Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(3): 6329–6349.
DOI: 10.3934/math.2026261
Received: 14 January 2026
Revised: 14 February 2026
Accepted: 03 March 2026
Published: 12 March 2026

Research article

Frailty-Augmented Logistic-Weighted Lomax Mixture regression for survival analysis

Mohieddine Rahmouni*

Applied College, King Faisal University, Al-Ahsa, Saudi Arabia

* **Correspondence:** Email: mrahmouni@kfu.edu.sa.

Abstract: We propose the Frailty-Augmented Logistic-Weighted Lomax Mixture (F-LLoM), a finite-mixture survival model that combines gamma frailty within components with covariate-dependent multinomial logistic mixing weights. After integrating out the frailty, each component follows a Lomax (Pareto II) distribution, enabling flexible modeling of heavy-tailed survival times and latent heterogeneity, while covariates affect both tier membership and within-tier hazard scales. We derive identifiability conditions, moment expressions, and tail properties, and develop a censoring-aware expectation–maximization (EM) algorithm for right-censored data. In simulation studies with sample sizes between 500 and 1000 and censoring rates up to 20%, the Bayesian information criterion consistently selected the true number of components, and the EM algorithm showed stable convergence with well-calibrated predictions. An application to the Rossi recidivism dataset illustrates the practical implementation of the model and favors a parsimonious specification. The proposed framework provides a practical and interpretable approach for analyzing heterogeneous and heavy-tailed survival data.

Keywords: frailty; Lomax distribution; finite-mixture; heavy-tailed distributions; logistic regression
Mathematics Subject Classification: 60E05, 62E15, 62F10

1. Introduction

Survival analysis plays a central role in biostatistics, reliability engineering, and risk management, where the primary objective is to model the time until an event such as death, system failure, or default occurs [1–3]. A fundamental challenge in this setting is unobserved heterogeneity: latent factors that influence event times but are not captured by observed covariates. Ignoring such heterogeneity may lead to biased parameter estimates, distorted hazard ratios, and reduced predictive accuracy. Classical approaches, including the Cox proportional hazards model [4] and standard parametric models such as the exponential or Weibull, implicitly assume that observed covariates sufficiently explain variability in event times. When latent structure is present, this assumption may result in model misspecification.

Two main strategies have been developed to address this limitation. Frailty models incorporate a random effect into the hazard function to capture unobserved individual-level variability [5]. When exponential hazards are combined with gamma frailty, the marginal distribution becomes Lomax (Pareto type II), a heavy-tailed form capable of accommodating extreme survival times [6]. Finite mixture models, in contrast, represent the population as a set of latent subgroups, each characterized by its own survival distribution and mixing probability [7]. Although both approaches enhance robustness, frailty models may be restrictive when heterogeneity is multi-tiered, while mixture models with fixed weights cannot adapt subgroup allocation to individual covariates.

Recent work has attempted to unify these perspectives. Rahmouni [8] introduced the doubly generalized exponential–geometric frailty (DGEGF) model, which represents the time to first failure as a geometric mixture of Lomax components. This model is a special case of the broader doubly generalized power series frailty (DGWPSF) framework [9], defined by

$$X \mid Z = z \sim \text{Weibull}(\beta, \theta z^{-1/\beta}), \qquad Z \sim \text{Gamma}(\lambda, \lambda),$$

with the latent count N following a geometric or shifted-Poisson distribution.

Under exponential hazards ($\beta = 1$) and focusing on the first failure ($k = 1$), the marginal density reduces to

$$f_Y(y) = (1-\eta)\,\theta\,\lambda^{\lambda+1} \sum_{n=1}^{\infty} n\,\eta^{n-1} \big(\lambda + n\theta y\big)^{-(\lambda+1)},$$

an infinite geometric mixture of Lomax-type components.

Despite its flexibility, this construction has important limitations. The infinite mixture over N increases computational burden. The geometric weights $(1-\eta)\eta^{n-1}$ are fixed and cannot depend on covariates, restricting individualized inference. Moreover, a single frailty parameter λ governs all mixture terms, preventing tier-specific heterogeneity. Consequently, the DGEGF and DGWPSF models are less suitable for applications requiring interpretable latent risk strata or covariate-driven prediction.

To address these limitations, we propose the *Frailty-Augmented Logistic-Weighted Lomax Mixture (F-LLoM)* model, a hierarchical regression framework that extends prior work along three dimensions. First, F-LLoM replaces the infinite geometric mixture with a finite mixture of M latent risk tiers, improving interpretability and computational tractability. Second, the mixing weights are specified as covariate-dependent probabilities,

$$\pi_m(x) = \frac{\exp(\alpha_m + x^\top \beta_m)}{\sum_{j=1}^{M} \exp(\alpha_j + x^\top \beta_j)},$$

allowing individualized tier allocation via multinomial logistic regression. Third, each tier possesses its own frailty parameter λ_m and covariate-dependent hazard scale

$$\theta_m(x) = \exp(\gamma_m + x^\top \delta_m),$$

thereby capturing both within-tier frailty-driven variability and between-tier covariate-driven heterogeneity. Together, these elements yield a flexible and interpretable framework for personalized survival modeling.

The motivation for F-LLoM arises from the need to model complex heterogeneity in real-world survival data. In reliability engineering, latent risk tiers facilitate proactive maintenance strategies that reduce downtime and prevent catastrophic failures [10]. In biomedical contexts, they enable clinically meaningful patient stratification. Heavy-tailed survival patterns, including long-lived outliers and abrupt failures, are common across domains and are not well captured by light-tailed models such as the exponential or Weibull. Because Lomax distributions naturally arise from exponential hazards with gamma frailty, the F-LLoM framework provides improved tail behavior relative to conventional models [11]. Unlike the Cox model, which assumes proportional hazards, or the DGEGF formulation with fixed geometric weights and a common frailty parameter [8], F-LLoM integrates covariate-dependent mixing with tier-specific frailty to achieve both predictive accuracy and structural interpretability.

In summary, F-LLoM is a finite mixture survival regression model with covariate-dependent tier weights and tier-specific Lomax components, designed to uncover latent risk structure while accommodating heavy-tailed event times. This paper develops the model, derives its theoretical properties, establishes identifiability conditions, and presents a censoring-aware expectation–maximization (EM) algorithm for right-censored data. Simulation studies and empirical applications demonstrate its predictive performance and interpretability relative to existing approaches.

The remainder of the paper is organized as follows. Section 2 derives the Lomax distribution via gamma frailty and introduces the F-LLoM model. Section 3 presents theoretical properties. Section 4 describes the EM estimation procedure. Section 5 reports simulation results. Section 6 provides an empirical application, and Section 7 concludes.

2. The F-LLoM model

The F-LLoM model is a finite mixture regression framework for survival analysis that combines covariate-driven latent risk stratification with component-specific unobserved heterogeneity. It extends the DGEGF formulation [8] by replacing the infinite geometric mixture with a finite and interpretable set of M latent risk tiers. Each tier is characterized by its own gamma frailty parameter, a covariate-dependent hazard scale, and multinomial logistic mixing probabilities.

Hierarchical structure. The F-LLoM model has a three-level hierarchical structure. First, covariates determine latent tier membership through multinomial logistic gating probabilities $\pi_m(x)$. Second, within each tier, covariates determine the component-specific hazard scale $\theta_m(x)$. Third, conditional on tier membership and gamma frailty, survival times follow a Lomax distribution obtained by integrating out the frailty term. This layered construction separates between-tier heterogeneity (via π_m) from within-tier hazard dynamics (via θ_m and λ_m), improving interpretability.

Figure 1 provides a schematic illustration of the hierarchical structure of the F-LLoM model.

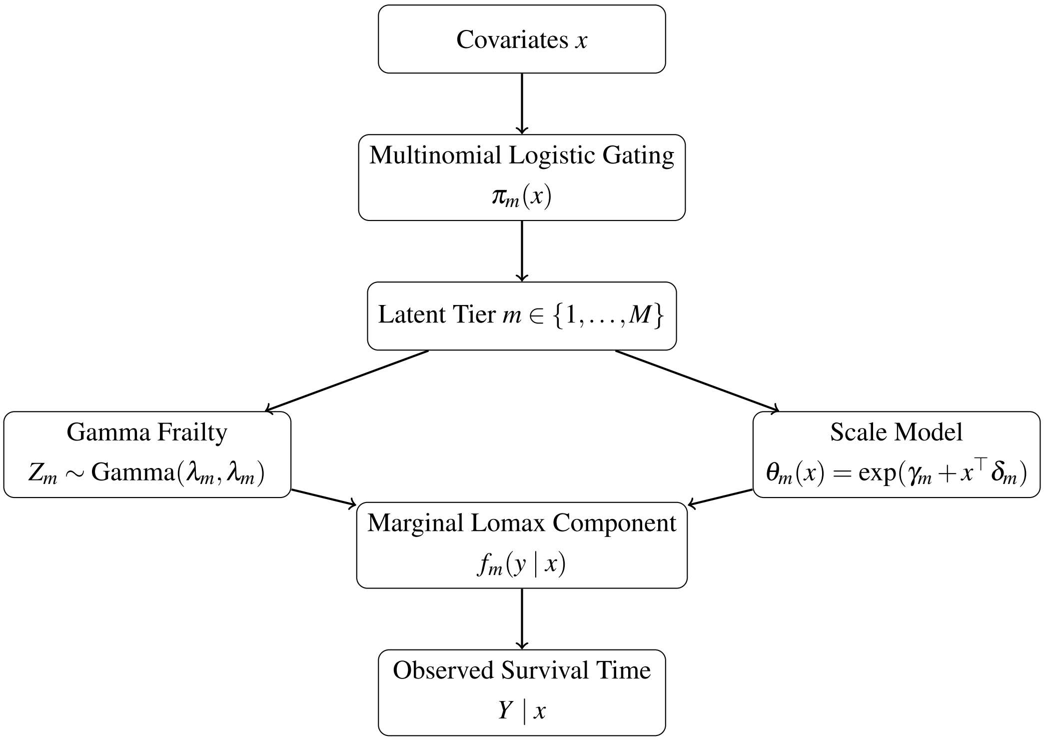


Figure 1. Hierarchical structure of the F-LLoM model. Covariates determine latent tier membership through logistic gating probabilities $\pi_m(x)$ and influence within-tier hazard scales $\theta_m(x)$. Integration over gamma frailty produces tier-specific Lomax components, which combine to form the observed survival distribution.

2.1. Lomax components via gamma frailty

For each tier $m = 1,\ldots,M$, unobserved heterogeneity is introduced through a gamma-distributed frailty

$$Z_m \sim \text{Gamma}(\lambda_m, \lambda_m),$$

parameterized to have $\mathbb{E}[Z_m] = 1$ and $\text{Var}(Z_m) = 1/\lambda_m$. Conditional on $Z_m = z$ and covariates x, the survival time follows an exponential distribution with hazard $z\,\theta_m(x)$, where $\theta_m(x) > 0$ is a covariate-dependent scale:

$$f_m(t \mid z,x) = z\,\theta_m(x)\exp\bigl(-z\,\theta_m(x)\,t\bigr), \qquad t \geq 0.$$

Marginalizing out the frailty yields a Lomax (Pareto II) distribution.

Theorem 1 (Gamma frailty $\Rightarrow$ Lomax)**.** *Let* $Z_m \sim \text{Gamma}(\lambda_m, \lambda_m)$ *and, conditional on* $(Z_m = z, x)$*, assume*

$$T \mid (Z_m = z, x) \sim \text{Exp}\bigl(z\,\theta_m(x)\bigr).$$

Then the marginal density of $T \mid x$ *is*

$$f_m(t \mid x) = \frac{\theta_m(x)\,\lambda_m^{\lambda_m+1}}{(\lambda_m + \theta_m(x)t)^{\lambda_m+1}}, \qquad t \geq 0,$$

that is, a Lomax (Pareto II) distribution with shape λ_m *and scale* $\kappa_m(x) = \lambda_m/\theta_m(x)$*.*

Proof. Conditional on $Z_m = z$, the survival function of T is

$$S_m(t \mid Z_m = z, x) = \exp\bigl(-z\,\theta_m(x)t\bigr).$$

Marginalizing over the Gamma frailty gives

$$S_m(t \mid x) = \mathbb{E}\left[e^{-Z_m \theta_m(x)t}\right].$$

For a Gamma(λ_m, λ_m) random variable, the Laplace transform is

$$\mathbb{E}\bigl[e^{-sZ_m}\bigr] = \left(\frac{\lambda_m}{\lambda_m + s}\right)^{\lambda_m}.$$

Setting $s = \theta_m(x)t$ yields

$$S_m(t \mid x) = \left(1 + \frac{\theta_m(x)t}{\lambda_m}\right)^{-\lambda_m}.$$

Differentiating this survival function gives the marginal density:

$$f_m(t \mid x) = -\frac{d}{dt}S_m(t \mid x) = \lambda_m \frac{\theta_m(x)}{\lambda_m + \theta_m(x)t}\left(1 + \frac{\theta_m(x)t}{\lambda_m}\right)^{-(\lambda_m+1)}.$$

Simplifying the expression gives the stated closed form:

$$f_m(t \mid x) = \frac{\theta_m(x)\,\lambda_m^{\lambda_m+1}}{(\lambda_m + \theta_m(x)t)^{\lambda_m+1}}.$$

Finally, writing

$$\lambda_m + \theta_m(x)t = \theta_m(x)\left(\frac{\lambda_m}{\theta_m(x)} + t\right) = \theta_m(x)\,\kappa_m(x)\left(1 + \frac{t}{\kappa_m(x)}\right),$$

with $\kappa_m(x) = \lambda_m/\theta_m(x)$, shows that this is exactly the Lomax density with shape λ_m and scale $\kappa_m(x)$. □

2.2. Finite mixture with covariate-dependent weights

The overall survival time Y is modeled as a finite mixture of the tier-specific Lomax components:

$$f_Y(y \mid x) = \sum_{m=1}^{M} \pi_m(x)\, f_m(y \mid x),$$

where $\pi_m(x) = \Pr(\text{Tier } m \mid x)$ are covariate-dependent mixing probabilities given by multinomial logistic regression:

$$\pi_m(x) = \frac{\exp(\alpha_m + x^\top \beta_m)}{\sum_{j=1}^{M}\exp(\alpha_j + x^\top \beta_j)}, \qquad m = 1, \ldots, M.$$

This multinomial logistic specification guarantees $\pi_m(x) > 0$ and $\sum_{m=1}^{M} \pi_m(x) = 1$ for all x, allowing personalized tier assignment in contrast to the fixed geometric weights of the DGEGF model [8].

The component-specific hazard scale is likewise modeled through a log-linear regression:

$$\theta_m(x) = \exp(\gamma_m + x^\top \delta_m),$$

so covariates influence both tier membership (via π_m) and within-tier hazard intensity (via θ_m). This dual dependence allows F-LLoM to capture nuanced interactions between observed risk factors and latent population structure.

Theorem 2. *The conditional density of the F-LLoM model is*

$$f_Y(y \mid x) = \sum_{m=1}^{M} \pi_m(x) \frac{\theta_m(x)\, \lambda_m^{\lambda_m+1}}{(\lambda_m + \theta_m(x) y)^{\lambda_m+1}}, \qquad y \geq 0,$$

and is a valid probability density function on $[0, \infty)$.

Proof. Each component density integrates to 1 by Theorem 1, and the mixing weights sum to 1, proving normalization. □

3. Theoretical properties

The F-LLoM model is supported by three theoretical properties: (i) identifiability of component-specific parameters, (ii) explicit expressions for component moments, and (iii) characterization of tail behaviour. Together, these results establish both statistical validity and practical interpretability, particularly in settings involving latent heterogeneity and heavy-tailed survival times.

3.1. Identifiability

A finite mixture model is identifiable when its parameters can be uniquely recovered (up to permutation of component labels) from the observed distribution. Identifiability is essential for meaningful estimation and for interpreting latent risk tiers.

Proposition 3 (Sufficient conditions for identifiability of F-LLoM)**.** *Consider the F-LLoM conditional density:*

$$f_Y(y \mid x) = \sum_{m=1}^{M} \pi_m(x) \frac{\theta_m(x)\, \lambda_m^{\lambda_m+1}}{(\lambda_m + \theta_m(x)\, y)^{\lambda_m+1}}, \qquad y \geq 0.$$

The model is identifiable up to label switching provided that

(i) the component-specific parameter vectors $(\lambda_m, \gamma_m, \delta_m)$ *are pairwise distinct so that the Lomax densities* $f_m(\cdot \mid x)$ *differ for some x, and*

(ii) the gating functions $\pi_m(x) = \exp(\alpha_m + x^\top \beta_m) / \sum_{j=1}^{M} \exp(\alpha_j + x^\top \beta_j)$ *vary non-degenerately with x (equivalently, the coefficient vectors* (α_m, β_m) *are not all identical).*

Sketch of proof. The result follows by combining two standard identifiability arguments. First, finite mixtures of Lomax (Pareto II) distributions are identifiable whenever the components have distinct shape–scale pairs $(\lambda_m, \theta_m(x))$ [12, 13]. Because both parameters enter the functional form, distinct pairs imply non-equivalent component densities. Second, when the mixing proportions vary non-degenerately with covariates, as in multinomial logistic regression, conditional variation in $\pi_m(x)$

across the covariate space enables recovery of the underlying latent components [14]. Combining these two facts yields identifiability of the full conditional model, up to the usual permutation of mixture labels. □

The above conditions ensure theoretical identifiability under standard finite-mixture assumptions. In finite samples, however, near-nonidentifiability may occur when both the mixing weights $\pi_m(x)$ and the component scales $\theta_m(x)$ depend on the same covariates. In such cases, partial compensating effects between gating and scale parameters may arise.

To mitigate such issues, practical diagnostics are advisable. These include examination of Hessian conditioning and standard errors, assessment of stability across multiple initializations, monitoring for extreme gating coefficients, and comparison of the Bayesian information criterion (BIC) and integrated complete likelihood (ICL) values to evaluate component separation. Such checks help distinguish theoretical identifiability from empirical estimability.

Under the sated conditions, latent tiers recovered by F-LLoM correspond to genuine subpopulations rather than parameter redundancies, subject to the usual considerations of finite-sample estimation.

3.2. Moments

Write the tier-m Lomax component in the standard parameterization with shape $\lambda_m > 0$ and scale $\kappa_m(x) = \lambda_m/\theta_m(x)$:

$$f_m(t \mid x) = \frac{\lambda_m}{\kappa_m(x)}\Big(1+\frac{t}{\kappa_m(x)}\Big)^{-(\lambda_m+1)}, \qquad t \geq 0.$$

Lemma 4 (Existence and formula of the r-th raw moment)**.** *For $r \geq 0$,*

$$\mathbb{E}\big[T^r \mid x, tier\ m\big] = \frac{\kappa_m(x)^r\,\Gamma(\lambda_m - r)\,\Gamma(1+r)}{\Gamma(\lambda_m)},$$

which is finite if and only if $r < \lambda_m$.

Proof. Direct calculation with the substitution $u = 1 + t/\kappa_m(x)$ yields a Beta integral:

$$\mathbb{E}[T^r] = \lambda_m\,\kappa_m(x)^r \int_1^\infty (u-1)^r u^{-(\lambda_m+1)}\,du = \kappa_m(x)^r \frac{\Gamma(r+1)\Gamma(\lambda_m - r)}{\Gamma(\lambda_m)}.$$

Finiteness follows from the pole of $\Gamma(\lambda_m - r)$ when $r \geq \lambda_m$. □

From Lemma 4 we obtain the familiar conditions:

$$\mathbb{E}[T \mid x, \text{tier } m] = \begin{cases} \dfrac{\lambda_m}{\theta_m(x)(\lambda_m - 1)}, & \lambda_m > 1, \\ \infty, & \lambda_m \leq 1, \end{cases}$$

and

$$\operatorname{Var}(T \mid x, \text{tier } m) = \begin{cases} \dfrac{\lambda_m^3}{\theta_m(x)^2(\lambda_m - 1)^2(\lambda_m - 2)}, & \lambda_m > 2, \\ \infty, & \lambda_m \leq 2. \end{cases}$$

Proposition 5 (Mixture moments at covariates x)**.** *Let $\pi_m(x)$ be the softmax weights, $\mu_m(x) = \mathbb{E}[T \mid x, tier\ m]$, and $v_m(x) = \mathrm{Var}(T \mid x, tier\ m)$ when finite. Then:*

(1) The mixture r-th moment exists iff every component with $\pi_m(x) > 0$ satisfies $\lambda_m > r$.
(2) If all active components satisfy $\lambda_m > 1$, then

$$\mathbb{E}[Y \mid x] = \sum_{m=1}^{M} \pi_m(x)\, \mu_m(x) = \sum_{m=1}^{M} \pi_m(x) \frac{\lambda_m}{\theta_m(x)(\lambda_m - 1)}.$$

(3) If all active components satisfy $\lambda_m > 2$, then

$$\mathrm{Var}(Y \mid x) = \sum_{m=1}^{M} \pi_m(x)\, v_m(x) + \sum_{m=1}^{M} \pi_m(x) \big(\mu_m(x) - \mathbb{E}[Y \mid x]\big)^2.$$

Proof. Immediate from the law of total expectation and the law of total variance: Mixture moments are convex combinations of component moments when those moments are finite; any infinite component moment renders the corresponding mixture moment infinite. □

Because both $\pi_m(x)$ and $\theta_m(x)$ depend on x, the conditional mean and variance vary nonlinearly with covariates. In particular, increasing weight on a tier with small λ_m (heavier tail) and large $\theta_m(x)$ (higher hazard) can substantially alter $\mathbb{E}[Y \mid x]$, even when individual regression effects are moderate.

3.3. *Tail behaviour and robustness to extremes*

The Lomax distribution has a heavy (power-law) tail, implying that extreme event times occur with higher probability than under light-tailed models such as the exponential or Weibull. For each F-LLoM component, the density and survival functions satisfy

$$f_m(y \mid x) \propto y^{-(\lambda_m+1)}, \qquad S_m(y \mid x) \propto y^{-\lambda_m} \quad (y \to \infty),$$

so the parameter λ_m directly controls tail thickness: Smaller values produce heavier tails.

In a finite mixture, the overall tail is governed by the component with the smallest shape parameter λ_m, because that component decays the slowest. Thus, letting $\lambda_{\min} = \min_m \lambda_m$, the mixture survival function satisfies

$$S_Y(y \mid x) \propto y^{-\lambda_{\min}}, \qquad y \to \infty,$$

with a constant determined by the mixing proportions $\pi_m(x)$ and scales $\theta_m(x)$ of the heaviest-tailed components.

Different tiers may therefore represent distinct tail regimes (e.g., low-risk strata with lighter tails and high-risk strata with heavier tails). Since both the scales $\theta_m(x)$ and mixing weights $\pi_m(x)$ depend on covariates, F-LLoM can model how extreme-event risk varies across covariate profiles.

4. Parameter estimation via EM algorithm

In the presence of right-censored data and latent tiers, direct maximization of the observed likelihood is intractable. We therefore employ the EM algorithm [7, 15], which is well suited to finite mixture models with latent classes and censoring. For subject $i = 1, \ldots, n$, we observe (y_i, c_i, x_i), where

$y_i > 0$ is the follow-up time, $c_i \in \{0,1\}$ indicates event occurrence, and $x_i \in \mathbb{R}^p$ denotes covariates. Let $z_{im} \in \{0,1\}$ indicate latent membership in tier m, with $\sum_{m=1}^{M} z_{im} = 1$. Right-censoring is assumed non-informative conditional on x_i.

The mixing probabilities follow a multinomial logit (softmax) model:

$$\pi_m(x) = \frac{\exp(\alpha_m + x^\top \beta_m)}{\sum_{j=1}^{M} \exp(\alpha_j + x^\top \beta_j)},$$

with $(\alpha_M, \beta_M) = (0,0)$ for identifiability [16].

Each component m follows a Lomax (Pareto II) distribution with shape $\lambda_m > 0$ and covariate-dependent rate $\theta_m(x) = \exp(\gamma_m + x^\top \delta_m) > 0$. The corresponding survival and density functions are

$$S_m(y \mid x) = \left(\frac{\lambda_m}{\lambda_m + \theta_m(x)\, y} \right)^{\lambda_m},$$

$$f_m(y \mid x) = \theta_m(x)\, \lambda_m^{\lambda_m+1} \left(\lambda_m + \theta_m(x)\, y\right)^{-(\lambda_m+1)}.$$

In log form,

$$\log f_m(y_i \mid x_i) = \log \theta_m(x_i) + (\lambda_m + 1) \log \lambda_m - (\lambda_m + 1) \log\left\{\lambda_m + \theta_m(x_i)\, y_i\right\},$$

$$\log S_m(y_i \mid x_i) = \lambda_m \log \lambda_m - \lambda_m \log\left\{\lambda_m + \theta_m(x_i)\, y_i\right\}.$$

The observed log-likelihood under right-censoring is

$$\ell(\Theta) = \sum_{i=1}^{n} \log \left[\sum_{m=1}^{M} \pi_m(x_i)\, f_m(y_i \mid x_i)^{c_i}\, S_m(y_i \mid x_i)^{1-c_i} \right].$$

Introducing the latent indicators z_{im} yields the complete-data log-likelihood:

$$\ell_c(\Theta) = \sum_{i=1}^{n} \sum_{m=1}^{M} z_{im} \Big\{ \log \pi_m(x_i) + c_i \log f_m(y_i \mid x_i) + (1 - c_i) \log S_m(y_i \mid x_i) \Big\}.$$

E-Step The E-step computes the posterior tier probabilities (responsibilities):

$$\gamma_{im} = \Pr(z_{im} = 1 \mid y_i, c_i, x_i; \Theta) = \frac{\pi_m(x_i)\, f_m(y_i \mid x_i)^{c_i}\, S_m(y_i \mid x_i)^{1-c_i}}{\sum_{j=1}^{M} \pi_j(x_i)\, f_j(y_i \mid x_i)^{c_i}\, S_j(y_i \mid x_i)^{1-c_i}}.$$

M-Step The M-step maximizes the expected complete-data log-likelihood:

$$Q(\Theta \mid \Theta^{(t)}) = \sum_{i=1}^{n} \sum_{m=1}^{M} \gamma_{im} \Big\{ \log \pi_m(x_i) + c_i \log f_m(y_i \mid x_i) + (1 - c_i) \log S_m(y_i \mid x_i) \Big\}.$$

(a) *Gating parameters* (α_m, β_m) are updated by solving a weighted multinomial logistic regression:

$$\max_{\{\alpha,\beta\}} \sum_{i=1}^{n} \sum_{m=1}^{M} \gamma_{im} \log \pi_m(x_i),$$

using standard methods such as IRLS or quasi-Newton optimization.

(b) *Component parameters* $(\lambda_m, \gamma_m, \delta_m)$ are updated by maximizing, for each m,

$$\ell_m = \sum_{i=1}^{n} \gamma_{im} \Big[c_i \log f_m(y_i \mid x_i) + (1 - c_i) \log S_m(y_i \mid x_i) \Big].$$

This optimization is performed numerically (e.g., via L-BFGS-B) over $(\gamma_m, \delta_m, \eta_m)$ for numerical stability, where λ_m is parameterized through η_m as described below. Initial values are obtained from preliminary fits within provisional clusters.

In the uncensored case ($c_i \equiv 1$), the algorithm reduces to standard mixture estimation. With censoring, the $\log S_m$ terms retain information from censored observations, improving both estimation efficiency and predictive calibration.

A Lomax component has a finite mean if and only if $\lambda_m > 1$. To avoid explicit inequality constraints, we adopt the following reparameterization:

$$\lambda_m = \begin{cases} \exp(\eta_m), & \text{default specification,} \\ 1 + \exp(\eta_m), & \text{finite-mean specification.} \end{cases}$$

The default specification permits fully heavy-tailed behavior, whereas the finite-mean specification enforces $\lambda_m > 1$ and ensures interpretability of conditional mean survival time. The chosen parameterization is stated explicitly in each application.

Practical implementation guidance

Although the F-LLoM framework is flexible, careful implementation enhances numerical stability and interpretability. We summarize several practical considerations for applied use.

Choice of the number of components M. Selection of M should balance fit and parsimony. Information criteria such as BIC and ICL are recommended, with preference given to the smallest M achieving stable improvement. Increases in log-likelihood alone should not justify additional components. Examination of posterior responsibilities can help detect redundant tiers.

Covariate scaling and preprocessing. Since covariates enter both the gating model $\pi_m(x)$ and the scale model $\theta_m(x)$, standardizing continuous predictors is recommended to improve numerical conditioning. Highly collinear predictors may inflate standard errors and reduce separation between gating and scale effects.

Initialization and local optima. As with most finite mixture models, the EM algorithm may converge to local maxima. Multiple random initializations are therefore recommended, with selection based on the highest attained log-likelihood.

Diagnosing weak identifiability. When both mixing weights and scale parameters depend on covariates, near-compensating effects may arise. Diagnostics include examining Hessian conditioning, monitoring unusually large gating coefficients, assessing sensitivity to initialization, and comparing BIC and ICL for evidence of weak separation.

Finite-mean specification. If interpretability of conditional mean survival time is important, the finite-mean parameterization ($\lambda_m = 1 + \exp(\eta_m)$) may be adopted. Otherwise, the default specification permits heavier tails when supported by the data.

These considerations improve numerical stability while preserving model flexibility.

5. Simulation study

We evaluate the ability of F-LLoM to recover latent structure across sample sizes $N \in \{500, 800, 1000\}$ and right-censoring rates CR $\in \{10\%, 15\%, 20\%\}$. Data are generated from a two-tier ($M = 2$) specification with covariate-dependent hazards and softmax gating:

$$\text{Tier 1 (low risk)}: \lambda_1 = 6.0, \quad \theta_1(x) = \exp(-1.0 - 0.8x),$$
$$\text{Tier 2 (high risk)}: \lambda_2 = 1.2, \quad \theta_2(x) = \exp(1.0 + 0.9x),$$

with gating logit $\log\{\pi_1(x)/\pi_2(x)\} = 1.2 - 0.9x$ and $x \sim \mathcal{N}(0,1)$. Independent right-censoring is applied conditional on x. For each scenario, we fit models with $M \in \{1, \ldots, 4\}$ using the censoring-aware EM algorithm (Section 4) with multiple random starts.

Model selection is performed using the Bayesian information criterion (BIC) and the integrated complete likelihood (ICL):

$$\mathrm{BIC}(M) = -2\,\ell(\widehat{\Theta}_M) + p_M \log n, \qquad \mathrm{ICL}(M) = \mathrm{BIC}(M) + 2H_M,$$

where p_M is the number of free parameters and $H_M = -\sum_{i,k} \widehat{\gamma}_{im} \log \widehat{\gamma}_{im}$ is the entropy penalty [17, 18]. Likelihood-ratio tests are generally inapplicable for finite mixtures; see [7].

Table 1 reports log-likelihood, BIC, and ICL across scenarios. In all settings, BIC selects the true $M = 2$ specification. The optimal BIC values range from 1161.085 at $N = 500$, CR $= 0.20$ to 2601.187 at $N = 1000$, and CR $= 0.10$. ICL, which incorporates an entropy penalty, also favors $M = 2$ in most scenarios but occasionally selects $M = 3$ or 4 under higher classification uncertainty (e.g., ICL = 3847.840 for $M = 3$ at $N = 1000$, CR $= 0.10$).

Table 2 reports estimated coefficients under the correctly specified $M = 2$ model. For clarity, results are grouped by sample size and censoring level. Parameter recovery improves with sample size. For example, at $N = 1000$ and CR $= 20\%$, the estimated component-scale intercepts $\log\theta \approx [1.055, -0.759]$ and covariate effects $\approx [0.937, -0.742]$ align closely with the true values (up to label switching). The estimated gating coefficients reproduce the intended tier allocation pattern. Estimated λ_m values remain within moderate ranges ($1.319 - 30.247$), with no boundary estimates observed.

Figure 2 shows that EM log-likelihood trajectories stabilize within approximately $20 - 30$ iterations and are consistent across random starts. Predictive calibration, evaluated using PIT histograms $U = S(T \mid X)$ (Figure 3), is approximately uniform across scenarios and improves with increasing N. Figure 4 displays BIC($M = 2$) across censoring levels and sample sizes. Absolute BIC values increase with N, while within each N they decrease as CR increases. In all settings, $M = 2$ remains BIC-optimal.

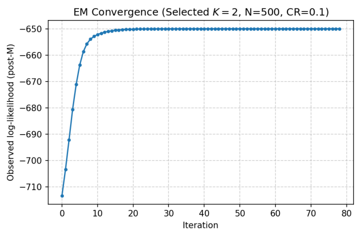
Table 1. Model-selection criteria across simulated scenarios.

Scenario	M	Log-lik	BIC	ICL
$N = 500, \mathrm{CR} = 0.10$	1	-705.511	1429.665	1429.665
	2	-649.926	**1349.570***	1572.206
	3	-644.355	1369.500	2058.497
	4	-643.597	1399.058	2193.361
$N = 500, \mathrm{CR} = 0.15$	1	-662.876	1344.395	1344.395
	2	-611.066	**1271.848***	1485.834
	3	-606.784	1294.357	1952.037
	4	-605.142	1322.147	2031.917
$N = 500, \mathrm{CR} = 0.20$	1	-601.915	1222.475	1222.475
	2	-555.684	**1161.085***	1383.692
	3	-552.215	1185.219	1865.249
	4	-550.487	1212.838	1986.408
$N = 800, \mathrm{CR} = 0.10$	1	-1093.195	2206.444	2206.444
	2	-992.907	**2039.291***	2654.207
	3	-991.119	2069.137	3259.278
	4	-988.839	2098.000	3148.554
$N = 800, \mathrm{CR} = 0.15$	1	-984.137	1988.327	1988.327
	2	-890.960	**1835.397***	2470.837
	3	-889.993	1866.885	2824.480
	4	-887.975	1896.272	3035.361
$N = 800, \mathrm{CR} = 0.20$	1	-917.554	1855.162	1855.162
	2	-824.332	**1702.142***	2319.047
	3	-823.601	1734.102	2668.785
	4	-822.436	1765.196	2947.055
$N = 1000, \mathrm{CR} = 0.10$	1	-1375.935	2772.593	2772.593
	2	-1272.962	**2601.187***	3087.882
	3	-1271.535	2632.872	3847.840
	4	-1267.519	2659.378	4162.070
$N = 1000, \mathrm{CR} = 0.15$	1	-1277.007	2574.738	2574.738
	2	-1175.280	**2405.822***	3124.256
	3	-1172.561	2434.924	3734.838
	4	-1168.386	2461.111	3871.588
$N = 1000, \mathrm{CR} = 0.20$	1	-1161.436	2343.596	2343.596
	2	-1066.077	**2187.415***	2690.571
	3	-1064.317	2218.434	3561.873
	4	-1059.169	2242.677	3635.158

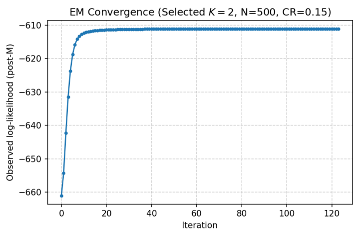
Table 2. Combined $M = 2$ coefficients across N and censoring.

Scenario N	CR	Tier	λ	Component scale ($\log\theta$) Intercept	x	Gating logit (vs. baseline) Intercept	x
500	0.10	1	5.285	-1.021	-0.683	1.539	-0.958
		2	13.744	1.419	0.533	0.000	0.000
500	0.15	1	4.182	-0.989	-0.701	1.591	-0.988
		2	30.247	1.436	0.513	0.000	0.000
500	0.20	1	3.283	-0.952	-0.727	1.598	-0.969
		2	11.293	1.429	0.566	0.000	0.000
800	0.10	1	2.425	0.506	1.169	-0.817	0.820
		2	7.101	-1.060	-0.882	0.000	0.000
800	0.15	1	2.195	0.490	1.170	-0.739	0.808
		2	7.330	-1.099	-0.930	0.000	0.000
800	0.20	1	2.493	0.495	1.183	-0.818	0.838
		2	7.429	-1.097	-0.925	0.000	0.000
1000	0.10	1	2.178	1.095	0.901	-1.905	1.147
		2	3.482	-0.782	-0.738	0.000	0.000
1000	0.15	1	1.319	0.776	1.079	-1.270	0.814
		2	7.566	-0.893	-0.809	0.000	0.000
1000	0.20	1	1.791	1.055	0.937	-1.922	1.181
		2	3.291	-0.759	-0.742	0.000	0.000

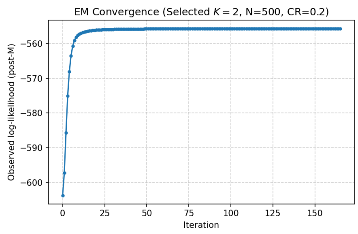
Note. Results are reported for the correctly specified $M = 2$ model across all sample sizes and censoring levels. Parameter recovery improves with increasing N. Estimates are reported by component; label switching does not affect interpretation of recovery accuracy.



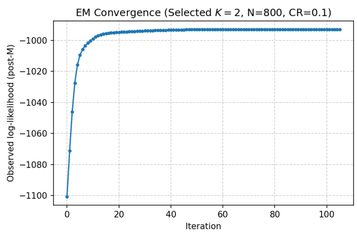
(a) $N = 500,\ \mathrm{CR} = 0.10$



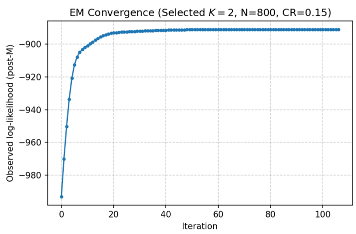
(b) $N = 500,\ \mathrm{CR} = 0.15$



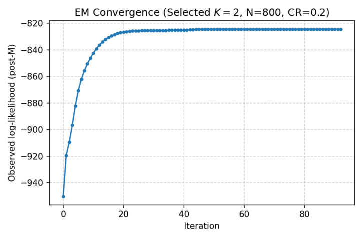
(c) $N = 500,\ \mathrm{CR} = 0.20$



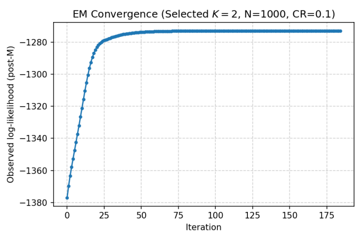
(d) $N = 800,\ \mathrm{CR} = 0.10$



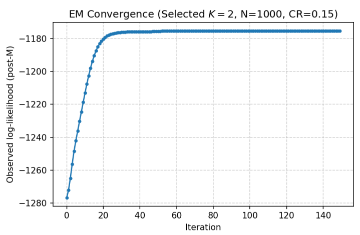
(e) $N = 800,\ \mathrm{CR} = 0.15$



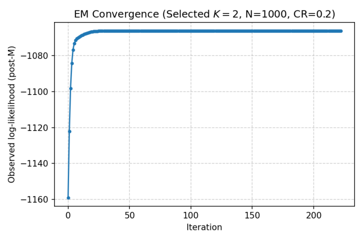
(f) $N = 800,\ \mathrm{CR} = 0.20$



(g) $N = 1000,\ \mathrm{CR} = 0.10$

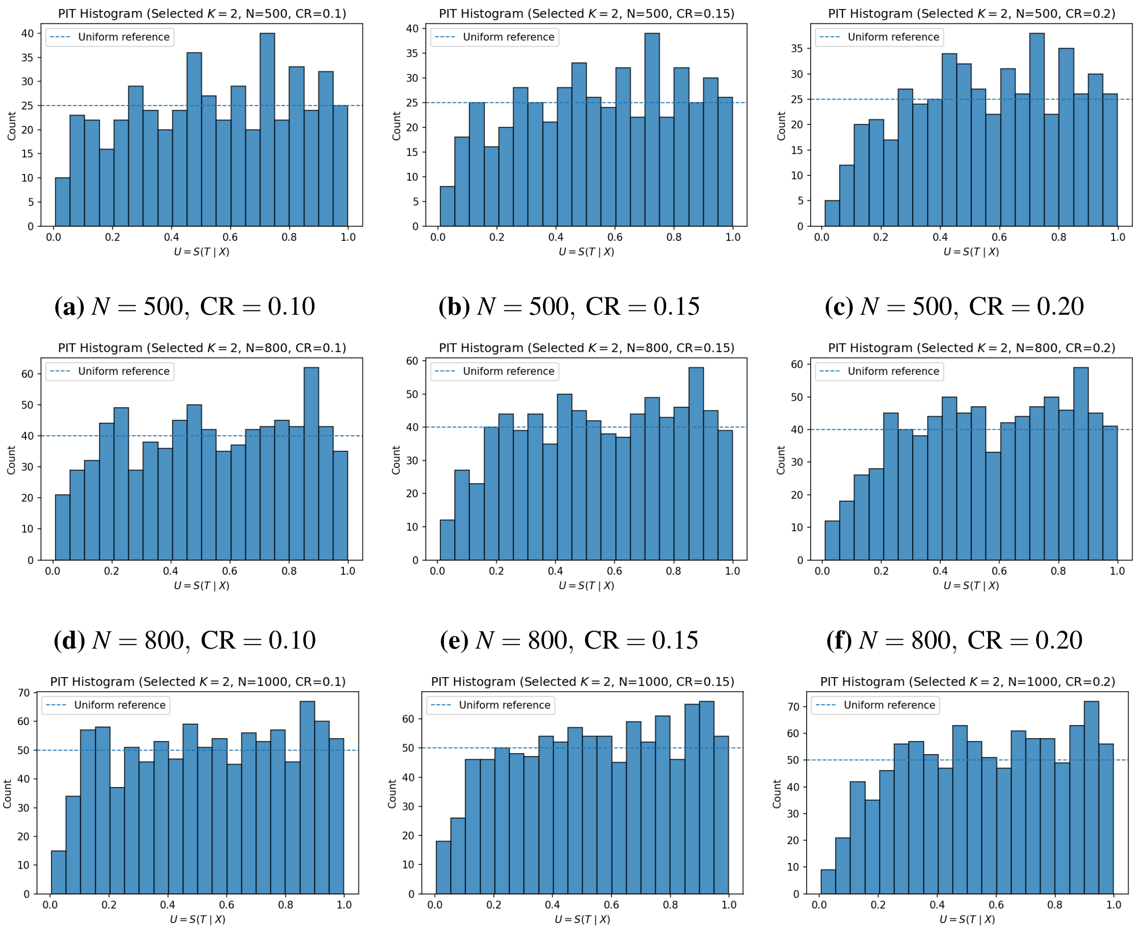


(h) $N = 1000,\ \mathrm{CR} = 0.15$



(i) $N = 1000,\ \mathrm{CR} = 0.20$

Figure 2. EM convergence for $M = 2$ across all (N, CR) scenarios.



(a) $N = 500,\ \mathrm{CR} = 0.10$ **(b)** $N = 500,\ \mathrm{CR} = 0.15$ **(c)** $N = 500,\ \mathrm{CR} = 0.20$

(d) $N = 800,\ \mathrm{CR} = 0.10$ **(e)** $N = 800,\ \mathrm{CR} = 0.15$ **(f)** $N = 800,\ \mathrm{CR} = 0.20$

(g) $N = 1000,\ \mathrm{CR} = 0.10$ **(h)** $N = 1000,\ \mathrm{CR} = 0.15$ **(i)** $N = 1000,\ \mathrm{CR} = 0.20$

Figure 3. PIT histograms $U = S(T \mid X)$ for $M = 2$ across all (N, CR) scenarios; near-uniformity improves with N.

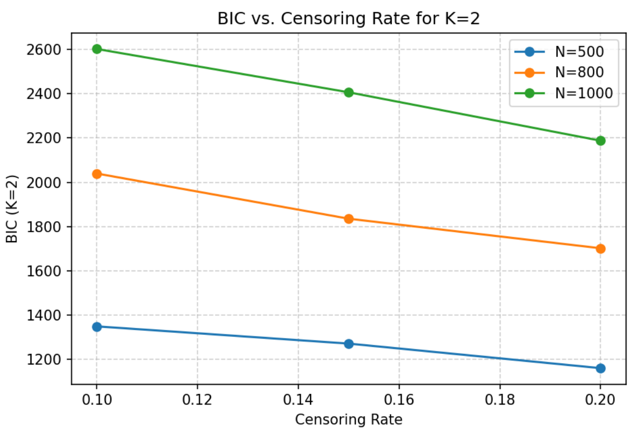


Figure 4. BIC for $M = 2$ across censoring rates and sample sizes.

Overall, the simulation results indicate that: (i) BIC selects the correct number of latent tiers ($M = 2$) across sample sizes and censoring levels; (ii) PIT diagnostics suggest adequate predictive calibration; and (iii) the censoring-aware EM algorithm converges stably under moderate censoring.

Computational considerations. All simulations and empirical analyses were implemented in Python using standard numerical optimization routines (L-BFGS-B) within the EM framework. Computational time increased approximately linearly with sample size n and with the number of mixture components M per EM iteration. For each fitted specification, multiple random initializations (typically 5–10 starts) were used to mitigate convergence to local maxima. The BIC-optimal solution was stable across initializations in all scenarios.

6. Application: time-to-rearrest in the Rossi recidivism dataset

We evaluate the proposed F-LLoM model using the widely studied Rossi recidivism dataset [19]. The dataset comprises $N = 432$ formerly incarcerated individuals followed for 52 weeks after release. During this period, 114 subjects (26.4%) were rearrested, while 318 (73.6%) were right-censored. Seven criminological covariates are available: financial aid (`fin`), age, race, work experience (`wexp`), marital status (`mar`), parole status (`paro`), and number of prior convictions (`prio`). Owing to its policy relevance and moderate censoring, this dataset has become a benchmark example in survival analysis.

For this application, we adopt the finite-mean parameterization $\lambda_m = 1 + \exp(\eta_m)$ to ensure interpretability of conditional mean survival times. F-LLoM models with $M \in \{1, 2, 3, 4\}$ components were estimated and compared using BIC.

Table 3. Model selection results for the F-LLoM models.

M	logLik	k	BIC	ΔBIC
1	-686.393	9	1427.4	0.0
2	-674.502	26	1506.8	79.4
3	-667.071	43	1595.1	167.7
4	-661.255	60	1686.6	259.2

The BIC increases monotonically with M, indicating no empirical support for additional latent components. The preferred specification is therefore the single-component model ($M = 1$).

Comparison with standard survival models

To provide empirical context, we additionally fitted a Cox proportional hazards model and a Weibull accelerated failure time (AFT) model using the same covariates. Table 4 summarizes the corresponding log-likelihood and BIC values.

Table 4. Comparison with standard survival models.

Model	logLik	k	BIC
Cox PH	-658.748	7	1359.97
Weibull AFT	-679.917	9	1414.45
F-LLoM ($M = 1$)	-686.393	9	1427.40

The Cox model yields the lowest BIC when computed from the partial likelihood. While partial-likelihood and full-likelihood BIC values are not directly comparable in a strict sense, this result suggests that a semiparametric specification provides an adequate description of the data. The Weibull AFT model serves as a fully parametric benchmark and exhibits fit comparable to the single-component F-LLoM specification. Overall, these comparisons indicate that the Rossi data do not display strong evidence of heavy-tailed or multimodal structure requiring mixture complexity, consistent with the BIC preference for $M = 1$ within the F-LLoM framework (Figure 5).

When $M = 1$, the multinomial gating mechanism becomes degenerate, with $\pi_1(x) \equiv 1$ for all covariate values. In this case, the gating coefficients have no effect on component allocation, likelihood evaluation, or predicted survival probabilities. They are included solely to maintain notational consistency with the general $M > 1$ formulation. All substantive covariate effects arise from the component-specific scale model.

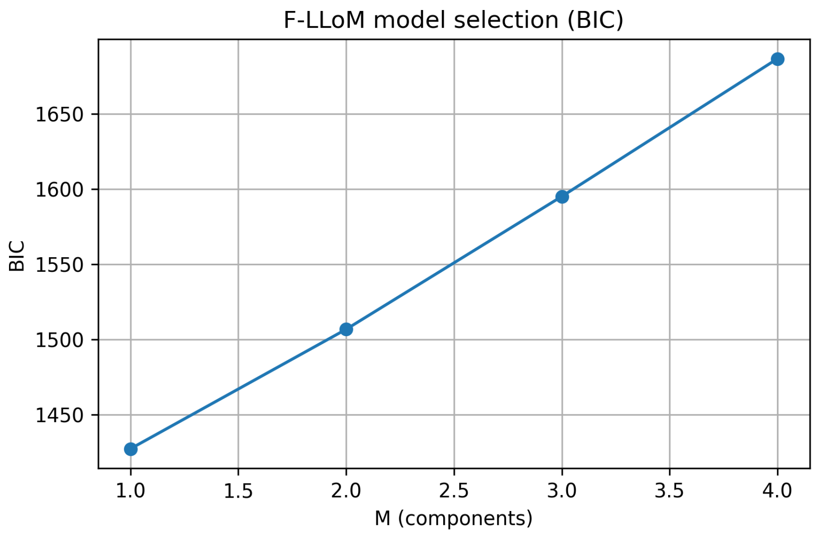


Figure 5. BIC values for F-LLoM models with $M = 1, \ldots, 4$.

Calibration and interpretation

Model calibration was assessed using the randomized probability integral transform (PIT). Under correct specification, PIT values are approximately uniformly distributed. Figure 6 shows no pronounced systematic deviations from uniformity for the selected $M = 1$ model, suggesting adequate predictive calibration.

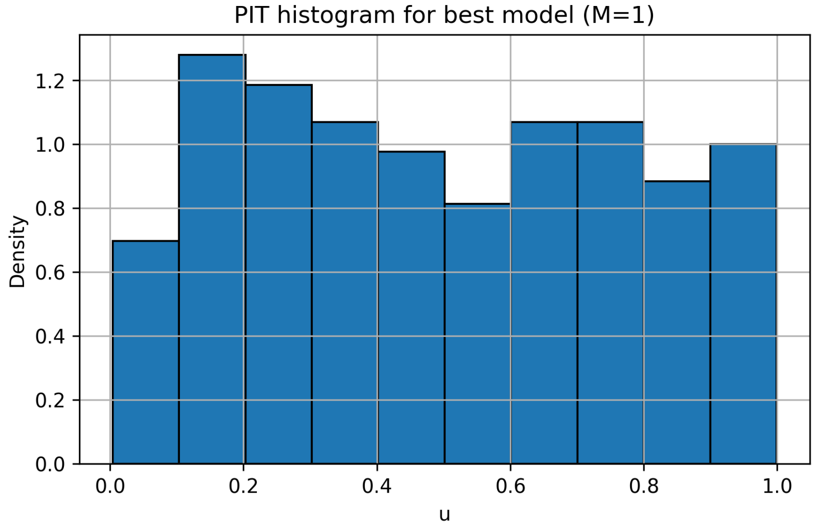


Figure 6. Randomized PIT histogram for the selected model ($M = 1$).

The estimated baseline parameters are $\hat{\lambda}_1 = 140.97$ and $\hat{\gamma}_1 = -5.264$ (Table 5). The large estimated value of $\hat{\lambda}_1$ suggests limited residual heterogeneity after accounting for observed covariates.

Table 5. Unified parameter estimates for the $M = 1$ F-LLoM model.

Parameter	Frailty/Baseline	Scale Effects δ_{1j}	Gating Effects β_{1j}
λ_1	140.968		
γ_1	-5.264		
fin		-0.182	-0.122
age		-0.339	0.078
race		0.100	0.245
wexp		-0.073	-0.063
mar		-0.137	-0.056
paro		-0.040	0.007
prio		0.247	-0.095
Intercept (α_1)			0.000

Note. Frailty parameters (λ_1, γ_1) define the Lomax component. Covariate effects δ_{1j} determine the scale parameter. Gating parameters are included for completeness but have no substantive role when $M = 1$.

Among the covariates, the number of prior convictions (`prio`, $\hat{\delta}_1 = 0.247$) is the strongest risk-increasing predictor, followed by race ($\hat{\delta}_1 = 0.100$). The most pronounced protective effects are associated with age ($\hat{\delta}_1 = -0.339$), financial aid ($\hat{\delta}_1 = -0.182$), and marital status ($\hat{\delta}_1 = -0.137$). These patterns are consistent with established findings in the criminological literature linking maturity, social stability, and economic support to reduced recidivism risk.

In summary, the F-LLoM model provides an interpretable and adequately calibrated fit to the Rossi dataset. The lack of BIC support for $M > 1$ and the large estimated frailty shape parameter suggest that observed covariates account for most of the heterogeneity in rearrest risk. At the same time, the unified framework retains the flexibility to accommodate latent tiers in applications where mixture structure is empirically supported.

7. Conclusions

We introduced the Frailty-Augmented Logistic-Weighted Lomax Mixture (F-LLoM) model, a finite mixture survival framework that integrates gamma frailty within components and covariate-dependent multinomial logistic mixing weights. By combining tier-specific Lomax components with covariate-driven allocation and scale effects, the model accommodates both latent heterogeneity and heavy-tailed survival behavior within a unified and interpretable structure.

We established sufficient identifiability conditions, derived explicit moment and tail properties, and developed a censoring-aware EM algorithm for right-censored data. Simulation results demonstrated reliable recovery of latent structure, stable convergence of the estimation procedure, and consistent component selection via BIC across moderate sample sizes and censoring levels. In the application to the Rossi recidivism dataset, the framework provided a well-calibrated and interpretable fit and favored a parsimonious single-component specification after incorporating covariates.

Overall, the F-LLoM model offers a theoretically grounded and computationally tractable approach for analyzing heterogeneous and potentially heavy-tailed survival data. The framework is flexible enough to recover latent risk tiers when supported by the data, while reducing to a simpler parametric structure when mixture complexity is not warranted.

Use of Generative-AI tools declaration

The author declares that the use of AI tools is limited to language editing and checking the logical consistency of certain mathematical arguments. All scientific content, proofs, results, and conclusions were developed and independently verified by the author, who takes full responsibility for the work.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU261110].

Conflict of interest

The author declares no conflict of interest in this paper.

References

1. M. E. Bakr, O. S. Balogun, A. A. El-Toony, A. M. Gadallah, Reliability analysis for unknown age class of lifetime distribution with real applications in medical science, *Symmetry*, **16** (2024), 1514. https://doi.org/10.3390/sym16111514 10.3390/sym16111514

2. K. J. Haganawiga, S. K. Pal, A. Sirohi, Suitable survival models for analyzing infant and child mortality data, *Stat. J. IAOS*, **40** (2024), 959–968. https://doi.org/10.3233/sji-240073

3. J. P. Klein, M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*, Springer New York, 2003. https://doi.org/10.1007/b97377

4. D. R. Cox, Regression models and life-tables, *J. R. Stat. Soc. Ser. B (Methodol.)*, **34** (1972), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

5. P. Hougaard, Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, **73** (1986), 387–396. https://doi.org/10.1093/biomet/73.2.387

6. A. Wienke, *Frailty models in survival analysis*, Chapman and Hall/CRC, 2010. https://doi.org/10.1201/9781420073911

7. G. McLachlan, D. Peel, *Finite mixture models*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc, 2000. https://doi.org/10.1002/0471721182

8. M. Rahmouni, The doubly generalized exponential-geometric frailty distribution, *AIMS Math.*, **10** (2025), 30384–30428. https://doi.org/10.3934/math.20251334

9. M. Rahmouni, The doubly generalized Weibull power series frailty distribution, *Int. J. Anal. Appl.*, **23** (2025), 255. https://doi.org/10.28924/2291-8639-23-2025-255

10. J. D. Kalbfleisch, R. L. Prentice, *The statistical analysis of failure time data*, John Wiley & Sons, Inc, 2002. https://doi.org/10.1002/9781118032985

11. K. S. Lomax, Business failures: another example of the analysis of failure data, *J. Amer. Stat. Assoc.*, **49** (1954), 847–852. https://doi.org/10.1080/01621459.1954.10501239

12. H. Teicher, Identifiability of finite mixtures, *Ann. Math. Statist.*, **34** (1963), 1265–1269. https://doi.org/10.1214/aoms/1177703862

13. S. J. Yakowitz, J. D. Spragins, On the identifiability of finite mixtures, *Ann. Math. Statist.*, **39** (1968), 209–214. https://doi.org/10.1214/aoms/1177698520

14. W. Yao, Label switching and its solutions for frequentist mixture models, *J. Stat. Comput. Simul.*, **85** (2013), 1000–1012. https://doi.org/10.1080/00949655.2013.859259

15. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc. Ser. B (Methodol.)*, **39** (1977), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

16. A. Agresti, *Categorical data analysis*, John Wiley & Sons, Inc, 2002. https://doi.org/10.1002/0471249688

17. G. Schwarz, Estimating the dimension of a model, *Ann. Statist.*, **6** (1978), 461–464. https://doi.org/10.1214/aos/1176344136

18. C. Biernacki, G. Celeux, G. Govaert, Assessing a mixture model for clustering with the integrated completed likelihood, *IEEE Trans. Pattern Anal. Mach. Intell.*, **22** (2000), 719–725. https://doi.org/10.1109/34.865189

19. P. H. Rossi, R. A. Berk, K. J. Lenihan, *Money, work, and crime: experimental evidence*, New York: Academic Press, Burlington, 1980.

Appendix

Additional empirical results for the F-LLoM model

This appendix reports supplementary estimation results for F-LLoM models with $M \in \{1,2,3,4\}$ components. The main text focuses on the BIC-optimal single-component model ($M = 1$). Results for higher-order mixtures are provided here for completeness and diagnostic comparison.

A.1. Model selection

Table A1 reproduces the log-likelihood, parameter count, and BIC values for all fitted models.

Table A1. Model selection results for F-LLoM models.

M	logLik	k	BIC	ΔBIC
1	-686.393	9	1427.4	0.0
2	-674.502	26	1506.8	79.4
3	-667.071	43	1595.1	167.7
4	-661.255	60	1686.6	259.2

The steady increase in BIC confirms that additional mixture components are not supported by the data. Despite higher log-likelihood values for $M > 1$, the complexity penalty outweighs any incremental improvement in fit.

A.2. Component-specific parameters

Tables A2–A4 summarize parameter estimates across models.

Table A2. Frailty shapes λ_m and baseline log-scales γ_m across models.

	$M = 1$	$M = 2$		$M = 3$		
Component	1	1	2	1	2	3
λ_m	140.968	142.036	141.732	141.993	140.500	142.047
γ_m	-5.264	-5.628	-4.082	-5.785	-4.710	-5.221

Table A3. Covariate effects δ_{mj} on the component-specific scale parameter.

	$M = 1$	$M = 2$		$M = 3$		
Covariate	1	1	2	1	2	3
`fin`	-0.182	-0.227	-0.898	-0.278	-1.440	-0.072
`age`	-0.339	-0.483	-0.499	-0.596	-0.361	-0.534
`race`	0.100	0.057	0.412	0.005	0.285	0.137
`wexp`	-0.073	0.060	0.014	0.079	0.015	-0.134
`mar`	-0.137	-0.186	-0.186	-0.177	-0.310	-0.057
`paro`	-0.040	-0.097	-0.390	-0.203	0.087	0.015
`prio`	0.247	0.306	0.136	0.334	2.024	0.253

Table A4. Gating parameters (α_m, β_m) across models.

	$M=1$	$M=2$		$M=3$		
Parameter	1	1	2	1	2	3
α_m	0.000	77.502	-77.512	160.208	-310.798	150.403
$\beta(\texttt{fin})$	-0.122	-51.454	51.379	83.107	-179.536	96.031
$\beta(\texttt{age})$	0.078	-42.633	42.876	-57.233	121.518	-64.128
$\beta(\texttt{race})$	0.245	-49.622	49.764	-44.107	71.461	-27.624
$\beta(\texttt{wexp})$	-0.063	47.132	-46.930	18.636	-41.629	22.859
$\beta(\texttt{mar})$	-0.056	-12.320	12.605	27.355	-61.507	34.017
$\beta(\texttt{paro})$	0.007	-60.798	61.031	-9.908	9.199	0.773
$\beta(\texttt{prio})$	-0.095	-1.409	1.728	36.135	-68.165	31.880

Frailty and baseline parameters. Across all specifications, the estimated frailty shape parameters remain large and stable. This indicates limited residual heterogeneity beyond observed covariates and explains why mixture extensions do not materially improve model fit.

Scale effects. For $M > 1$, covariate effects exhibit instability across components, particularly for `fin` and `prio`. This variability reflects over-parameterization rather than meaningful latent segmentation, consistent with the BIC results.

Gating parameters. For $M > 1$, gating coefficients become extremely large in magnitude, indicating near-separation and unstable class allocation. This behavior confirms that the data do not support meaningful latent tiers and that mixture extensions primarily introduce numerical instability.

A.3. Robustness and stability

Re-estimation across subgroups defined by race, parole status, and financial aid yielded consistent qualitative conclusions. In every subsample, $M = 1$ remained BIC-optimal. Parameter signs and relative magnitudes were stable.

Alternative preprocessing of covariates (e.g., centering and standardization) produced nearly identical results. Across 20 random EM initializations, the $M = 1$ model consistently converged to the same optimum, while $M > 1$ models frequently exhibited near-separated gating solutions and sensitivity to initialization.

Overall, these supplementary analyses reinforce the conclusion that the Rossi dataset does not exhibit latent mixture structure beyond covariate-driven heterogeneity.