartheta}^{m(K)}_{M_0+1}(\tau_0)) \geq PL^m_{M_0+1}(\boldsymbol{\vartheta}^{m(1)}_{M_0+1}(\tau_0)) + o_p(1)$, $\tau^{(K)} - \tau_0 = o_p(1)$ and $2\{PL^m_n(\boldsymbol{\vartheta}^{m(1)}_{M_0+1}(\tau_0)) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} = \omega_{n,m} + o_p(1)$, we have $2\{PL^m_n(\boldsymbol{\vartheta}^{m(K)}_{M_0+1}(\tau_0)) - L_{0,n}(\hat{\boldsymbol{\vartheta}}_{M_0})\} = \omega_{n,m} + o_p(1)$ holds for all m. □

B Quadratic approximation of the log-likelihood function

This appendix derives a Le Cam's differentiable in quadratic mean(DQM)-type of expansion that is useful in proving lemma 1 - 3 of KS2018 under the finite mixture model of normal panel regression. Liu and Shao (2003) develop a DQM expansion under the loss of identifiability in terms of generalized score function. Lemma 1 follow the Lemma 4 and 5 of Kasahara and Shimotsu (2019).

Assume for now there are two component in the finite mixture model:

$$f_2(\boldsymbol{w}; \vartheta_2) = \alpha f(\boldsymbol{w}; \theta_1) + (1-\alpha) f(\boldsymbol{w}; \theta_2).$$

For any given α, the re-parameterized parameters: $\psi_\alpha = (\eta^\top, \lambda^\top)$, where $\eta = \alpha\theta_1 + (1-\alpha)\theta_2$ and $\lambda = \theta_1 - \theta_2$. Consider the one-to-one reparameterization:

$$\begin{pmatrix} \nu \\ \lambda \end{pmatrix} = \begin{pmatrix} \alpha\theta_1 + (1-\alpha)\theta_2 \\ \theta_1 - \theta_2 \end{pmatrix}, \text{ therefore } \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \nu + (1-\alpha)\lambda \\ \nu - \alpha\lambda \end{pmatrix}.$$

Collect the reparameterized parameters as $\psi = (\nu^\top, \lambda^\top)^\top$. The reparameterized parameters:

$$g(\boldsymbol{w}; \boldsymbol{\psi}, \alpha) = \alpha f(\boldsymbol{w}; \nu + (1-\alpha)\lambda)) + (1-\alpha) f(\boldsymbol{w}; \nu - \alpha\lambda). \tag{57}$$

Denote the density ratio by

$$l(\boldsymbol{w};\boldsymbol{\psi},\alpha)=\frac{g(\boldsymbol{w};\boldsymbol{\psi},\alpha)}{g(\boldsymbol{w};\boldsymbol{\psi}^{*},\alpha)} \tag{58}$$

Definition 1 ($O_{p\epsilon}(a_n)$ and $o_{p\epsilon}(a_n)$)**.** *(1)For a sequence of $X_{n\epsilon}$ indexed by $n=1,\dots,$ and ϵ, we write $X_{n\epsilon}$ if for any $\Delta>0$, there exist $\epsilon>0$ and $M,n_0<\infty$ such that $\mathbb{P}(|X_{n\epsilon}/a_n|\le M)\ge 1-\Delta$ for all $n>n_0$. (2)We write $X_{n\epsilon}=o_{p\epsilon}(1)$ if for any $\Delta_1,\Delta_2>0$, there exists $\epsilon>0$, there exist $\epsilon>0$ and n_0 such that $\mathbb{P}(|X_{n\epsilon}/a_n|\le\Delta_1)\ge 1-\Delta_2$ for all $n>n_0$.*

Write the likelihood under null hypothesis parameters as:

$$L_{n,0}(\boldsymbol{\vartheta})=\sum_{i=1}^{N}\log f(\boldsymbol{W}_i,\boldsymbol{\vartheta})$$

Lemma 1. *Under assumption 1 and 2 hold and $\boldsymbol{X}$ given $\boldsymbol{Z}$ has the density of $f(x|z;\gamma,\mu,\sigma)$ defined in (2). Let $L_n(\psi,\alpha)=\sum_{i=1}^{n}\log g(X_i|Z_i;\psi,\alpha)$ with $g(X_i|Z_i;\psi,\alpha)$ defined in equation (57). For $\alpha\in(0,1)$, define $s(\boldsymbol{w})$ and $t(\boldsymbol{\psi},\alpha)$ as the score function and the collection of parameters. In addition, let $\mathcal{N}_\epsilon=\{\vartheta_2\in\Theta_{\vartheta_2}:|t(\psi,\alpha)|<\epsilon\}$ and $\boldsymbol{\mathcal{I}}=E[s(X,Z)s(X,Z)^\top]$. Then for $\epsilon_\sigma\in(0,1)$ and any $\delta>0$, we have (a) $\sup_{\vartheta_2\in A_{n\epsilon}}|t(\boldsymbol{\psi},\alpha)|=O_{p\epsilon}(n^{-\frac{1}{2}})$, (b) $\sup_{\vartheta_2\in A_{n\epsilon}(\delta)}\left|L_n(\boldsymbol{\psi},\alpha)-L_n(\boldsymbol{\psi}^*,\alpha)-\sqrt{n}t(\boldsymbol{\psi},\alpha)^\top\nu_n(s(x,z))+nt(\boldsymbol{\psi},\alpha)^\top\boldsymbol{\mathcal{I}}t(\boldsymbol{\psi},\alpha)/2\right|=o_{p\epsilon}(1)$. where $A_{n\epsilon}(\delta)=\{\varepsilon_2\in\mathcal{N}_\epsilon:L_n(\boldsymbol{\psi},\alpha)-L_n(\boldsymbol{\psi}^*,\alpha)\ge-\delta\}$.*

Proof of Lemma 1. First we show that $l(\boldsymbol{w},\boldsymbol{\psi},\alpha)=g(\boldsymbol{w},\boldsymbol{\psi},\alpha)/g(\boldsymbol{w},\boldsymbol{\psi}^*,\alpha)$ with $\boldsymbol{w}=(y,\boldsymbol{x}^\top,\boldsymbol{z}^\top)^\top$. In this expansion, $l(\boldsymbol{w},\boldsymbol{\psi},\alpha)$ plays the role of $l(\boldsymbol{w},\boldsymbol{\vartheta})$ and $\boldsymbol{t}(\boldsymbol{w},\alpha)$ plays the role of $\boldsymbol{t}(\vartheta)$. Observe that $\boldsymbol{t}(\boldsymbol{w},\alpha)$ defined in equation (13) satisfies $\boldsymbol{t}(\boldsymbol{w},\alpha)=0$ if and only if $\psi=\psi^*$ because $\boldsymbol{\lambda}=0$ if and only if $\theta_1=\theta_2$. We expand $l(\boldsymbol{w};\boldsymbol{\psi},\alpha)-1$ five times with respect to $\boldsymbol{\psi}$ and the show that the expansion satisfies Assumption 8. □

Define

$$\nu(\boldsymbol{w},\vartheta_2)=(\nabla_\psi g(\boldsymbol{w},\boldsymbol{\psi},\alpha)^\top,\nabla_{\psi^{\otimes 2}}g(\boldsymbol{w},\boldsymbol{\psi},\alpha)^\top)\top/g(\boldsymbol{w},\boldsymbol{\psi}^*,\alpha). \tag{59}$$

Note that equation (59) satisfies $\mathbb{E}(\nu(\boldsymbol{w},\vartheta_2))=0$. In order to apply Lemma 3 to $l(\boldsymbol{w},\boldsymbol{\psi},\boldsymbol{\alpha})$, we first show

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{N}_\epsilon}|P_n[\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)]^\top-\mathbb{E}[\boldsymbol{\nu}(\boldsymbol{W},\boldsymbol{\vartheta}_2)\boldsymbol{\nu}(\boldsymbol{W},\boldsymbol{\vartheta}_2)]^\top|=o_p(1) \tag{60}$$

$$\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)\Rightarrow\boldsymbol{W}(\boldsymbol{\vartheta_2}), \tag{61}$$

$\boldsymbol{W}(\boldsymbol{\vartheta_2})$ is a mean-zero continuous Gaussian process with $\mathbb{E}[\boldsymbol{W}(\boldsymbol{\vartheta_2})\boldsymbol{W}(\boldsymbol{\vartheta_2})^\top]=\mathbb{E}[\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)^\top]$. Equation (60) holds because $[\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)\boldsymbol{\nu}(\boldsymbol{w},\boldsymbol{\vartheta}_2)^\top]$ satisfies a uniform law of large numbers(see Lemma 2.4 of Newey and McFadden (1994)) because $\boldsymbol{\vartheta}(\boldsymbol{w},\boldsymbol{\vartheta_2})$ is continuous in $\boldsymbol{\vartheta_2}$ and $\mathbb{E}\sup_{\boldsymbol{\vartheta_2}\in\mathcal{N}_\epsilon}[\boldsymbol{\vartheta}(\boldsymbol{w},\boldsymbol{\vartheta_2})\boldsymbol{\vartheta}(\boldsymbol{w},\boldsymbol{\vartheta_2})^\top]<\infty$ from the property of the normal density and Assumption

2. Equation (61) follows from Theorem 10.2 of Pollard (1990) if (i) $\boldsymbol{\Theta}_{\boldsymbol{\vartheta}_2}$ is totally bounded, (ii) the finite dimensional distribution of $\nu_n(\boldsymbol{w}, \boldsymbol{\vartheta}_2)$ converges to the distribution of $\boldsymbol{W}(\boldsymbol{\vartheta_2})$, and (iii) $\{\nu_n(\boldsymbol{w}, \boldsymbol{\vartheta_2})]) : n \geq 1\}$ is stochastically equicontinuous. Condition (i) holds because $\boldsymbol{\Theta}_{\vartheta_2}$ is compact in the Euclidean space. Condition (ii) follows from Assumption 2 and the multivariat CLT. Condition (iii) holds Theorem 2 of Andrews (1994) because $v(\boldsymbol{w}, \boldsymbol{\vartheta}_2)$ is Lipschitz continuous in $\boldsymbol{\vartheta}_2$.

Note that the $(p+1)$-th order Taylor expansion of $g(\boldsymbol{\psi})$ around $\boldsymbol{\psi} = \boldsymbol{\psi}^*$ is given by

$$g(\boldsymbol{\psi}) = g(\boldsymbol{\psi^*}) + \sum_{j=1}^{p} \frac{1}{j!} \nabla_{(\boldsymbol{\psi}^{\otimes j})^\top} g(\boldsymbol{\psi^*})(\boldsymbol{\psi} - \boldsymbol{\psi}^*)^{\otimes j} + \frac{1}{(p+1)!} \nabla_{(\boldsymbol{\psi}^{\otimes p+1})^\top} g(\bar{\boldsymbol{\psi}})(\boldsymbol{\psi} - \boldsymbol{\psi}^*)^{\otimes (p+1)},$$

where $\bar{\boldsymbol{\psi}}$ lies between $\boldsymbol{\psi}$ and $\boldsymbol{\psi}^*$, and $\bar{\boldsymbol{\psi}}$ may differ from element to element of $\nabla_{(\boldsymbol{\psi}^{\otimes(p+1)})^\top} g(\bar{\boldsymbol{\psi}})$.

Let g^* and ∇g^* denote $g(\boldsymbol{w}; \boldsymbol{\psi}, \alpha)$ and $\nabla g(\boldsymbol{w}; \boldsymbol{\psi}, \alpha)$, and let $\nabla \bar{g}$ denote $\nabla g(\boldsymbol{w}; \boldsymbol{\psi}, \alpha)$. Let $\dot{\boldsymbol{\psi}} := \boldsymbol{\psi} - \boldsymbol{\psi}^*$ and $\dot{\boldsymbol{\eta}} := \boldsymbol{\eta} - \boldsymbol{\eta}^*$. Expanding $l(\boldsymbol{w}; \boldsymbol{\psi}, \alpha)$ around $\boldsymbol{\psi}^*$ while fixing α and using Lemma 4, we can write $l(\boldsymbol{w}; \boldsymbol{\psi}, \alpha) - 1$ as

$$l(\boldsymbol{w}; \boldsymbol{\psi}, \alpha) - 1 = \boldsymbol{s}(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) + r(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}),$$

where

$$\boldsymbol{s}(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) := \frac{\nabla_\eta^\top g^*}{g^*} \dot{\boldsymbol{\eta}} + \frac{1}{2!} \frac{\nabla_{\boldsymbol{\lambda}^{\otimes 2}}^\top g^*}{g^*} \boldsymbol{\lambda}^{\otimes 2},$$

and

$$r(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) = \frac{1}{2!} \frac{\nabla_{\boldsymbol{\eta}^{\otimes 2}} g^*}{g^*} \dot{\boldsymbol{\eta}}^{\otimes 2} + \frac{1}{3!} \frac{\nabla_{\boldsymbol{\psi}^{\otimes 3}} g^*}{g^*} \dot{\boldsymbol{\psi}}^{\otimes 3} + \frac{1}{4!} \sum_{p=0}^{3} \binom{4}{p} \frac{\nabla_{\boldsymbol{\eta}^{\otimes p} \boldsymbol{\lambda}^{\otimes (4-p)}} g^*}{g^*} \boldsymbol{\eta}^{\otimes p} \boldsymbol{\lambda}^{\otimes (4-p)}, \tag{62}$$

where $s(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ is the leading term in the expansion. We first show $s(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) = \boldsymbol{t}(\boldsymbol{\psi}, \alpha)\boldsymbol{s}(\boldsymbol{w})$ score and $\boldsymbol{t}(\boldsymbol{\vartheta})$ defined in (12) and (13). Let f^* and ∇f^* denote $f(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$ and $\nabla f(y|\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\gamma}^*, \boldsymbol{\theta}^*)$. The first term of $s(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ is $\frac{\nabla_\eta^\top g^*}{g^*} \dot{\boldsymbol{\eta}} = \frac{\nabla_{(\boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top} f^*}{f} \dot{\boldsymbol{\eta}}$. Using the result from Lemma 4 , the second term of $s(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ can be written as $\alpha(1-\alpha)\frac{1}{2!} \frac{\nabla_{\boldsymbol{\theta}^{\otimes 2}}^\top f^*}{f^*} \boldsymbol{\theta}^{\otimes 2}$, and hence $s(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) = \boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{s}(\boldsymbol{w})$.

From (60) and (61) and the property of normal density, $\boldsymbol{s}(\boldsymbol{w})$ satisfies Assumption 8 (a) (b) (e). We need to show that $r(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda})$ satisfies Assumption 8 (c) (d). Following the proof Kasahara and Shimotsu (2019) Lemma 1, we can show that $r(\boldsymbol{w}; \boldsymbol{\eta}, \boldsymbol{\lambda}) = \boldsymbol{\xi}(\boldsymbol{w}; \boldsymbol{\vartheta}) O(|\boldsymbol{\psi} - \boldsymbol{\psi}^*||\boldsymbol{t}(\boldsymbol{\psi}, \alpha)|)$ where $\sup_{\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon} |\boldsymbol{\xi}(\boldsymbol{w}; \boldsymbol{\vartheta})| \leq \sup_{\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon} |\boldsymbol{\nu}(\boldsymbol{w}; \boldsymbol{\vartheta})|$, where $\sup_{\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon}$ is defined in (59).

First, we show that $\frac{1}{2!} \frac{\nabla_{\boldsymbol{\eta}^{\otimes 2}} g^*}{g^*} \dot{\boldsymbol{\eta}}^{\otimes 2}$ is written as $\frac{\nabla_{\boldsymbol{\eta}^{\otimes 2}} g^*}{g^*} O(|\dot{\boldsymbol{\eta}}|^2)$. Then we write $\frac{1}{3!} \frac{\nabla_{\boldsymbol{\psi}^{\otimes 3}} g^*}{g^*} \dot{\boldsymbol{\psi}}^{\otimes 3} = \frac{1}{3!} \sum_{p=0}^{3} \binom{3}{p} \frac{\nabla_{\boldsymbol{\eta}^{\otimes p} \boldsymbol{\lambda}^{\otimes (3-p)}} g^*}{g^*} \boldsymbol{\eta}^{\otimes p} \boldsymbol{\lambda}^{\otimes (3-p)}$. The terms with $p = 1$ are written as $(\nabla_{\boldsymbol{\psi}^{\otimes 3}}) O(|\dot{\boldsymbol{\eta}}|) O(|\dot{\boldsymbol{\lambda}}|)$. For the term with $p > 1$, we have $\nabla_{\boldsymbol{\lambda}^{\otimes 3}} g^* = 0$ and $\nabla_{\boldsymbol{\lambda}^{\otimes 2}} g^* = 0$ by Lemma 4. Therefore, we can write $\frac{1}{3!} \frac{\nabla_{\boldsymbol{\psi}^{\otimes 3}} g^*}{g^*} \dot{\boldsymbol{\psi}}^{\otimes 3}$ as $\boldsymbol{\nu}(\boldsymbol{w}; \boldsymbol{\vartheta}_2) O(|\boldsymbol{\psi} - \boldsymbol{\psi}^*||\boldsymbol{t}(\boldsymbol{\psi}, \alpha)|)$. For the last term of (62), we use a similar arguement. Therefore the Assumption 8 holds.

Remark 1. *I am uncertain whether this proof that Assumption 8 holds is clear.*

C Quadratic expansion under singular Fisher information matrix

Let $\boldsymbol{\vartheta}$ be a parameter vector, and let $g(\boldsymbol{w}, \boldsymbol{\theta})$ denote the density of $\boldsymbol{w}$. Let $L_n(\boldsymbol{\vartheta}) := \sum_{i=1}^n \log g(\boldsymbol{w}, \boldsymbol{\theta})$ denote the log-likelihood function. Split $\boldsymbol{\vartheta} = (\boldsymbol{\psi}^\top, \boldsymbol{\pi}^\top)^\top$, and write $L_n(\boldsymbol{\vartheta}) := L_n(\boldsymbol{\psi}, \boldsymbol{\pi})$.

We establish a general quadratic expansion that expresses $L_n(\boldsymbol{\psi}, \boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*, \boldsymbol{\pi})$ as a quadratic function of $\boldsymbol{t(\vartheta)}$ for $\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon$. Denote the density ratio by:

$$l(\boldsymbol{w}; \boldsymbol{\vartheta}) := \frac{g(\boldsymbol{w}; \boldsymbol{\psi}, \boldsymbol{\pi})}{g(\boldsymbol{w}; \boldsymbol{\psi}^*, \boldsymbol{\pi})}, \tag{63}$$

so that $L_n(\boldsymbol{\psi}, \boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*, \boldsymbol{\pi}) = \sum_{i=1}^n \log l(\boldsymbol{W}_i, \boldsymbol{\vartheta})$.

Let $\boldsymbol{\pi}$ denote the part of parameters that is not identified under the null. Denote the true parameter value of $\boldsymbol{\psi}$ by $\boldsymbol{\psi}^*$, and denote the set of $(\boldsymbol{\psi}, \boldsymbol{\pi})$ corresponding to the null hypothesis by $\Upsilon^* = \{(\boldsymbol{\psi}, \boldsymbol{\pi}) \in \Theta : \boldsymbol{\psi} = \boldsymbol{\psi}^*\}$. For $\epsilon > 0$, define a neighborhood of Υ^* by $\mathcal{N}_\epsilon = \{\boldsymbol{\vartheta} \in \Theta : |\boldsymbol{t}(\boldsymbol{\vartheta})| < \epsilon\}$. We assume that $l(\boldsymbol{w}, \boldsymbol{\vartheta})$ can be expanded around $l(\boldsymbol{w}, \boldsymbol{\vartheta}^*) = 1$ as follows.

Assumption 8. $l(y, \vartheta) - 1$ *admits an expansion*

$$l(\boldsymbol{w}, \vartheta) - 1 = t(\vartheta)^\top s(\boldsymbol{w}, \pi) + r(\boldsymbol{w}, \vartheta),$$

where $s(\boldsymbol{w}, \pi)$ *and* $r(\boldsymbol{w}, \vartheta)$ *satisfy for some* $C \in (0, \infty)$ *and* $\epsilon > 0$. *(a)* $\mathrm{E} \sup_{\pi \in \Theta_\pi} |s(Y, \pi)| < C$, *(b)* $\sup_{\pi \in \Theta_\pi} |P_n(s(y, \pi) s(y, \pi)^\top) - \boldsymbol{\mathcal{I}}_\pi|$, *(c)* $E[\sup_{\vartheta \in \mathcal{N}_\epsilon} [r(Y; \vartheta)/(|t(\vartheta)||\psi - \psi^*|)^2] < \infty$, *(d)* $\sup_{\vartheta \in \mathcal{N}_\epsilon} [r(Y; \vartheta)/(|t(\vartheta)||\psi - \psi^*|)] = O_p(1)$, *(e)* $\sup_{\pi \in \Theta_\pi} |\nu_n(s(y; \pi))| = O_p(1)$.

The assumption 8 follows the Assumption 6 in Kasahara and Shimotsu (2019).

Check if the likelihood satisfies the assumption 8. Apply the Taylor expansion to the likelihood ratio:

$$l(\boldsymbol{w}, \vartheta) - 1 \approx \frac{\partial l(y, \vartheta^*)}{\partial \vartheta}(\vartheta - \vartheta^*) + \frac{\partial^2 l(y, \vartheta^*)}{\partial \vartheta^2}(\vartheta - \vartheta^*)/2 \tag{64}$$

First look at the first order derivative of the $l(y, \vartheta)$,

We first establish an expansion $L_n(\psi, \pi)$ in a neighborhood $\mathcal{N}_{c/\sqrt{n}}$ that holds for any $c > 0$.

Lemma 2. *Suppose that Assumption 8 (a) - (d) holds. Then for all* $c > 0$,

$$\sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} |L_n(\boldsymbol{\psi}, \boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*, \boldsymbol{\pi}) - \sqrt{n} \boldsymbol{t}(\boldsymbol{\vartheta})^\top \nu_n(\boldsymbol{s}(\boldsymbol{w}, \boldsymbol{\pi})) + \boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}} \boldsymbol{t}(\boldsymbol{\vartheta})| = o_p(1).$$

Proof of Lemma 2. Define $h(\boldsymbol{w}, \boldsymbol{\vartheta}) := \sqrt{l(\boldsymbol{w}, \boldsymbol{\vartheta})} - 1$. By applying the Taylor expansion of $2\log(1+x) = 2x - x^2(1+o(1))$ for small x, we have, uniformly for $\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}$,

$$L_n(\boldsymbol{\psi}, \boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*, \boldsymbol{\pi}) = 2\sum_{i=1}^{n} \log(1 + h(\boldsymbol{w}_i, \boldsymbol{\vartheta})) = nP_n(2h(\boldsymbol{w}, \boldsymbol{\vartheta}) - [1 + o_p(1)]h(\boldsymbol{w}, \boldsymbol{\vartheta})^2). \quad (65)$$

The stated result holds if we show

$$\sup_{\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}} |nP_n\Big(h(\boldsymbol{w}_i, \boldsymbol{\vartheta}))^2\Big) - n\boldsymbol{t}(\boldsymbol{\vartheta})^\top \nu_n(\boldsymbol{s}(\boldsymbol{w}))/4| = o_p(1) \quad (66)$$

$$\sup_{\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}} |nP_n\Big(h(\boldsymbol{w}_i, \boldsymbol{\vartheta}))\Big) - \sqrt{n}\boldsymbol{t}(\boldsymbol{\vartheta})^\top \nu_n(\boldsymbol{s}(\boldsymbol{w}))/2 + n\boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}} \boldsymbol{t}(\boldsymbol{\vartheta})/8| = o_p(1), \quad (67)$$

because the right hand side of (65) equals $\sqrt{n}\boldsymbol{t}(\boldsymbol{\vartheta})^\top \nu_n(\boldsymbol{s}(\boldsymbol{w})) - n\boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}} \boldsymbol{t}(\boldsymbol{\vartheta})/2$ uniformly in $\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}$.

To show equation (66), write $4P_n\Big(h(\boldsymbol{w}_i, \boldsymbol{\vartheta}))^2\Big) = P_n\left(\frac{4(l(\boldsymbol{w},\vartheta)-1)^2}{(\sqrt{l(\boldsymbol{w},\vartheta)}+1)^2}\right) = P_n(l(\boldsymbol{w}, \vartheta) - 1)^2 - P_n\left((l(\boldsymbol{w}, \vartheta)-1)^3 \frac{\sqrt{l(\boldsymbol{w},\vartheta)}+3}{(\sqrt{l(\boldsymbol{w},\vartheta)}+1)^3}\right)$. If follows from Assumption 8(a) - (c) and $(\mathbb{E}(XY))^2 \le \mathbb{E}|X|^2\mathbb{E}|Y|^2$ that, uniformly for $\boldsymbol{\vartheta}$,

$$\begin{aligned} P_n(l(\boldsymbol{w}, \vartheta) - 1)^2 &= \boldsymbol{t}(\boldsymbol{\vartheta})^\top P_n\big(\boldsymbol{s}(\boldsymbol{w})\boldsymbol{s}(\boldsymbol{w}, \boldsymbol{\vartheta})^\top\big)\boldsymbol{t}(\boldsymbol{\vartheta}) + 2\boldsymbol{t}(\boldsymbol{\vartheta})^\top P_n\big(\boldsymbol{s}(\boldsymbol{w}, \boldsymbol{\vartheta})\boldsymbol{r}(\boldsymbol{w}, \boldsymbol{\vartheta})\big) + P_n(\boldsymbol{r}(\boldsymbol{w}))^2 \\ &= (1 + o_p(1))\boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}} \boldsymbol{t}(\boldsymbol{\vartheta}) + O_p(|\boldsymbol{t}(\boldsymbol{\vartheta})|^2|\boldsymbol{\psi} - \boldsymbol{\psi}^*|). \end{aligned} \quad (68)$$

Therefore, we have that $P_n(l(\boldsymbol{w}, \vartheta) - 1)^2 = \boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}} \boldsymbol{t}(\boldsymbol{\vartheta}) + o_p(n^{-1})$. Note that, if the data are random variables with $\max_{1\le i\le n} \mathbb{E}|X_i|^q < C$ for some $q > 0$ and $C < \infty$, then we have $\max_{1\le i\le n} |X_i| = o_p(n^{1/q})$. Therefore, from Assumption 8(a) (c), we have

$$\max_{1\le i\le n} \sup_{\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}} |l(\boldsymbol{W}_i, \boldsymbol{\vartheta}) - 1| = \max_{1\le i\le n} \sup_{\boldsymbol{\vartheta} \in \mathcal{N}_{c/\sqrt{n}}} |\boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{s}(\boldsymbol{W}_i, \boldsymbol{\pi}) + r(\boldsymbol{W}_i, \boldsymbol{\vartheta})| = o_p(1). \quad (69)$$

Therefore, $P_n\left((l(\boldsymbol{w}, \vartheta) - 1)^3 \frac{\sqrt{l(\boldsymbol{w},\vartheta)}+3}{(\sqrt{l(\boldsymbol{w},\vartheta)}+1)^3}\right) = o_p(1)$, and (66) follows.

We proceed to show (67). Consider the following expansion of $h(\boldsymbol{w}, \boldsymbol{\vartheta})$:

$$h(\boldsymbol{w}, \boldsymbol{\vartheta}) = (l(\boldsymbol{w}, \boldsymbol{\vartheta}) - 1)/2 - h(\boldsymbol{w}, \boldsymbol{\vartheta})^2/2 = (\boldsymbol{t}(\boldsymbol{\vartheta})^\top)s(\boldsymbol{w}, \boldsymbol{\pi}) + r(\boldsymbol{w}, \boldsymbol{\vartheta})/2 - h(\boldsymbol{w}, \boldsymbol{\vartheta})^2/2. \quad (70)$$

Then (67) follows from (66), (70), Assumption 8(d), and the stated result follows. □

The Lemma 3 expands $L_n(\boldsymbol{\psi}, \boldsymbol{\pi})$ in $A_{n\epsilon} := \{\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon : L_n(\boldsymbol{\psi}, \boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*, \boldsymbol{\pi})\} \ge -\delta$ for $\delta \in (0, \infty)$.

This lemma is useful for deriving

Lemma 3. *Suppose that Assumption 8 holds. Then for any $\delta > 0$, (a) $\sup_{\vartheta in A_{n\epsilon}(\delta)} |\boldsymbol{t}(\boldsymbol{\vartheta})| = O_{p\epsilon}(n^{-1/2})$; (b) $\sup_{\boldsymbol{\vartheta}\in A_{n\epsilon}(\delta} |L_n(\boldsymbol{\psi},\boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*,\boldsymbol{\pi}) - \sqrt{n}\boldsymbol{t}(\boldsymbol{\vartheta})^\top \nu_n(\boldsymbol{s}(\boldsymbol{w},\boldsymbol{\pi})) + n\boldsymbol{t}\boldsymbol{\vartheta}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}\boldsymbol{\vartheta}/2| = o_{p\epsilon}(1)$.*

Proof of Lemma 3. For part (a), applying the inequality $\log(1+x) \leq x$ to the log-likelihood ratio function and with (70) give:

$$L_n(\boldsymbol{\psi},\boldsymbol{\pi}) - L_n(\boldsymbol{\psi}^*,\boldsymbol{\pi}) = 2\sum_{i=1}^{n}\log(1+h(\boldsymbol{W}_i,\boldsymbol{\vartheta})) \leq 2nP_n(h(\boldsymbol{y},\boldsymbol{\vartheta})) = \sqrt{n}\nu_n(l(\boldsymbol{y},\boldsymbol{\vartheta})) - nP_n(h(\boldsymbol{y},\boldsymbol{\vartheta})^2). \tag{71}$$

We derived a lower bound on $P_n(h(\boldsymbol{y},\boldsymbol{\vartheta})^2)$. Observe that $h(\boldsymbol{y},\boldsymbol{\vartheta})^2 = l(\boldsymbol{y},\boldsymbol{\vartheta})^2/(\sqrt{l(\boldsymbol{y},\boldsymbol{\vartheta})}+1)^2 \geq \mathbb{1}\{l(\boldsymbol{y},\boldsymbol{\vartheta}) \leq \kappa\}(l(\boldsymbol{y},\boldsymbol{\vartheta})-1)^2/(\sqrt{\kappa}+1)^2$ for any $\kappa > 1$. Therefore,

$$\begin{aligned} P_n(h(\boldsymbol{y},\boldsymbol{\vartheta})^2) &\geq (\sqrt{\kappa}+1)^{-2} P_n(\mathbb{1}\{l(\boldsymbol{y},\boldsymbol{\vartheta}) \leq \kappa\})(l(\boldsymbol{y},\boldsymbol{\vartheta}-1)^2) \\ &\geq (\sqrt{\kappa}+1)^{-2}\Big[P_n(l(\boldsymbol{y},\boldsymbol{\vartheta})-1)^2)) - P_n(\mathbb{1}\{l(\boldsymbol{y},\boldsymbol{\vartheta}) < \kappa\})(l(\boldsymbol{y},\boldsymbol{\vartheta}-1)^2)\Big]. \end{aligned} \tag{72}$$

Let $B := \sup_{\boldsymbol{\vartheta}\in\mathcal{N}_\epsilon} |l(\boldsymbol{y},\boldsymbol{\vartheta}) - 1|$. From Assumption 8(a),(c), we have $\mathbb{E}B^2 < \infty$, and hence $\lim_{\kappa\to\infty}\sup_{\vartheta\in\mathcal{N}_\epsilon} P_n(\mathbb{1}\{l(\boldsymbol{y},\boldsymbol{\vartheta}) > \kappa\}(l(\boldsymbol{y},\boldsymbol{\vartheta})-1)^2) \leq \lim_{\kappa\to\infty} P_n(\mathbb{1}\{B+1 > \kappa\}B^2)$ almost surely. Let $\tau = (\sqrt{\kappa}+1)^{-2}/2$. By choosing κ sufficiently large, it follows from (68), uniformly for $\boldsymbol{\vartheta} \in \mathcal{N}_\epsilon$,

$$P_n(h(\boldsymbol{y},\boldsymbol{\vartheta})^2) \geq \tau(1+o_p(1))\big(\boldsymbol{t}(\boldsymbol{\vartheta})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\pi}}\boldsymbol{t}(\boldsymbol{\vartheta})\big) + O_p(|\boldsymbol{t}(\boldsymbol{\vartheta})|^2|\boldsymbol{\psi}-\boldsymbol{\psi}^*|). \tag{73}$$

Because $\sqrt{n}\nu_n(l(\boldsymbol{w},\boldsymbol{\vartheta})) = \sqrt{n}\boldsymbol{t}(\boldsymbol{\vartheta})^\top[\nu_n(\boldsymbol{s}(\boldsymbol{w},\boldsymbol{\vartheta})) + O_p(1)]$ from Assumption 8(c)(e), and the fact $\boldsymbol{\psi}-\boldsymbol{\psi}^* \to 0$ if $\boldsymbol{t}(\boldsymbol{\vartheta}) \to 0$, we obtain the following results: for any $\Delta > 0$, there exist $\epsilon > 0$ and $M, n_0 < \infty$ such that

$$\Pr\left(\inf_{\vartheta\in\mathcal{N}_\epsilon}(|\boldsymbol{T}_n|M - (\tau/2)|\boldsymbol{T}_n|^2 + M) \geq 0\right) \geq 1-\Delta, \text{for all} n > n_0. \tag{74}$$

Rearranging the terms inside $\Pr(\cdot)$ gives $\sup_{\boldsymbol{\vartheta}\in\mathcal{N}_\epsilon}(|\boldsymbol{T}_n| - (M/\tau))^2 \leq 2M/\tau + (M/\tau)^2$, and part (a) follows. Part (b) follows part(a) and Lemma 2. □

D Auxiliary results and their proofs

Lemma 4. *Suppose that $g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha)$ is defined as (8), where $\boldsymbol{\psi}_\alpha = (\boldsymbol{\eta}^\top,\boldsymbol{\lambda}^\top)^\top$. Let g^* and ∇g^* denote $g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha)$ and $\nabla g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha)$ evaluated at $(\boldsymbol{\psi}_\alpha,\alpha)$, respectively. Let ∇f^* denote $f(\boldsymbol{w};\boldsymbol{\gamma}^*,\boldsymbol{\theta}^*)$ The fololowing statements hold. (a) For $l = 0,1,\ldots$, $\nabla_{\boldsymbol{\lambda}\boldsymbol{\eta}^{\otimes l}} g^* = 0$. (b) $\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} g^* = \alpha(1-\alpha)\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} f^*$.*

Proof of Lemma 4. Recall that

$$g(\boldsymbol{w};\boldsymbol{\psi}_\alpha,\alpha) = \alpha f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}+(1-\alpha)\boldsymbol{\lambda}) + (1-\alpha) f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda}).$$

First we show that for $l=0$ holds for (a), $\nabla_{\boldsymbol{\lambda}} g^* = \alpha(1-\alpha)\nabla_{\boldsymbol{\theta}} f^* - \alpha(1-\alpha)\nabla_{\boldsymbol{\theta}} f^* = 0$. For $l>0$, by Fubini's Theorem, we have

$$\begin{aligned}
\nabla_{\boldsymbol{\lambda}\boldsymbol{\eta}^{\otimes l}} g = &\ \nabla_{\boldsymbol{\lambda}}\Big(\alpha \nabla_{(\boldsymbol{\gamma},\boldsymbol{\theta})^{\otimes l}} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}+(1-\alpha)\boldsymbol{\lambda}) + (1-\alpha)\nabla_{(\boldsymbol{\gamma},\boldsymbol{\theta})^{\otimes l}} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda})\Big|_{\boldsymbol{\nu}=\boldsymbol{\theta}^*,\boldsymbol{\lambda}=\boldsymbol{0}}\Big) \\
= &\ \Big(\alpha(1-\alpha)\nabla_{(\boldsymbol{\gamma}^{\otimes l},\boldsymbol{\theta}^{\otimes l+1})} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}+(1-\alpha)\boldsymbol{\lambda}) - \alpha(1-\alpha)\nabla_{(\boldsymbol{\gamma}^{\otimes l},\boldsymbol{\theta}^{\otimes l+1})} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda})\Big|_{\boldsymbol{\nu}=\boldsymbol{\theta}^*,\boldsymbol{\lambda}=\boldsymbol{0}}\Big) \\
= &\ 0.
\end{aligned} \tag{75}$$

To show part (b), note that

$$\begin{aligned}
\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} g &= \nabla_{\boldsymbol{\lambda}}\Big(\alpha(1-\alpha)\nabla_{\boldsymbol{\lambda}^\top} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha(1-\alpha)\boldsymbol{\lambda}) + (1-\alpha)\nabla_{\boldsymbol{\lambda}^\top} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda})\Big) \\
&= \alpha(1-\alpha)^2\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}+\alpha^2(1-\alpha)\boldsymbol{\lambda}) + (1-\alpha)\nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} f(\boldsymbol{w};\boldsymbol{\gamma},\boldsymbol{\nu}-\alpha\boldsymbol{\lambda})\big|_{\boldsymbol{\nu}=\boldsymbol{\theta}^*,\boldsymbol{\lambda}=\boldsymbol{0}} \\
&= \nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top} f^*.
\end{aligned}$$

□

Lemma 5. *If $T\geq 2$, the Fisher information matrix for finite normal mixture panel regression model is positive definite.*

Proof of Lemma 5. As mentioned in Kasahara and Shimotsu (2015), the normal mixture model with cross-sectional data suffers from degeneracy of Fisher information matrix under this reparameterization, and thus cannot be approximated by the local expansion. When $T=1$, the regression model is reduced to the cross-sectional regression. Then as shown in D.1, the score functions are collinear. Then $\boldsymbol{\mathcal{I}}$ is not positive definite.

For the normal panel regression with $T\geq 2$, the score functions for 2-component finite mixture normal panel regression is defined by equation (12). Due to the orthogonality of the Hermite polynomials of different orders, the score functions are linear independent. As a result, $Es_i s_i^\top$ has full rank and $\boldsymbol{\mathcal{I}} = Es_i s_i^\top < \infty$. $\boldsymbol{\mathcal{I}}$ is finite and positive definite. □

D.1 Score function for testing $H_0: m=1$ against $H_A: m=2$

$H^b(\cdot)$ is defined as the b-th order Hermite polynomial. $H^1(t)=t$, $H^2(t)=t^2-1$, $H^3(t)=t^3-3t$, and $H^4(t)=t^4-6t^2+3$. As shown in Kasahara et al. (2015) supplement material, the derivative of $\{\frac{1}{\sigma}\phi(\frac{t}{\sigma})\}$ is

$$\frac{\nabla_{\mu^m}\nabla_{(\sigma^2)^\ell}\{\frac{1}{\sigma}\phi(\frac{t}{\sigma})\}}{\{\frac{1}{\sigma}\phi(\frac{t}{\sigma})\}} = \left(\frac{1}{2}\right)^\ell (\frac{1}{\sigma})^{m+2\ell} H^{m+2\ell}(\frac{t}{\sigma}). \tag{76}$$

Let $f^* = f(\boldsymbol{w};\gamma^*,\theta^*)$, $\nabla f^* = \nabla f(\boldsymbol{w};\gamma^*,\theta^*)$, $H_{i,t}^{b*} = \frac{1}{\sigma^* b!}H^b(\frac{y_{it}-\boldsymbol{x}_{it}^\top\boldsymbol{\beta}^*-\boldsymbol{z}_{it}^t op\boldsymbol{\gamma}^*-\mu^*}{\sigma^*})$..

$$\begin{aligned}\nabla_\mu f^* &= f^*\sum_{t=1}^T \frac{1}{\sigma}H_{i,t}^{1*}; \nabla_{\sigma^2} f^* = f^*\sum_{t=1}^T \frac{1}{2}\frac{1}{\sigma^2}H_{it}^{2*};\\ \nabla_{\boldsymbol{\beta}} f^* &= f^*\sum_{t=1}^T \frac{1}{\sigma}H_{it}^{1*}\boldsymbol{x}_{it}; \nabla_{\boldsymbol{\gamma}} f^* = f^*\sum_{t=1}^T \frac{1}{\sigma}H_{it}^{1*}\boldsymbol{z}_{it}.\end{aligned} \tag{77}$$

Define the score functions in a similar way to Kasahara and Shimotsu (2012),

$$s_{\eta i} = \begin{pmatrix} s_{\mu i}\\ s_{\beta i}\\ s_{\sigma i}\\ s_{\gamma i}\end{pmatrix} = \begin{pmatrix}\sum_{t=1}^T H_{i,t}^{1*}\\ \sum_{t=1}^T H_{i,t}^{1*}x_{it}\\ \sum_{t=1}^T H_{i,t}^{2*}\\ \sum_{t=1}^T H_{i,t}^{1*}z_{it}\end{pmatrix}, s_{\lambda_{\mu\sigma}i} = \begin{pmatrix} s_{\lambda_{\mu\mu i}}\\ s_{\lambda_{\sigma\sigma i}}\\ s_{\lambda_{\mu\sigma i}}\\ s_{\lambda_{\mu\beta i}}\\ s_{\lambda_{\sigma\beta i}}\end{pmatrix} = \begin{pmatrix}\sum_{t=1}^T H_{i,t}^{2*} + \frac{1}{2}\sum_{t=1}^T\sum_{s\neq t} H_{1,i,t}^{1*}H_{i,s}^{1*}\\ 3\sum_{t=1}^T H_{i,t}^{4*} + \frac{1}{2}\sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{2*}H_{i,t}^{2*}\\ 3\sum_{t=1}^T H_{i,t}^{3*} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}H_{i,s}^{2*}\\ 2\sum_{t=1}^T H_{i,t}^{2*}x_{it} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it}H_{i,s}^{1*}\\ 3\sum_{t=1}^T H_{i,t}^{3*}x_{it} + 2\sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it}H_{i,s}^{2*}\end{pmatrix},$$

$$s_{\lambda_\beta i} = \begin{pmatrix}\sum_{t=1}^T H_{i,t}^{2*}x_{it,1}^2 + \frac{1}{2}\sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,1}H_{i,s}^{1*}x_{is,1}\\ \vdots\\ \sum_{t=1}^T H_{i,t}^{2*}x_{it,q}^2 + \frac{1}{2}\sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,q}H_{i,s}^{1*}x_{is,q}\\ 2\sum_{t=1}^T H_{i,t}^{2*}x_{it,1}x_{it,2} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,1}H_{i,s}^{1*}x_{is,2}\\ \vdots\\ 2\sum_{t=1}^T H_{i,t}^{2*}x_{it,1}x_{it,q} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,1}H_{i,s}^{1*}x_{is,q}\\ 2\sum_{t=1}^T H_{i,t}^{2*}x_{it,2}x_{it,3} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,2}H_{i,s}^{1*}x_{is,3}\\ \vdots\\ 2\sum_{t=1}^T H_{i,t}^{2*}x_{it,q-1}x_{it,q} + \sum_{t=1}^T\sum_{s\neq t} H_{i,t}^{1*}x_{it,q-1}H_{i,s}^{1*}x_{is,q}\end{pmatrix}. \tag{78}$$

When $T = 1$, the score functions are as follow:

$$s_{\eta i} = \begin{pmatrix} H_i^{1*} \\ H_i^{1*} x_i \\ H_i^{2*} \\ H_i^{1*} z_i \end{pmatrix}, \; s_{\lambda_{\mu\sigma} i} = \begin{pmatrix} H_i^{2*} \\ 3H_i^{4*} \\ 3H_i^{3*} \\ 2H_i^{2*} x_i \\ 3H_i^{3*} x_i \end{pmatrix}, \text{ and } s_{\lambda_\beta i} = \begin{pmatrix} H_i^{2*} x_{i,1}^2 \\ \vdots \\ H_i^{2*} x_{i,q}^2 \\ 2H_i^{2*} x_{i,1} x_{i,2} \\ \vdots \\ 2H_i^{2*} x_{i,1} x_{i,q} \\ 2H_i^{2*} x_{i,2} x_{i,3} \\ \vdots \\ 2H_i^{2*} x_{i,q-1} x_{i,q} \end{pmatrix}.$$

Notice that $s_{\sigma i}$ and $s_{\lambda_{\mu\mu}}$ are perfect collinear, the Fisher information matrix is therefore singular under this reparameterization for data with $T = 1$.

D.2 Score function for testing $H_0 : m = M_0$ against $H_A : m = M_0 + 1$

The derivative of the reparameterized density w.r.t λ at ψ_τ^{h*} is zero similar to testing homogeneity case. With the constraint $\pi^{M_0} = 1 - \sum_{j=1}^{M_0-1} \pi^j$. The score functions $s_{\eta i}$'s contain the first order derivatives w.r.t π's γ and ν at ψ_τ^{h*}:

$$\begin{aligned} \nabla_{\pi^j} l^h(\boldsymbol{w}; \psi_\tau^{h*}, \tau) &= \frac{f(\boldsymbol{w}; \gamma^*, \theta_0^{j*}) - f(\boldsymbol{w}; \gamma^*, \theta_0^{M_0*})}{\sum_{j=1}^{M_0} \alpha_0^{j*} f(\boldsymbol{w}; \gamma^*, \theta_0^{j*})}; \\ \nabla_\gamma l^h(\boldsymbol{w}; \psi_\tau^{h*}, \tau) &= \frac{\sum_{j=1}^{M_0} \alpha_0^{j*} \nabla_\gamma f(\boldsymbol{w}; \gamma^*, \theta_0^{j*})}{\sum_{j=1}^{M_0} \alpha_0^{j*} f(\boldsymbol{w}; \gamma^*, \theta_0^{j*})}; \\ \nabla_\nu l^h(\boldsymbol{w}; \psi_\tau^{h*}, \tau) &= \frac{\nabla_\theta f(\boldsymbol{w}; \gamma^*, \theta_0^{h*})}{\sum_{j=1}^{M_0} \alpha_0^{j*} f(\boldsymbol{w}; \gamma^*, \theta_0^{j*})}. \end{aligned} \tag{79}$$

Define $H_{j,i,t}^{b*}$ as an abridged expression for $\frac{1}{b!} \frac{1}{\sigma_0^*} H^b(\frac{y_{it} - \mu_0^{j*} - x_{it}' \beta_0^{j*} - z_{it}' \gamma^*}{\sigma_0^{j*}})$.
Define the weight w_i^{j*} as

$$w_i^{j*} = \frac{\alpha_0^{j*} f(\{\boldsymbol{W}_{it}\}_{t=1}^T; \gamma^*, \theta_0^{j*})}{f_{M_0}(\{\boldsymbol{W}_{it}\}_{t=1}^T; \vartheta_{M_0}^*)}, j = 1, \ldots, M_0, \tag{80}$$

where $f_{M_0}(\{\boldsymbol{W}_{it}\}_{t=1}^T; \vartheta_{M_0}^*)$ is defined by equation (20). As shown in section D.2, the score functions are:

$$s_{\alpha i} = \begin{pmatrix} \frac{f(\boldsymbol{w}|\theta_0^{1*}) - f(\boldsymbol{w}|\theta_0^{M_0*})}{\sum_l \alpha_0^{l*} f(\boldsymbol{w}|\theta_0^{l*})} \\ \vdots \\ \frac{f(\boldsymbol{w}|\theta_0^{M_0-1*}) - f(\boldsymbol{w}|\theta_0^{M_0*})}{\sum_l \alpha_0^{l*} f(\{\boldsymbol{W}_{it}^*\}_{t=1}^T|\theta_0^{l*})} \end{pmatrix}, s_{\mu i} = \begin{pmatrix} w_i^{1*} \sum_{t=1}^T H_{1,i,t}^{1*} \\ \vdots \\ w_i^{M_0*} \sum_{t=1}^T H_{M_0,i,t}^{1*} \end{pmatrix},$$

$$s_{\beta_i} = \begin{pmatrix} w_i^{1*} \sum_{t=1}^T H_{1,i,t}^{1*} x_{it} \\ \vdots \\ w_i^{M_0*} \sum_{t=1}^T H_{M_0,i,t}^{1*} x_{it} \end{pmatrix}, s_{\sigma i} = \begin{pmatrix} w_i^{1*} \sum_{t=1}^T H_{1,i,t}^{2*} \\ \vdots \\ w_i^{M_0*} \sum_{t=1}^T H_{M_0,i,t}^{2*} \end{pmatrix}, s_{\gamma i} = \begin{pmatrix} w_i^{1*} \sum_{t=1}^T H_{1,i,t}^{1*} z_{it} \\ \vdots \\ w_i^{M_0*} \sum_{t=1}^T H_{M_0,i,t}^{1*} z_{it} \end{pmatrix}; \tag{81}$$

and $\tilde{S}_{\lambda,\eta} = ((S_{\lambda,\eta}^1)^\top, \ldots, (S_{\lambda,\eta}^{M_0})^\top)^\top \sim N(0, \tilde{\boldsymbol{\mathcal{I}}}_{\lambda,\eta})$. $\tilde{S}_{\lambda,\eta}$ is an $\mathbb{R}^{M_0 q_\lambda}$-vector, and $(S_{\lambda,\eta}^h) \in \mathbb{R}^{q_\lambda}$, $q_\lambda = (q+2)(q+1)/2$. Define $S_\lambda^h = n^{-1/2} \sum_{i=1}^N s_{\lambda i}^h$. In addition, define $\boldsymbol{\mathcal{I}}_{\lambda,\eta}^h = E[S_{\lambda,\eta}^h (S_{\lambda,\eta}^h)']$. Define $W_{\lambda,\eta}^h = (\boldsymbol{\mathcal{I}}_{\lambda,\eta}^h)^{-1} S_{\lambda,\eta}^h$, and $W_{\lambda,\eta}^h \sim N(0, (I_{\lambda,\eta}^h)^{-1})$.

$$s_{\lambda_{\mu\sigma} i}^h = w_i^{h*} \begin{pmatrix} \sum_{t=1}^T H_{h,i,t}^{2*} + \frac{1}{2} \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} H_{h,i,s}^{1*} \\ 3 \sum_{t=1}^T H_{h,i,t}^{4*} + \frac{1}{2} \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{2*} H_{h,i,t}^{2*} \\ 3 \sum_{t=1}^T H_{h,i,t}^{3*} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} H_{h,i,s}^{2*} \\ 2 \sum_{t=1}^T H_{h,i,t}^{2*} x_{it} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it} H_{h,i,s}^{1*} \\ 3 \sum_{t=1}^T H_{h,i,t}^{3*} x_{it} + 2 \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it} H_{h,i,s}^{2*} \end{pmatrix},$$

$$s_{\lambda_\beta i}^h = w_i^{h*} \begin{pmatrix} \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,1}^2 + \frac{1}{2} \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,1} H_{h,i,s}^{1*} x_{is,1} \\ \vdots \\ \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,q}^2 + \frac{1}{2} \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,q} H_{h,i,s}^{1*} x_{is,q} \\ 2 \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,1} x_{it,2} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,1} H_{h,i,s}^{1*} x_{is,2} \\ \vdots \\ 2 \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,1} x_{it,q} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,1} H_{h,i,s}^{1*} x_{is,q} \\ 2 \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,2} x_{it,3} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,2} H_{h,i,s}^{1*} x_{is,3} \\ \vdots \\ 2 \sum_{t=1}^T H_{h,i,t}^{2*} x_{it,q-1} x_{it,q} + \sum_{t=1}^T \sum_{s \neq t} H_{h,i,t}^{1*} x_{it,q-1} H_{h,i,s}^{1*} x_{is,q} \end{pmatrix}. \tag{82}$$

E Other tables

Table 9: Parameter specification for null models with $M_0 = 2, 3, 4$

	$M_0 = 2$
N	$\{100, 500\}$
T	$\{2, 5, 10\}$
α	$\{(0.5, 0.5); (0.2, 0.8)\}$
μ	$\{(-1, 1), (-0.5, 0.5)\}$
σ	$\{(1, 1), (1.5, 0, 75)\}$
	$M_0 = 3$
N	$\{100, 500\}$
T	$\{2, 10\}$
α	$\{(1/3, 1/3, 1/3); (0.25, 0.5, 0.25)\}$
μ	$\{(-4, 0, 4); (-4, 0, 5); (-5, 0, 5); (-4, 0, 6); (-5, 0, 6); (-6, 0, 6)\}$
σ	$\{(1, 1, 1); (0.75, 1.5, 0.75)\}$
	$M_0 = 4$
N	$\{100, 500\}$
T	$\{2, 10\}$
α	$\{(0.25, 0.25, 0.25, 0.25)\}$
μ	$\{(-4, -1, 1, 4); (-5, -1, 1, 5); (-6, -2, 2, 6); (-6, -1, 2, 5); (-5, 0, 2, 4); (-6, 0, 2, 4)\}$
σ	$\{(1, 1, 1, 1); (1, 0.75, 0.5, 0.25)\}$
a_n	$(0.05, 0.1, 0.15, 0.2, 0.3, 0.4)$

Table 10: Criteria using simulation for Japanese producer in Machine industry ($10\%, 5\%, 1\%$)

	$H_0 : M_0 = 1$			$H_0 : M_0 = 2$			$H_0 : M_0 = 3$			$H_0 : M_0 = 4$			$H_0 : M_0 = 5$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	4.61	5.99	9.21	5.74	7.25	10.48	6.40	7.87	11.04	7.12	8.54	11.48	-	-	-
2	4.13	5.72	10.60	5.14	7.52	11.06	6.18	7.95	11.22	6.60	8.60	11.88	8.32	10.93	27.07
3	4.06	5.58	10.09	5.54	7.15	10.91	6.16	8.33	11.72	6.74	8.42	12.35	7.23	8.54	12.15
4	3.94	5.52	10.20	5.50	6.98	10.79	6.39	8.11	11.96	6.56	8.75	12.04	7.45	9.11	11.63
5	3.83	5.54	9.95	5.30	6.78	10.19	6.45	8.00	11.51	7.24	9.43	13.79	7.90	9.80	15.29

T represent panel length of each model. The table presents the simulated critical values for models of panel length 1 to 5.

Table 11: Criteria using simulation for Chilean producer in Machinery industry, except electrical($10\%, 5\%, 1\%$)

T	$H_0: M_0 = 1$			$H_0: M_0 = 2$			$H_0: M_0 = 3$			$H_0: M_0 = 4$			$H_0: M_0 = 5$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	4.61	5.99	9.21	5.73	7.08	10.28	-	-	-	-	-	-	-	-	-
2	4.13	5.43	9.4	5.45	7.18	10.79	6.14	7.98	11.23	6.75	8.72	11.15	7.5	9	12.32
3	3.82	5.32	9.5	5.36	7.18	10.22	6.2	8.01	11.23	6.97	8.71	12.19	7.12	8.68	12.19
4	3.83	5.31	9.36	5.31	7.35	10.19	6.26	7.85	12.18	6.82	8.45	11.61	7.11	8.53	12.02
5	3.82	5.32	9.31	5.31	7.19	10.61	6.29	7.85	11.38	6.83	8.46	11.65	7.05	8.76	11.65

T represent panel length of each model. The table presents the simulated critical values for models of panel length 1 to 5.

Table 12: Descriptive statistics for Japanses producer revenue share of intermediate material

Industry	Number of observations	Number of firms	Mean	Standard deviation
Wood Industry	18	12	-1.89	0.08
Chemical	1770	223	-1.18	0.5
Ceramics	597	89	-1.14	0.5
Other	522	108	-1.05	0.46
Food	138	146	-1.36	0.55
Othermetal	389	57	-0.75	0.35
Machine	2643	244	-0.78	0.48
Textile	481	83	-0.99	0.46
Paper	239	70	-0.7	0.24
Petro	25	7	-0.48	0.22
Steel	644	67	-0.54	0.2
Electronics	2422	250	-0.93	0.56
Transportation equipment	1002	144	-0.6	0.26
Precision instrument	504	58	-0.85	0.44
Metal product	345	90	-0.88	0.41
Plastic	235	23	-1.08	0.24

Table 13: Descriptive statistics for Chilean producer revenue share of intermediate material

Industry	Number of observations	Number of firms	Mean	Standard deviation
Wood products, except furniture	5148	1056	-0.5	0.36
Machinery, except electrical	2410	542	-0.78	0.44
Manufacture of furniture and fixtures	1990	454	-0.6	0.38
Transport equipment	1557	333	-0.69	0.48
Other chemicals	2792	376	-0.68	0.41
Printing and publishing	3021	461	-0.76	0.41
Other non-metallic mineral products	1824	328	-0.66	0.37
Fabricated metal products	5943	1156	-0.66	0.4
Textiles	5593	944	-0.58	0.37
Beverages	1704	277	-0.63	0.43
Machinery electric	943	207	-0.73	0.44
Paper and products	1029	189	-0.6	0.34
Wearing apparel, except footwear	4834	985	-0.57	0.37
Other manufactured products	893	198	-0.78	0.45
Food products	19375	2904	-0.38	0.29
Industrial chemicals	1018	194	-0.6	0.49
Footwear, except rubber or plastic	2233	373	-0.45	0.3
Non-ferrous metals	510	105	-0.54	0.45
Plastic products	3071	560	-0.61	0.36
Professional and scientific equipment	287	42	-0.85	0.44
Rubber products	875	136	-0.67	0.36
Manufacture of pottery, china and earthenware	206	43	-0.9	0.38
Leather products	842	144	-0.49	0.31
Iron and steel	562	148	-0.63	0.35
Animal feeds, etc	1059	184	-0.48	0.38
Tobacco	60	6	-1.12	0.89
Glass and products	327	52	-0.81	0.38
Misc. petroleum and coal products	229	37	-0.52	0.3
Petroleum refineries	80	11	-0.47	0.28

Table 14: Estimated LR for Japanese producer(T = 1)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$
Chemical	39.119***	4.971	-
Ceramics	1.898	-	-
Other	29.802***	2.267	-
Food	40.396***	11.700	-
Othermetal	31.324***	0.411	-
Textile	15.857***	1.185	-
Paper	5.485	-	-
Steel	15.221***	0.007	-
Electronics	122.642***	24.106	-
Transportation equipment	53.170***	19.220	-
Precision instrument	1.314	-	-
Metal product	9.373	-	-
Plastic	0.036	-	-

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.

Table 15: Estimated LR for Japanese producer(T = 2)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$	$H_0 : M_0 = 4$	$H_0 : M_0 = 5$
Chemical	226.387***	95.997***	95.445***	61.846***	33.470***
Ceramics	47.393***	19.939***	19.938***	9.257**	7.714
Other	118.481***	39.467***	29.780***	21.311***	6.454
Food	138.022***	96.038***	60.252***	42.777***	45.893***
Othermetal	21.011***	16.365***	7.532	-	-
Textile	65.113***	38.819***	22.184***	28.098***	14.090
Paper	40.466***	25.543	-	-	-
Steel	56.517***	12.050***	7.961*	2.776	-
Electronics	380.111***	146.134***	60.844***	48.359***	48.386***
Transportation equipment	171.980***	72.937***	51.851***	43.700***	41.408***
Precision instrument	27.780***	17.896***	13.987	-	-
Metal product	60.267***	32.172***	25.641***	19.194***	14.914***
Plastic	8.578**	19.782	-	-	-

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.

Table 16: Estimated LR for Japanese producer(T = 3)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$	$H_0 : M_0 = 4$	$H_0 : M_0 = 5$
Chemical	403.339***	194.093***	156.425***	113.422***	56.085***
Ceramics	91.292***	37.443***	36.815***	18.171***	14.047***
Other	192.819***	79.602***	50.880***	40.071***	15.492***
Food	241.678***	149.299***	117.284***	68.486***	68.738***
Othermetal	45.648***	37.073***	13.827***	8.808**	4.189
Textile	106.283***	65.965***	41.037***	33.179***	33.435***
Paper	69.602***	60.818***	37.951***	32.226***	21.972***
Steel	93.357***	28.474***	17.638***	6.991	-
Electronics	594.899***	246.828***	115.358***	102.100***	70.146***
Transportation equipment	296.068***	114.279***	105.383***	79.653***	69.531***
Precision instrument	60.226***	40.543***	28.038***	20.395***	17.378***
Metal product	111.281***	54.729***	40.948***	45.652***	24.149***
Plastic	19.777***	37.905***	17.842***	8.309**	3.064

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.

Table 17: Estimated LR for Chilean producer (T = 1)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$	$H_0 : M_0 = 4$
Wood products, except furniture	103.921***	22.883***	1.560	-
Machinery, except electrical	20.692***	0.880	-	-
Manufacture of furniture and fixtures, except primarily of metal	70.846***	1.559	-	-
Transport equipment	29.692***	5.579	-	-
Other chemicals	44.039	-	-	-
Printing and publishing	29.264***	1.151	-	-
Other non-metallic mineral products	37.482***	14.975***	1.339	-
Fabricated metal products	56.302***	8.206**	0.282	-
Textiles	94.435***	6.973*	5.475	-
Beverages	18.656***	6.749*	2.738	-
Paper and products	7.185*	2.060	-	-
Wearing apparel, except footwear	100.962***	4.568	-	-
Other manufactured products	6.737**	6.658*	1.305	-
Food products	359.541***	17.317***	2.496	-
Industrial chemicals	39.871***	10.223**	10.894**	3.018
Footwear, except rubber or plastic	42.393***	1.175	-	-
Plastic products	37.128***	8.802**	2.211	-
Animal feeds, etc	21.843**	2.440	-	-

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.

Table 18: Estimated LR for Chilean producer (T = 2)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$	$H_0 : M_0 = 4$	$H_0 : M_0 = 5$
Wood products, except furniture	186.880***	77.620***	31.323***	10.757**	3.219
Machinery, except electrical	83.806***	45.110***	18.532***	7.565*	2.932
Manufacture of furniture and fixtures, except primarily of metal	40.654***	22.990***	14.365***	9.141	-
Transport equipment	91.800***	22.827***	25.528***	8.512*	1.449
Other chemicals	110.303***	78.811***	25.833***	9.203*	10.592**
Printing and publishing	82.394***	49.556***	27.482***	7.673*	4.545
Other non-metallic mineral products	82.204***	42.037***	8.455	-	-
Fabricated metal products	213.535***	83.586***	25.287***	21.073***	12.655***
Textiles	239.855***	100.305***	37.081***	25.336***	20.301***
Beverages	25.391***	47.697***	6.319	-	-
Paper and products	33.963***	10.476**	3.936	-	-
Wearing apparel, except footwear	181.034***	72.182***	33.352***	8.833**	3.594
Other manufactured products	19.953***	7.878**	2.975	-	-
Food products	615.382***	361.304***	135.129***	73.252***	59.098***
Industrial chemicals	110.883***	27.371***	18.611***	5.613	-
Footwear, except rubber or plastic	92.782***	37.579***	17.387***	6.373	-
Plastic products	152.800***	62.834***	10.302**	3.242	-
Animal feeds, etc	73.527***	21.743***	12.148***	5.942	-

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.

Table 19: Estimated LR for Chilean producer (T = 3)

Industry	$H_0 : M_0 = 1$	$H_0 : M_0 = 2$	$H_0 : M_0 = 3$	$H_0 : M_0 = 4$	$H_0 : M_0 = 5$
Wood products, except furniture	297.521***	91.542***	39.790***	18.305***	17.992***
Machinery, except electrical	122.470***	54.215***	22.497***	14.706***	15.632***
Manufacture of furniture and fixtures, except primarily of metal	132.014***	76.074***	28.846***	17.551***	5.921
Transport equipment	139.085***	44.002***	33.166***	8.706**	4.025
Other chemicals	191.602***	120.986***	68.998***	25.401***	10.884**
Printing and publishing	167.840***	79.424***	42.941***	14.213***	10.327**
Other non-metallic mineral products	143.145***	80.821***	11.399***	7.398**	5.437
Fabricated metal products	376.798***	99.785***	41.379***	27.263***	36.049***
Textiles	366.744***	173.421***	65.305***	38.531***	40.812***
Beverages	45.504***	51.135***	10.587**	5.928	-
Paper and products	54.351***	28.462***	4.467	-	-
Wearing apparel, except footwear	310.464***	126.692***	59.537***	11.045**	12.253**
Other manufactured products	29.896***	23.569***	16.518***	2.239	-
Food products	995.387***	705.907***	222.406***	104.126***	82.330***
Industrial chemicals	145.085***	43.978***	19.462***	10.085**	3.062
Footwear, except rubber or plastic	192.388***	85.106***	25.934***	18.155***	7.143*
Plastic products	217.270***	90.007***	21.242***	9.314**	5.909
Animal feeds, etc	107.763***	33.931***	15.365***	7.818*	6.467

The estimation is based on revenue share of intermediate material. * indicate the result is significant at 10% level. ** indicate the result is significant at 5% level. *** indicate the result is significant at 1% level.