Estimating the Number of Components in Panel Data Finite Mixture Regression Models with an Application to Production Function Heterogeneity

Yu Hao*
Faculty of Business and Economics
The University of Hong Kong
haoyu@hku.hk

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

June 2026

Abstract

This paper develops likelihood-based methods for selecting the number of components M_0 in panel data finite mixture regression models. The panel structure ($T \geq 2$) eliminates the higher-order degeneracy of cross-sectional normal mixtures, enabling a tractable likelihood ratio test (LRT). We derive the LRT's asymptotic null distribution, propose a sequential bootstrap testing procedure, and establish BIC consistency; AIC is inconsistent. Applied to Chilean manufacturing plants, our headline specification uses the T=10 panel without correlated-random-effects augmentation and allows a two-component normal mixture residual (K_ϵ=2). It selects $\hat{M}$=4, 5, 2 latent technology types in Metal Products, Food Products, and Textiles, respectively. Per-type GNR point estimates reveal substantial heterogeneity in materials output elasticities, with Food spanning roughly 0.60–0.75 and Textiles 0.56–0.62 in the clean updated full-GNR fits; Food's mixture-weighted aggregate exceeds the pooled M=1 confidence interval. The heterogeneity resides in materials-intensity levels ($\beta_{0,j}$), not input-substitution curvature, consistent with Hicks-neutral technology differences that pooled estimators cannot recover.

*Address for correspondence: Yu (Jasmine) Hao, Faculty of Business and Economics, The University of Hong Kong. We thank the editor, Stéphane Bonhomme, and two anonymous referees whose insightful comments and suggestions substantially improved this paper. We are also grateful to Chun Pang Chow, Vadim Marmer, Kevin Song, Bin Chen, and participants at the 2019 IAAE Conference for their helpful discussions. Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) is gratefully acknowledged.

Keywords: finite mixture models, panel data, likelihood ratio test, Bayesian Information Criterion, production function heterogeneity, technology types

JEL Classification: C12, C23, D24, L60

1 Introduction

Finite mixture models provide a flexible and natural framework for representing unobserved heterogeneity across a finite number of latent classes. These models have found extensive applications across disciplines since the seminal two-component normal mixture model of Pearson (1894). In economics, they have been particularly influential in capturing unobserved individual-specific effects across labour economics, health economics, and industrial organization, with panel data settings featuring prominently.[1] Comprehensive surveys of the theoretical underpinnings and practical applications are provided by Titterington et al. (1985), Lindsay (1995), and McLachlan and Peel (2004).

A critical aspect in the application of finite mixture models is determining the appropriate number of components, which often corresponds to the number of latent individual types or abilities in economic models. Selecting an incorrect number of components leads to estimation biases, invalid identification, or spurious heterogeneity. Reliable statistical procedures for selecting M_0 are therefore essential for empirical analysis.

This paper investigates the statistical properties and practical performance of methods for determining the number of components in *panel data finite mixture regression models*, including the likelihood ratio test (LRT) and the Akaike and Bayesian Information Criteria (AIC and BIC). Panel data finite mixture models offer significant advantages over their cross-sectional counterparts: repeated observations per unit eliminate the higher-order degeneracy that makes the LRT intractable in cross-sectional normal mixtures. The existing literature on selecting the number of components in panel data finite mixture regression models is limited, particularly when the component-specific regression errors are normally distributed (a commonly assumed baseline model in regression analyses).

Specifically, this paper develops the LRT and BIC methods for consistently selecting the number of components in panel data finite mixture regression models and applies these methods to examine production function heterogeneity. We first consider a baseline scenario where regression errors are conditionally independent and normally distributed across periods, conditional on the latent type. We also analyze cases where regression errors follow more flexible, component-specific normal mixtures. Theoretical results in the main text concern the static (conditional-independence) model. We additionally consider

[1] For instance, Heckman and Singer (1984) utilize finite mixture models as an alternative method for handling unobserved heterogeneity in unemployment duration analysis. Similarly, Keane and Wolpin (1997) and Cameron and Heckman (1998) employ these models to analyze dynamic decision-making regarding schooling and occupational choices in the presence of unobserved heterogeneity in human capital. In health economics, Deb and Trivedi (1997) propose a finite mixture negative binomial model to capture the unobserved dispersion in healthcare utilization among elderly populations. Additionally, finite mixture models are extensively used in consumer segmentation within industrial organization, as demonstrated by Kamakura and Russell (1989) and Andrews and Currim (2003). Kasahara and Shimotsu (2009) and Hu and Shum (2012) establish the conditions under which finite mixture models are non-parametrically identified using panel data.

a nonparametric lower bound on M_0 via the rank test of Kleibergen and Paap (2006) and Kasahara and Shimotsu (2014).

Testing for the number of components in normal mixture regression models remains challenging. Standard asymptotic regularity conditions fail due to parameter non-identifiability, singular Fisher information, and boundary parameter issues. Numerous studies have addressed the likelihood ratio test (LRT) for finite mixture models (e.g., Ghosh and Sen, 1985; Chernoff and Lander, 1995; Lemdani and Pons, 1997; Chen and Chen, 2001, 2003; Chen et al., 2004; Garel, 2001, 2005; Chen et al., 2014) and derived its asymptotic distribution through Gaussian processes (Dacunha-Castelle and Gassiat, 1999; Liu and Shao, 2003; Zhu and Zhang, 2004; Azaïs et al., 2009). These approaches fail in *cross-sectional* normal regression models due to: (i) infinite Fisher information for testing, (ii) unbounded log-likelihood function, and (iii) linear dependence between density derivatives for mean and variance parameters.[2]

To the best of our knowledge, whether the aforementioned problems (i)–(iii) of the cross-sectional normal mixture still arise in panel data normal finite mixture models or their extensions remains unknown in the literature. No likelihood-based test has yet been developed for testing the null hypothesis of an M_0-component model against an alternative $(M_0 + 1)$-component model for $M_0 \geq 1$ in panel normal regression mixture models with conditionally independent errors.

We show that the higher-order degeneracy of problem (iii) disappears in panel normal mixture models with conditional independence. The problems related to (i) and (ii) persist. Consequently, the existing approaches in Dacunha-Castelle and Gassiat (1999), Liu and Shao (2003), Zhu and Zhang (2004), and Azaïs et al. (2009) do not directly apply to this class of panel normal mixture models. We impose bounds on the component-specific variance parameters to address unboundedness and on mixing proportions to address the infinite Fisher information issue. We then analyze the asymptotic distribution of the LRT using reparameterization orthogonal to the direction in which the Fisher information matrix is singular. The likelihood ratio of an $(M_0 + 1)$-component model against the M_0-component model is approximated with local quadratic-form expansion with squares and cross-products of the reparameterized parameters. We demonstrate that the asymptotic null distributions of the LRT statistic are characterized by the maximum of M_0 random variables. Building on the LRT tests, we propose a sequential hypothesis testing approach for consistently estimating the number of components.

This paper makes four contributions. First, we analyze the likelihood ratio test (LRT) for determining the number of components in panel data finite mixture regression models, where the regression errors follow component-specific normal mixtures. While Kasahara

[2]Chen and Li (2009), Chen et al. (2012), and Kasahara and Shimotsu (2015) analyze the LRT asymptotics for univariate models, with Kasahara and Shimotsu (2019) extending to multivariate cases. Additionally, Amengual et al. (2022) develop a score-type test for cross-sectional normal mixtures.

and Shimotsu (2015) and Kasahara and Shimotsu (2019) study the LRT for cross-sectional univariate and multivariate normal mixture regression models, respectively, we show that the asymptotic distribution of the LRT statistic in panel data differs due to the absence of higher-order dependencies when repeated outcome measurements are available under conditional independence.[3]

Second, the log-likelihood function of normal mixture models is unbounded (Hartigan, 1985), but whether this issue arises in panel data remains unexplored. We show that the likelihood ratio test (LRT) statistic is also unbounded in panel normal mixture models with conditionally independent errors when the time dimension is finite, though this issue diminishes as the time dimension grows. This unboundedness may lead to excessive rejection rates in the LRT, which we address by imposing explicit bounds on the component-specific variance parameters.

Third, we establish the consistency of the Bayesian Information Criterion (BIC) and the inconsistency of the Akaike Information Criterion (AIC) for selecting the true number of components in panel data finite mixture models. Our consistency result generalizes Keribin (2000) by relaxing its restrictive condition (P2) through higher-order rank conditions and utilizing a generalized form of Le Cam's differentiability in quadratic mean (DQM) framework (Liu and Shao, 2003; Kasahara and Shimotsu, 2018).

Fourth, we empirically analyze production technology heterogeneity using panel data from Chilean manufacturing plants in a two-stage procedure. In the first stage we select the number of latent technology types $\hat{M}$ using the sequential bootstrap LRT and BIC under three specifications: Cobb-Douglas, Translog (linear), and the Translog (exact). Our headline empirical specification is the T=10 no-CRE Translog (exact) model with a two-component within-type mixture error (K_ϵ=2), which selects $\hat{M}$=4, 5, 2 for Metal, Food, and Textiles. Normal-error and CRE specifications are used as diagnostics and robustness checks: Normal errors often require additional technology components to absorb residual non-normality, while CRE absorbs persistent plant-level heterogeneity through Mundlak controls. In the second stage, we apply Gandhi et al. (2020)'s two-step procedure conditional on the headline $\hat{M}$. This recovers type-specific elasticities of capital, labour, and materials; monotonicity projections enforce regularity conditions of production theory. The empirical analysis reveals within-industry variation in output elasticities of material inputs and in TFP persistence; constant-returns-to-scale holds within

[3]From a technical perspective, our panel data finite mixture models with normally distributed errors constitute a special case of the multivariate normal framework studied by Kasahara and Shimotsu (2019). The primary distinction lies in our assumption of conditional independence within each mixture component, which facilitates improved identification of model parameters. We show that this structure yields asymptotic null distributions for the likelihood ratio test that differ from those in Kasahara and Shimotsu (2019), mainly because our panel data setting does not involve higher-order singularities. Our extension to models with component-specific normal mixture errors is related to, but distinct from, standard multivariate normal models.

each type. This contrasts sharply with standard production function estimation methods, which typically impose homogeneous coefficients across plants (Olley and Pakes, 1996a; Levinsohn and Petrin, 2003a; Ackerberg et al., 2015). These findings point to the importance of accounting for unobserved plant heterogeneity beyond Hicks-neutral technological differences (Li and Sasaki, 2017; Doraszelski and Jaumandreu, 2018; Balat et al., 2019; Kasahara et al., 2022). Selecting the correct $\hat{M}$ is prerequisite: GNR assignments from an under-fitted mixture contaminate per-type moments. For Food Products, where the updated Translog (exact) Mixture selector gives $\hat{M}$=5, the pooled M=1 estimate cannot capture the high-materials-intensity types documented in the per-type GNR estimates (Table 5).

Our sequential testing approach is also related to panel grouping methods that classify units into latent groups without parametric distributional assumptions. The identification and estimation of such latent group structures in panel data has received attention in recent studies (Kasahara and Shimotsu, 2009; Lin and Ng, 2012; Bonhomme and Manresa, 2015; Ando and Bai, 2016; Su et al., 2016; Lu and Su, 2017). Finite mixture modeling provides a practical, model-based approach to determining unobserved group structures, where selecting the number of groups is a prerequisite for classifying each individual's group membership. We estimate the number of groups in panel data regression models by applying our proposed sequential hypothesis testing approach.

This paper closely relates to Kasahara et al. (2022), who nonparametrically identify and estimate the full production function under latent technology type heterogeneity — recovering per-type capital, labour, and materials elasticities and TFP dynamics — but *assume the number of types M_0 is known*. Our paper fills the logically prior gap: we develop LRT and BIC methods for selecting M_0 itself, a precondition for the Kasahara et al. (2022) identification strategy since an incorrect M propagates misclassified type assignments into all downstream parameter estimates. Two further distinctions hold. First, our empirical approach is *parametric* (Cobb-Douglas and Translog first-order conditions) while Kasahara et al. (2022) is fully nonparametric. Second, our scope is narrower: we identify materials-elasticity heterogeneity via the GNR share equation, not the full production structure; we therefore cannot address heterogeneity in predetermined-input elasticities or Hicks-neutral technological change. The two papers are complementary: ours answers "how many types?", enabling Kasahara et al. (2022) to answer "what are their parameters?"

The rest of this paper is organized as follows. In Section 2, we introduce several classes of panel data finite mixture regression models studied in this paper, along with empirical examples. In Section 4.2, we discuss the failure of regularity conditions when analyzing panel data finite mixture normal regression models. Section 4.3 analyzes the consistency of Maximum Likelihood Estimation (MLE). Section 5.1 analyzes the likelihood ratio test (LRT) for testing $H_0 : M_0 = 1$ against $H_1 : M_0 = 2$, while Section 5.2 considers the LRT for testing $H_0 : M_0 = M$ against $H_1 : M_0 = M + 1$ for $M \geq 2$. Section 5.3 develops a sequential

hypothesis testing procedure. Section 5.4 discusses the estimation of a lower bound for the number of components using the rank test. Section 6 presents simulation results, while Section 7 presents empirical analysis.

In what follows, all limits are taken as $n \to \infty$ unless otherwise stated. Let $:=$ denote "equals by definition." For a $k \times 1$ vector $\boldsymbol{a}$ and a function $f(\boldsymbol{a})$, let $\nabla_{\boldsymbol{a}} f(\boldsymbol{a})$ denote the $k \times 1$ vector of partial derivatives $(\partial/\partial \boldsymbol{a}) f(\boldsymbol{a})$, and let $\nabla_{\boldsymbol{a}\boldsymbol{a}^\top} f(\boldsymbol{a})$ denote the $k \times k$ matrix of second partial derivatives $(\partial/\partial \boldsymbol{a} \partial \boldsymbol{a}^\top) f(\boldsymbol{a})$. Let $||\cdot||$ denote the Euclidean norm. We adopt the convention that capitalized letters (e.g. $\boldsymbol{W}$) represent random variables and their lowercase counterparts (e.g. $\boldsymbol{w}$) denote evaluation points. Let $\phi(t) := (2\pi)^{-1/2} \exp(-t^2/2)$ denote the standard normal density.

2 Heteroskedastic Finite Mixture Panel Regression Model

We consider finite mixture regression models with panel data, where the panel length $T \geq 2$ is fixed, and the number of cross-sectional observations n tends to infinity. Let M_0 denote the true number of components in the data-generating process, while M represents a candidate number of components in our model specification. Throughout this paper, we use asterisks to denote true parameter values; for instance, $\boldsymbol{\vartheta}^*_{M_0}$ represents the true parameter vector.

For each unit i, we specify the conditional probability density function of $\{y_{it}\}_{t=1}^T$ given $\{\boldsymbol{x}_{it}\}_{t=1}^T$ as:

$$g_M(\{y_{it}\}_{t=1}^T \mid \{\boldsymbol{x}_{it}\}_{t=1}^T; \boldsymbol{\vartheta}_M) = \sum_{j=1}^M \alpha_j f_j(\{y_{it}\}_{t=1}^T \mid \{\boldsymbol{x}_{it}\}_{t=1}^T; \boldsymbol{\theta}_j), \tag{1}$$

where $y_{it} \in \mathbb{R}$ and $\boldsymbol{x}_{it} \in \mathbb{R}^q$; α_j represents the population proportion of the j-th component, and $f_j(\{y_{it}\}_{t=1}^T \mid \{\boldsymbol{x}_{it}\}_{t=1}^T; \boldsymbol{\theta}_j)$ is the component-specific conditional density function with parameter vector $\boldsymbol{\theta}_j$. We collect the parameters into a vector as $\boldsymbol{\vartheta}_M := (\boldsymbol{\alpha}^\top, \boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_M^\top)^\top \in \Theta_{\boldsymbol{\vartheta}_M}$, with $\boldsymbol{\alpha}^\top := (\alpha_1, \ldots, \alpha_{M-1})^\top$ and $\alpha_M = 1 - \sum_{j=1}^{M-1} \alpha_j$. The q-dimensional vector $\boldsymbol{x}_{it}$ may include both time-varying and time-invariant regressors. It can also accommodate correlated random effects specifications (Mundlak, 1978; Chamberlain, 1984) by including individual-specific averages or the complete sequence $\{\boldsymbol{x}_{it}\}_{t=1}^T$; see Table B2 for CRE sensitivity.

The component-specific density $f_j(\{y_{it}\}_{t=1}^T \mid \{\boldsymbol{x}_{it}\}_{t=1}^T; \boldsymbol{\theta}_j)$ in (1) assumes that outcomes $\{y_{it}\}_{t=1}^T$ are conditionally independent over time given covariates and component membership as:

$$f_j(\{y_{it}\}_{t=1}^T \mid \{\boldsymbol{x}_{it}\}_{t=1}^T; \boldsymbol{\theta}_j) = \prod_{t=1}^T f_{jt}(y_{it} \mid \boldsymbol{x}_{it}; \boldsymbol{\theta}_j), \tag{2}$$

where the period-specific density takes the form:

$$f_{jt}(y_{it} \mid \boldsymbol{x}_{it}; \boldsymbol{\theta}_j) = \sum_{k=1}^{K_\epsilon} \tau_{jk} \frac{1}{\sigma_j} \phi\left(\frac{y_{it} - \mu_{jk} - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_j}{\sigma_j}\right). \tag{3}$$

When $K_\epsilon = 1$, the density in (3) is normal with $\boldsymbol{\theta}_j := (\mu_j, \sigma_j^2, \boldsymbol{\beta}_j^\top)^\top \in \Theta_{\boldsymbol{\theta}}$, where $\sigma_j^2 \in \mathbb{R}_{++}$ and $\boldsymbol{\beta}_j \in \Theta_{\boldsymbol{\beta}}$. When $K_\epsilon \geq 2$, it is a finite mixture of normals with $\boldsymbol{\theta}_j := (\boldsymbol{\tau}_j, \boldsymbol{\mu}_j, \sigma_j^2, \boldsymbol{\beta}_j^\top)^\top \in \Theta_{\boldsymbol{\theta}}$, where $\boldsymbol{\tau}_j = (\tau_{j1}, \ldots, \tau_{j,K_\epsilon-1})^\top$ with $\tau_{jK_\epsilon} := 1 - \sum_{k=1}^{K_\epsilon-1} \tau_{jk}$, and $\boldsymbol{\mu}_j = (\mu_{j1}, \ldots, \mu_{jK_\epsilon})^\top$ satisfies $\mu_{j1} < \cdots < \mu_{jK_\epsilon}$. The true values of τ_{jk} and μ_{jk} are denoted τ_{jk}^* and μ_{jk}^*, following the asterisk convention above. In dynamic panel AR(1) models (Remark 1), the initial period $t = 1$ has separate sub-component means $\mu_{1,jk}$ collected in $\boldsymbol{\mu}_{1,j} = (\mu_{1,j1}, \ldots, \mu_{1,jK_\epsilon})^\top$, which may differ from the stationary means μ_{jk} ($t \geq 2$).

We assume that the number of sub-components $K_\epsilon \geq 1$ within each mixture component is known to the researcher and that all true mixing proportions are strictly positive.[4]

Assumption 1. *K_ϵ is known to the researcher. In case of $K_\epsilon \geq 2$, $\tau_{jk}^* > 0$, $\mu_{jk} \neq \mu_{j\ell}$, and $\mu_{1,jk} \neq \mu_{1,j\ell}$ if $k \neq \ell$ for $j = 1, \ldots, M$ and $k, \ell = 1, \ldots, K_\epsilon$.*

The true number of components, M_0, is defined as the smallest integer M such that the density admits a representation of the form (1).

3 Production Function Heterogeneity

We use Chilean manufacturing data to ask: *how many distinct production technologies are required to explain observed dispersion in material revenue shares?* We investigate unobserved heterogeneity in output elasticities of the flexible material input, beyond Hicks-neutral productivity. The identifying restriction shared by all three specifications is the materials first-order condition, which eliminates Hicks-neutral productivity ω_{it} from the regression.

3.1 Data and common setup

Panel data structure. We observe a panel of N plants over T years. For plant i at time t the observable vector is $(Y_{it}, V_{it}, L_{it}, K_{it}, Z_{it})$, where Y_{it} is gross output, V_{it} intermediate materials, L_{it} labour, K_{it} capital, and Z_{it} a vector of additional plant or industry characteristics. Lower-case variables denote natural logs: $y_{it} = \ln Y_{it}$, $v_{it} = \ln V_{it}$, etc. Let P_t^Y be the output price, P_t^V the materials price index, and P_t^L the wage rate.

[4] Testing the number of components K_ϵ simultaneously with the number of components M is certainly an important issue but beyond the scope of this paper, and left for future research.

Timing and decision structure. Capital K_{it} is predetermined at the start of period t. The Hicks-neutral productivity shock ω_{it} is realised before the firm chooses flexible inputs. Following Gandhi et al. (2020), material is chosen after ω_{it} is observed but before an *ex post* shock to output ϵ_{it} is observed.

We formalise this timing protocol as follows.

Assumption 2 (Timing and Scalar Unobservable)**.** *(i)* ***Scalar unobservable:*** *The Hicks-neutral productivity ω_{it} is a scalar, realised and observed by the firm at the beginning of period t, before any variable input decisions are made.*

(ii) ***Predetermined capital, labour, and other plant characteristics:*** *L_{it}, K_{it}, and Z_{it} are chosen at $t-1$ and is not adjustable within period t.*

(iii) ***Flexible inputs and unanticipated shocks:*** *Material V_{it} is chosen after ω_{it} is observed, conditional on $(\omega_{it}, L_{it}, K_{it}, Z_{it}, P_t^V, P_t^L, P_t^Y)$. An i.i.d. unanticipated ex post shock ϵ_{it} is then realised, independent of the firm's full information set: $\epsilon_{it} \perp\!\!\!\perp (\omega_{it}, V_{it}, L_{it}, K_{it}, Z_{it}, P_t^V, P_t^L, P_t^Y)$. The distribution of ϵ_{it} may be different across latent types.*

Assumption 2 follows the standard timing protocol in the production function literature (Olley and Pakes, 1996b; Levinsohn and Petrin, 2003b; Gandhi et al., 2020). Under Assumption 2(ii)–(iii), the predetermined inputs $(\ell_{it}, k_{it}, z_{it})$ are determined at or before $t-1$.[5]

Latent Technology Types. Let $D_i \in \{1, 2, \ldots, M\}$ denote the latent technology type of plant i, which determines the production function parameters and shock processes. Define the regressor vector $\boldsymbol{x}_{it} = (k_{it}, \ell_{it}, z_{it})^\top$.

3.2 Production function specification with Hicks-neutral technology

We consider three production function specifications: Cobb-Douglas, Translog (linear), and Translog (exact). The Translog (exact) uses the exact first-order condition derived from the Translog production function, estimated with $K_\epsilon = 2$ mixture errors.

3.2.1 Cobb-Douglas

Under Cobb-Douglas technology,

$$y_{it} = \beta_{0,j} + \beta_{K,j} k_{it} + \beta_{L,j} \ell_{it} + \beta_{V,j} v_{it} + \omega_{it} + \epsilon_{it}. \tag{4}$$

[5] Alternatively, one can assume L_{it} is also a flexible input with differential timing relative to V_{it}, as in Ackerberg et al. (2015); this variant is left for future work.

The first-order condition for profit-maximising material choice yields[6]

$$s_{it}^V := \ln\left(\frac{P_t^V V_{it}}{P_t^Y Y_{it}}\right) = \ln\beta_{V,j} + \ln\mathbb{E}[e^{\epsilon_{it}}|D_i = j] - \epsilon_{it} =: c_{V,j} + u_{it}, \tag{5}$$

where $c_{V,j} := \ln\beta_{V,j} + \ln\mathbb{E}[e^{\epsilon_{it}}|D_i = j]$ is a type-specific constant and $u_{it} := -\epsilon_{it}$. Here, ϵ_{it} (and hence u_{it}) is serially independent over time under Assumption 2(iii), which is the standard timing assumption maintained throughout the proxy-variable production function literature (Olley and Pakes, 1996a; Levinsohn and Petrin, 2003a; Ackerberg et al., 2015; Gandhi et al., 2020).

To allow for heterogeneous technologies, we consider the finite mixture model:

$$s_{it}^V = c_{V,j} + u_{it}, \quad j \sim \text{Discrete}(\alpha_1, \ldots, \alpha_M), \tag{6}$$

where u_{it} follows either a normal or normal mixture distribution, and the number of types M is selected using the methods in Sections 5.1–5.3. The Cobb-Douglas restriction that material shares are constant within each type (up to the industry shifters $z_{it}'\gamma$) is testable against the Translog alternative (Section 3.2.2). The Cobb-Douglas results serve as a benchmark; the Translog (linear) and the Translog (exact) specification are our preferred specifications.

3.2.2 Translog (linear approximation)

The Translog specification allows output elasticities to vary with input levels:

$$\begin{aligned} y_{it} = {} & \beta_{0,j} + \beta_{K,j}k_{it} + \beta_{L,j}\ell_{it} + \beta_{V,j}v_{it} + \beta_{KK,j}k_{it}^2 + \beta_{LL,j}\ell_{it}^2 + \beta_{VV,j}v_{it}^2 \\ & + \beta_{KL,j}k_{it}\ell_{it} + \beta_{KV,j}k_{it}v_{it} + \beta_{LV,j}\ell_{it}v_{it} + \omega_{it} + \epsilon_{it}. \end{aligned} \tag{7}$$

The first-order condition yields

$$s_{it}^V = \ln\left(\beta_{V,j} + \beta_{VK,j}k_{it} + \beta_{VL,j}\ell_{it} + \beta_{VV,j}v_{it}\right) + \ln\mathbb{E}[e^{\epsilon_{it}}|D_i = j] - \epsilon_{it}. \tag{8}$$

Estimating the exact first-order condition in (8) directly within a finite mixture framework presents severe computational challenges. The argument inside the logarithm must be

[6] Following Gandhi et al. (2020), this specification is obtained from a static expected profit maximisation problem:

$$\max_{V_{it}\geq 0} \Pi_{it}(V_{it}; j) = \max_{V_{it}\geq 0}\left\{P_t^Y B_{it,j} V_{it}^{\beta_{V,j}}\mathbb{E}[e^{\epsilon_{it}} \mid D_i = j] - P_t^V V_{it}\right\}$$

where $B_{it,j} := \exp(\beta_{0,j})K_{it}^{\beta_{K,j}}L_{it}^{\beta_{L,j}}\exp(\omega_{it})$.

strictly positive for all observations across all latent types j during numerical optimization. This global positivity constraint is notoriously difficult to impose in a mixture model. As a result, coefficient estimates are highly sensitive to the extreme realized values of k_{it}, ℓ_{it}, and v_{it}.

To circumvent these numerical instabilities, we consider an empirical specification (the *linear approximation*) based on the first-order Taylor expansion of (8) around the sample means $(\bar{k}, \bar{\ell}, \bar{v})$:

$$s_{it}^{V} = c_{V,j} + \beta_{VK,j}^{*}\tilde{k}_{it} + \beta_{VL,j}^{*}\tilde{\ell}_{it} + \beta_{VV,j}^{*}\tilde{v}_{it} + u_{it}, \tag{9}$$

where $\beta_{VX,j}^{*} := \beta_{VX,j}/\bar{\beta}_{V,j}$ with $\bar{\beta}_{V,j} := \beta_{V,j} + \beta_{VK,j}\bar{k} + \beta_{VL,j}\bar{\ell} + \beta_{VV,j}\bar{v}$ for $X \in \{K, L, V\}$, $\tilde{x}_{it} := x_{it} - \bar{x}$ are sample-demeaned inputs, and $c_{V,j} := \ln\bar{\beta}_{V,j} + \ln\mathbb{E}[e^{\epsilon_{it}}|D_i = j]$ is the log material elasticity evaluated at the sample means. The error term u_{it} captures the ex-post shock $-\epsilon_{it}$ and the linearization approximation error.

This specification is more flexible than the Cobb-Douglas baseline. It allows technology types to differ not only in average material intensity but also in how material shares respond to variations in capital, labour, and material inputs. The coefficient $\beta_{VK,j}^{*}$ is the normalized interaction parameter from the Translog production function (7).[7] It measures how the output elasticity of materials varies with capital intensity. A negative $\beta_{VK,j}^{*}$ indicates that capital-intensive plants use materials less intensively at the margin, consistent with capital–material substitutability.

This specification has two limitations. Computing the full set of Translog parameters requires estimating the remaining production function coefficients (including $\beta_{K,j}$, $\beta_{KK,j}$, and $\beta_{KL,j}$) via the second step of the Gandhi et al. (2020) non-parametric framework under latent heterogeneity. This sequential estimation is beyond the scope of the current paper; see Kasahara et al. (2022) for a related implementation. Second, the linearization residual is potentially heteroskedastic and correlated with the regressors due to an approximation error. We address this approximation concern through the *Translog (exact)* specification below.

3.2.3 Translog (exact)

The Translog (exact) specification estimates the first-order condition (8) directly within the EM framework, bypassing the linear approximation. Following the same grand-mean centring used in the linear approximation, we set $\tilde{x}_{it} := x_{it} - \bar{x}$ for $x \in \{k, \ell, v\}$. The material elasticity function $\varepsilon_{M,j}(\tilde{k}, \tilde{\ell}, \tilde{v}) = \beta_{0,j} + \beta_{VK,j}\tilde{k}_{it} + \beta_{VL,j}\tilde{\ell}_{it} + \beta_{VV,j}\tilde{v}_{it}$ then enters the share equation through its logarithm:

$$s_{it}^{V} = \ln\big(\beta_{0,j} + \beta_{VK,j}\tilde{k}_{it} + \beta_{VL,j}\tilde{\ell}_{it} + \beta_{VV,j}\tilde{v}_{it}\big) + u_{jit}, \tag{10}$$

[7] By Young's theorem on the symmetry of second-order partial derivatives, $\beta_{VK,j} = \beta_{KV,j}$.

where u_{jit} follows a K_ϵ-component mean-zero normal mixture (with K_ϵ=2 in the baseline). The intercept $\beta_{0,j} \equiv \varepsilon_{M,j}(0,0,0)$ is the material output elasticity evaluated at the industry grand means $(\bar{k}, \bar{\ell}, \bar{v})$, i.e., the average elasticity for type-j plants at a representative input bundle. Grand-mean centring plays two roles here. First, it ensures $\beta_{0,j}$ is identified from within-sample variation rather than an extrapolation to zero log inputs—a point far outside the data range—so parameter estimates are numerically stable across industries. Second, it directly links the exact and linear-approximation parameters: $c_{V,j} \approx \ln \beta_{0,j}$ and $\beta^*_{VX,j} \approx \beta_{VX,j}/\beta_{0,j}$, confirming that both specifications recover the same economic object (the material elasticity surface) while differing only in whether the log transformation is applied before or after the Taylor expansion.

4 Regularity Failures and Restricted Maximum Likelihood Estimation

4.1 Data-Generating Process

We consider a random sample of n observations $\{\boldsymbol{W}_i\}_{i=1}^n$, where $\boldsymbol{W}_i = \{(Y_{it}, \boldsymbol{X}_{it}^\top)^\top\}_{t=1}^T$ is drawn from an M_0-component density $g_{M_0}(\boldsymbol{w}; \boldsymbol{\vartheta}^*_{M_0})$ defined in equation (11):

$$g_{M_0}(\boldsymbol{w}; \boldsymbol{\vartheta}^*_{M_0}) = \sum_{j=1}^{M_0} \alpha^*_j f_j(\boldsymbol{w}; \boldsymbol{\theta}^*_j), \tag{11}$$

where $\boldsymbol{\vartheta}^*_{M_0} = (\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top}, \ldots, \boldsymbol{\theta}_{M_0}^{*\top}, \alpha^*_1, \ldots, \alpha^*_{M_0-1})^\top \in \Theta_{\boldsymbol{\vartheta}_{M_0}}$ and $\alpha^*_{M_0} = 1 - \sum_{j=1}^{M_0-1} \alpha^*_j$. For identification, we assume $\mu^*_1 < \mu^*_2 < \cdots < \mu^*_{M_0}$ in the true parameters.

In Section 3, M_0 represents the number of latent technology types, reflecting plant-level heterogeneity in output elasticities. The central inferential goal is to determine M_0; the next subsection examines why standard regularity conditions for this testing problem fail in the normal mixture setting.

4.2 The Testing Problem and Regularity Failures

To examine the failure of the regularity conditions of the likelihood ratio test statistic (LRTS) in panel finite mixture models, consider testing the null hypothesis $H_0 : M_0 = 1$ against the alternative hypothesis $H_1 : M_0 = 2$, where $\boldsymbol{W}_i = \{(Y_{it}, \boldsymbol{X}_{it}^\top)^\top\}_{t=1}^T$ is drawn from a true one-component density $f(\boldsymbol{w}; \boldsymbol{\theta}^*) = \prod_{t=1}^T f(y_t \mid \boldsymbol{x}_t; \boldsymbol{\theta}^*)$.

The two-component model

$$g_2(\boldsymbol{w}; \vartheta_2) = \alpha f(\boldsymbol{w}; \boldsymbol{\theta}_1) + (1-\alpha) f(\boldsymbol{w}; \boldsymbol{\theta}_2) \tag{12}$$

can generate the true one-component density in two cases: (1) $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$ with $\alpha \in (0,1)$, and (2) $\alpha \in \{0,1\}$ with arbitrary $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. Consequently, $H_0 : M_0 = 1$ can be partitioned into two sub-hypotheses: $H_{01} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ and $H_{02} : \alpha(1-\alpha) = 0$. The regularity conditions of the LRTS fail under both sub-hypotheses: under H_{01}, α is not identified and the Fisher information matrix for the other parameters becomes singular; under H_{02}, α is on the boundary of the parameter space and either $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$ is not identified.

Three Fundamental Challenges in Normal Mixtures Analyzing the asymptotic distribution of the LRTS for *cross-sectional* normal mixtures (i.e., with $T = 1$) is notoriously challenging due to three fundamental problems (cf., Chen and Li, 2009):

(i) **Infinite Fisher Information:** The Fisher information for testing H_{02} is not finite.

(ii) **Unbounded Likelihood:** The log-likelihood function is unbounded (Hartigan, 1985).

(iii) **Linear Dependence of Derivatives:** The first-order derivative of $g_2(\boldsymbol{w}; \boldsymbol{\vartheta}_2)$ with respect to σ_j^2 is linearly dependent on its second-order derivative with respect to μ_j.

A central question is whether problems (i)–(iii) persist in panel normal mixture models with $T \geq 2$. We now examine each problem systematically.

Problem (i): Infinite Fisher Information Persists The issue of infinite Fisher information for testing $H_{02} : \alpha = 0$ arises in panel normal mixture models. The score for testing $H_{02} : \alpha = 0$ takes the form

$$\left.\frac{\partial \log g_2(\boldsymbol{W}; \alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \alpha}\right|_{\alpha=0, \boldsymbol{\theta}_2=\boldsymbol{\theta}^*} = \frac{f(\boldsymbol{W}; \boldsymbol{\theta}_1)}{f(\boldsymbol{W}; \boldsymbol{\theta}^*)} - 1, \tag{13}$$

where $f(\boldsymbol{W}; \boldsymbol{\theta}) = \prod_{t=1}^{T} \phi((Y_t - \mu - \boldsymbol{X}_t^\top \boldsymbol{\beta})/\sigma)/\sigma$. When $\sigma_1^2 > 2\sigma^{*2}$, $\mathbb{E}[\{f(\boldsymbol{W}; \boldsymbol{\theta}_1)/f(\boldsymbol{W}; \boldsymbol{\theta}^*) - 1\}^2] = \infty$. This infinite Fisher information also arises for the normal mixture density (2)-(3); see Proposition 5.

Implication: We focus on testing H_{01} by restricting $\alpha \in [c_1, 1 - c_1]$ for some $c_1 > 0$, sidestepping the infinite Fisher information under H_{02}. Appendix A.1 treats H_{02} under $\sigma_1^2 \leq 2\sigma^{*2} - c$ for some $c > 0$ following Andrews (2001).

Problem (ii): Unbounded Likelihood Persists The LRTS in normal mixture models with panel data becomes unbounded as the sample size n goes to ∞. Define the likelihood ratio statistic with respect to the true parameter under H_0 as

$$LR_n^*(\boldsymbol{\vartheta}_2) := 2\left\{\sum_{i=1}^{n} \log g_2(\boldsymbol{W}_i; \boldsymbol{\vartheta}_2) - \sum_{i=1}^{n} \log f(\boldsymbol{W}_i; \boldsymbol{\theta}^*)\right\}, \tag{14}$$

where g_2 is the density of the two-component finite mixture distribution in (1) with $M = 2$ and $\boldsymbol{\theta}^*$ is the true parameter value under H_0.

Proposition 1. *Suppose that the true model is described by the one-component model with $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ and that the parameter space for σ_j^2 is $\Theta_{\sigma^2} = \mathbb{R}_{++}$. Then, for any positive constant $\kappa > 0$, when the component-specific density function is given by* (2) *with* (3)*,*

$$\Pr\left(\sup_{\boldsymbol{\vartheta}_2 \in \Theta_{\boldsymbol{\vartheta}_2}} LR_n^*(\boldsymbol{\vartheta}_2) < \kappa\right) \to 0 \quad \text{as } n \to \infty. \tag{15}$$

In particular, the probability decays at exponential rate $\exp(-C_T n^{1/T})$ for positive constants C_T depending on κ and T (Online Appendix Lemma 2).

Remark 1 (Static vs. dynamic panel)**.** *Proposition 1 concerns the static panel, where all T periods share a common variance σ_j^2. Taking $\sigma_j^2 \to 0$ generates cross-product terms $\sum_{s \neq t} H_t^{k*} H_s^{k*}$ across periods, which slow the divergence rate to $O_p(n^{1-T/2})$—bounded for $T \geq 2$ (Online Appendix Lemma 2). In the dynamic panel AR(1) model, the initial-period variance $\sigma_{1,j}^2$ is separate from the transition variance σ_j^2 ($t \geq 2$). Taking $\sigma_{1,j}^2 \to 0$ while holding σ_j^2 fixed eliminates the compensating cross-period terms, so the likelihood ratio is unbounded as in the cross-sectional case. The non-compactness assumption on Θ_{σ^2} is therefore essential for this result; see Section 4.3 for the compactness restriction used in practice.*

Problem (iii): Linear Dependence Resolved for $T \geq 2$ Consider the two-component panel mixture model without covariates:

$$\begin{aligned} g_2(\{y_t\}_{t=1}^T; \alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \alpha f(\{y_t\}_{t=1}^T; \boldsymbol{\theta}_1) + (1-\alpha) f(\{y_t\}_{t=1}^T; \boldsymbol{\theta}_2) \quad \text{with} \\ f(\{y_t\}_{t=1}^T; \boldsymbol{\theta}_j) &= \prod_{t=1}^T \frac{1}{\sigma_j} \phi\left(\frac{y_t - \mu_j}{\sigma_j}\right) \quad \text{for } j = 1, 2, \end{aligned} \tag{16}$$

where $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2)^\top$. When evaluated under the null hypothesis $H_{01} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$ with $\alpha(1-\alpha) \neq 0$, the first-order derivative of $g_2(\{y_t\}_{t=1}^T; \alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ with respect to σ_j^2 is *not* linearly dependent on its second-order derivative with respect to μ_j.

Proposition 2. *For any $j \in \{1, 2\}$, any $a, b \in \mathbb{R}$, and any $\bar{\alpha} \in [c, 1-c]$ for some constant $c > 0$, when $T \geq 2$:*

$$\Pr\left(\left.\frac{\partial^2 g_2(\{Y_t\}_{t=1}^T; \alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{(\partial \mu_j)^2}\right|_{(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\bar{\alpha}, \boldsymbol{\theta}^*, \boldsymbol{\theta}^*)} = a + b \left.\frac{\partial g_2(\{Y_t\}_{t=1}^T; \alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \sigma_j^2}\right|_{(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\bar{\alpha}, \boldsymbol{\theta}^*, \boldsymbol{\theta}^*)}\right) = 0. \tag{17}$$

Implication: The panel mixture model (1) with (2)-(3) is *strongly identifiable* (Chen, 1995): the best convergence rate for the mixing distribution is $n^{-1/4}$ (vs. $n^{-1/8}$ for cross-sectional normal mixtures; Kasahara and Shimotsu, 2015). The key is the product structure of the component density across time periods under conditional independence.

The Remaining Challenge: Non-Identification of α Under H_{01} The standard asymptotic analysis breaks down in testing $H_{01} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$ because α is not identified under H_{01}. The first-order derivatives at the true value $\boldsymbol{\vartheta}_2^* = (\alpha, (\boldsymbol{\theta}^*)^\top, (\boldsymbol{\theta}^*)^\top)^\top$ are also linearly dependent:

$$\nabla_{\boldsymbol{\theta}_1} \log g_2(\boldsymbol{w}; \boldsymbol{\vartheta}_2^*) = \frac{\alpha}{1-\alpha} \nabla_{\boldsymbol{\theta}_2} \log g_2(\boldsymbol{w}; \boldsymbol{\vartheta}_2^*). \tag{18}$$

Consequently, we derive the asymptotic distribution of the LRTS using a fourth-order Taylor series approximation for the log-likelihood function.

4.3 Restricted MLE and Consistency

To address unboundedness, following Hathaway (1985), we restrict the parameter space so that component-specific variances are bounded away from zero relative to the unconditional variance: $\min_j \left(\frac{\sigma_j^2}{\sum_{k=1}^M \alpha_k \sigma_k^2} \right) \geq c_2$ for some constant $c_2 > 0$. This constraint enforces a positive lower bound on the component variances when the unconditional variance is strictly positive with $\sum_{k=1}^M \alpha_k \sigma_k^2 \geq c_3 > 0$. We also assume that α_j is bounded away from 0 and 1 by a small constant $c_1 > 0$ so that $\boldsymbol{\alpha} \in \Theta_{\boldsymbol{\alpha}, c_1} := \{\boldsymbol{\alpha} : \alpha_j \in [c_1, 1-c_1], \sum_{j=1}^{M-1} \alpha_j \leq 1-c_1\}$. This sidesteps the infinite Fisher information under H_{02} and allows us to focus on H_{01}. Let $\boldsymbol{c} = (c_1, c_2, c_3)^\top$, and define a restricted parameter space for $\boldsymbol{\vartheta}_M$ for a model with the component-specific density function (2) with (3) by

$$\bar{\Theta}_{\boldsymbol{\vartheta}_M}(\boldsymbol{c}) := \{\boldsymbol{\vartheta}_M \in \Theta_{\boldsymbol{\vartheta}_M} : \boldsymbol{\alpha} \in \Theta_{\boldsymbol{\alpha}, c_1},\ \min_j \sigma_j^2 \geq c_2 \sum_{k=1}^M \alpha_k \sigma_k^2,\ \sum_{k=1}^M \alpha_k \sigma_k^2 \geq c_3\}.$$

We consider the MLE of $\boldsymbol{\vartheta}_M$ over the restricted parameter space $\bar{\Theta}_{\boldsymbol{\vartheta}_M}(\boldsymbol{c})$ for the M-component model:

$$\ell_n^M(\widehat{\boldsymbol{\vartheta}}_M) = \sup_{\boldsymbol{\vartheta}_M \in \bar{\Theta}_{\boldsymbol{\vartheta}_M}(\boldsymbol{c})} \ell_n^M(\boldsymbol{\vartheta}_M), \tag{19}$$

where $\ell_n^M(\boldsymbol{\vartheta}_M) = \sum_{i=1}^n \log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M)$ is the log-likelihood function. The MLE $\widehat{\boldsymbol{\vartheta}}_M$ may not be unique when M exceeds the true number of components.

For $M > M_0$, define the set of parameter values that yield the true M_0-component

density as

$$\Theta^*_{\vartheta_M} := \{\vartheta_M \in \Theta_{\vartheta_M} : g_M(\boldsymbol{w}; \vartheta_M) = g_{M_0}(\boldsymbol{w}; \vartheta^*_{M_0}) \text{ for all } \boldsymbol{w} \in \mathcal{W}\}.$$

For instance, if $M_0 = 1$ with true density $f(\boldsymbol{w}; \boldsymbol{\theta}^*)$, then

$$\Theta^*_{\vartheta_2} = \{(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta_{\vartheta_2} : (\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*) \text{ or } (\alpha = 1, \boldsymbol{\theta}_1 = \boldsymbol{\theta}^*) \text{ or } (\alpha = 0, \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*)\},$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are component-specific parameters. Under the restricted parameter space used for the LRT, the relevant part of this set is the component-merging representation $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}^*$ with α bounded away from zero and one. The following proposition establishes the consistency of the MLE in over-specified mixture models.

Assumption 3. *(a) $T \geq 2$ and $\boldsymbol{X}$ has finite second moments and $Pr(\boldsymbol{X}^\top \boldsymbol{\beta}_j \neq \boldsymbol{X}^\top \boldsymbol{\beta}^*_j) > 0$ for $\boldsymbol{\beta}_j \neq \boldsymbol{\beta}^*_j$ for $j = 1, 2, ..., M_0$. (b) For finite $M > M_0$, $\Theta^*_{\vartheta_M} \cap \bar{\Theta}_{\vartheta_M}(\boldsymbol{c})$ is non-empty. (c) $\bar{\Theta}_{\vartheta_M}(\boldsymbol{c})$ is compact.*

Proposition 3. *Suppose that Assumptions 1 and 3 hold and that the data are generated from M_0-component models with the component-specific density function* (2) *with* (3)*. Then, for any finite $M > M_0$, $\inf_{\vartheta_M \in \Theta^*_{\vartheta_M}} ||\widehat{\vartheta}_M - \vartheta_M|| \to 0$ almost surely.*

Thus, $\widehat{\vartheta}_M$ converges almost surely to a parameter set that reproduces the true density $g_{M_0}(\boldsymbol{w}; \boldsymbol{\theta}^*_{M_0})$ within the overspecified M-component family.

5 Testing for the Number of Components

5.1 Likelihood ratio test for $H_0 : M_0 = 1$ against $H_1 : M_0 = 2$

We first consider testing $H_0 : M_0 = 1$ against $H_1 : M_0 = 2$. Let $\widehat{\boldsymbol{\theta}}_0$ be the one-component MLE and define the LRTS as

$$LR_n := 2\left\{\ell^2_n(\hat{\vartheta}_2) - \ell^1_n(\hat{\boldsymbol{\theta}}_0)\right\}, \tag{20}$$

where $\hat{\vartheta}_2$ is the MLE under the restricted parameter space defined by (19).

The asymptotic distribution of the LRTS is non-standard: the Fisher information matrix is singular under the null hypothesis. We reparameterize $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ following Kasahara and Shimotsu (2012):

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\nu} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \\ \alpha\boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2 \end{pmatrix}, \tag{21}$$

where $\boldsymbol{\lambda}$ captures deviations from the one-component model such that $\boldsymbol{\lambda} = \mathbf{0}$ under H_0.

Standard quadratic approximation fails because $\nabla_{\boldsymbol{\lambda}} \log g_2(\boldsymbol{w}; \alpha, \boldsymbol{\nu} + (1-\alpha)\boldsymbol{\lambda}, \boldsymbol{\nu} - \alpha\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\mathbf{0}} = \mathbf{0}$. Identification therefore relies on second-order derivatives with respect to $\boldsymbol{\lambda}$.

Define the score function $\boldsymbol{s}(\boldsymbol{W}) = (\boldsymbol{s}_{\boldsymbol{\nu}}(\boldsymbol{W})^\top, \boldsymbol{s}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\boldsymbol{W})^\top)^\top$, where $\boldsymbol{s}_{\boldsymbol{\nu}}(\boldsymbol{W}) := \nabla_{\boldsymbol{\theta}} f^*/f^*$ and $\boldsymbol{s}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\boldsymbol{W}) := \frac{1}{2}\widetilde{\text{vech}}_2(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}^\top} f^*/f^*)$ with $f^* = f(\boldsymbol{W}; \boldsymbol{\theta}^*)$. The function $\boldsymbol{s}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\boldsymbol{w})$ comprises second-order derivatives of the log-likelihood and serves as the score for identifying $\boldsymbol{\lambda}$. Explicit expressions via Hermite polynomials are given in Appendix A.3.2. That appendix also establishes non-singularity of $\boldsymbol{\mathcal{I}} := \mathbb{E}[\boldsymbol{s}(\boldsymbol{W})\boldsymbol{s}(\boldsymbol{W})^\top]$.

Define the information matrix partition and related quantities:

$$\boldsymbol{\mathcal{I}} = \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\boldsymbol{\nu}} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\nu}\boldsymbol{\lambda}} \\ \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\nu}} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \end{pmatrix}, \quad \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}} := \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\lambda}} - \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\nu}}\boldsymbol{\mathcal{I}}_{\boldsymbol{\nu}}^{-1}\boldsymbol{\mathcal{I}}_{\boldsymbol{\nu}\boldsymbol{\lambda}}, \quad \boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\nu}} := (\boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}})^{-1}\boldsymbol{S}_{\boldsymbol{\lambda},\boldsymbol{\nu}}, \tag{22}$$

where $\boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\nu}} \sim \mathcal{N}(\mathbf{0}, (\boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}})^{-1})$. Here $\boldsymbol{S}_{\boldsymbol{\lambda},\boldsymbol{\nu}} := n^{-1/2}\sum_{i=1}^n \boldsymbol{s}_{\boldsymbol{\lambda},i} - \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda}\boldsymbol{\nu}}\boldsymbol{\mathcal{I}}_{\boldsymbol{\nu}}^{-1} n^{-1/2}\sum_{i=1}^n \boldsymbol{s}_{\boldsymbol{\nu},i}$ is the score for $\boldsymbol{\lambda}$ projected orthogonal to $\boldsymbol{\nu}$.

Let $\widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}$ be the projection of $\boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ onto the cone $\Lambda_{\boldsymbol{\lambda}} := \{v(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \mathbb{R}^q\}$, where

$$v(\boldsymbol{\lambda}) = (\lambda_1^2, \lambda_2\lambda_1, \lambda_2^2, \ldots, \lambda_q^2)^\top \tag{23}$$

contains unique elements of $\boldsymbol{\lambda}\boldsymbol{\lambda}^\top$. Formally,

$$\widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}} = \underset{\boldsymbol{t}_{\boldsymbol{\lambda}} \in \Lambda_{\boldsymbol{\lambda}}}{\arg\min} (\boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\nu}})^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}} (\boldsymbol{t}_{\boldsymbol{\lambda}} - \boldsymbol{G}_{\boldsymbol{\lambda},\boldsymbol{\nu}}). \tag{24}$$

Assumption 4. *(a) $\boldsymbol{X}$ has finite 9th moments. (b) $\mathbb{E}[\boldsymbol{U}\boldsymbol{U}^\top]$ is non-singular, where $\boldsymbol{U} = [1, \boldsymbol{X}^\top]^\top$. (c) $\boldsymbol{\mathcal{I}}$ is non-singular. (d) The true parameter $\boldsymbol{\theta}^*$ is interior to Θ.*

Assumption MOD-1 (Panel-covariate joint identifying variation)**.** *Let $\boldsymbol{U}_t := (1, \boldsymbol{X}_t^\top)^\top \in \mathbb{R}^{1+q}$ for each $t \in \{1, \ldots, T\}$. The $(1+q)^2 \times (1+q)^2$ matrix*

$$\mathbb{E}\big[\, (\boldsymbol{U}_t \otimes \boldsymbol{U}_s)\,(\boldsymbol{U}_t \otimes \boldsymbol{U}_s)^\top \,\big]$$

is non-singular for every distinct pair $t \neq s$. Equivalently, no non-trivial polynomial in $(\boldsymbol{X}_t, \boldsymbol{X}_s)$ of degree at most $(1,1)$ vanishes almost surely.

Proposition 4. *Suppose data are generated under $H_0 : M_0 = 1$. For the model with $K_\epsilon = 1$, suppose Assumptions 1-4 hold. For the model with $K_\epsilon = 2$, suppose Assumptions 1, 3, 4 (a)(b)(d), and MOD-1 hold. Then, for models (1) with the component density function (2) with (3), $LR_n \xrightarrow{d} \widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}} \widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}$.*

Lemma 1. *Suppose Assumptions 1, 3, and 4 hold. Then, under $H_0 : M_0 = 1$, for $\alpha \in [c_1, 1-c_1]$: (a) for any $\delta > 0$, $\limsup_{n\to\infty} \Pr(\sup_{\boldsymbol{\Psi}:\|\boldsymbol{\Psi}-\boldsymbol{\Psi}^*\| \leq \kappa} |R_n(\boldsymbol{\Psi}, \alpha)| > \delta(1 + \|\sqrt{n}\boldsymbol{t}(\boldsymbol{\Psi}, \alpha)\|^2)) \to 0$ as*

$\kappa \to 0$; *(b)* $\boldsymbol{S}_n \xrightarrow{d} N(0, \boldsymbol{\mathcal{I}})$; *(c)* $\boldsymbol{\mathcal{I}}_n \xrightarrow{p} \boldsymbol{\mathcal{I}}$ *with* $\boldsymbol{\mathcal{I}}$ *finite and non-singular; (d)* $\boldsymbol{t}(\hat{\psi}_\alpha, \alpha) = O_p(n^{-1/2})$. *Proof in Online Appendix.*

Lemma 2 (Companion to Lemma 1 for $K_\epsilon = 2$)**.** *Suppose Assumptions 1 (with* $K_\epsilon = 2$*), 3, 4 (a)(b)(d), and MOD-1 hold. Then conclusions (a), (b), and (d) of Lemma 1 hold, and (c) holds with* $\boldsymbol{\mathcal{I}} = \mathbb{E}[\boldsymbol{s}(\boldsymbol{W})\boldsymbol{s}(\boldsymbol{W})^\top]$ *finite and positive definite for model* (1) *with* (2)–(3)*. Proof in Online Appendix.*

Assumption MOD-1 holds in our empirical specifications (continuous panel covariates with non-degenerate joint distribution across periods); Remark 2 of the Online Appendix extends the non-singularity argument to $K_\epsilon \geq 3$ via iterated phase-peeling.

5.2 Likelihood ratio test for $H_0 : M_0 = M$ against $H_1 : M_0 = M + 1$

Testing $H_0 : M_0 = M$ in the $(M+1)$-component model requires addressing two scenarios: (i) two components have identical parameters, or (ii) one component has zero mixing proportion. We partition $H_0 = H_{01} \cup H_{02}$, where H_{01} requires $\boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1}$ for some $h \in \{1, \ldots, M\}$ with $\alpha_j > 0$ for all j; H_{02} requires $\alpha_h = 0$ for some h.

Issue with testing H_{02}: When $\alpha_h = 0$, the component-specific parameter $\boldsymbol{\theta}_h$ is unidentified, leading to infinite Fisher information unless a priori restrictions are imposed on component variances.

Proposition 5. $\sup_{\vartheta_{M_0+1} \in \Theta^*_{\boldsymbol{\vartheta}_{M_0+1}, 2h}} \mathbb{E}[\{\nabla_{\alpha_h} \log g_{M_0+1}(\boldsymbol{W}, \vartheta_{M_0+1})\}^2] < \infty$ *if and only if* $\max\{\sigma^2 : \sigma \in \Theta_\sigma\} < 2\max\{\sigma_1^{2*}, \ldots, \sigma_{M_0}^{2*}\}$ *for* $g_{M_0+1}(\boldsymbol{w}, \vartheta_{M_0+1})$ *in (1) with the component-specific density function* (2) *with* (3)*.*

Because the variance restriction in Proposition 5 is difficult to justify and enforce, we focus on testing H_{01}.

Testing H_{01}: Partition $H_{01} = \cup_{h=1}^{M_0} H_{0,1h}$, where $H_{0,1h} : \boldsymbol{\theta}_h = \boldsymbol{\theta}_{h+1}$ with ordered means $\mu_1 < \cdots < \mu_h = \mu_{h+1} < \cdots < \mu_{M_0+1}$. Each h corresponds to one way of describing the M_0-component null model in the (M_0+1)-component space.

We construct local MLEs $\widehat{\boldsymbol{\vartheta}}^h_{M_0+1}$ by restricting the (M_0+1)-component estimator to a neighborhood $\widehat{\boldsymbol{\Psi}}^*_h$ that contains parameters satisfying $H_{0,1h}$ but excludes other $H_{0,1l}$ for $l \neq h$. The formal definition of $\boldsymbol{\Psi}^*_h$ and partition $\Theta^*_{\boldsymbol{\theta},h}$ are in Appendix A.3. Define the local LRTS:

$$LR_n^{M_0,h} := 2\left\{\ell_n^{M_0+1}(\widehat{\boldsymbol{\vartheta}}^h_{M_0+1}) - \ell_n^{M_0}(\widehat{\boldsymbol{\vartheta}}_{M_0})\right\} \quad \text{for } h = 1, \ldots, M_0.$$

Because $\cup_{h=1}^{M_0} \boldsymbol{\Psi}_h^* = \bar{\Theta}_{\boldsymbol{\vartheta}_{M_0+1}}(\boldsymbol{c})$, the global LRTS is the maximum of local LRTS:

$$LR_n^{M_0} := 2\left\{\ell_n^{M_0+1}(\hat{\boldsymbol{\vartheta}}_{M_0+1}) - \ell_n^{M_0}(\widehat{\boldsymbol{\vartheta}}_{M_0})\right\} = \max\{LR_n^{M_0,1}, \ldots, LR_n^{M_0,M_0}\}. \tag{25}$$

Each $LR_n^{M_0,h}$ converges to $(\widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}}^h \widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h$, analogous to the one-component case. Thus, the asymptotic distribution of $LR_n^{M_0}$ is the maximum of M_0 such random variables.

Assumption 5. *(a) Assumption 4(a)(b) hold. (b) $\boldsymbol{\mathcal{I}}$ (the M_0-component score Fisher information; Online Appendix eq. (47)) is finite and non-singular. (c) The true parameter $\boldsymbol{\vartheta}_{M_0}^*$ is interior to $\bar{\Theta}_{\boldsymbol{\vartheta}_{M_0}}(\boldsymbol{c})$.*

Proposition 6. *Suppose Assumptions 1, 3, and 5 hold. Under $H_0 : M_0 = M$, for models (1) with component density function* (2) *with* (3)*,*

$$LR_n^{M_0} \xrightarrow{d} \max_{h=1,\ldots,M_0} \left\{(\widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}}^h \widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h\right\},$$

where $(\widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h)^\top \boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}}^h \widehat{\boldsymbol{t}}_{\boldsymbol{\lambda}}^h$ is the component-h projected quadratic form (Online Appendix eq. (48)).

Implementation: In practice, we use parametric bootstrap to obtain p-values for testing $H_0 : M_0 = M$ against $H_1 : M_0 = M + 1$ (see, e.g., Kasahara and Shimotsu, 2019). The parametric bootstrap is preferred because it circumvents the need to compute $\boldsymbol{\mathcal{I}}_{\boldsymbol{\lambda},\boldsymbol{\nu}}$ and to project onto the cone $\Lambda_{\boldsymbol{\lambda}}$, both of which are computationally demanding when q is moderate.

Regularized estimation. The theoretical results above are derived for the restricted MLE over $\bar{\Theta}_{\boldsymbol{\vartheta}_M}(\boldsymbol{c})$. In estimation, the EM algorithm maximises the penalised log-likelihood $\ell_n^{\text{pen}}(\boldsymbol{\vartheta}) := \ell_n(\boldsymbol{\vartheta}) + p(\boldsymbol{\vartheta})$ over $\bar{\Theta}_{\boldsymbol{\vartheta}_M}(\boldsymbol{c})$, where $p(\boldsymbol{\vartheta})$ collects the log-prior kernels from the following regularisation devices:

(i) **Variance floor:** $\sigma_j^2 \geq \delta\, \hat{s}^2$ with $\delta = 0.01$, where $\hat{s}^2 = \text{Var}(y)$ is the pooled sample variance.

(ii) **Penalty on mixing proportions:** a logarithmic penalty $(a_\alpha - 1)\sum_{j=1}^{M} \log \alpha_j$ on $(\alpha_1, \ldots, \alpha_M)$ with $a_\alpha = 1.5$, which discourages mixing proportions α_j from collapsing to zero.

(iii) **Penalty on sub-component weights:** a logarithmic penalty $(a_\tau - 1)\sum_{k=1}^{K_\epsilon} \log \tau_{jk}$ on $(\tau_{j1}, \ldots, \tau_{jK_\epsilon})$ with $a_\tau = 1.1$, applied per type.

Only the variance has a hard lower bound (item (i)); the mixing proportions α_j and the sub-component weights τ_{jk} are kept away from the simplex boundary by the logarithmic penalties in items (ii)–(iii) alone. The penalty derivatives diverge as $\alpha_j \to 0$ or $\tau_{jk} \to 0$, providing smooth boundary repulsion that obviates a separate hard bound while preserving asymptotic equivalence to the unconstrained MLE on the interior of the parameter

space. The explicit penalty $p(\boldsymbol{\vartheta})$ is given in equation (76) (Remark 1). Since $p(\boldsymbol{\vartheta}) = O(1)$ at any interior point of the parameter space, the penalised MLE is first-order asymptotically equivalent to the restricted MLE; see the appendix for a formal argument.

The following proposition establishes that the parametric bootstrap is asymptotically valid.

Proposition 7 (Bootstrap validity of the parametric bootstrap LRT)**.** *Suppose Assumptions 1 and 3 hold, the number of bootstrap replications satisfies $B_n \to \infty$ as $n \to \infty$, and the true number of components is M_0.*

(a) ***Testing*** $H_0 \colon M_0 = 1$ ***vs.*** $H_1 \colon M_0 = 2$**.** *Suppose Assumption 4 holds and, when $K_\epsilon = 2$, Assumption MOD-1 also holds. Let F_1 denote the CDF of the limiting null distribution of LR_n in Proposition 4. Generate bootstrap samples $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_n^*$ i.i.d. from the estimated one-component density $g_1(\cdot\,; \hat{\boldsymbol{\vartheta}}_1)$, and let LR_n^* denote the LRT statistic computed from the bootstrap sample. Then, at every continuity point x of F_1,*

$$P^*(LR_n^* \le x \mid \{\boldsymbol{W}_i\}_{i=1}^n) - F_1(x) \xrightarrow{p} 0.$$

(b) ***Testing*** $H_0 \colon M_0 = M$ ***vs.*** $H_1 \colon M_0 = M + 1$ ***for*** $M \ge 2$**.** *Suppose additionally that Assumption 5 holds. Let F_M denote the CDF of the limiting null distribution of $LR_n^{M_0}$ in Proposition 6. Generate bootstrap samples $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_n^*$ i.i.d. from $g_M(\cdot\,; \hat{\boldsymbol{\vartheta}}_M)$. Then, at every continuity point x of F_M,*

$$P^*(LR_n^{*,M} \le x \mid \{\boldsymbol{W}_i\}_{i=1}^n) - F_M(x) \xrightarrow{p} 0.$$

(c) ***Consistency of bootstrap sequential testing.*** *Under the conditions of Proposition 8, and with $B_n q_n \to \infty$ so that simulated bootstrap quantiles are consistent in the tail, the bootstrap sequential estimator $\widehat{M}_{\mathrm{LRT}}$ using bootstrap critical values satisfies $\widehat{M}_{\mathrm{LRT}} - M_0 = o_p(1)$.*

The proof in Appendix A.6 follows the local-null stability argument developed in this paper, combined with the bootstrap-continuity template in Lehmann and Romano (2005, Theorem 15.4.2) (see also Kasahara and Shimotsu (2019, Proposition 6 and Lemma 12)). Remark 1 in the appendix verifies that the penalised bootstrap used in implementation is first-order asymptotically equivalent.

5.3 Sequential hypothesis testing

To estimate the number of components, we sequentially test $H_0 : M_0 = M$ against $H_1 : M_0 = M + 1$ for $M = 1, 2, \ldots, \overline{M}$. Here $\overline{M}$ is the upper bound for the number of components and is assumed to be known and larger than M_0. The first value of M that leads to a

nonrejection of H_0 gives our estimate for M_0. Robin and Smith (2000) develop a similar sequential hypothesis test for estimating the rank of a matrix.

For $M = 1, \ldots, \overline{M}$, let $c^M_{1-q_n}$ denote the $100(1-q_n)$ percentile of the cumulative distribution function of a random variable $\max\{(\widehat{\boldsymbol{t}}^1_{\boldsymbol{\lambda}})^\top \boldsymbol{\mathcal{I}}^1_{\boldsymbol{\lambda},\boldsymbol{\nu}} \widehat{\boldsymbol{t}}^1_{\boldsymbol{\lambda}}, \ldots, (\widehat{\boldsymbol{t}}^M_{\boldsymbol{\lambda}})^\top \boldsymbol{\mathcal{I}}^M_{\boldsymbol{\lambda},\boldsymbol{\nu}} \widehat{\boldsymbol{t}}^M_{\boldsymbol{\lambda}}\}$ for testing $H_0 : M_0 = M$ in Proposition 6. Let $\widehat{c}^M_{1-q_n}$ be a consistent estimator of $c^M_{1-q_n}$. Then, our estimator based on sequential hypothesis testing (SHT, hereafter) is defined as

$$\hat{M}_{\text{LRT}} = \min_{M \in \{1,\ldots,\bar{M}\}} \{M : LR^m_n \geq \widehat{c}^m_{1-q_n} \text{ for all } m < M, \text{ and } LR^M_n < \widehat{c}^M_{1-q_n}\}. \tag{26}$$

The estimator $\hat{M}_{\text{LRT}}$ depends on the choice of the significance level q_n. The following proposition states that $\hat{M}_{\text{LRT}}$ converge to M_0 in probability as $n \to \infty$ when $-n^{-1}\log q_n = o(1)$ and $q_n = o(1)$.

Let

$$Q^M_n(\boldsymbol{\vartheta}_M) := n^{-1}\sum_{i=1}^n \log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M) \text{ and } Q^M(\boldsymbol{\vartheta}_M) := \mathbb{E}[\log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M)].$$

Assumption 6. *For $M = 1, \ldots, M_0$, (a) $Q^M(\boldsymbol{\vartheta}_M)$ has a unique maximum at $\boldsymbol{\vartheta}^*_M$ in $\Theta_{\boldsymbol{\vartheta}_M}$; (b) $\Theta_{\boldsymbol{\vartheta}_M}$ is compact; (c) $\boldsymbol{\vartheta}^*_M$ is interior to $\Theta_{\boldsymbol{\vartheta}_M}$; (d) $B^M(\boldsymbol{\vartheta}^*_M) := \mathbb{E}\left\{\nabla_{\boldsymbol{\vartheta}_M} \log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M) \nabla_{\boldsymbol{\vartheta}^\top_M} \log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M)\right\}$ is non-singular; (e) $A^M(\boldsymbol{\vartheta}^*_M) := \mathbb{E}\left\{\nabla_{\boldsymbol{\vartheta}_M \boldsymbol{\vartheta}^\top_M} \log g_M(\boldsymbol{W}_i; \boldsymbol{\vartheta}_M)\right\}$ has a constant rank in some open neighborhood of $\boldsymbol{\vartheta}^*_M$; and (f) $Q^{M+1}(\boldsymbol{\vartheta}^*_{M+1}) - Q^M(\boldsymbol{\vartheta}^*_M) > 0$ for $M = 1, \ldots, M_0 - 1$.*

Proposition 8. *Suppose that $M_0 < \overline{M}$ and Assumptions 1, 4, 5, and 6 hold. When $K_c = 2$, suppose Assumption MOD-1 also holds. If we choose q_n such that $-n^{-1}\log q_n = o(1)$ and $q_n = o(1)$, then $\hat{M}_{LRT} - M_0 = o_p(1)$.*

Assumptions 6(a)–(e) ensure the consistency and asymptotic normality of $\widehat{\boldsymbol{\vartheta}}_M$ for $M \leq M_0$; conditions (c)–(e) correspond to Assumption A6 of White (1982). Assumption 6(f) requires that the Kullback–Leibler information criterion strictly decreases as M increases for $M < M_0$.

Selecting the significance level q_n in finite samples is challenging. The formal analysis of the optimal choice of q_n is left for future research. We recommend presenting results across conventional levels (1%, 5%, 10%).

We show the inconsistency of AIC and consistency of BIC in Appendix A.9.

5.4 A Nonparametric Alternative: Rank-Based Lower Bound

The methods in Sections 5.1–5.3 rely on parametric specifications of the component densities $f(\boldsymbol{w};\boldsymbol{\theta}_j)$. As a robustness check, we briefly outline a nonparametric approach that estimates a *lower bound* for M_0 without imposing distributional assumptions.

The rank test method. Following Kasahara and Shimotsu (2014), we partition the support of each Y_t into mutually exclusive subsets $\Delta_t = \{\delta_1^t, \ldots, \delta_{|\Delta_t|}^t\}$ and construct bivariate probability matrices $\boldsymbol{P}_k$ (equation 100) whose rank identifies the lower bound of M under the conditional independence structure of model (1). The key identity is $\boldsymbol{P}_k = \sum_{j=1}^{M} \alpha^j \boldsymbol{p}_k^j (\boldsymbol{q}_k^j)^\top$, which implies $\text{rank}(\boldsymbol{P}_k) \leq M$ with equality under suitable rank conditions on the component-specific probability vectors.

Test statistic and implementation. The rank-test statistic of Kleibergen and Paap (2006) (Proposition A.7) provides a χ^2 test for $\text{rank}(\boldsymbol{P}_k) \leq r$. When $T \geq 3$, we aggregate evidence across different choices of $k \in \{1, \ldots, T\}$ using $\text{max-rk}(r) = \max\{\text{rk}^1(r), \ldots, \text{rk}^T(r)\}$ and $\text{ave-rk}(r)$ statistics. Bootstrap p-values are obtained via the Bayesian bootstrap, and a sequential testing procedure (see Kasahara and Shimotsu (2014) Section 3.2 and Appendix A.11) yields $\widehat{M}_{\text{rank}}$, a consistent estimator of a lower bound for M_0.

This method complements our parametric LRT by providing a specification-free diagnostic.

6 Simulation

This section evaluates the performance of five selection criteria (LRT, AIC, BIC, ave-rk, max-rk) in estimating the number of components. All estimation uses the default regularisation of Section 5.2: a variance floor $\sigma_j^2 \geq 0.01\,\hat{s}^2$; an α-penalty ($a_\alpha = 1.5$) on mixing proportions α_j; and a τ-penalty ($a_\tau = 1.1$) on sub-component weights τ_{jk}. Parameters are estimated by penalised MLE using the EM algorithm (Appendix A.8).

Appendix B.1 reports bootstrap LRT size and power under controlled DGPs (Table B1).

DGPs are calibrated to Chilean manufacturing data using M_0-component estimates for each reported industry. The current interim Monte Carlo table reports the completed Metal Products and partially completed Textiles designs, both calibrated at M_0=2 under the Translog (exact, q=3) specification with inputs $[\ln K, \ln L, \ln M]$. These simulation calibrations are distinct from the empirical headline counts $(4, 5, 2)$ used in the second-stage application in Section 7.

Table 1 reports under-selection ($\hat{M} < M_0$), correct selection ($\hat{M} = M_0$), and over-selection ($\hat{M} > M_0$) rates using the $T = 10$ panel. Panel A uses a normal-error DGP

($K_\epsilon = 1$); Panel B uses a mixture-error DGP ($K_\epsilon = 2$). Each row shows performance under both correctly-specified and misspecified error densities. The current interim version reports the completed Metal Products mixture-estimation replications and the completed Textiles Panel A mixture-estimation replications; the remaining cells are left blank until their reruns finish.

Table 1: Monte Carlo selection accuracy (%), $T = 10$

	Panel A: $K_\epsilon = 1$ **(Normal) DGP**						**Panel B:** $K_\epsilon = 2$ **(Mixture) DGP**					
	Metal (381)			**Textiles (321)**			**Metal (381)**			**Textiles (321)**		
Method	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$
Normal error estimation ($K_\epsilon = 1$)												
AIC												
BIC												
LR ($q_n = 0.01$)												
LR ($q_n = 0.05$)												
LR ($q_n = 0.10$)												
Mixture error estimation ($K_\epsilon = 2$)												
AIC	0	84.5	15.5	0	42.5	57.5	0	78.5	21.5			
BIC	0	100	0	0	59	41	0	91.5	8.5			
LR ($q_n = 0.01$)	0	99.5	0.5	53	47	0	25	75	0			
LR ($q_n = 0.05$)	0	95	5	39	60.5	0.5	0	100	0			
LR ($q_n = 0.10$)	0	88.5	11.5	6.5	90.5	3	0	100	0			
ave-rk (0.05)	100	0	0									
max-rk (0.05)	100	0	0									

Notes: Metal Panel A and Panel B and Textiles Panel A under Mixture error estimation use completed 200-replication reruns. Normal error estimation and Textiles Panel B cells are intentionally left blank until the corresponding reruns finish. Computation uses $n_{\text{sim}} = 200$, $B = 199$ bootstrap draws per LRT step, 8 DGP-calibration batches with 15 EM random starts per batch, 15 EM random starts for each simulated-data fit, 15 EM random starts for each LRT fit, maximum 500 LRT EM iterations, tolerance 10^{-8}, 30 short-run screening iterations, 32 parallel workers, and seed 42000. The three LR rows apply different thresholds to the same simulated draws and stepwise bootstrap p-values. DGPs use the *gnr_nonlinear* ($K_\epsilon = 1$ and $K_\epsilon = 2$) specification with X=[ln K, ln L, ln M], calibrated to M_0-component empirical estimates: Metal $n = 157$, $M_0 = 2$, Textiles $n = 162$, $M_0 = 2$; the calibrated DGP parameter estimates ($\hat{\alpha}_j, \hat{\beta}_{0j}, \hat{\beta}_{Kj}, \hat{\beta}_{Lj}, \hat{\beta}_{Mj}, \hat{\sigma}_j, \hat{\tau}_{1j}, \hat{\mu}_{1j}, \hat{\mu}_{2j}$) are reported in Table 2. The empirical headline $\hat{M} = (4, 5, 2)$ is selected separately in Table 4. BIC uses N (plants) as sample size (Leroux, 1992). "</=/>": % selecting $\hat{M}$ below/equal/above M_0.

Table 2: Calibrated DGP parameter estimates used in Table 1

Industry	Panel	Type	$\hat{\alpha}_j$	$\hat{\beta}_{0j}$	$\hat{\beta}_{Kj}$	$\hat{\beta}_{Lj}$	$\hat{\beta}_{Mj}$	$\hat{\sigma}_j$	$\hat{\tau}_{1j}$	$\hat{\mu}_{1j}$	$\hat{\mu}_{2j}$
Metal	A	1	0.3773	0.5254	0.0205	-0.2360	0.1540	0.3516	—	—	—
Metal	A	2	0.6227	0.5449	-0.0932	-0.1077	0.1930	0.1768	—	—	—
Metal	B	1	0.5803	0.5868	-0.0531	-0.1262	0.1550	0.1705	0.9603	-0.0191	0.4627
Metal	B	2	0.4197	0.4493	-0.0404	-0.1339	0.1506	0.2402	0.0210	-1.4203	0.0305
Textiles	A	1	0.5025	0.4957	-0.0449	-0.1639	0.2080	0.1920	—	—	—
Textiles	A	2	0.4975	0.6646	-0.0404	-0.1822	0.1989	0.1928	—	—	—
Textiles	B	1	0.5116	0.6621	-0.0445	-0.1773	0.1999	0.1781	0.9711	-0.0142	0.4756
Textiles	B	2	0.4884	0.4944	-0.0437	-0.1568	0.1991	0.1725	0.0137	-0.6637	0.0092

Notes: Parameters define the calibrated data-generating processes for Table 1. Panel A uses case 5 with a normal error ($K_\epsilon = 1$); Panel B uses case 6 with a two-component normal-mixture error ($K_\epsilon = 2$). The nonlinear mean is $\log(\beta_{0j} + \beta_{Kj}\tilde{k} + \beta_{Lj}\tilde{\ell} + \beta_{Mj}\tilde{m})$. For Panel B, $\tau_{2j} = 1 - \tau_{1j}$; both sub-component means (μ_{1j}, μ_{2j}) are shown. Inputs are grand-mean centred.

The completed mixture-estimation rows show that selection accuracy is sensitive to both the DGP and the penalty implicit in the criterion. For Metal Products, BIC is exact under the normal-error DGP and correct in 91.5% of mixture-DGP replications; LR at $q_n = 0.05$ is also highly accurate, with 95% correct selection in Panel A and 100% in Panel B. For Textiles Panel A, the more conservative LR threshold under-selects frequently, while LR at $q_n = 0.10$ selects M_0 in 90.5% of replications. Textiles Panel B and the Normal-error estimation rows remain blank until the corresponding reruns finish.

7 Empirical Application: Chilean Manufacturing Plants

We apply the finite-mixture production-function framework to plant-level panel data from Chilean manufacturing. The analysis covers three industries with at least 150 observed plants: Food Products (CIIU 311), Fabricated Metal Products (CIIU 381), and Textiles (CIIU 321). For each industry we construct two balanced panels: a short panel spanning 1992–1996 (T=5, used in Section B.2.3) and a long panel spanning 1987–1996 (T=10, the main analysis). A sequential bootstrap LRT selects $\hat{M}$; conditional on $\hat{M}$, per-type production elasticities are estimated via the Gandhi–Navarro–Rivers (GNR) two-step procedure (Gandhi et al., 2020). Throughout this section, N denotes the number of plants (the cross-sectional unit count); the theoretical sections use n for the same quantity.

7.1 Data and Descriptive Evidence

The data are drawn from Chile's Annual National Industrial Survey (ENIA). All monetary variables are deflated to 1986 pesos. We use the same three industries as Gandhi et al. (2020) and Kasahara and Rodrigue (2008): Food Products (CIIU 311, a material-intensive sector), Fabricated Metal Products (CIIU 381, capital-intensive), and Textiles (CIIU 321).

Table 3 presents summary statistics for the T=5 balanced panel (1992–1996). The key observation is **substantial within-industry dispersion in the materials revenue share** $s_{it} = M_{it}/Y_{it}$, with standard deviations of 0.14–0.20 and interquartile ranges of 0.16–0.25, large relative to means of 0.52–0.65. Figure 1 plots the firm-level time average $\bar{s}_i=(1/T)\sum_t s_{it}$ for the T=10 panel: averaging over years removes transitory shocks, so the remaining dispersion (σ=0.16–0.29 across industries) reflects permanent differences in material intensity across firms. Under a single homogeneous technology (M=1), this permanent variation would be absent. The observed spread motivates a model with multiple latent technology types that differ in their materials elasticity.

Table 3: Summary Statistics by Industry (T=5 panel, 1992–1996)

	Fabricated Metal (CIIU 381)	**Food Products (CIIU 311)**	**Textiles (CIIU 321)**
Panel A: Sample characteristics (1992–1996)			
Observations	1,389	5,146	1,180
Plants (N)	279	1,030	237
Avg. years/plant	4.98	5.00	4.98
Panel B: Materials revenue share $s_{it} = M_{it}/Y_{it}$			
Mean (SD)	0.516 (0.186)	0.654 (0.143)	0.568 (0.198)
25th–75th pct	[0.389, 0.635]	[0.577, 0.741]	[0.439, 0.693]
Panel C: Other variables (log levels; mean with SD)			
$\ln Y_{it}$ (gross output)	13.29 (1.37)	13.01 (1.63)	13.35 (1.34)
$\ln K_{it}$ (capital)	11.55 (1.65)	11.00 (2.03)	11.86 (1.65)
$\ln L_{it}$ (labour cost)	11.49 (1.27)	10.83 (1.41)	11.30 (1.37)
T=10 panel: N	157	669	162

Notes: Chilean ENIA. Monetary variables deflated to 1986 pesos. Standard deviations in parentheses. Panel A covers all plants observed at least once in 1992–1996 (unbalanced; Metal: N=279; Food: 1,030; Textiles: 237). The balanced T=5 panel used for model selection in Appendix B.2.3 has N=242 (Metal), 862 (Food), and 203 (Textiles) plants. Bottom row: number of plants in the long (T=10, 1987–1996) balanced panel.

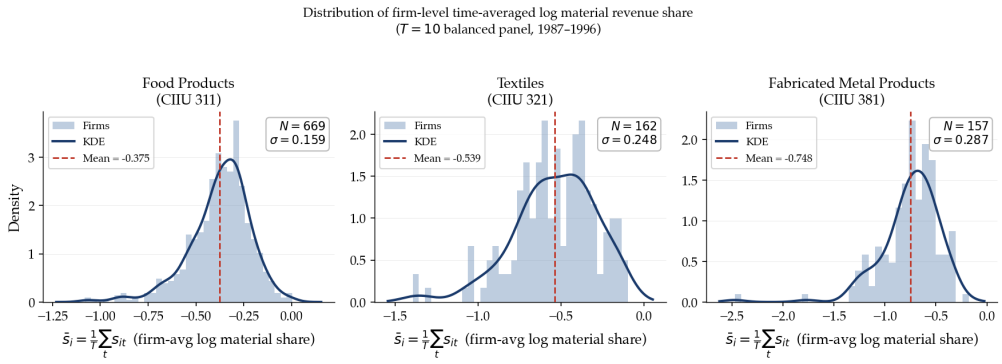

Figure 1: Distribution of firm-level time-averaged log materials revenue share $\bar{s}_i = \frac{1}{T}\sum_{t=1}^{T} s_{it}$, $s_{it} = \ln(M_{it}/Y_{it})$, across plants within each industry (T=10 balanced panel, 1987–1996; one observation per firm). Solid curve: kernel density estimate. The dispersed, non-Gaussian distributions reflect permanent unobserved heterogeneity across firms and motivate a finite-mixture model with multiple latent technology types.

Log inputs $(\tilde{k}_{it}, \tilde{\ell}_{it}, \tilde{m}_{it})$ are grand-mean centred within each industry–panel cell. Four-digit CIIU sub-industry dummies enter as common (across-type) regressors in the GNR Step 1 share equation.

7.2 Headline Specification

The **headline specification** is the T=10 Translog (exact) model with a two-component normal mixture residual (K_ϵ=2) and no Correlated Random Effects (CRE) augmentation. The mixture-error specification is the main empirical object because it lets within-type residuals have fat tails or latent shock heterogeneity, rather than forcing the top-level technology mixture to absorb all non-normal residual variation. The no-CRE specification is the main text benchmark because it remains within the finite-mixture likelihood framework studied in Sections 5.1–5.3; CRE is reported as an appendix robustness exercise. Under the GNR timing protocol (Gandhi et al., 2020), Hicks-neutral productivity ω_{it} cancels from the materials share first-order condition, so the share-equation system is free of endogeneity bias. Per-type coefficients $(\beta_{0,j}, \beta_{MM,j}, \beta_{LM,j}, \beta_{KM,j})$ are estimated by EM via the exact specification (10), yielding a linear materials elasticity surface $\bar{\varepsilon}_{M,j}(\tilde{m}, \tilde{\ell}, \tilde{k}) = \beta_{0,j} + \beta_{MM,j}\tilde{m} + \beta_{LM,j}\tilde{\ell} + \beta_{KM,j}\tilde{k}$. The intercept $\beta_{0,j} \equiv \bar{\varepsilon}_{M,j}(0,0,0)$ is the material output elasticity at the industry grand means, the key measure of materials intensity for type j. This quantity feeds into the GNR Step 2 (cubic-Markov GMM) to recover capital and labour elasticities. Robustness to Normal errors, CRE augmentation, and winsorisation is reported in Appendix B.2. CRE (Mundlak) results appear in Online Appendix Table B2 as a robustness check rather than the baseline. The theoretical reason is specific

to finite mixture models: under the Mundlak (1978) device, the unit-specific average $\bar{x}_i$ is a continuous random variable, so conditioning on it converts the component error from a discrete finite mixture into a mixture convolved with a continuous random effect. This falls outside the finite mixture framework on which the asymptotic theory of Sections 5.1–5.3 rests; establishing LRT consistency under the convolved structure is an open theoretical problem that we leave for future work.

7.3 Model Selection: Number of Latent Technology Types

Table 4 reports the selected number of types $\hat{M}$ from the sequential bootstrap LRT under the headline Translog (exact, K_ϵ=2) specification, alongside Cobb-Douglas and Translog (linear) alternatives, each estimated under both Normal and Mixture error structures. The sequential test uses B=199 parametric bootstrap replicates per LRT step at the 5% nominal level, tests $M = 1$ through $M_{\max} = 10$, and uses 15 EM random starts per fit. The H_0 null uses unconstrained EM; H_1 applies a Dirichlet penalty $a_\alpha \sum_j \log \alpha_j + a_\tau \sum_{j,k} \log \tau_{jk}$ (a_α=1.5, a_τ=1.1, σ^2 floor $0.01\widehat{\text{Var}}(y)$) to prevent degenerate entropy-inflated components from sharpening the type partition.

Three patterns emerge. First, the headline Translog (exact) Mixture model selects $\hat{M}$=4 in Metal, $\hat{M}$=5 in Food, and $\hat{M}$=2 in Textiles at T=10, using the raw (no-winsor) data and the sequential LRT threshold; the selected triple is unchanged when the stopping threshold is tightened from 5% to 1%. Second, less flexible functional forms tend to select larger $\hat{M}$. At T=10, Normal Cobb-Douglas selects 8–9, suggesting that type-heterogeneous curvature is absorbed by extra mixture components when the mean function is too rigid. Third, under-selection ($\hat{M}$<2) is negligible across all Mixture specifications at T=10, consistent with meaningful latent technology heterogeneity in all three industries.

The Normal–Mixture comparison clarifies what the selected types represent. When the residual distribution is restricted to be Normal, the model can use extra top-level components to fit residual tails, skewness, or other within-type shock heterogeneity. Allowing K_ϵ=2 separates this residual heterogeneity from technology heterogeneity and gives the headline $\hat{M}$ used for the GNR second stage. CRE has a different interpretation: it absorbs persistent plant-level heterogeneity through time averages of inputs, and therefore can reduce the number of residual latent types needed in some specifications. Because the CRE likelihood falls outside the finite-mixture structure covered by our main asymptotic theory, we treat it as an appendix robustness check rather than the headline result. Input winsorisation, CRE augmentation, and the K_ϵ=3 check are documented in Appendix B.2.1 (Table B2).

Table 4: Sequential bootstrap LRT $\hat{M}$, headline no-CRE specifications. Three functional forms × {Normal, Mixture} error structures; B=199, 5% level. Bold: preferred Mixture specification at each flexibility level.

Panel A: $T = 10$ (main results, $N_{Food} = 669$, $N_{Metal} = 157$, $N_{Tex} = 162$)

Spec (per-type q)	**Error dist.**	**Metal**	**Food**	**Textiles**
CD (q=0)	Normal	8	9	8
CD (q=0)	**Mixture**	**2**	**6**	**4**
Translog (linear, q=3)	Normal	7	9	7
Translog (linear, q=3)	**Mixture**	**2**	**2**	**3**
Translog (exact, q=3)	Normal	4	6	2
Translog (exact, q=3)	**Mixture**	**4**	**5**	**2**

Panel B: $T = 5$ robustness (see Appendix B.2.3; $N_{Food} = 862$, $N_{Metal} = 242$, $N_{Tex} = 203$)

Spec	**Error dist.**	**Metal**	**Food**	**Textiles**
Translog (exact, q=3)	Normal	2	8	2
Translog (exact, q=3)	**Mixture**	**4**	**6**	**4**

Notes. All specifications include 4-digit CIIU sub-industry dummies as common (across-type) regressors. Nonlinear uses the inside-log design with a Linear-Mixture warm start. Each LRT step uses $B = 199$ parametric bootstrap replications and 15 EM random starts per fit (SQUAREM acceleration with 30 short-run screening iterations); AIC/BIC are computed over $M = 1, \ldots, 11$. The Dirichlet-penalised H_1 avoids degenerate solutions; H_0 is always unpenalised. Raw (no-winsor) data. Less flexible functional forms (CD, Linear) push selectors toward larger $\hat{M}$ because they absorb type-heterogeneous curvature via extra components. Panel B shows no-CRE T=5 robustness. The bolded Translog (exact) rows of Panel A (Metal: 4, Food: 5, Textiles: 2) define the headline type counts used for the second-stage GNR point estimates. The headline row is selected by the sequential stopping rule: reject $H_0 : M = k$ until the first non-rejection. In the stopping step, the bootstrap LRT statistics and 5% critical values are Metal $H_0 : M$=4 (55.53 < 248.78), Food $H_0 : M$=5 (125.08 < 289.71), and Textiles $H_0 : M$=2 (201.42 < 247.01); the same stopping values obtain at the 1% threshold. Winsorisation, CRE, and Normal/Mixture error comparisons are in Appendix B.2.1 (Table B2). Robustness to three-component within-type errors (K_ϵ=3) is reported in Appendix B.2.2 within Table B2.

7.4 Latent Technology Types: Materials Elasticity Surface

The EM-estimated materials coefficients $\beta_{M,j}$ capture each type's materials elasticity at the industry-mean input bundle ($\tilde{m}$=$\tilde{\ell}$=$\tilde{k}$=0; Appendix Table B6). Figure 2 plots $\beta_{M,j}$ point estimates for the headline second-stage GNR fits.[8]

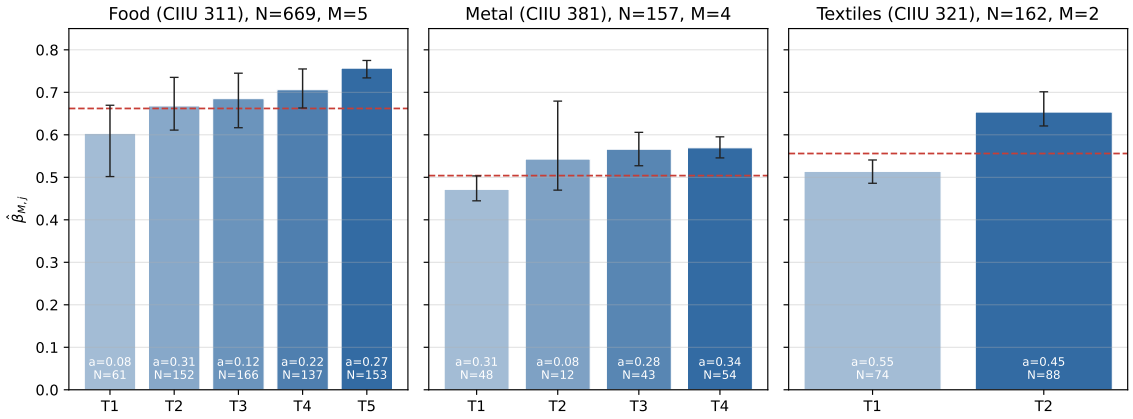

Figure 2: Per-type materials coefficient $\hat{\beta}_{M,j}$ (Translog at grand-mean-centred inputs). Dashed line: pooled M=1 estimate. Error bars are centered 95% firm-cluster nonparametric-bootstrap percentile spreads (B=500), obtained by resampling firms with replacement, re-estimating the fixed-$\hat{M}$ mixture, and recomputing the GNR point estimates. Baseline: Translog (exact, K_ϵ=2), T=10, no CRE.

Three patterns are visible. First, **heterogeneity is concentrated in the materials coefficient** $\beta_{M,j}$: across types within a cell, $\beta_{M,j}$ varies by 0.14–0.24 (a 27–37% range around the pooled mean). The full Translog coefficient vector, including curvature and interaction terms, is reported in Appendix Table B3. Latent technology types are identified primarily by their *level* of materials intensity rather than by a uniform sign pattern in substitution curvature. Second, the updated point estimates remain economically close to the pooled GNR benchmarks at the aggregate level while exposing within-industry heterogeneity that the pooled estimator hides. Third, the monotonicity projection keeps the implied plant-level input elasticities non-negative even when the underlying Translog curvature coefficients vary across types.

7.5 Production Elasticities and Returns to Scale

Conditioning on the type assignments from Step 1, the GNR Step 2 cubic-Markov GMM recovers capital and labour elasticities for each type. Figure 3 plots output elasticities ($\bar{\varepsilon}_K$,

[8] Full per-type point estimates are in Appendix B.2.4.

$\bar{\varepsilon}_L$) and returns to scale ($\widehat{\text{RTS}}$) across types. The dashed lines mark the pooled M=1 GNR point estimates.

Two features of Figure 3 warrant comment. First, capital elasticities are small relative to materials elasticities in the updated point estimates. Capital plays a secondary role in technology-type differentiation; types separate primarily along materials intensity $\beta_{M,j}$. Second, all three headline industries show economically visible cross-type dispersion, with Metal more imprecise because the four-type partition includes one small cell. The $\hat{\alpha}$-weighted aggregates remain close to constant returns, while the per-type bars reveal materials-intensity differences that the pooled M=1 estimator hides.

Table 5 quantifies the pooled benchmarks and the $\hat{\alpha}$-weighted aggregate $\bar{\varepsilon}_k^{\text{agg}} = \sum_j \hat{\alpha}_j \bar{\varepsilon}_{k,j}$ as a point estimate.

7.6 Discussion

Materials elasticity. The pooled M=1 materials elasticity lies at 0.50–0.66 across industries at T=10 (Table 5), consistent with the GNR Chilean and Colombian benchmarks (Gandhi et al., 2020). The per-type distribution in Figure 2 reveals meaningful heterogeneity across all three headline second-stage fits: Food spans $\beta_{0,j} \in [0.603, 0.757]$, Metal spans [0.472, 0.569], and Textiles spans [0.514, 0.653]. The $\hat{\alpha}$-weighted aggregate $\bar{\varepsilon}_M^{\text{agg}}$ is 0.697 in Food, 0.540 in Metal, and 0.585 in Textiles. Food lies above the pooled Python and `gnrprod` R confidence intervals, driven by high-materials-intensity types with substantial mixture weights; Metal and Textiles remain closer to the pooled benchmarks. In all three industries, the novelty of the mixture model lies not in whether the aggregate matches the pooled estimate, but in revealing that the pooled estimate conceals substantial within-industry heterogeneity in materials intensity that a single-technology estimator cannot recover.

Capital and labour elasticities. The pooled M=1 capital elasticity excludes zero in five of six cells (Table 5); the sole exception is Metal T=10 with the Python GNR implementation (CI [−0.092, +0.278]), reflecting finite-sample imprecision in the cubic-Markov GMM step at small N. The $\hat{\alpha}$-weighted aggregate capital elasticity is modest in the updated point estimates (+0.159 for Food, +0.162 for Metal, and +0.198 for Textiles). The monotonicity projection constrains $\varepsilon_K \geq 0$ at every plant-year observation; aggregate values at the grand mean reflect each type's input mix relative to the industry centre. Across all plant-year observations, the SLSQP projection reduces the fraction with any violated monotonicity condition ($\bar{\varepsilon}_K < 0$, $\bar{\varepsilon}_L < 0$, or $\bar{\varepsilon}_M < 0$) from up to 21% (unconstrained) to 0.0% (constrained) for every type and industry, confirming that the production function satisfies the basic input-monotonicity requirements of production theory at the individual plant level. Labour elasticity is positive in all pooled cells and in all updated aggregate

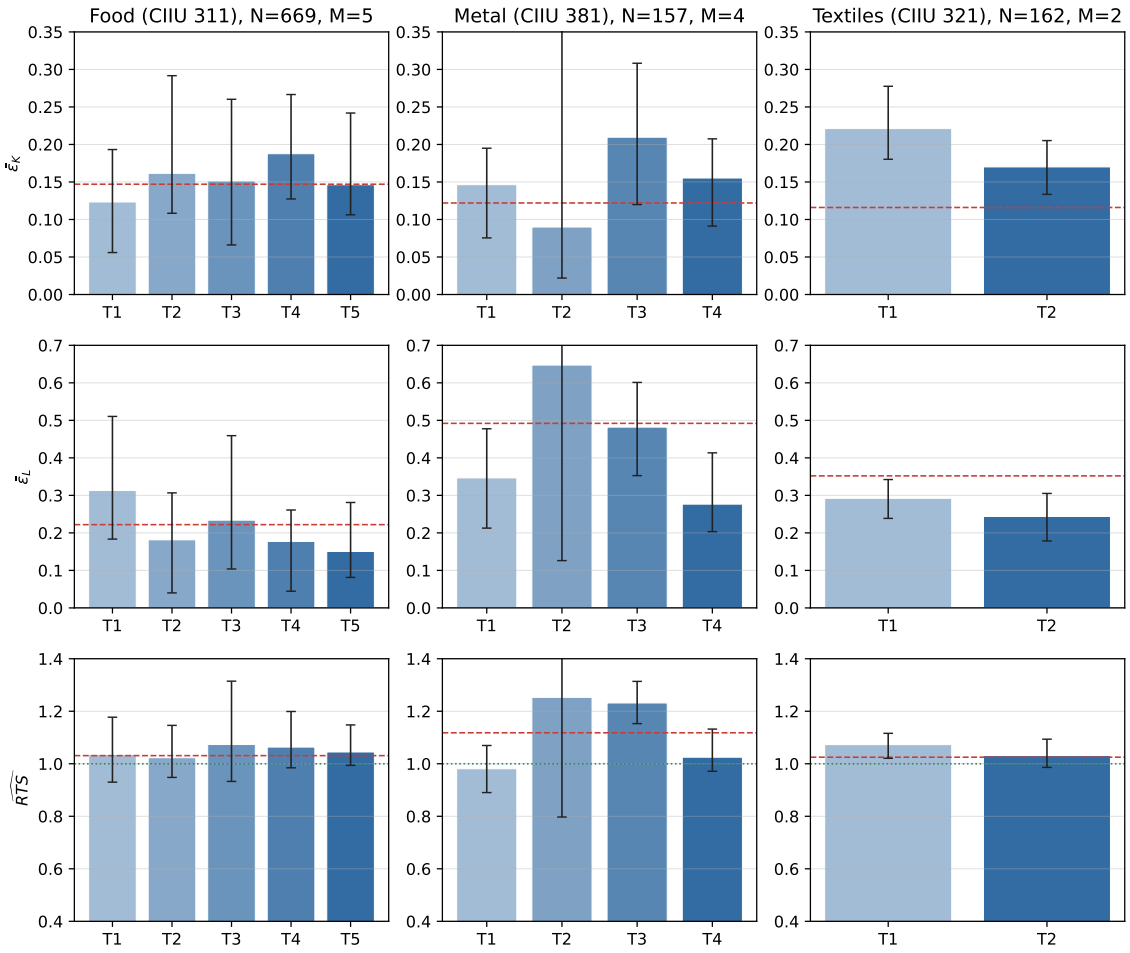

Figure 3: Per-type GNR output elasticities and returns to scale (T=10, Translog (exact, K_ϵ=2), no CRE). Each column is one industry; rows show $\bar{\varepsilon}_K$, $\bar{\varepsilon}_L$, and $\widehat{\text{RTS}}$. Bars: data point estimates. Error bars are centered 95% firm-cluster nonparametric-bootstrap percentile spreads (B=500) from the same fixed-$\hat{M}$ resampling-and-reestimation procedure as in Figure 2. For readability, a few very long upper Metal whiskers are truncated by the plotting range; intervals are computed from the full bootstrap distribution. Red dashed line: pooled M=1 Python GNR estimate. Green dotted line in bottom row: constant returns ($\widehat{\text{RTS}}$=1). Elasticities evaluated at the industry-wide grand mean. The GNR monotonicity projection enforces $\varepsilon_K \geq 0$ at every plant-year observation, so per-plant capital elasticities are non-negative by construction.

Table 5: GNR output elasticities at sample mean (T=10): pooled benchmarks and headline mixture aggregate estimates. † CI contains zero.

Source	$\bar{\varepsilon}_K$	$\bar{\varepsilon}_L$	$\bar{\varepsilon}_M$	$\widehat{\text{RTS}}$
Food CIIU 311, T=10, N=669, $\hat{M}$=5				
Pool M=1, Python GNR	+0.147	+0.222	+0.662	+1.031
95% CI	[+0.089, +0.281]	[+0.186, +0.591]	[+0.654, +0.671]	[+0.952, +1.488]
Pool M=1, `gnrprod` R	+0.105	+0.256	+0.662	+1.023
95% CI	[+0.090, +0.123]	[+0.219, +0.287]	[+0.654, +0.670]	—
Headline mixture aggregate	+0.159	+0.189	+0.697	+1.045
95% CI	[+0.142, +0.180]	[+0.156, +0.231]	[+0.689, +0.705]	[+1.007, +1.087]
Metal CIIU 381, T=10, N=157, $\hat{M}$=4				
Pool M=1, Python GNR	+0.122	+0.492	+0.504	+1.118
95% CI	[−0.092, +0.278]†	[+0.198, +0.733]	[+0.483, +0.528]	[+0.669, +1.415]
Pool M=1, `gnrprod` R	+0.129	+0.500	+0.504	+1.133
95% CI	[+0.084, +0.175]	[+0.393, +0.563]	[+0.483, +0.528]	—
Headline mixture aggregate	+0.162	+0.385	+0.540	+1.087
95% CI	[+0.127, +0.850]	[+0.312, +0.631]	[+0.519, +0.561]	[+1.021, +2.013]
Textiles CIIU 321, T=10, N=162, $\hat{M}$=2				
Pool M=1, Python GNR	+0.116	+0.352	+0.556	+1.025
95% CI	[+0.054, +0.172]	[+0.179, +0.473]	[+0.535, +0.580]	[+0.841, +1.173]
Pool M=1, `gnrprod` R	+0.114	+0.374	+0.556	+1.044
95% CI	[+0.075, +0.142]	[+0.304, +0.431]	[+0.536, +0.579]	—
Headline mixture aggregate	+0.198	+0.271	+0.585	+1.054
95% CI	[+0.169, +0.232]	[+0.221, +0.314]	[+0.564, +0.606]	[+1.018, +1.091]

Notes. Elasticities are evaluated at the grand-mean-centred input mean ($\tilde{m}$=$\tilde{\ell}$=$\tilde{k}$=0). Raw (no-winsor) data. Pool CIs are standalone parametric bootstraps: Python 11-parameter GNR polynomial (`gnr_per_type_exact`, B=500) and `gnrprod` R (Jin, 2023) (B=200); `gnrprod` R does not output a joint RTS CI (—) because the reported marginal elasticity intervals do not retain the joint bootstrap draws needed to form a valid interval for $\varepsilon_K + \varepsilon_L + \varepsilon_M$. Headline mixture aggregates are $\sum_j \hat{\alpha}_j \bar{\varepsilon}_{x,j}$. Their CIs are centered 95% firm-cluster nonparametric-bootstrap percentile spreads (B=500), obtained by resampling firms, re-estimating the fixed-$\hat{M}$ mixture, recomputing per-type GNR estimates, and aggregating within each draw. Point estimates are from `results/gnr_headline_point_20260702/per_type_point.csv`.

point estimates. The Python GNR and `gnrprod` R pool estimates agree closely on all four elasticities, confirming that the pooled M=1 finding is not an artefact of the software implementation.

Returns to scale. The pooled RTS lies in $[1.023, 1.133]$ across all six industry–implementation pairs, consistent with approximately constant returns. The per-type RTS distribution (Figure 3, bottom row) shows modest dispersion in point estimates. The $\hat{\alpha}$-weighted aggregate RTS is close to constant returns in the clean updated point estimates: 1.045 for Food and 1.054 for Textiles. The Metal aggregate RTS is also close to constant returns at 1.087, though its per-type estimates are noisier because the smallest Metal type contains only 12 plants.

Productivity persistence. The Step 2 AR(1) persistence parameter $\hat{\rho}$ ranges from 0.52 to 0.88 across the updated point estimates. This persistent productivity is consistent with the broader literature (Gandhi et al., 2020) and is not a failure of AR(1) identification. Type-level variation in $\hat{\rho}$ (Table B6 in the Online Appendix) reflects genuine heterogeneity in the speed at which idiosyncratic productivity shocks dissipate across latent technology groups. A pooled M=1 estimator averages over these types, masking differences in the speed at which plant-level productivity shocks dissipate across latent technology groups.

Why recovering $\hat{M}$ matters economically. The model-selection step is not a preliminary formality but a substantive input into the economic findings. Setting M=1—the conventional single-technology assumption—collapses all within-industry variation in the materials share into a single intercept and curvature, erasing the $\beta_{0,j}$ spread documented above. Setting M too high risks decomposing sampling noise into spurious additional types, inflating apparent heterogeneity. The sequential bootstrap LRT procedure, which controls asymptotic size at level q_n while exploiting the degeneracy structure unique to panel normal mixtures (Section 5.1), provides the first theoretically justified criterion for this choice. The updated headline selections, $\hat{M}$=4, 5, 2 under the T=10 no-CRE Mixture baseline, therefore reflect genuine within-industry heterogeneity after allowing the residual distribution itself to be flexible, not an artefact of an ad-hoc selection rule. More broadly, the econometric contribution (consistent $\hat{M}$ selection under panel degeneracy conditions) and the economic finding (Hicks-neutral technology differences concentrated in materials-intensity levels) are inseparable: without a principled $\hat{M}$, the per-type GNR estimates inherit the arbitrariness of the assumed number of types.

8 Conclusion

This paper develops likelihood-based methods for selecting the number of components in panel data finite mixture regression models. We show that panel data eliminates the higher-order degeneracy of cross-sectional normal mixtures while unbounded likelihood and infinite Fisher information persist. We derive the asymptotic null distribution of the LRT statistic and establish BIC consistency and AIC inconsistency. We also propose a sequential testing procedure for consistent component selection.

In the empirical application to Chilean manufacturing, the headline T=10 no-CRE Translog (exact) specification with mixture errors (K_ϵ=2) selects $\hat{M}$=4, 5, 2 types in Metal, Food, and Textiles (Table 4). This specification is deliberately flexible about within-type residual heterogeneity, so the selected top-level components are interpreted as latent technology types rather than as a device for fitting non-normal shocks. Type heterogeneity is concentrated in the level of materials intensity (the intercept $\beta_{0,j}$), not in curvature or factor-substitution slopes, a form of heterogeneity that single-technology estimators cannot detect. The updated full-GNR point estimates preserve this interpretation: Food's mixture-weighted materials elasticity exceeds the pooled M=1 confidence interval, while Metal and Textiles remain closer to the pooled benchmark but display meaningful type-level dispersion. Returns to scale are close to constant in the updated point estimates, consistent with competitive output markets. Type-level deviations in both materials intensity and TFP dynamics are invisible to aggregate estimators.

Proofs and additional simulation and empirical results are provided in the Online Appendix.

Several limitations point toward future work. First, our identification strategy uses only the materials first-order condition and therefore cannot recover elasticities of predetermined inputs such as capital. Extending to a full factor-augmenting model (Doraszelski and Jaumandreu, 2018; Li and Sasaki, 2017) would require a separate identification argument. Second, our theoretical results maintain T fixed while $n \to \infty$. Studying the large-T regime, where time-averaged Mundlak controls may be dispensed with, is an open question. Third, the number of within-type error components K_ϵ is treated as fixed at two; a joint test for M and K_ϵ would sharpen the model-selection framework.

Data Availability Statement

The Chilean Annual National Industrial Survey (ENIA) data used in this study are publicly available from the Instituto Nacional de Estadísticas (INE), Chile. Anonymized plant-level panels are included in the replication package deposited at Zenodo (DOI to be assigned upon acceptance). All code necessary to reproduce the results is included in the replication package.

References

[1] Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

[2] Amengual, D., Bei, X., Carrasco, M., and Sentana, E. (2022). Score-type tests for normal mixtures. Working Papers wp2022_2213, CEMFI.

[3] Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191.

[4] Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69:683–734.

[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] Azaïs, J.-M., Gassiat, E., and Mercadier, C. (2009). The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM: Probability and Statistics*, 13:301—327.

[7] Balat, J., Brambilla, I., and Sasaki, Y. (2019). Heterogeneous firms: Skilled-labor productivity and the destination of exports.

[8] Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

[9] 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.

[10] Chamberlain, G. (1984). Panel data. In Griliches†, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition.

[11] Chen, H. and Chen, J. (2001). The likelihood ratio test for homogeneity in finite mixture models. *Canadian Journal of Statistics*, 29:201–215.

[12] Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, 13:351–365.

[13] Chen, H., Chen, J., and Kalbfleisch, J. D. (2004). Testing for a finite mixture model with two components. *Journal of the Royal Statistical Society, Series B*, 66:95–115.

[14] Chen, J. (1995). Optimal rate of convergence for finite mixture models. *Annals of Statistics*, 23(1):221–233.

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

[16] Chen, J., Li, P., and Fu, Y. (2012). Inference on the order of a normal mixture. *Journal of the American Statistical Association*, 107(499):1096–1105.

[17] Chen, X., Ponomareva, M., and Tamer, E. (2014). Likelihood inference in some finite mixture models. *Journal of Econometrics*, 182(1):87–99.

[18] Chernoff, H. and Lander, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *Journal of Statistical Planning and Inference*, 43:19–40.

[19] Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *Annals of Statistics*, 27:1178–1209.

[20] 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.

[21] Doraszelski, U. and Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, 126(3):1027–1084.

[22] Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

[23] Garel, B. (2001). Likelihood ratio test for univariate Gaussian mixture. *Journal of Statistical Planning and Inference*, 96:325–350.

[24] Garel, B. (2005). Asymptotic theory of the likelihood ratio test for the identification of a mixture. *Journal of Statistical Planning and Inference*, 131:271–296.

[25] Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. In Le Cam, L. and Olshen, R., editors, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 789–806. Belmont, CA: Wadsworth.

[26] 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.

[27] Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Annals of Statistics*, 13(2):795–800.

[28] 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.

[29] Hu, Y. and Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1):32–44.

[30] Jin, D. (2023). `gnrprod`: Gandhi–Navarro–Rivers gross output production function estimation. R package version 1.1.2.

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

[32] Kasahara, H. and Rodrigue, J. (2008). Does the use of imported intermediates increase productivity? plant-level evidence. *Journal of Development Economics*, 87(1):106–118.

[33] Kasahara, H., Schrimpf, P., and Suzuki, M. (2022). Identification and estimation of production function with unobserved heterogeneity. Technical report, ESRI Discussion Paper Series No.368.

[34] Kasahara, H. and Shimotsu, K. (2009). Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica*, 77(1):135–175.

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

[36] Kasahara, H. and Shimotsu, K. (2014). Non-parametric identification and estimation of the number of components in multivariate mixtures. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(1):97–111.

[37] 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.

[38] Kasahara, H. and Shimotsu, K. (2018). Testing the number of regimes in markov regime switching models.

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

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

[41] Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhya Series A*, 62:49–62.

[42] Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126.

[43] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, third edition edition.

[44] Lemdani, M. and Pons, O. (1997). Likelihood ratio tests for genetic linkage. *Statistics and Probability Letters*, 33:15–22.

[45] Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *The Annals of Statistics*, 20:1350–1360.

[46] Levinsohn, J. and Petrin, A. (2003a). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, pages 317–341.

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

[48] Li, T. and Sasaki, Y. (2017). Constructive identification of heterogeneous elasticities in the Cobb-Douglas production function.

[49] Lin, C.-C. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1:42–55.

[50] Lindsay, B. G. (1995). Mixture Models: Theory, Geometry and Applications NSF-CBMS Regional Conference Series in Probability and Statistics. *Source: NSF-CBMS Regional Conference Series in Probability and Statistics*, 5:1–163.

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

[52] Lu, X. and Su, L. (2017). Determining the number of groups in latent panel structures with an application to income and democracy. *Quantitative Economics*, 8(3):729–760.

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

[54] Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

[55] Olley, G. S. and Pakes, A. (1996a). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 6(64):1263–1297.

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

[57] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. (A.)*, 185:71–110.

[58] Robin, J.-M. and Smith, R. J. (2000). Tests of rank. *Econometric Theory*, 16(2):151–175.

[59] Su, L., Zhentao, S., and Phillips, P. (2016). Identifying latent structures in panel data. *Econometrica*, 84:2215–2264.

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

[61] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

[62] Zhu, H.-T. and Zhang, H. (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society, Series B*, 66:3–16.