Shahid Chamran University of Ahvaz

Journal of Advanced Mathematical Modeling
ISSN (print): 2251-8088, ISSN (on-line): 2645-6141
Vol. x No. x (2026), pp. xx-xx.
DOI: x No.

ROBUST MODEL-BASED REGRESSION CLUSTERING FOR TIME SERIES VIA SKEW-LAPLACE POLYNOMIAL REGRESSION

ALI SHADROKH*, SAMIRA SAADATI AND PARVIZ NASIRI

Communicated by:

Abstract. This study proposes a model-based clustering framework for time series by incorporating Laplace and Skew-Laplace distributions to address non-normality, skewness, and heavy-tailedness in empirical data. The proposed approach assumes that cluster observations follow a polynomial regression characterized by a temporal Vandermonde design matrix. Model parameters are estimated using both Frequentist and Bayesian frameworks. Simulation results demonstrate that while the Skew-Laplace model performs comparably to Normal and Laplace models under symmetric conditions, it significantly outperforms them when the data exhibit skewness and heavy tails. In such scenarios, symmetric distributions typically overestimate variance to compensate for their lack of flexibility, whereas the Skew-Laplace model provides a robust and efficient fit. Furthermore, the Bayesian approach exhibits superior performance over frequentist methods, particularly in clusters with limited sample sizes. These findings underscore the effectiveness of the Skew-Laplace model as a comprehensive framework for complex time series analysis, offering high adaptability across diverse data structures.

MSC(2020): 00A09, 54C40.

Keywords: Time series clustering, Model-based clustering, Polynomial regression, Skew-Laplace distribution, MCMC and EM algorithms.

Received: dd mmmm yyyy, Accepted: dd mmmm yyyy.

*Corresponding author.

1. Introduction

Clustering is one of the fundamental methods in time series analysis. In this approach, a specific set of time series is partitioned into several groups based on distinct temporal patterns. From a statistical learning perspective, cluster analysis falls under the category of unsupervised learning. The primary goal of clustering is to form partitions or clusters of observations such that members within each cluster exhibit maximum similarity based on a similarity or dissimilarity criterion, while showing minimal similarity to members of other groups.

Generally, clustering methods can be classified into four main categories based on their underlying mechanisms: hierarchical, centroid-based, density-based, and model-based approaches. Furthermore, all clustering algorithms can be broadly categorized into hard (crisp) and soft (overlapping) clustering. In hard clustering, each observation is assigned deterministically to a single cluster; in contrast, soft clustering allows an observation to belong to multiple clusters with varying degrees of membership.

In the context of time series clustering, the model-based approach ([6, 16, 25]) possesses two fundamental advantages compared to other methods. First, if the selected model accurately describes the temporal patterns in the time series, this approach typically yields more precise results than other methods ([27, 35]) Unlike symmetric robust models such as Student-t or Pearson VII mixtures, the proposed framework simultaneously accounts for both heavy tails and distributional asymmetry (skewness), which are frequently observed in real-world data. Furthermore, while our previous work focused on supervised multilevel structures [14], the current study develops an unsupervised model-based clustering framework, addressing the inherent challenges of latent group membership estimation. Second, it enables future prediction through the use of the estimated model.

In model-based clustering, it is assumed that observations are samples from a finite mixture distribution [26]. A finite mixture distribution is a convex combination of k component densities with distinct characteristics. In this framework, each density, referred to as a component, corresponds to a separate cluster in the data. The fundamental assumption is that each observation is generated from one of the components with a certain probability. Consequently, only in model-based clustering are the principles of probability distributions and statistical inference tools employed to model the hidden

structure of the data, whereas in other methods, the distribution of observations plays no role in the grouping process.

This fundamental distinction has significant implications. Unlike methods that rely solely on distance structures, model-based clustering elevates the analytical framework from mere grouping to statistical inference. Specifically, this approach allows for the determination of the optimal number of clusters using standard statistical evaluation metrics, such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC). Thus, model parameters are determined not by subjective intuition but through quantitative criteria. Consequently, model-based clustering provides a robust statistical tool for analyzing complex data by directly relying on the theoretical foundations of probability distributions.

2. Theoretical Foundations

A K-component mixture distribution is defined by the probability density function $f(x, \mathbf{\Psi}) = \sum_{k=1}^{K} \pi_k f_k(x, \boldsymbol{\psi}_k)$,where π_k are the mixing proportions satisfying $\sum_{k=1}^{K} \pi_k = 1$.

In the present study, we assume that n independent time series of length m follow a mixture polynomial regression model with K components. For each observation j in the i-th series, assuming its membership to the k-th cluster, the model is formulated as:

$$x_{ij} = \sum_{k=1}^{K} \omega_{il}(T'_{kj}\boldsymbol{\beta}_k + \sigma_k \varepsilon_{ij}), \quad \varepsilon_{ij} \sim f_k(\mu = 0, \sigma^2 = 1, \boldsymbol{\theta}_k) \tag{2.1}$$

Where $\boldsymbol{\beta}_k$ is the vector of coefficients, σ_k is the scale parameter, ω_{ik} is the latent cluster assignment variable, and $\mathbf{T}_{kj}$ is the j-th row of the temporal Vandermonde Matrix in the k-th component. Here, $\boldsymbol{\theta}_k$ denotes a vector of shape parameters whose dimension and composition are determined by the specific choice of the component density f_k. While $\boldsymbol{\theta}_k$ is null for location-scale families such as the Normal or Laplace distributions, it captures additional distributional features (e.g., skewness or degrees of freedom) when more flexible densities like the Skew-Laplace ($\boldsymbol{\theta}_k$ = skewness parameter in f_k) or Skew-t ($\boldsymbol{\theta}_k$ = skewness parameter and df_k in f_k) are employed, thereby providing a unified notation for the model's structural parameters.

In contrast to traditional models, real-world data often exhibit pronounced skewness and heavy tails; the use of Gaussian distributions under such conditions typically results in diminished clustering accuracy ([22, 28]). To address this, Farokhi and Shadrokh

([14]) demonstrated that employing multilevel models based on Skew-Normal and Skew-Laplace distributions is an effective strategy for achieving robust estimates in the presence of heterogeneity and skewness in correlated data.

Although robust mixture models based on Student-t, Pearson Type VII, and Laplace distributions have been developed ([36, 37, 34]), their primary focus has remained on independent and identically distributed (i.i.d.) data, leaving a notable gap in the systematic modeling of time series. Furthermore, Farokhi and Shadrokh ([15]) introduced multilevel models based on the Skew-Slash distribution, demonstrating the substantial potential of this family in modeling heavy-tailed data and ensuring robustness against outliers. Among skew-symmetric distributions, the Skew-Normal distribution often exhibits low efficiency when dealing with extreme deviations due to its limited skewness range and the computational challenges associated with MLE estimation ([4, 33]). In contrast, the Skew-Laplace family provides greater flexibility in capturing skewness and heavy tails, serving as a robust alternative in the presence of outliers. Consequently, the objective of this research is to extend model-based clustering to the domain of time series by employing the Skew-Laplace distribution to simultaneously account for temporal dependence and non-normality.

The remainder of this paper is organized as follows: Section 2.1 provides a concise overview of the Laplace and Skew-Laplace distributions. In Section 3, we develop the model-based clustering framework using these distributions for time series and present the parameter estimation procedures through both Frequentist and Bayesian approaches. Section 4 evaluates the performance of the proposed models relative to the traditional Normal model under both skewed and symmetric conditions via a comprehensive simulation study. The efficacy of these models is further assessed on real-world datasets in Section 5, including cluster variance analysis and parameter estimation results. Finally, Section 6 is dedicated to the interpretation of the findings, a detailed discussion, and the concluding remarks.

2.1. **Laplace and Skew-Laplace Distributions.** This section first examines the properties and stochastic representation of the symmetric Laplace distribution. Subsequently, by extending this framework to the asymmetric case, the Skew-Laplace distribution is introduced, and its hierarchical structure is detailed to facilitate the parameter estimation

process. This statistical foundation provides the necessary basis for implementing robust clustering in the subsequent sections.

2.1.1. *Laplace Distribution.* A random variable X follows a Laplace distribution with location parameter μ and scale parameter σ if its probability density function is given by:

$$f(x \mid \mu, \sigma) = \frac{1}{2\sigma} \exp\left\{-\frac{|x-\mu|}{\sigma}\right\}, \quad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. This distribution, typically denoted by $L(\mu, \sigma)$, possesses a mean $E(X) = \mu$ and a variance $Var(X) = 2\sigma^2$. According to Andrews and Mallows ([2]), the stochastic representation of a Laplace random variable can be defined as:

$$\begin{cases} X_i \mid V_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2v_i}\right) \\ V_i \sim I\Gamma\left(1, \frac{1}{2}\right) \end{cases} \tag{2.2}$$

Here, $I\Gamma(\cdot \mid a, b)$ denotes the Inverse-Gamma[1] distribution. Consequently, the marginal distribution of X_i is given by $L\left(\mu, \frac{\sigma}{\sqrt{2}}\right)$. Defining $S_i = \left(\frac{x_i-\mu}{\frac{\sigma}{\sqrt{2}}}\right)^2$, it follows from Equation (2.2) that:

$$\begin{aligned} f(x_i, v_i \mid \mu, \sigma) &= f(x_i \mid v_i, \mu, \sigma) f(v_i) = \frac{1}{2\sqrt{\pi\sigma^2}} v_i^{-3/2} \exp\left\{-\left(v_i S_i + \frac{1}{2v_i}\right)\right\} \\ &= IG(S_i, 1) \times L\left(\mu, \frac{\sigma}{\sqrt{2}}\right) = f(v_i \mid x_i, \mu, \sigma) f(x_i \mid \mu, \sigma) \end{aligned} \tag{2.3}$$

where $IG(S_i, 1)$ represents the Inverse-Gaussian [2] distribution. Consequently, based on the expectations of the IG distribution, we obtain:

$$N_{1i} = \mathbb{E}(v_i \mid x_i) = \frac{1}{\sqrt{S_i}} = \frac{\sigma}{\sqrt{2}|x_i - \mu|}, \qquad N_{2i} = \mathbb{E}(v_i^{-1} \mid x_i) = 1 + \sqrt{2}\left|\frac{x_i - \mu}{\sigma}\right|$$

2.1.2. *Skew-Laplace Distribution.* Arslan ([3]) introduced a simplified version of the Skew-Laplace distribution with the following probability density function:

$$f(x) = \frac{1}{2\sigma a} \exp\left\{\frac{\gamma}{\sigma^2}(x-\mu) - a\left|\frac{x-\mu}{\sigma}\right|\right\},$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\gamma \in \mathbb{R}$ is the skewness parameter, with $a = \sqrt{1 + \gamma^2/\sigma^2}$. This distribution, denoted as $SL(\mu, \sigma, \gamma)$, is more practical than other Skew-Laplace variants due to its computationally efficient

[1] $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp\left(-\frac{\beta}{x}\right)$

[2] $f(x) = \frac{b^{1/2} \exp\{\sqrt{ab}\}}{\sqrt{2\pi}} x^{-3/2} \exp\left\{-\frac{1}{2}\left(ax + \frac{b}{x}\right)\right\}$

density function. The $SL(\mu, \sigma, \gamma)$ distribution reduces to the symmetric $L(\mu, \sigma)$ when $\gamma = 0$, exhibiting right-skewness for $\gamma > 0$ and left-skewness for $\gamma < 0$. For a random variable $X \sim SL(\mu, \sigma, \gamma)$, the mean and variance are given by:

$$E(X) = \mu + 2\gamma, \quad Var(X) = 2(\sigma^2 + 2\gamma^2).$$

The stochastic representation of the Skew-Laplace distribution, as proposed by Arslan ([3]), is expressed as:

$$\begin{cases} X_i \mid v_i \sim \mathcal{N}\left(\mu + v_i^{-1}\gamma, \frac{\sigma^2}{v_i}\right) \\ V_i \sim I\Gamma\left(1, \frac{1}{2}\right) \end{cases} \tag{2.4}$$

where the marginal distribution of X_i is $SL(\mu, \sigma, \gamma)$. Accordingly, the joint density can be written as:

$$f(x_i, v_i) = \frac{v_i^{-3/2}}{2\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} v_i (x_i - \mu)^2 + \frac{\gamma}{\sigma^2}(x_i - \mu) - \frac{v_i^{-1}}{2}\left(\frac{\gamma^2}{\sigma^2} + 1\right)\right\}. \tag{2.5}$$

By assuming $s_i = \left(\frac{x_i - \mu}{\sigma}\right)^2$ and $\alpha = \sqrt{\gamma^2/\sigma^2 + 1}$, the conditional density of V_i is obtained as:

$$f(v_i \mid x_i) = \frac{\alpha}{\sqrt{2\pi}} v_i^{-3/2} \exp\left\{\alpha\sqrt{s_i} - \frac{1}{2}(s_i v_i + \alpha^2/v_i)\right\},$$

which follows an Inverse-Gaussian distribution, $IG(s_i, \alpha^2)$. Consequently, the conditional expectations are:

$$M_{1i} = \mathbb{E}(V_i \mid x_i) = \frac{\sqrt{\gamma^2 + \sigma^2}}{|x_i - \mu|}, \qquad M_{2i} = \mathbb{E}(V_i^{-1} \mid x_i) = \frac{\sigma^2}{\gamma^2 + \sigma^2} + \frac{|x_i - \mu|}{\sqrt{\gamma^2 + \sigma^2}}$$

In light of the stochastic representations in Equations (2.2) and (2.4), X is treated as the observed variable and V as the latent variable, such that the pair (X, V) constitutes the complete data. Accordingly, the complete-data likelihood function is constructed based on the joint densities (2.3) and (2.5), and the maximum likelihood estimates of the parameters are derived using the EM algorithm.

3. Model-Based Clustering for Time Series using Laplace and Skew-Laplace Distributions: Frequentist and Bayesian Approaches

For brevity in modeling, the density of the k-th cluster ($k = 1, \ldots, K$) is denoted by $f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k)$. Finally, the parameter estimation for each of the Laplace and Skew-Laplace models is separately calculated via the EM algorithm.

3.1. **Frequentist Approach for Time Series Clustering with Laplace and Skew Laplace Errors.** Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ consist of n independent time series generated from the following mixture model:

$$f(\mathbf{x}_i \mid \boldsymbol{T}, \boldsymbol{\Psi}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k), \qquad i = 1, \ldots, n \tag{3.1}$$

where $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_K, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K)$ represents the set of all model parameters. Furthermore, $\boldsymbol{\psi}_k = (\boldsymbol{\beta}_k, \sigma_k, \boldsymbol{\theta}_k)$ denotes the parameter vector for the k-th cluster density, $f_k(\cdot)$; such that $\boldsymbol{\beta}_k$ is the vector of regression coefficients, σ_k is the scale parameter, and $\boldsymbol{\theta}_k$ is the vector of other parameters (e.g., the skewness parameter) in the k-th component. If the latent variable Z_{ik} represents the membership status of time series $\mathbf{x}_i$ to the k-th cluster in model(3.1), then the complete-data likelihood function can be calculated as follows:

$$\mathcal{L}_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\pi_k f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k)\right)^{Z_{ik}} \tag{3.2}$$

The primary difference between the proposed model (3.1) and the classical model (assuming independence) lies in the definition of the densities $f_k(\cdot)$. Here, $f_k(\cdot)$ is equivalent to the time series distribution in the k-th cluster. To compute it, assuming $Z_{ik} = 1$, the j-th observation from the i-th time series at time t_j (i.e., x_{ij}) is generated according to the polynomial mixture regression model in the k-th cluster; meaning that:

$$Z_{ik} = \begin{cases} 1 & x_{ij} = \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k + \sigma_k \varepsilon_{ij} \equiv x_{ij} \in f_k \\ 0 & x_{ij} \neq \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k + \sigma_k \varepsilon_{ij} \equiv x_{ij} \notin f_k \end{cases}$$

It is evident that if the time series $\mathbf{x}_i$ belongs to the k-th cluster, then for every $j = 1, 2, \ldots, m$, the observation x_{ij} will also belong to this cluster. Accordingly, $f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k)$ can be calculated as follows:

$$f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k) = \prod_{j=1}^{m} \left(f(x_{ij} \mid \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k, \boldsymbol{I}\sigma_k, \boldsymbol{\theta}_k)\right) \tag{3.3}$$

In calculating the density function (3.3), the property of independent residuals is utilized; i.e., assuming $\varepsilon_{i1}, \ldots, \varepsilon_{im} \overset{\text{iid}}{\sim} f(\mu = 0, \sigma = 1, \boldsymbol{\theta})$ we have:

$$f(\boldsymbol{\varepsilon}_i, \boldsymbol{\theta}) = \prod_{j=1}^{m} \left(f(\varepsilon_{ij} \mid 0, 1, \boldsymbol{\theta})\right) \xrightarrow{\varepsilon_{ij} = \frac{1}{\sigma}(x_{ij} - \boldsymbol{T}'_j\boldsymbol{\beta})} f(\mathbf{x}_i \mid \boldsymbol{T}, \boldsymbol{\psi}) = \prod_{j=1}^{m} \left(f(x_{ij} \mid \boldsymbol{T}'_j\boldsymbol{\beta}, \boldsymbol{I}\sigma, \boldsymbol{\theta})\right)$$

Now, by substituting (3.3) in to the complete-data likelihood function (3.2), the complete data likelihood function for the proposed model is rewritten as follows:

$$\begin{aligned}\mathcal{L}_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T}) &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left(\pi_k f_k(\mathbf{x}_i \mid \boldsymbol{T}_k, \boldsymbol{\psi}_k)\right)^{Z_{ik}} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left(\pi_k \prod_{j=1}^{m}\left(f_{kij}(x_{ij} \mid \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k, \sigma_k^2, \boldsymbol{\theta}_k)\right)\right)^{Z_{ik}}\end{aligned}$$

Consequently, the complete-data log-likelihood function is calculated as:

$$\begin{aligned}\ell_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T}) &= \log\left(\mathcal{L}_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T})\right) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik}\log(\pi_k) + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{K} Z_{ik}\log\left(f_{kij}(x_{ij} \mid \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k, \sigma_k^2, \boldsymbol{\theta}_k)\right)\end{aligned}$$

On the other hand, by substituting the joint distributions (2.3) and (2.5) instead of $f_{kij}(x_{ij} \mid \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k, \sigma_k^2, \boldsymbol{\theta}_k)$ in to the complete-data likelihood and log-likelihood functions, the following relations are obtained:

$$\pi^{L}(\boldsymbol{\Theta}, \boldsymbol{v} \mid \boldsymbol{x}) \propto \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\pi_k \prod_{j=1}^{m}\left\{\frac{1}{2\sqrt{\pi\sigma_k^2}} v_{ij}^{-\frac{3}{2}} \exp\left\{-\frac{v_{ij}}{\sigma_k^2}(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k)^2 - \frac{1}{2v_{ij}}\right\}\right\}\right\}^{Z_{ik}}$$

$$\begin{aligned}\pi^{SL}(\boldsymbol{\Theta}, \boldsymbol{v} \mid \boldsymbol{x}) \propto \prod_{i=1}^{n}\prod_{k=1}^{K}\Bigg\{\pi_k \prod_{j=1}^{m}\Bigg\{&\frac{v_{ij}^{-\frac{3}{2}}}{\sqrt{\sigma_k^2}} \exp\Bigg\{-\frac{v_{ij}}{2\sigma_k^2}(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k)^2 + \frac{\gamma_k}{\sigma_k^2}(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k) \\ &- \frac{v_{ij}^{-1}}{2}\left(\frac{\gamma_k^2}{\sigma_k^2} + 1\right)\Bigg\}\Bigg\}\Bigg\}^{Z_{ik}}\end{aligned}$$

(3.4)

$$\begin{aligned}\ell_c^{L}(\boldsymbol{\Psi} \mid \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{Z}) &= \log\left(\mathcal{L}_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{V})\right) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik}\log(\pi_k) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{K}\left\{-\frac{1}{2}Z_{ik}\log(\sigma_k^2) - \frac{3}{2}Z_{ik}\log(v_{ij}) - \frac{1}{\sigma_k^2}Z_{ik}v_{ij}(x_{ij} - T_{kj}\beta_k)^2\right. \\ &\quad \left. - \frac{1}{2}Z_{ik}v_{ij}^{-1}\right\}\end{aligned}$$

(3.5)

$$\begin{aligned}\ell_c^{SL}(\boldsymbol{\Psi} \mid \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{Z}) &= \log\left(\mathcal{L}_c(\boldsymbol{\Psi} \mid \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{V})\right) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} Z_{ik}\log(\pi_k) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{K}\Bigg\{-\frac{1}{2}Z_{ik}\log(\sigma_k^2) - \frac{3}{2}Z_{ik}\log(v_{ij}) - \frac{1}{2\sigma_k^2}Z_{ik}v_{ij}(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k)^2 \\ &\quad + \frac{\gamma_k}{\sigma_k^2}Z_{ik}(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k) - \frac{1}{2}Z_{ik}v_{ij}^{-1}\left(\frac{\gamma_k^2}{\sigma_k^2} + 1\right)\Bigg\}\end{aligned}$$

The EM algorithm proceeds by iteratively maximizing the expected complete-data log-likelihood. Let $\boldsymbol{\Psi}^{(r)}$ denote the parameter estimates at iteration r. The E-step involves computing the Q-function, defined as the conditional expectation of the complete-data log-likelihood $\ell_c(\boldsymbol{\Psi})$ given the observed data and current parameter estimates: $Q(\boldsymbol{\Psi} \mid \boldsymbol{\Psi}^{(r)}) = \mathbb{E}\left[\ell_c(\boldsymbol{\Psi}) \mid \boldsymbol{X}, \boldsymbol{\Psi}^{(r)}\right]$. For the SL model, this requires evaluating the conditional expectations of the latent membership variables Z_{ik} and the stochastic representation variables v_{ij}, which are used to update the parameters in the subsequent M-step through the maximization of the Q-function.

Note that the complete-data log-likelihood functions in relations (3.4) and (3.5) generally include two latent variables, Z_{ik} and v_{ij}, and their combinations. In the first step of the EM algorithm, the mathematical expectations of the terms in (3.4) and (3.5) that depend on these two variables are computed. In fact, during the expectation step (E-step), the mathematical expectation of the complete-data log-likelihood function must be calculated with respect to the latent variables, conditional on the observed data:

$$\begin{aligned}Q^L(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(r)}) &= \mathrm{E}_{\boldsymbol{\Psi}^{(r)}}\left\{\mathcal{L}_c^L(\boldsymbol{\Psi} \mid \mathbf{x}, \boldsymbol{v}^{(r)}, Z^{(r)}) \mid \mathbf{x}_i\right\} \\ &\propto \sum_{i=1}^{n}\sum_{k=1}^{K}\hat{Z}_{ik}^{(r)}\log(\pi_k) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{K}\Bigg\{-\frac{1}{2}\hat{Z}_{ik}^{(r)}\log(\sigma_k^{2(r)}) - \frac{3}{2}\mathrm{E}\left(Z_{ik}\log(v_{ij}) \mid x_{ij}\right) \\ &\quad - \frac{1}{\sigma_k^{2(r)}}\hat{N}_{1ijk}^{(r)}\left(x_{ij} - \boldsymbol{T}'_{kj}\boldsymbol{\beta}_k^{(r)}\right)^2 - \frac{1}{2}\hat{N}_{2ijk}^{(r)}\Bigg\}\end{aligned}$$

$$
\begin{aligned}
Q^{SL}(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(r)}) &= \mathrm{E}_{\boldsymbol{\Psi}^{(r)}} \left\{ \mathcal{L}_c^{SL}(\boldsymbol{\Psi} \mid \mathbf{x}, \boldsymbol{v}^{(r)}, Z^{(r)}) \mid \mathbf{x}_i \right\} \\
&\propto \sum_{i=1}^{n} \sum_{k=1}^{K} \hat{Z}_{ik}^{(r)} \log(\pi_k) \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{K} \Bigg\{ -\frac{1}{2} \hat{Z}_{ik}^{(r)} \log\left(\sigma_k^{2(r)}\right) - \frac{3}{2} \mathrm{E}\left(Z_{ik} \log(v_{ij}) \mid x_{ij}\right) \\
&\quad - \frac{1}{2\sigma_k^{2(r)}} \widehat{M}_{1ijk}^{(r)} \left(x_{ij} - \boldsymbol{T}'_{kj} \boldsymbol{\beta}_k^{(r)}\right)^2 + \frac{\gamma_k^{(r)}}{\sigma_k^{2(r)}} \hat{Z}_{ik}^{(r)} \left(x_{ij} - \boldsymbol{T}'_{kj} \boldsymbol{\beta}_k^{(r)}\right) \\
&\quad - \frac{1}{2} \widehat{M}_{2ijk}^{(r)} \left(\frac{\gamma_k^{2(r)}}{\sigma_k^{2(r)}} + 1 \right) \Bigg\}
\end{aligned}
$$

where:

$$
\hat{z}_{ik}^{(r)} = \mathrm{E}(Z_{ik} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \frac{\pi_k^{(r)} f_k^L(\mathbf{x}_i \mid \mathbf{t}, \boldsymbol{\psi}_k^{(r)})}{\sum_{h=1}^{K} \pi_h^{(r)} f_h^L(\mathbf{x}_i \mid \mathbf{t}, \boldsymbol{\psi}_h^{(r)})} \quad \text{(for model } L\text{)}
$$

$$
\hat{Z}_{ik}^{(r)} = \mathrm{E}(Z_{ik} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \frac{\pi_k^{(r)} f_k^{SL}(\mathbf{x}_i \mid \mathbf{t}, \boldsymbol{\psi}_k^{(r)})}{\sum_{h=1}^{K} \pi_h^{(r)} f_h^{SL}(\mathbf{x}_i \mid \mathbf{t}, \boldsymbol{\psi}_h^{(r)})} \quad \text{(for model } SL\text{)}
$$

$$
\widehat{N}_{1ijk}^{(r)} = \mathrm{E}(Z_{ik} v_{ij} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \hat{Z}_{ik}^{(r)} \left(\frac{\sigma_k^{(r)}}{\sqrt{2} |x_{ij} - \mathbf{T}'_{kj} \boldsymbol{\beta}_k^{(r)}|} \right)
$$

$$
\widehat{N}_{2ijk}^{(r)} = \mathrm{E}(Z_{ik} v_{ij}^{-1} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \hat{Z}_{ik}^{(r)} \left(1 + \sqrt{2} \left| \frac{x_{ij} - \mathbf{T}'_{kj} \boldsymbol{\beta}_k^{(r)}}{\sigma_k^{(r)}} \right| \right)
$$

$$
\widehat{M}_{1ijk}^{(r)} = \mathrm{E}(Z_{ik} v_{ij} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \hat{Z}_{ik}^{(r)} \left(\frac{\sqrt{\gamma_k^{2(r)} + \sigma_k^{2(r)}}}{|x_{ij} - \mathbf{T}'_{kj} \boldsymbol{\beta}_k^{(r)}|} \right)
$$

$$
\widehat{M}_{2ijk}^{(r)} = \mathrm{E}(Z_{ik} v_{ij}^{-1} \mid \mathbf{t}, \mathbf{x}_i, \boldsymbol{\Psi}^{(r)}) = \hat{Z}_{ik}^{(r)} \left(\frac{\sigma_k^{2(r)}}{\gamma_k^{2(r)} + \sigma_k^{2(r)}} + \frac{|x_{ij} - \mathbf{T}'_{kj} \boldsymbol{\beta}_k^{(r)}|}{\sqrt{\gamma_k^{2(r)} + \sigma_k^{2(r)}}} \right)
$$

Finally, in the second step of the EM algorithm, the mathematical expectation of the complete-data log-likelihood function must be maximized with respect to the model parameters.

Accordingly, the parameter estimates for both models at the $(r+1)$-th iteration of the EM algorithm are obtained by differentiating the Q function and solving the corresponding equations as follows:

a) Laplace Distribution

$$\pi_k^{(r+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{Z}_{ik}^{(r)}$$

$$\boldsymbol{\beta}_k^{(r+1)} = \left(\boldsymbol{T}'_k\left(\sum_{i=1}^{n}\widehat{N}_{1ik}^{(r)}\right)\boldsymbol{T}_k\right)^{-1}\boldsymbol{T}'_k\sum_{i=1}^{n}\widehat{N}_{1ik}^{(r)}\mathbf{x}_i$$

$$\sigma_k^{2(t+1)} = \frac{\sum_{i=1}^{n}\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r+1)}\right)'\widehat{N}_{1ik}^{(r)}\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r+1)}\right)}{m\sum_{i=1}^{n}\hat{Z}_{ik}^{(r)}}$$

b) Skew-Laplace Distribution The updated estimates for the parameters in the $(r+1)$-th iteration are given by:

$$\pi_k^{(r+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{Z}_{ik}^{(r)}$$

$$\boldsymbol{\beta}_k^{(r+1)} = \left(\boldsymbol{T}'_k\left(\sum_{i=1}^{n}\widehat{\boldsymbol{M}}_{1ik}^{(r)}\right)\boldsymbol{T}_k\right)^{-1}\boldsymbol{T}'_k\sum_{i=1}^{n}\left(\widehat{\boldsymbol{M}}_{1ik}^{(r)}\mathbf{x}_i - \gamma_k^{(r)}\widehat{\boldsymbol{Z}}_{ik}^{*(r)}\right)$$

$$\sigma_k^{2(r+1)} = \frac{\sum_{i=1}^{n}\left\{\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r+1)}\right)'\widehat{\boldsymbol{M}}_{1ik}^{(r)}\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r+1)}\right) - 2\gamma_k^{(r)}\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r+1)}\right)'\widehat{\boldsymbol{Z}}_{ik}^{*(r)} + \gamma_k^{2(r)}\boldsymbol{J}'_m\widehat{\boldsymbol{M}}_{2ik}^{(r)}\right\}}{m\sum_{i=1}^{n}\hat{Z}_{ik}^{(r)}}$$

$$\gamma_k^{(t+1)} = \frac{\sum_{i=1}^{n}\left(\mathbf{x}_i - \boldsymbol{T}_k\boldsymbol{\beta}_k^{(r)}\right)'\widehat{\boldsymbol{Z}}_{ik}^{*(r)}}{\sum_{i=1}^{n}\boldsymbol{J}'_m\widehat{\boldsymbol{M}}_{2ik}^{(r)}}$$

Where the following matrix definitions hold:
$\widehat{\boldsymbol{N}}_{1ik}^{(r)} = \boldsymbol{diag}\left(\widehat{N}_{1i1k}^{(r)},\ldots,\widehat{N}_{1imk}^{(r)}\right)'$ The diagonal matrix derived from the mathematical expectations of the Laplace latent variable.
$\widehat{\boldsymbol{M}}_{1ik}^{(r)} = diag\left(\widehat{M}_{1i1k}^{(r)},\ldots,\widehat{M}_{1imk}^{(r)}\right)'$: The diagonal matrix related to the Skew-Laplace model.

$\widehat{\boldsymbol{M}}_{2ikl}^{(r)} = \left(\widehat{M}_{2i1kl}^{(r)}, \ldots, \widehat{M}_{2imkl}^{(r)}\right)'$: The vector of mathematical expectations of the inverse latent variable.
$\widehat{\boldsymbol{Z}}_{ik}^{*(r)} = \left(\hat{Z}_{ik}^{(r)}, \hat{Z}_{ik}^{(r)}, \ldots, \hat{Z}_{ik}^{(r)}\right)' = rep(\hat{Z}_{ik}^{(r)}, m)$: A vector of length m consisting of the clustering weights.
$\boldsymbol{J}_m = (1, \ldots, 1)'$: A vector of length m consisting of 1's.

It is worth noting that the EM algorithm is guaranteed to converge to a stationary point, typically a local maximum, as the likelihood function for the Skew-Laplace mixture model is bounded from above [11]. Furthermore, under standard regularity conditions, the resulting MLEs are consistent and asymptotically normal. Regarding the computational efficiency, the time complexity of each EM iteration is primarily dominated by the matrix inversions, such as $(T_k'[\sum_{i=1}^{n} \hat{z}_{ik}]T_k)^{-1}$. Since the polynomial degree p is typically small relative to the number of series n and time length m, the computational burden remains manageable, with an overall complexity of approximately $O(T \cdot n \cdot k \cdot p^3)$ per iteration. Furthermore, to address the memory requirements for storing expectation matrices, the algorithm is implemented using on-the-fly computations to minimize the storage footprint.

3.2. **Bayesian Approach for Time Series Clustering with Laplace and Skew-Laplace Errors.** Choosing appropriate prior distributions and determining the values of hyperparameters is considered the most critical stage in the Bayesian approach, as these choices directly influence the structure of the posterior distribution and the subsequent inferences. In this study, the chosen prior distributions (Normal, Inverse-Gamma, Gamma, and Dirichlet) are all proper, which ensures that the resulting posterior distribution is valid. Moreover, according to the Bernstein-von Mises theorem, these Bayes estimators are asymptotically equivalent to the MLEs as the sample size $n \to \infty$, inheriting their consistency and asymptotic normality. Marin et al. ([23]), in their study on the challenges of the Bayesian approach in mixture models, demonstrated that due to the complexity and lack of a closed-form expression for the posterior distribution, sampling parameters using Markov Chain Monte Carlo (MCMC) methods is often difficult and time-consuming. Furthermore, Roeder and Wasserman ([32]) and Diebolt and Robert ([12, 13]) argued that in mixture models, it is not possible to consider all prior distributions as non-informative; this is because the posterior distribution will not necessarily be proper if no observations are assigned to certain components.

In light of the aforementioned considerations, this study assumes independent prior distributions for the parameters to perform Bayesian inference. Additionally, to facilitate the implementation of Monte Carlo algorithms, conjugate priors are utilized so that the full conditional posterior distributions of the parameters possess a known, closed-form structure. Accordingly, following Gelman et al. ([18]) and Richardson and Green ([31]), Normal, Inverse-Gamma, and Dirichlet prior distributions are considered for the regression coefficients, the scale parameter, and the mixture weights, respectively. Furthermore, a hierarchical Bayesian approach is employed for the hyperparameter of the Inverse-Gamma distribution. This structure, while enhancing the model's flexibility in estimating the scale parameter, reduces the sensitivity of the results to the selection of fixed values for hyperparameters. This approach allows the model to provide more accurate estimates of dispersion in small clusters by sharing information across clusters. Finally, the prior distribution for the skewness parameter is also assumed to be Normal, which, in addition to the advantage of conjugacy in the Skew-Laplace distribution structure, is consistent with classical concepts of skewness.

$$\boldsymbol{\beta}_k \sim N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad \sigma_k^2 | b \sim I\Gamma(\alpha, b), \quad b \sim \Gamma(\lambda_1, \lambda_2), \quad \gamma_k \sim N(\zeta_k, \tau_k^2), \quad \boldsymbol{\pi} \sim \text{Dirichlet}(h, h, \ldots, h).$$

where $\boldsymbol{\Sigma}_k$ is the prior covariance matrix reflecting initial uncertainty, typically set as $\sigma^2_{prior}\mathbf{I}$ for a non-informative setup.

The complete-data posteriors for both models are given in (3.6) and (3.7):

$$\begin{aligned}
&\pi(\boldsymbol{\Theta}, \mathbf{v} \mid \mathbf{x}) \\
&\qquad \propto \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\pi_k \prod_{j=1}^{m}\left\{\frac{1}{2\sqrt{\pi\sigma_k^2}} v_{ij}^{-\frac{3}{2}} \exp\left\{-\frac{v_{ij}}{\sigma_k^2}(x_{ij} - \mathbf{T}'_{kj}\boldsymbol{\beta}_k)^2 - \frac{1}{2v_{ij}}\right\}\right\}\right\}^{z_{ik}} \\
&\qquad \times \prod_{k=1}^{K}\left\{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k)\right\} \times b^{\alpha}(\sigma_k^2)^{-(\alpha+1)} \exp\left[-\frac{b}{\sigma_k^2}\right] \pi_k^{h-1}\right\} \\
&\qquad \times b^{\lambda_1 - 1} \exp\{-\lambda_2 b\} \qquad (3.6)
\end{aligned}$$

$$
\begin{aligned}
&\pi(\boldsymbol{\Theta}, \mathbf{v} \mid \boldsymbol{x}) \\
&\quad \propto \prod_{i=1}^{n}\prod_{k=1}^{K}\left\{\pi_k \prod_{j=1}^{m}\left\{\frac{v_{ij}^{-\frac{3}{2}}}{\sqrt{\sigma_k^2}}\exp\left\{-\frac{v_{ij}}{2\sigma_k^2}(x_{ij}-\mathbf{T}'_{kj}\boldsymbol{\beta}_k)^2+\frac{\gamma_k}{\sigma_k^2}(x_{ij}-\mathbf{T}'_{kj}\boldsymbol{\beta}_k)\right.\right.\right. \\
&\quad \left.\left.\left.-\frac{v_{ij}^{-1}}{2}\left(\frac{\gamma_k^2}{\sigma_k^2}+1\right)\right\}\right\}\right\}^{z_{ik}} \\
&\quad \times \prod_{k=1}^{K}\left\{\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_k-\boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\beta}_k-\boldsymbol{\mu}_k)\right\}\times b^{\varrho}(\sigma_j^2)^{-(\varrho+1)}\exp\left\{-\frac{b}{\sigma_k^2}\right\}\right. \\
&\quad \left.\times \exp\left\{-\frac{1}{2\tau_k^2}(\gamma_k-\zeta_k)^2\right\}\pi_k^{h-1}\right\} b^{\lambda_1-1}\exp\{-\lambda_2 b\} \qquad (3.7)
\end{aligned}
$$

As observed, the joint posterior distributions of both models in (3.6) and (3.7) do not have a known form; therefore, to compute posterior quantities, MCMC algorithms such as the Gibbs sampler must be employed, which requires knowledge of the full conditional posterior distributions of the parameters and latent variables. These distributions, after algebraic manipulations, are derived separately for the Laplace and Skew-Laplace models as follows:

3.2.1. *Full Conditional Posterior Distributions of Parameters and Variables in the Laplace Mixture Model.* The full conditional posterior distributions for the Laplace mixture model are derived as follows:

- **Posterior distribution of the indicator variables ($\boldsymbol{Z}_i$):**

$$
\begin{aligned}
\pi(\boldsymbol{Z}_i \mid \text{others}) &= \prod_{k=1}^{K}\left\{\pi_k\prod_{j=1}^{m}\left\{\frac{1}{2\sqrt{\pi\sigma_k^2}}v_{ij}^{-\frac{3}{2}}\exp\left[-\frac{v_{ij}}{\sigma_k^2}(x_{ij}-\boldsymbol{T}'_{kj}\boldsymbol{\beta}_k)^2-\frac{1}{2v_{ij}}\right]\right\}\right\}^{Z_{ik}} \\
&= \prod_{k=1}^{K}\left\{\pi_k\prod_{j=1}^{m}\{f(x_{ij},v_{ij}\mid\boldsymbol{\beta}_k,\sigma_k^2)\}\right\}^{Z_{ik}} \equiv \prod_{k=1}^{K}\left\{\pi_k\prod_{j=1}^{m}\left\{L\left(x_{ij}\mid\boldsymbol{\beta}_k,\frac{\sigma_k^2}{\sqrt{2}}\right)\right\}\right\}^{Z_{ik}}
\end{aligned}
$$

- **Posterior distribution of the auxiliary variables (v_{ij}):**

$$
\begin{aligned}
\pi(v_{ij}\mid\text{others}) &= \left\{v_{ij}^{-\frac{3}{2}}\exp\left\{-v_{ij}\left(\frac{x_{ij}-\mathbf{T}'_{kj}\boldsymbol{\beta}_k}{\sigma_j}\right)^2-\frac{1}{2v_{ij}}\right\}\right\}^{z_{ik}} \\
&= v_{ij}^{-\frac{3}{2}}\exp\left\{-\frac{1}{2}\left(v_{ij}S_{ijk}+\frac{1}{v_{ij}}\right)\right\}\sim \text{IG}(S_{ijk},1)
\end{aligned}
$$

- **Posterior distribution of the mixture weights ($\boldsymbol{\pi}$):**

$$\pi(\boldsymbol{\pi} \mid \text{others}) \propto \prod_{k=1}^{K} \pi_k^{(h+\sum_{i=1}^{n} Z_{ik})-1} \sim \text{Dirichlet}(n_1 + h, n_2 + h, \ldots, n_K + h)$$

- **Posterior distribution of the hyperparameter (b):**

$$\pi(b \mid \text{others}) \propto b^{(K\alpha+\lambda_1)-1} \exp\left\{ -\left(\lambda_2 + \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right) b \right\} \sim \Gamma\left(K\alpha + \lambda_1, \lambda_2 + \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)$$

- **Posterior distribution of the regression coefficients ($\boldsymbol{\beta}_k$):**

$$\pi(\boldsymbol{\beta}_k \mid \text{others}) = N_{p_k}\left(\sum_{i=1}^{n} \boldsymbol{F}_{1ik}^{-1} \boldsymbol{F}_{2ik}, \sum_{i=1}^{n} \boldsymbol{F}_{1ik}^{-1} \right)$$

- **Posterior distribution of the scale parameter (σ_k^2):**

$$\pi(\sigma_k^2 \mid \text{others}) \sim \mathcal{I}\Gamma\left(\alpha + \frac{mn_k}{2}, b + \sum_{i=1}^{n} \mathrm{B}_{ik} \right)$$

3.2.2. *Full Conditional Posterior Distributions of Parameters and Variables in the Skew-Laplace Mixture Model.* For the Skew-Laplace mixture model, the conditional distributions are modified as follows:

- **Posterior distribution of the indicator variables ($\boldsymbol{Z}_i$):**

$$\pi(\boldsymbol{Z}_i \mid \text{others}) = \prod_{k=1}^{K} \left\{ \pi_k SL_m(\mathbf{x}_i \mid \boldsymbol{\beta}_k, \boldsymbol{I}\sigma_k^2, \gamma_k) \right\}^{Z_{ik}} = \mathcal{M}\text{ultynomial}(1, \pi_1^*, \pi_2^*, \ldots, \pi_g^*)$$

- **Posterior distribution of the auxiliary variables (v_{ij}):** In this model, the conditional distribution of v_{ij} is updated as:

$$\pi(v_{ij} \mid \text{others}, Z_{ik} = 1) = v_{ij}^{-\frac{3}{2}} \exp\left\{ -\frac{1}{2}\left(v_{ij} S_{ijk} + \frac{\alpha_k}{v_{ij}} \right) \right\} = \text{IG}(S_{ijk}, \alpha_k),$$

- **Posterior distribution of the mixture weights ($\boldsymbol{\pi}$):**

$$\pi(\boldsymbol{\pi} \mid \text{others}) \propto \prod_{k=1}^{K} \left\{ \pi_k^{(h+\sum_{i=1}^{n} Z_{ik})-1} \right\} = \text{Dirichlet}(n_1 + h, n_2 + h, \ldots, n_k + h)$$

- **Posterior distribution of the hyperparameter (b):**

$$\pi(b \mid \text{others}) \sim \Gamma\left(K\varrho + \lambda_1, \lambda_2 + \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)$$

- **Posterior distribution of the regression coefficients ($\boldsymbol{\beta}_k$):**

$$\pi(\boldsymbol{\beta}_k \mid \text{others}) = N_{p_k}\left(\sum_{i=1}^{n} \boldsymbol{F}_{1ik}^{-1}\boldsymbol{F}_{2ik}, \sum_{i=1}^{n} \boldsymbol{F}_{1ik}^{-1}\right)$$

- **Posterior distribution of the skewness parameter (γ_k):** The posterior distribution for the skewness parameter of each cluster is derived as Normal:

$$\pi(\gamma_k \mid \text{others}) = N(\mathbb{M}_k, \mathbb{S}_k)$$

- **Posterior distribution of the scale parameter (σ_k^2):**

$$\pi(\sigma_k^2 \mid \text{others}) \sim \mathcal{I}\Gamma\left(\varrho + \frac{mn_k}{2}, b + \sum_{i=1}^{n} \mathrm{B}_{ik}\right)$$

The parameters for each of the full conditional posterior distributions for both models are separately defined in Table 1.

To mitigate challenge of non-informative priors in finite mixture models, we adopt an Empirical Bayes (EB) approach for specifying the priors $\boldsymbol{\beta}_k$ and γ_k. In complete Bayes, prior information must be independent of data information, and data cannot be used to estimate the hyperparameters of prior distributions. Therefore, using EB means violating the principle of complete Bayes. However, EB allows us to use sample information to estimate hyperparameters in cases where non-informative priors cannot be used or when we are dealing with complex models. The advantages of EB include algorithmic stability, faster convergence, and prevention of Improper posteriors in complex and high-dimensional models. Specifically, we perform an initial clustering using dynamic time warping (DTW) and k-means clustering, dividing the n time series into K clusters. Now, assuming that the observations in each cluster are normal, the hyperparameters can be estimated by the maximum likelihood estimation (MLE) method. For example, the prior distribution of the regression coefficients $\boldsymbol{\beta}_k$ can be estimated as follows:

$$\boldsymbol{\mu}_k = \widehat{\boldsymbol{\beta}}_k = (\mathbf{T}_k'\mathbf{T}_k)^{-1}\mathbf{T}_k'\mathbf{x}_k^*, \quad \boldsymbol{\Sigma}_k = cov(\widehat{\boldsymbol{\beta}}_k) = (\mathbf{T}_k'\mathbf{T}_k)^{-1}\sigma_k^2$$

TABLE 1. Comparison between Model of Laplace and Model of Skew Laplace

Model of Laplace	**Model of Skew Laplace**
$\pi_j^* = \frac{\pi_k L_m(\mathbf{x}_i\|\boldsymbol{\beta}_k,\mathbf{I}\sigma_k^2/\sqrt{2})}{\sum_{h=1}^K \pi_h L_m(\mathbf{x}_i\|\boldsymbol{\beta}_h,\mathbf{I}\sigma_h^2/\sqrt{2})}$	$\pi_j^* = \frac{\pi_k SL_m(\mathbf{x}_i\|\boldsymbol{\beta}_k,\mathbf{I}\sigma_k^2,\gamma_k)}{\sum_{h=1}^K\{\pi_h SL_m(\mathbf{x}_i\|\boldsymbol{\beta}_h,\mathbf{I}\sigma_h^2,\gamma_h)\}}$
$S_{ijk} = \left(\frac{x_{ij}-\mathbf{T}'_{kj}\boldsymbol{\beta}_k}{\sigma_k/\sqrt{2}}\right)^2$	$S_{ijk} = \left(\frac{x_{ij}-\mathbf{T}'_{kj}\boldsymbol{\beta}_k}{\sigma_k^2}\right)^2$
	$\alpha_k = \sqrt{\gamma_k^2/\sigma_k^2+1}$
$\mathbf{F}_{1ik} = 2\mathbf{T}'_k\frac{Z_{ik}\mathbf{V}_i}{\sigma_k^2}\mathbf{T}_k + \frac{\boldsymbol{\Sigma}_k^{-1}}{n}$	$\mathbf{F}_{1ik} = \mathbf{T}'_k\frac{Z_{ik}\mathbf{V}_{ik}}{\sigma_k^2}\mathbf{T}_k + \frac{\boldsymbol{\Sigma}_k^{-1}}{n}$
$\mathbf{F}_{2ik} = 2\mathbf{x}'_i\frac{Z_{ik}\mathbf{V}_i}{\sigma_k^2}\mathbf{T}_k + \frac{\boldsymbol{\mu}'_k\boldsymbol{\Sigma}_k^{-1}}{n}$	$\mathbf{F}_{2ik} = \mathbf{x}'_i\frac{Z_{ik}\mathbf{V}_{ik}}{\sigma_k^2}\mathbf{T}_k - \frac{\gamma_k}{\sigma_k^2}\mathbf{J}'_m Z_{ik}\mathbf{T}_k + \frac{\boldsymbol{\mu}'_k\boldsymbol{\Sigma}_k^{-1}}{n}$
$A_{ik} = \frac{m}{2}Z_{ik}$	$A_{ik} = \frac{m}{2}Z_{ik}$
$B_{ik} = (\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k)'Z_{ik}\mathbf{V}_i(\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k)$	$B_{ik} = (\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k)'\frac{Z_{ik}\mathbf{V}_i}{2}(\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k) + \gamma_k^2\mathbf{J}'_m Z_{ik}\mathbf{V}_i^{-1} - \gamma_k Z_{ik}\mathbf{I}'_m(\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k)$
$n_k = \sum_{i=1}^n Z_{ik}$	$n_k = \sum_{i=1}^n Z_{ik}$
$h_i = w_i Z_{ij}(y_i - \mu_j)(y_i - \mu_j)'$	$h_i = w_i Z_{ij}(y_i - \mu_j)(y_i - \mu_j)'$
	$\mathbb{M}_k = \frac{\left[\mathbf{J}'_m\sum_{i=1}^n Z_{ik}(\mathbf{x}_i - \mathbf{T}_k\boldsymbol{\beta}_k) + \zeta_k/\tau_k\right]}{\left[\mathbf{J}'_m\sum_{i=1}^n Z_{ik}\mathbf{V}_i^{-1} + 1/\tau_k\right]}$
	$\mathbb{S}_k = \left[\frac{\mathbf{J}'_m}{\sigma_k^2}\sum_{i=1}^n Z_{ik}\mathbf{V}_{ik}^{-1} + 1/\tau_k^2\right]^{-1}$

Or, the prior hyperparameters of skewness in cluster k can be considered as follows:

$$\zeta_k = \gamma_k^{(0)} = G_1 = \frac{m^2}{(m-1)(m-2)} \times \frac{\frac{1}{m}\sum_{j=1}^m (x_{kj}^* - \bar{x}_k^*)^2}{\left[\frac{1}{m-1}\sum_{j=1}^m (x_{kj}^* - \bar{x}_k^*)^2\right]^{\frac{3}{2}}}$$

$$\tau_k^2 = VAR(G_1) = \frac{6m(m-1)}{(m-2)(m+1)(m+3)}$$

Where, $\mathbf{x}_k^*$ is the time series mean of the k-th cluster, x_{kj}^* is its value at time j, and $\bar{x}_k^*$ is the mean of the entire cluster.

Following Richardson and Green ([31]) and Bernardi et al ([7]), other model hyperparameters were also considered as follows:

$$\alpha = 2, \quad \lambda_1 = 0.2, \quad \lambda_2 = \frac{100\lambda_1}{\alpha R^2}, \quad h > 1, \quad R = Max([\mathbf{X}]_{m\times n}) - Min([\mathbf{X}]_{m\times n})$$

In a similar way, the initial values of the parameters in the EM algorithm can be determined, or the initial values of the parameters in the MCMC algorithm can be selected from the prior distribution of the parameters.

3.3. **Determining the Number of Clusters and Regression Order.** In the frame work of mixture models, model selection criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are commonly employed. Unlike classical mixture regression models, two parameters must be tuned here: the number of clusters K and the degree of the polynomials P. These two criteria can be defined as follows:

$$AIC = 2v(K,p) - 2\ell(\widehat{\mathbf{\Psi}})$$
$$BIC = v(K,p)\log(nm) - 2\ell(\widehat{\mathbf{\Psi}})$$

where $\widehat{\mathbf{\Psi}}$ is the estimated parameter vector and $v(K,p)$ represents the number of free parameters in the model. In the proposed model, the number of free parameters is equal to the sum of the mixture weights, polynomial regression coefficients, variances, and the skewness parameter:

$$\nu(K,p) = (K-1) + K(p+1) + 2K$$

It should be noted that the effective number of parameters in mixture models is inherently more complex than a simple summation due to the latent variable allocation process; therefore, the provided $\nu(K,p)$ is treated as an upper bound on the model's complexity.

In practice, the maximum values K_{max} and p_{max} are first determined. Subsequently, the proposed models are fitted for all values of $p \in \{1,\ldots,p_{max}\}$ and $K \in \{1,\ldots,K_{max}\}$. Finally, the AIC and BIC scores are calculated, and the pair (K,p) that yields the minimum value is selected as the optimal solution (or the preferred model).

While criteria such as WAIC or DIC are often preferred for Bayesian models, we employ BIC here as a consistent and computationally efficient benchmark for comparing models across both the Frequentist and Bayesian frameworks, ensuring a unified approach to selecting the number of clusters K and polynomial degree p.

4. Simulation

In this section, a comprehensive simulation study is designed to evaluate the performance and capabilities of the proposed models based on "Laplace" and "Skew-Laplace" distributions within a mixture modeling framework. The objective of this study is to compare the accuracy of the estimates obtained from these models against the classical "Normal" model.

In this simulation scenario, a dataset consisting of 60 time series ($n = 60$), each with a length of 60 ($m = 60$), is generated. It is assumed that these observations are distributed across two distinct clusters ($K = 2$) with equal weights ($\pi_k = 1/2$). The latent structure of the data in each cluster is defined based on polynomial regression models of different degrees; specifically, time series belonging to the first cluster follow a degree-5 polynomial regression ($p_1 = 5$), while those in the second cluster follow a degree-3 polynomial regression ($p_2 = 3$). The true values of the model parameters, including the regression coefficients (β_{kl}), scale parameters (σ_k^2), and skewness parameters (γ_k) for each cluster, are reported in Table 2.

Figure 1 displays the time series plots of a simulated sample from the Skew-Laplace distribution, based on the parameters reported in Table 2. The red dashed lines in this figure illustrate the mean of the time series within each cluster. In this study, the process of sampling and model fitting was repeated 1,000 times. Finally, the Root Mean Square Error (RMSE) of the estimators, the Mean Clustering Accuracy, the Mean Misclassification Rate, and the Mean Adjusted Rand Index were calculated.

To evaluate the performance and compare the Skew-Laplace (SL), Laplace (L), and Normal (N) models, 5 different simulation scenarios are considered based on the following structure:

$$x_{ij} = \mathbf{T}_j'\boldsymbol{\beta}_k + \sigma_k\varepsilon_{ij}$$

Scenario 1: $\varepsilon_{ij} \sim N(0, 1)$

Scenario 2: $\varepsilon_{ij} \sim L(0, 1)$

TABLE 2. True parameter values used for the simulation study.

Clusters	π_k	p_k	β_{kl}	σ_k^2	γ_k
Cluster 1	1/2	5	$\beta_{10} = 15.00000$	2	-3
			$\beta_{11} = 1.079640$		
			$\beta_{12} = -0.01572353$		
			$\beta_{13} = 0.00138712$		
			$\beta_{14} = -5.500484e-05$		
			$\beta_{15} = 5.64805e-07$		
Cluster 2	1/2	3	$\beta_{20} = 64.08820$	3	3
			$\beta_{21} = -2.493231$		
			$\beta_{22} = 0.09689376$		
			$\beta_{23} = -0.001053774$		

Scenario 3: $\varepsilon_{ij} \sim SL(0, 1, \gamma_k)$

Scenario 4: $\varepsilon_{ij} \sim 0.05 * Unif(-6, -3) + 0.9 * N(0, 1) + 0.05 * Unif(3, 6)$

Scenario 5: $\varepsilon_{ij} \sim 0.9 * N(0, 1) + 0.1 * Unif(3, 6)$

Scenario 6: $\varepsilon_{ij} \sim 0.1 * Unif(-6, -3) + 0.9 * N(0, 1)$ To assess the convergence of the Markov Chain in posterior sampling, trace and autocorrelation plots were employed. As illustrated in Figure 1 for Scenario 3, these plots effectively demonstrate the convergence of the Markov Chain generated from the posterior distribution. For the simulation, a total of 5,000 iterations were conducted, with a burn-in period of 1,000 iterations.

After discarding the first 1,000 samples as the burn-in period, convergence was assessed for all model parameters using the split-$\hat{R}$ diagnostic, as summarized in Table 3. Given that the values for all parameters are below the threshold of 1.01, it can be concluded that the Markov chains have successfully converged to their target posterior distribution.

Furthermore, regarding the convergence of the EM algorithm, various criteria such as the parameter difference $|\mathbf{\Psi}^{(r+1)} - \mathbf{\Psi}^{(r)}| \leq \varepsilon$ or the log-likelihood change $|\ell(\mathbf{\Psi}^{(r+1)}) -$

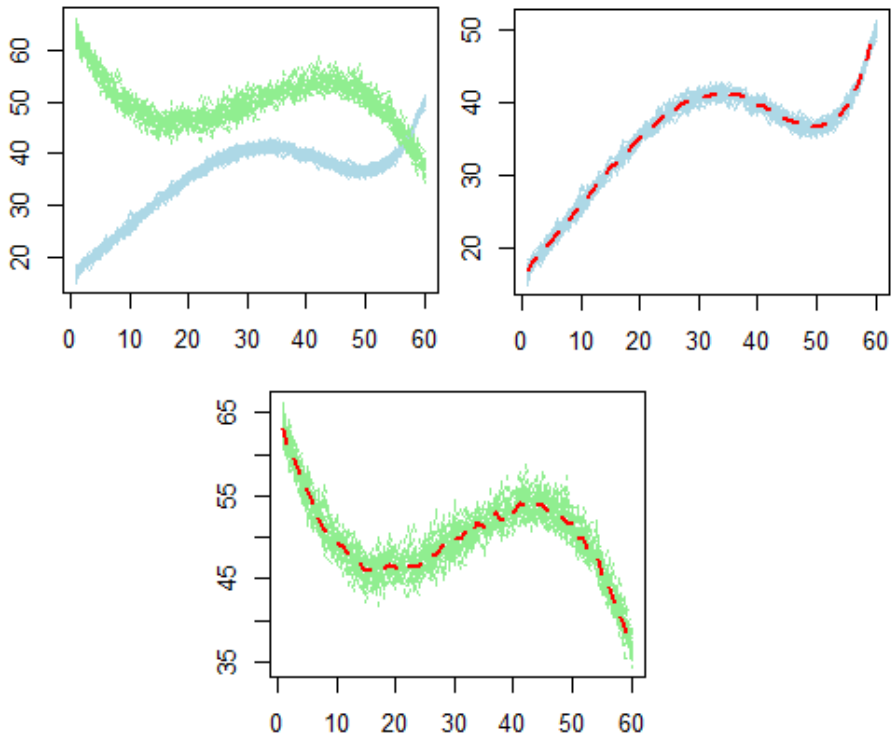


FIGURE 1. Time series plot of a simulated sample separated by different clusters, generated from a Skew-Laplace distribution.

TABLE 3. Value of $split - \widehat{R}$ of parameters

parameter	$split - \widehat{R}$	*parameter*	$split - \widehat{R}$
β_{10}	1.0000	β_{22}	0.9999
β_{11}	0.9999	β_{23}	0.9999
β_{12}	1.0000	σ_1^2	0.9999
β_{13}	0.9998	σ_2^2	0.9997
β_{14}	0.9998	γ_1	0.9999
β_{15}	0.9997	γ_2	1.0000
β_{20}	0.9999	π_1	0.9999
β_{21}	0.9999		

$\ell(\mathbf{\Psi}^{(r)})| \leq \varepsilon$ can be utilized ([9]). In the present study, the former criterion was adopted with a convergence threshold of $\epsilon = 10^{-5}$. It should be noted that the EM algorithm

was also run with different initial values to avoid getting stuck in local optima; for this purpose, in addition to the normal ml estimates obtained through an initial clustering (DTW+kmeans), the initial values of the Markov chains of the parameters were also used.

Based on the results of the first scenario in Table 4, where the distributional structure of the data is normal in both behavioral clusters, the normal model—as the true data-generating model—yields the lowest RMSE and thus the best estimation performance. This finding is theoretically expected, since when the model's assumed distribution fully conforms to the data-generating mechanism, the statistical efficiency of the estimates is maximized. Nevertheless, the results indicate that the other two models, and particularly the skew-Laplace model, do not exhibit a severe deterioration in performance in this scenario and continue to maintain acceptable accuracy. The relative superiority of the skew-Laplace model over the Laplace model in this setting can be attributed to its greater flexibility in capturing slight deviations from perfect symmetry, as well as its capacity to represent relatively heavier tails. Consequently, even under conditions in which the data appear to be normally distributed, the skew-Laplace model can provide a satisfactory fit, albeit with a minor difference from the true (normal) model.

The right-hand side of Table 4 demonstrates that when the data follow a symmetric but heavy-tailed distribution (Scenario 2), the normal model exhibits weak performance. The primary reason for this is that the presence of outlying observations, or the greater probability mass in the tail regions, compels the normal model to inflate the estimated variance in order to accommodate these data points. As a result, the central portion of the distribution is not adequately represented, and the overall model fit deteriorates. Since the data in this scenario are generated from a Laplace distribution, it is natural for the Laplace model to perform best and record the lowest estimation error. Nonetheless, the close performance of the skew-Laplace model to that of the Laplace model is a noteworthy point, because the Laplace distribution can be regarded as a special case of the skew-Laplace family. Hence, even under symmetry, the skew-Laplace model is able to reproduce the heavy-tailed behavior of the data without a significant loss of accuracy. This result confirms that the skew-Laplace distribution is not exclusively suitable for skewed data; rather, it also exhibits high efficiency in symmetric settings characterized by tail kurtosis.

TABLE 4. RMSE values of parameters of all three models (Bayesian and frequentist approaches)

Cluster	**PAR**	**Scenario 1 (Normal distribution)**						**Scenario 2 (Laplace distribution)**					
		$m = 50$						$m = 50$					
		Frequentist			Bayesian			Frequentist			Bayesian		
		N	L	SL	N	L	SL	N	L	SL	N	L	SL
1	β_{01}	0.1806	0.5438	0.3951	0.1132	0.4860	0.2710	2.1781	0.1851	0.1925	1.1171	0.1244	0.1853
	β_{11}	0.2215	0.5251	0.4165	0.1208	0.4920	0.2818	2.3209	0.1724	0.2045	1.1455	0.0855	0.1713
	β_{21}	0.2289	0.5620	0.4447	0.1303	0.4648	0.2914	2.0829	0.2040	0.2026	1.7534	0.1045	0.1712
	β_{31}	0.1806	0.5432	0.3609	0.1048	0.5096	0.2528	2.2293	0.1833	0.2004	1.4738	0.0922	0.1759
	β_{41}	0.2013	0.5692	0.4333	0.1675	0.4565	0.2612	1.8713	0.2230	0.1968	1.6768	0.1407	0.1827
	β_{51}	0.1934	0.5125	0.3800	0.1314	0.5042	0.2994	2.2937	0.2201	0.2075	1.6623	0.1007	0.1718
	σ_1^2	0.1801	0.5147	0.3846	0.1373	0.5004	0.3212	2.0631	0.2004	0.1938	1.1135	0.1453	0.1753
	γ_1	–	–	0.3870	–	–	0.2833	–	–	0.1984	–	–	0.1866
	π_1	0.2093	0.5588	0.3546	0.1478	0.4780	0.2459	1.8469	0.2054	0.2023	1.5082	0.0879	0.1816
2	β_{02}	0.2256	0.5617	0.3503	0.1492	0.4662	0.2773	2.3215	0.2032	0.2129	1.3108	0.1677	0.1875
	β_{12}	0.1986	0.5220	0.3714	0.1547	0.5052	0.2880	2.1586	0.1716	0.1963	1.7850	0.1578	0.1813
	β_{22}	0.2016	0.5482	0.3523	0.1223	0.5044	0.2697	2.3511	0.2124	0.2021	1.5162	0.1066	0.1743
	β_{32}	0.1731	0.5279	0.4373	0.1074	0.4616	0.3155	1.9414	0.1709	0.2139	1.3169	0.1050	0.1715
	σ_2^2	0.2152	0.5794	0.4334	0.1094	0.5035	0.2724	2.4166	0.2102	0.2029	1.5594	0.1246	0.1810
	γ_2	–	–	0.3751	–	–	0.2309	–	–	0.2134	–	–	0.1862
$m = 100$													
1	β_{01}	0.0745	0.4565	0.2309	0.0004	0.4541	0.2302	1.1098	0.0144	0.2139	0.3398	0.0135	0.1771
	β_{11}	0.0678	0.4562	0.2309	0.0004	0.4550	0.2305	1.1043	0.0186	0.1766	0.6767	0.0139	0.1964
	β_{21}	0.0805	0.4561	0.2308	0.0016	0.4520	0.2296	1.1052	0.0169	0.1847	0.3769	0.0069	0.1982
	β_{31}	0.0343	0.4562	0.2308	0.0000	0.4515	0.2300	1.1097	0.0231	0.2028	0.6632	0.0000	0.1867
	β_{41}	0.0512	0.4560	0.2308	0.0017	0.4527	0.2298	1.0935	0.0232	0.1739	1.1130	0.0136	0.1718
	β_{51}	0.0048	0.4565	0.2309	0.0005	0.4481	0.2299	1.1080	0.0183	0.1882	0.9028	0.0135	0.1812
	σ_1^2	0.0619	0.4565	0.2308	0.0008	0.4533	0.2291	1.1125	0.0132	0.2137	0.7751	0.0117	0.1726
	γ_1	–	–	0.2308	–	–	0.2294	–	–	0.1823	–	–	0.1907
	π_1	0.0415	0.4560	0.2308	0.0005	0.4528	0.2291	1.1906	0.0116	0.2088	1.0910	0.0076	0.2105
2	β_{02}	0.0733	0.4564	0.2309	0.0003	0.4509	0.2299	1.2300	0.0172	0.1775	0.9602	0.0133	0.2107
	β_{12}	0.0493	0.4564	0.2308	0.0015	0.4538	0.2300	1.2563	0.0151	0.1885	0.5812	0.0139	0.2063
	β_{22}	0.0727	0.4564	0.2296	0.0017	0.4524	0.2296	1.2260	0.0125	0.1867	0.4764	0.0090	0.1940
	β_{32}	0.0131	0.4562	0.2309	0.0016	0.4465	0.2300	1.2307	0.0138	0.2119	0.3111	0.0130	0.1779
	σ_2^2	0.0663	0.4563	0.2308	0.0017	0.4504	0.2299	1.3804	0.0140	0.2153	0.9122	0.0045	0.2129
	γ_2	–	–	0.2308	–	–	0.2308	–	–	0.2007	–	–	0.2155

The left-hand side of Table 5 shows that when the data simultaneously exhibit skewness and tail thickness (Scenario 3), the performance of the normal and Laplace models declines considerably. This arises from the fact that the normal model can neither capture skewness nor withstand heavy tails, while the Laplace model, although more appropriate than the normal one for heavy-tailed data, suffers from misspecification when faced with asymmetric data due to its symmetry assumption. In contrast, the skew-Laplace model, by virtue of its simultaneous capacity to model both asymmetry and heavy tails, achieves the best performance in this scenario. Therefore, when the data possess both skewness and tail kurtosis, employing a more flexible model such as the skew-Laplace not only constitutes a suitable choice but is indeed the optimal one.

In Scenarios 4 through 6, with the aim of evaluating the robustness of the models under more realistic conditions, the simulated residuals were generated not from a pure distribution but from a combination of normal data and outlying observations. In Scenario 4, 90% of the data were generated from a standard normal distribution, 5% of the outlying points were located in the interval $(-6, -3)$, and the remaining 5% in the interval $(3, 6)$. This structure gave rise to a symmetric distribution with heavy tails on both sides, which behaviorally resembles the Laplace distribution. The results of this scenario (the fourth), presented on the right-hand side of Table 5, indicated that the pattern of model fit is similar to that of Scenario 2; namely, the normal model performed more poorly than the Laplace and skew-Laplace models, whereas the latter two models provided a better fit because of their greater ability to cope with heavy tails. Nevertheless, the important difference between these two scenarios lies in the fact that in Scenario 2 the data were generated directly from a Laplace distribution, whereas in Scenario 4 the Laplace-like behavior results from the contamination of normal data with outlying observations.

In Scenario 5, 90% of the data were generated from a standard normal distribution and 10% of the outlying points from the negative side of the distribution, namely the interval $(-6, -3)$. As a result, the resulting distribution exhibited negative skewness and a heavy left tail. Likewise, in Scenario 6, by considering 10% of outlying points in the interval $(3, 6)$ and 90% standard normal data, a distribution with positive skewness and a heavy right tail was obtained. The results of both scenarios, presented in Table6, were similar to those of Scenario 3; that is, the skew-Laplace model outperformed the normal and Laplace models, since it is the only model capable of modeling asymmetry

TABLE 5. RMSE values of parameters of all three models (Bayesian and frequentist approaches)

		Scenario 3 (Skew Laplace distribution)						**Scenario 4 (Outlier contamination)**					
		$m = 50$						$m = 50$					
Cluster	**PAR**	Frequentist			Bayesian			Frequentist			Bayesian		
		N	L	SL	N	L	SL	N	L	SL	N	L	SL
1	β_{01}	2.1524	1.2237	0.1307	2.0078	1.1302	0.0947	1.1262	0.0680	0.1838	1.0246	0.0387	0.1577
	β_{11}	2.3056	1.0666	0.1512	1.9694	1.0800	0.0730	1.2295	0.0497	0.1519	0.7729	0.0472	0.1502
	β_{21}	2.3051	1.9326	0.1450	1.9633	1.0069	0.0805	1.2654	0.0467	0.1839	0.7750	0.0325	0.1522
	β_{31}	2.3384	1.5511	0.1563	1.9740	0.9975	0.1067	1.0333	0.0517	0.1468	1.0185	0.0469	0.1455
	β_{41}	2.3465	1.4922	0.1515	1.9984	1.0232	0.0799	1.8342	0.0599	0.1669	0.8540	0.0468	0.1193
	β_{51}	2.5974	1.4079	0.1289	1.9609	0.9352	0.1100	0.9563	0.0783	0.1698	0.9183	0.0503	0.1664
	σ_1^2	2.4844	1.4817	0.1302	2.0123	1.1760	0.1062	1.4755	0.0449	0.1062	0.7916	0.0426	0.0907
	γ_1	–	–	0.1485	–	–	0.0652	–	–	0.1707	–	–	0.1254
	π_1	2.5193	1.8887	0.1340	2.0199	1.2036	0.1106	1.4518	0.0421	0.1240	0.9031	0.0384	0.0968
2	β_{02}	2.1851	1.0218	0.1289	2.0103	0.9665	0.0923	1.5768	0.0523	0.1338	0.9409	0.0471	0.1277
	β_{12}	2.5609	1.6904	0.1344	2.0013	0.9427	0.1078	1.0296	0.0729	0.1244	0.8425	0.0419	0.1211
	β_{22}	2.3816	1.3088	0.1399	1.9567	0.9835	0.1148	1.2772	0.0546	0.1072	1.1418	0.0486	0.1042
	β_{32}	2.3231	1.6973	0.1501	1.9650	1.1252	0.0777	1.2732	0.0550	0.0970	0.7845	0.0375	0.0672
	σ_2^2	2.3770	1.8138	0.1207	1.9602	0.9700	0.0957	1.8515	0.0561	0.1683	1.4993	0.0447	0.0761
	γ_2	–	–	0.1339	–	–	0.0739	–	–	0.7960	–	–	0.4485
		$m = 100$						$m = 100$					
1	β_{01}	2.0078	0.9632	0.0224	1.3884	0.8079	0.0091	0.9043	0.0354	0.1332	0.4896	0.0246	0.0846
	β_{11}	1.6699	0.8153	0.0107	1.2162	0.7830	0.0105	0.7605	0.0390	0.0879	0.5237	0.0241	0.0857
	β_{21}	1.9326	0.6058	0.0173	1.7514	0.5966	0.0085	0.7643	0.0317	0.1303	0.3959	0.0282	0.0964
	β_{31}	1.9513	0.6018	0.0127	1.5511	0.5736	0.0059	0.9138	0.0380	0.0921	0.4386	0.0284	0.0712
	β_{41}	1.7983	0.7601	0.0172	1.5268	0.5713	0.0067	0.6715	0.0288	0.1149	0.6533	0.0174	0.0658
	β_{51}	1.9609	0.7353	0.0091	1.6657	0.5817	0.0085	0.8281	0.0424	0.1514	0.4910	0.0250	0.0907
	σ_1^2	1.4817	0.8428	0.0222	1.4521	0.4715	0.0108	0.7495	0.0373	0.0872	0.4456	0.0256	0.0581
	γ_1	–	–	0.0130	–	–	0.0049	–	–	0.0877	–	–	0.0626
	π_1	1.8887	0.8010	0.0117	1.5399	0.5324	0.0076	0.8194	0.0289	0.0870	0.4880	0.0242	0.0506
2	β_{02}	1.8235	0.6575	0.0237	1.1959	0.6336	0.0119	0.7949	0.0378	0.0835	0.4775	0.0285	0.0629
	β_{12}	1.6904	0.9427	0.0106	1.6225	0.7407	0.0091	0.7313	0.0370	0.0955	0.5007	0.0195	0.0894
	β_{22}	1.9567	0.9529	0.0143	1.7029	0.8306	0.0120	0.6758	0.0479	0.0763	0.5464	0.0263	0.0571
	β_{32}	1.8705	1.0331	0.0127	1.6973	0.8825	0.0109	0.7370	0.0306	0.0614	0.4691	0.0248	0.0564
	σ_2^2	1.8138	0.9700	0.0144	1.6043	0.7830	0.0132	0.8939	0.0358	0.0709	0.3895	0.0220	0.0660
	γ_2	–	–	0.0187	–	–	0.0081	–	–	0.1219	–	–	0.0696

and tail thickness simultaneously. Nevertheless, as in Scenario 4, it should be noted that in Scenario 3 the data were simulated directly from a skew-Laplace distribution, whereas in Scenarios 5 and 6 the skew-Laplace-like behavior was generated through the combination of a normal distribution with one-sided outlying points.

Overall, the results of Scenarios 4 through 6 demonstrate that the adequate performance of the Laplace and skew-Laplace models is not confined to cases in which the data are generated exactly from these distributions. Rather, whenever the data—even as a result of contamination with outlying points—possess characteristics such as symmetric heavy tails or asymmetry accompanied by heavy tails, these models still provide a satisfactory fit. This finding highlights the high flexibility of the skew-Laplace model and its suitability as an appropriate framework for modeling contaminated, heavy-tailed, and asymmetric data.

Based on the results obtained, it is evident that the Bayesian approach outperforms its frequentist counterpart across all three models. This superiority is reflected in consistently lower RMSE values for the estimated parameters, indicating higher estimation accuracy and more efficient use of the available information through the incorporation of prior knowledge.

Moreover, increasing the sample size (m) from 60 to 120 leads to a noticeable reduction in the RMSE values of the parameters for all three models under both approaches. This decline in estimation error as the sample size grows is a clear indication of the consistency of the parameter estimators, a highly desirable property in statistical methodology. It should be noted that in this study, the number of simulated time series remains constant ($n = 60$), while the length of each time series (m) is increased from 60 to 120.

Table 7 presents the average Accuracy, Mean Misclassification Rate (MCR), and Adjusted Rand Index (ARI) across all scenarios, categorized by Bayesian and frequentist approaches for various values of m. Given that the simulation and model-fitting process was repeated 1,000 times for each scenario, sample size, and approach, the reported values represent the mean of these indicators. The results substantiate the efficacy of the proposed Laplace and Skew-Laplace distribution-based models in handling non-normal conditions, specifically demonstrating superior performance over the classical Normal model in the presence of outliers and skewness. Furthermore, the stability of the performance metrics and the marginal improvement in accuracy as the sample size increases underscore

TABLE 6. RMSE values of parameters of all three models (Bayesian and frequentist approaches)

		Scenario 5 (0.10 outlier + 0.9 Normal)						**Scenario 6 (0.9 Normal + 0.10 outlier)**					
		$m = 50$						$m = 50$					
Cluster	**PAR**	Frequentist			Bayesian			Frequentist			Bayesian		
		N	L	SL	N	L	SL	N	L	SL	N	L	SL
1	β_{01}	2.3370	1.1497	0.1494	2.3172	1.1359	0.1422	2.3172	1.1339	0.1570	2.2980	1.1228	0.1488
	β_{11}	2.4956	1.1603	0.1698	2.4770	1.1539	0.1620	2.5161	1.1755	0.1778	2.4961	1.1679	0.1550
	β_{21}	2.4990	1.7898	0.1633	2.4811	1.7744	0.1563	2.5189	1.7741	0.1707	2.4623	1.7606	0.1502
	β_{31}	2.1341	1.4654	0.1520	2.1165	1.4505	0.1451	2.1543	1.4804	0.1518	2.1368	1.4362	0.1453
	β_{41}	2.5476	1.6038	0.1407	2.5287	1.5902	0.1328	2.5474	1.6180	0.1397	2.5270	1.6040	0.1328
	β_{51}	2.4282	1.5182	0.1482	2.4080	1.5033	0.1404	2.4271	1.5341	0.1556	2.4089	1.4903	0.1466
	σ_1^2	2.6833	1.5751	0.1411	2.6650	1.5603	0.1342	2.6853	1.5601	0.1411	2.6634	1.5468	0.1335
	γ_1	–	–	0.0797	–	–	0.0730	–	–	0.0865	–	–	0.0659
	π_1	2.3068	1.8108	0.1516	2.2867	1.7970	0.1441	2.3264	1.8260	0.1588	2.2666	1.8109	0.1508
2	β_{02}	2.1952	1.1113	0.1470	2.1779	1.0974	0.1402	2.2148	1.1247	0.1402	2.1971	1.0837	0.1336
	β_{12}	2.3591	1.7821	0.1526	2.3416	1.7679	0.1448	2.3371	1.7801	0.1595	2.3220	1.7688	0.1518
	β_{22}	2.1948	1.2371	0.1589	2.1761	1.2214	0.1517	2.1760	1.2525	0.1514	2.1571	1.2090	0.1450
	β_{32}	2.1321	1.6104	0.1454	2.1117	1.5967	0.1382	2.1118	1.5957	0.1524	2.0926	1.5846	0.1449
	σ_2^2	2.5772	1.7873	0.1379	2.5596	1.7730	0.1312	2.5986	1.8027	0.1380	2.5769	1.7586	0.1303
	γ_2	–	–	0.0857	–	–	0.0784	–	–	0.0791	–	–	0.0786
		$m = 100$						$m = 100$					
1	β_{01}	1.8452	1.0581	0.0888	1.8297	1.0494	0.0838	1.8622	1.0566	0.0825	1.8133	1.0496	0.0783
	β_{11}	2.1554	1.1467	0.0668	2.1392	1.1454	0.0617	2.1380	1.1549	0.0724	2.1226	1.1378	0.0671
	β_{21}	1.8341	1.0975	0.0866	1.8180	1.0890	0.0811	1.8502	1.0970	0.0925	1.8014	1.0896	0.0760
	β_{31}	1.8195	0.9164	0.1175	1.8045	0.9074	0.1125	1.8202	0.9146	0.1184	1.8020	0.9084	0.1113
	β_{41}	2.1858	1.1242	0.0796	2.1706	1.1152	0.0741	2.1699	1.1325	0.0734	2.1562	1.1225	0.0689
	β_{51}	1.7873	0.8534	0.1214	1.7711	0.8457	0.1159	1.7869	0.8616	0.1209	1.7685	0.8386	0.1150
	σ_1^2	2.2037	1.2642	0.1002	2.1875	1.2566	0.0950	2.2209	1.2718	0.1007	2.2038	1.2487	0.0938
	γ_1	–	–	0.0648	–	–	0.0594	–	–	0.0589	–	–	0.0539
	π_1	2.2088	1.2920	0.1225	2.1936	1.2842	0.1168	2.2249	1.2837	0.1282	2.1779	1.2761	0.1113
2	β_{02}	1.9872	1.0544	0.1034	1.9703	1.0458	0.0983	2.0054	1.0536	0.1089	1.9546	1.0468	0.1042
	β_{12}	1.8293	1.0388	0.1141	1.8138	1.0310	0.1083	1.8465	1.0382	0.1137	1.8282	1.0305	0.1085
	β_{22}	2.1427	0.9054	0.1213	2.1258	0.8967	0.1157	2.1250	0.8970	0.1273	2.1108	0.8893	0.1215
	β_{32}	1.7792	1.0427	0.0767	1.7644	1.0344	0.0715	1.7977	1.0346	0.0703	1.7794	1.0273	0.0659
	σ_2^2	1.9128	0.8838	0.0900	1.8963	0.8751	0.0843	1.8961	0.8825	0.0845	1.8819	0.8756	0.0785
	γ_2	–	–	0.0736	–	–	0.0682	–	–	0.0736	–	–	0.0718

the consistency and reliability of the developed models in accurately reconstructing the underlying cluster structure.

5. Real Data Application

In this section, the `toydataset` from the `SaMUraiS` software package is utilized to evaluate the performance of the proposed models. It is worth noting that this package is available for installation via GitHub from the personal page of Faicel Chamroukhi, a prominent researcher in the field of time series clustering[3]. The `SaMUraiS` package provides a comprehensive collection of diverse, high-dimensional time series datasets suitable for various applications, including clustering, classification, and analysis of variance.

Specifically, the `toydataset` consists of 30 time series, each containing 350 observations, designed as a benchmark for testing the flexibility of functional clustering methods. This dataset is particularly appropriate for our study as it contains distinct underlying patterns with varying degrees of complexity, allowing us to test the ability of the polynomial mixture models to capture different structural trends and distributional characteristics in a controlled yet realistic setting.

Accordingly, the Normal, Laplace, and Skew-Laplace models were fitted under both Frequentist and Bayesian approaches, considering a maximum number of clusters $K_{max} = 5$ and a maximum polynomial degree $p_{max} = 10$. Model selection was conducted based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

The results indicated that the minimum AIC and BIC values for all models across both approaches were achieved at $K = 3$ clusters with polynomial degrees $p = (1, 0, 1)$. In other words, the polynomial degrees for the first, second, and third clusters were estimated as 1, 0, and 1, respectively. Based on this, three clusters were formed, each containing 10 time series. Notably, every 10 consecutive time series were correctly assigned to a single cluster, demonstrating the perfect alignment of the model with the true underlying structure of the data.

The model selection indices for the three-cluster solution are summarized for both approaches in Table 8. Empirical findings suggest that in both Frequentist and Bayesian frameworks, the Skew-Laplace model outperforms the Laplace and Normal models by providing the best fit to the data. Furthermore, as observed in the simulation study, the

[3] `devtools::install_github("fchamroukhi/SaMuRaiS")`

TABLE 7. Clustering quality indicators

Scenario	**Model**	**Approach**	$m = 60$			$m = 120$		
			Acc	MCR	ARI	Acc	MCR	ARI
Scenario 1	Normal	Frequentist	0.95	0.05	0.90	0.96	0.04	0.92
		Bayesian	0.97	0.03	0.90	0.98	0.02	0.93
	Laplace	Frequentist	0.88	0.12	0.84	0.87	0.13	0.84
		Bayesian	0.90	0.10	0.84	0.92	0.08	0.89
	Skew-Laplace	Frequentist	0.92	0.08	0.89	0.93	0.07	0.90
		Bayesian	0.94	0.06	0.90	0.95	0.05	0.90
Scenario 2	Normal	Frequentist	0.78	0.22	0.76	0.79	0.21	0.77
		Bayesian	0.82	0.18	0.80	0.84	0.16	0.81
	Laplace	Frequentist	0.95	0.05	0.90	0.96	0.04	0.91
		Bayesian	0.98	0.02	0.93	0.98	0.02	0.93
	Skew-Laplace	Frequentist	0.92	0.08	0.88	0.93	0.07	0.89
		Bayesian	0.93	0.07	0.89	0.93	0.07	0.89
Scenario 3	Normal	Frequentist	0.71	0.29	0.70	0.73	0.27	0.72
		Bayesian	0.73	0.27	0.72	0.74	0.26	0.73
	Laplace	Frequentist	0.74	0.26	0.73	0.76	0.24	0.75
		Bayesian	0.78	0.22	0.76	0.78	0.22	0.76
	Skew-Laplace	Frequentist	0.96	0.04	0.91	0.97	0.03	0.92
		Bayesian	0.98	0.02	0.93	0.98	0.02	0.93
Scenario 4	Normal	Frequentist	0.78	0.22	0.76	0.80	0.20	0.78
		Bayesian	0.81	0.19	0.79	0.84	0.16	0.81
	Laplace	Frequentist	0.93	0.07	0.89	0.94	0.06	0.90
		Bayesian	0.95	0.05	0.90	0.96	0.04	0.91
	Skew-Laplace	Frequentist	0.90	0.10	0.86	0.90	0.10	0.86
		Bayesian	0.92	0.08	0.88	0.93	0.07	0.89
Scenario 5	Normal	Frequentist	0.72	0.28	0.71	0.75	0.25	0.74
		Bayesian	0.73	0.27	0.72	0.78	0.22	0.76
	Laplace	Frequentist	0.77	0.23	0.75	0.80	0.20	0.78
		Bayesian	0.81	0.19	0.79	0.84	0.16	0.81
	Skew-Laplace	Frequentist	0.95	0.05	0.90	0.95	0.05	0.90
		Bayesian	0.96	0.04	0.91	0.97	0.03	0.92
Scenario 6	Normal	Frequentist	0.70	0.30	0.70	0.71	0.29	0.70
		Bayesian	0.71	0.29	0.70	0.73	0.27	0.72
	Laplace	Frequentist	0.75	0.25	0.74	0.83	0.17	0.80
		Bayesian	0.79	0.21	0.77	0.85	0.15	0.82
	Skew-Laplace	Frequentist	0.96	0.04	0.91	0.96	0.04	0.91
		Bayesian	0.96	0.04	0.91	0.98	0.02	0.93

Bayesian approach provides a superior fit compared to the Frequentist approach for all three models, emphasizing the higher efficiency of Bayesian methods in handling complex clustering structures.

From an interpretative perspective, the clustering results reveal distinct structural differences among the three identified groups. Cluster 1 is characterized by a steady upward trend, Cluster 2 exhibits a more stable mean with higher volatility, and Cluster 3 shows a clear declining trend. The ability of the Skew-Laplace mixture model to accurately separate these groups, while accounting for the skewness and heavy tails of the residuals, confirms its superiority over the standard Normal mixture model in real-world scenarios.

TABLE 8. Model selection criteria for both Frequentist and Bayesian approaches.

Approach	**Criterion**	**Normal**	**Laplace**	**Skew-Laplace**
Frequentist	$-LogL$	471.6511	374.9947	371.6897
	AIC	963.3022	769.9895	769.3793
	BIC	1106.513	913.1998	955.5528
Bayesian	$-LogL$	354.1602	343.3763	293.0657
	AIC	728.3204	706.7525	612.1313
	BIC	871.5308	849.9629	798.3048

5.1. **Parameter Estimation and Variance Analysis of Clusters.** Table 9 presents the estimated parameters for the three models under both Frequentist and Bayesian frameworks. A comparative analysis of these estimates reveals a notable inflation of the variance (σ^2) in the Normal model relative to the Laplace and Skew-Laplace models. This phenomenon can be attributed to a *spurious inflation of variance*, where the Normal distribution, unable to accommodate the inherent skewness and heavy-tailed structure of the data, is compelled to expand its dispersion parameter to achieve an acceptable fit.

This limitation is also evident in the Laplace model, albeit to a lesser extent. While the Laplace distribution inherently possesses heavier tails than the Normal distribution and thus maintains better control over data dispersion, its fundamental symmetry still

poses a significant challenge when dealing with skewed data. Consequently, to optimize the fit under skewed scenarios, the Laplace model also exhibits a tendency to increase its estimated variance, resulting in values that remain consistently higher than those of the Skew-Laplace model. In contrast, by explicitly incorporating the skewness parameter (γ), the Skew-Laplace mixture model directly captures the distributional asymmetry. This allows for a significantly more parsimonious fit, preventing any spurious increase in variance and providing a more precise characterization of the cluster dispersions.

TABLE 9. Parameter estimates for all three models (Frequentist and Bayesian).

Parameter	**Frequentist**			**Bayesian**		
	Normal	**Laplace**	**Skew-L.**	**Normal**	**Laplace**	**Skew-L.**
β_{01}	5.5004	5.9833	6.7729	5.4580	5.9412	6.4518
β_{11}	0.0044	0.0032	0.0025	0.0030	0.0022	0.0018
σ_1^2	0.8689	0.2869	0.2158	0.8568	0.2604	0.2126
γ_1	0.0000	0.0000	-0.3934	0.0000	0.0000	-0.3908
π_1	0.3700	0.3500	0.3200	0.3500	0.3200	0.3300
β_{02}	5.4895	5.0873	4.9283	5.3772	5.0618	4.8032
σ_2^2	0.7254	0.1963	0.1606	0.7123	0.1631	0.1494
γ_2	0.0000	0.0000	0.3035	0.0000	0.0000	0.2880
π_2	0.2900	0.3600	0.3300	0.3600	0.3400	0.3200
β_{03}	7.2428	6.7934	7.1267	6.8197	6.7909	6.6057
β_{13}	-0.0202	-0.0147	-0.0109	-0.0102	-0.0101	-0.0098
σ_3^2	0.3945	0.2566	0.1503	0.3788	0.2525	0.1298
γ_3	0.0000	0.0000	0.0753	0.0000	0.0000	0.0707

6. Conclusion and Future Directions

6.1. **Discussion.** The primary objective of this study was to evaluate the performance of a model-based clustering framework for time series across diverse residual distribution scenarios. The simulation results provide empirical evidence of the robustness of

the proposed models when subjected to non-normal conditions, specifically skewness and heavy-tailedness.

In the first scenario, where the data followed a Normal distribution, the Normal model exhibited the optimal performance, yielding the lowest RMSE and the highest clustering quality (accuracy, misclassification rate, and adjusted Rand index). This result aligns with classical statistical expectations, where the model architecture perfectly matches the underlying data-generating process. Conversely, in the second and third scenarios—characterized by heavy tails (Laplace) and skewness—the efficacy of the Normal model diminished significantly. Our findings indicate that the Laplace and Skew-Laplace models offer superior robustness compared to traditional normal-based methods by effectively reconstructing the mean structure despite the presence of asymmetry and outliers. Notably, the Skew-Laplace model demonstrated a remarkable capacity to maintain estimation accuracy even in highly asymmetric environments.

A consistent finding across all tested scenarios was the performance advantage of the Bayesian approach over the frequentist framework. The Bayesian estimators not only reduced RMSE more effectively but also exhibited greater stability in preserving clustering accuracy. Furthermore, the observed improvement in estimation as the series length (m) increased from 60 to 120 corroborates the principle of statistical convergence, suggesting that additional temporal information enhances the precision of cluster parameter estimation.

Despite these advantages, certain limitations must be acknowledged. The current framework assumes independent, white-noise-like residuals. While this is a standard assumption in many time-series applications, it may not fully account for financial data characterized by conditional heteroscedasticity. In such instances, integrating the model with volatility-modeling components. Additionally, while the polynomial bases employed here are efficient for capturing smooth temporal trends, they may lack the flexibility required to characterize abrupt structural breaks.

6.2. **Conclusion.** In conclusion, this study demonstrates that incorporating heavy-tailed and skewed distributions into a model-based clustering framework significantly improves robustness for time-series data compared to standard normal models. The Bayesian approach proved to be a more reliable tool for recovering cluster structures in non-ideal

conditions. Future research could extend this framework by addressing the identified limitations. Specifically, integrating conditional heteroscedasticity models could broaden the applicability to financial contexts. Moreover, to enhance the flexibility of the temporal modeling, future studies could explore the incorporation of regime-switching mechanisms within each cluster, allowing the model to adapt to structural changes or shifts in the underlying data-generating process over time.

$$\mathbf{x}_1, \ldots, \mathbf{x}_n \sim f(\mathbf{x}_i|\boldsymbol{T}, \boldsymbol{\Psi}) = \sum_{k=1}^{K} \pi_k \sum_{l=1}^{L} \eta_{kl} f_{kl}(x_{ij}|\boldsymbol{T}'_{klj}\boldsymbol{\beta}_{kl}, \boldsymbol{\theta}_{kl})$$

References

[1] Adrian, A., Maitra, R. and Hurn, M. (2016). *Developments in model-based clustering*. New York: Springer.

[2] Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B*, 36(1), pp. 99-102.

[3] Arslan, O. (2009). An alternative multivariate skew-Laplace distribution. *Statistical Papers*, 51(4), pp. 771-787.

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), pp. 171-178.

[5] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society: Series B*, 61(3), pp. 579-602.

[6] Banfield, J.D. and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49(3), pp. 803-821.

[7] Bernardi, M., Maruotti, A., Lea, P. (2012). Skew mixture models for loss distributions: A Bayesian approach. The American Statistician, 53, 357–361.

[8] Browne, R.P. and McNicholas, P.D. (2015). A mixture of generalized hyperbolic distributions. *Canadian Journal of Statistics*, 43(2), pp. 176-198.

[9] Bohning, H., et al. (1994). On the convergence of the EM algorithm for mixture models. *Computational Statistics & Data Analysis*, 20(3), pp. 450-465.

[10] Chamroukhi, F. (2010). *Bayesian and maximum likelihood approaches for model-based curve clustering*. Ph.D. Dissertation, Université de Technologie de Compiègne.

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

[12] Diebolt, J. and Robert, C. P. (1990). Bayesian estimation of finite mixture distributions. *Journal of the Royal Statistical Society: Series B*, 2(1).

[13] Diebolt, J. and Robert, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B*, 56(2), pp. 363-375.

[14] Farokhi, A. and Shadrokh, A. (2018). Bayesian Estimation in Skew-Normal and Skew-Laplace Linear Multilevel Models. *Afghanistan Quarterly*, 11(2), pp. 144-165.

[15] Farokhi, A. and Shadrokh, A. (2016). Bayesian Analysis of Multilevel Models with the Skew-Slash Distribution. *International Journal of Applied Mathematics and Statistics*, 55(1), pp. 1-11.

[16] Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), pp. 611-631.

[17] Gaffney, S.J. (2004). *Probabilistic curve clustering using mixture models*. Ph.D. Dissertation, University of California, Irvine.

[18] Gelman, A., Jakulin, A., Pittau, M.G. and Su, Y.S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), pp. 1360-1383.

[19] Gui, R. (2003). *Model-based clustering of time series*. Ph.D. Dissertation, University of British Columbia.

[20] Hebrial, C., Samé, A., Chamroukhi, F., Govaert, G. and Aknin, P. (2010). A latent class model for diagnosing railway track circuits. *IFAC Symposium on Fault Detection*.

[21] Keeney, S. (2013). *Developments in model-based clustering of time series*. Ph.D. Dissertation, University of Dublin.

[22] Lee, S. and McLachlan, G.J. (2013). On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. *arXiv preprint arXiv:1305.1095*.

[23] Marin, J.M., Mengersen, K. and Robert, C.P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, 25, pp. 459-507.

[24] McLachlan, G.J. and Krishnan, T. (1997). *The EM algorithm and extensions*. New York: John Wiley & Sons.

[25] McLachlan, G. J. and Krishnan, T. (1998). *The EM algorithm and extensions*. Updated edition. New York: John Wiley & Sons.

[26] McLachlan, G.J. and Peel, D. (2000). *Finite mixture models*. New York: John Wiley & Sons.

[27] Paparrizos, J., Kang, U., Antunes, N., Tsay, P.S., Palpanas, T. and Franklin, M.J. (2025). Time-Series Clustering: A Comprehensive Study. *Proceedings of the VLDB Endowment*, 18.

[28] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27(7), pp. 859-870.

[29] Picard, F., Robin, S., Lavielle, M., Vaisse, C. and Daudin, J.J. (2007). A statistical approach for array CGH data segmentation. *BMC Bioinformatics*, 6(1), pp. 1-14.

[30] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*. 2nd ed. New York: Springer.

[31] Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series B*, 59(4), pp. 731-792.

[32] Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of Normals. *Journal of the American Statistical Association*, 92(439), pp. 894-902.

[33] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew-normal and skew-t distributions. *Biometrika*, 93(2), pp. 421-433.

[34] Song, W., Yao, W. and Xing, Y. (2014). Robust mixture regression model based on Laplace distribution. *Computational Statistics & Data Analysis*, 71, pp. 128-140.

[35] Xiang, Y., Gubian, S., Martin, F., Lanni, P. and Rivier, A. (2004). Generalized simulated annealing for efficient global optimization: the GenSA package. *The R Journal*.

[36] Yao, W., Wei, Y. and Yu, C. (2014). Robust mixture regression using the t-distribution. *Computational Statistics & Data Analysis*, 71, pp. 116-127.

[37] Zhang, J. (2013). Robust mixture regression modeling based on Pearson type VII distribution. *Communications in Statistics*, 42(10).

Ali Shadrokh
Department of Statistics,
Faculty of Mathematical Sciences and Computer,
Payame Noor University of Tehran,
Tehran, Iran
Email:A.shadrokh@pnu.ac.ir

Samira Saadat
Department of Statistics,
Faculty of Mathematical Sciences and Computer,
Payame Noor University of Tehran,
Tehran, Iran
Email: samira.saadati231@gmail.com

Parviz Nasiri
Department of Statistics,
Faculty of Mathematical Sciences and Computer,
Payame Noor University of Tehran,
Tehran, Iran
Email: pnasiri45@pnu.ac.ir



©2026 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonComertial 4.0 International (CC BY-NC 4.0 license) (http://creativecommons.org/licenses/by-nc/4.0/).