http://communications.science.ankara.edu.tr
ISSN: 1303-5991 E-ISSN: 2618-6470
Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.
Volume 75, Number 2, Pages 211-227 (2026)
https://doi.org/10.31801/cfsuasmas.1677165
Research Article

Robust Parameter Estimation in PLMs with Skewed and Heavy-tailed Error Distributions

Fatma Zehra Doğru[1,*], Olçay Arslan[2]

[1]Department of Data Science and Analytics, Faculty of Arts and Science, Giresun University, Giresun, Türkiye
[2]Department of Statistics, Faculty of Science, Ankara University, Ankara, Türkiye

ABSTRACT

Partially linear models (PLMs) provide a flexible extension of classical linear models by incorporating parametric and nonparametric components, making them suitable for complex data structures often encountered in economics and biometrics. Traditionally, these models assume normally distributed errors, an assumption that may not hold in many practical scenarios. To address this, previous studies introduced the skew normal distribution in PLMs (PLM-SN) to account for asymmetry. However, such data sets may also exhibit heavy tails and skewness. To capture both characteristics, we propose a new approach: modeling the error structure using the skewed Laplace normal distribution (PLM-SLN), which maintains the same number of parameters as PLM-SN but better accommodates heavy-tailed behavior. An efficient estimation procedure using an expectation-conditional maximization (ECM) algorithm is developed to compute maximum likelihood estimates. The proposed method is validated using simulation experiments and applied to an empirical dataset to highlight its practical utility.

Keywords: ECM algorithm, maximum likelihood estimation, PLM-SLN, PLM-SN
2020 Mathematics Subject Classification: 62J02, 62F10, 62F35

***Corresponding Author:** Fatma Zehra Doğru
fatma.dogru@giresun.edu.tr (F. Z. Doğru) 0000-0001-8220-2375 https://ror.org/05szaq822
oarslan@ankara.edu.tr (O. Arslan) 0000-0002-7067-4997 https://ror.org/01wntqw50

Research Article; Received: April 22, 2025; **Accepted:** September 9, 2025
Cite this article as: Doğru, F. Z. and Arslan, O. (2026). Robust parameter estimation in PLMs with skewed and heavy-tailed error distributions. *Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.*, 75 (2), 211-227.

This is an open-access article under the Creative Commons Attribution Non-Commercial 4.0 International (CC BY-NC 4.0) License.

1 Introduction

Partially linear models (PLMs), belonging to the broader family of semiparametric models, have gained considerable attention due to their capacity to model both linear and nonlinear associations between explanatory variables and the response. This combination enables more flexible representations of the relationship between covariates and the response variable. Several applications have shown the effectiveness of PLMs in modeling asymmetric or dependent data. For instance, [25] applied PLMs to positive and skewed outcomes,

while [22] investigated diagnostic tools and estimation techniques for PLMs under autoregressive symmetric error structures.

To better capture deviations from normality, various extensions have been introduced in the literature. [18] proposed PLMs with t-distributed errors, allowing for symmetric heavy-tailed behavior. For asymmetric data, [12] adopted the skew-normal distribution, originally introduced by [2] within the PLM framework called PLM-SN. They also presented maximum likelihood (ML) estimation and influence diagnostics, based on the work of [26]. More recently, [13] extended these ideas by considering autoregressive skew normal errors.

In a related work, [11] applied the heteroscedastic PLM-SN errors to analyze ragweed pollen concentration data. Their study showed that incorporating skewness can improve model fit in environmental datasets with asymmetric features. However, the SN distribution may still fall short when data exhibit both skewness and heavy tails.

To address this limitation, [4] proposed a class of PLMs incorporating two-piece scale mixture normal errors, originally developed by [20], which offer additional flexibility at the cost of increased model complexity. Similarly, skewed-t distributions are highly flexible but computationally more intensive due to the estimation of extra parameters.

In this context, the skew Laplace normal (SLN) distribution, proposed by [14], offers a promising alternative. It effectively captures heavy-tailed and skewed behavior without increasing the number of parameters and is a member of the more general skew exponential power family [3, 6]. Moreover, its hierarchical and stochastic forms make it highly compatible with EM-based estimation frameworks.

The SLN distribution has demonstrated considerable versatility across various statistical modeling applications. It has been employed in regression settings through mixture models [7], as well as in unified estimation frameworks that simultaneously model location, scale, and skewness parameters [8]. Further, its utility has been shown in the context of modeling heteroscedastic data [10]. More recently, [9] investigated shape mixture extensions of the SLN distribution, enhancing its flexibility for complex data structures. Motivated by these developments, the present work introduces a PLM-SLN, aiming to capture both asymmetry and heavy-tailedness within a unified modeling framework.

The rest of the paper is organized as follows: Section 2 outlines the key characteristics of the SLN distribution. Section 3 presents the structure of the PLM-SLN model. Section 4 describes the estimation process using the ECM algorithm. In Section 5, we evaluate the model through simulations and a real-world application. Section 6 concludes the paper with a summary and discussion of future directions.

2 Skew Laplace Normal Distribution

We begin by formally introducing the SLN distribution, which blends features of both the Laplace and skew normal distributions, making it suitable for modeling data that exhibits skewness and heavy tails.

Definition 1. *Let Y be a continuous random variable characterized by a location parameter $\mu \in \mathbb{R}$, scale $\sigma^2 > 0$, and skewness parameter $\lambda \in \mathbb{R}$. Then, Y follows a SLN distribution, presented $Y \sim SLN(\mu, \sigma, \lambda)$, if its probability density function (pdf) is defined by:*

$$f(y) = 2 f_L(y; \mu, \sigma) \Phi\left(\lambda \frac{y - \mu}{\sigma}\right), \quad -\infty < y < \infty, \tag{1}$$

where $f_L(y;\mu,\sigma)$ is the Laplace density defined as:

$$f_L(y;\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|y-\mu|}{\sigma}\right),$$

and $\Phi(\cdot)$ represents the cumulative distribution function (cdf) of the standard normal distribution.

Definition 2 (Stochastic Representation)**.** *An equivalent representation of the SLN distribution can be obtained as a scale mixture of skew-normal distributions. Let $Z \sim SN(0,1,\lambda)$ be a standard skew-normal random variable, and suppose V is an independent positive variable with density*

$$f_V(v) = v^{-3}\exp\left(-(2v^2)^{-1}\right), v > 0.$$

Then the transformed variable

$$Y = \mu + \sigma\frac{Z}{V},$$

follows an SLN distribution. Alternatively, using two independent standard normals Z_1 and Z_2, one may write:

$$Y = \mu + \sigma\left(\frac{\lambda|Z_1|}{\sqrt{V^2(V^2+\lambda^2)}} + \frac{Z_2}{\sqrt{V^2+\lambda^2}}\right),$$

where Z_1 and Z_2 are independent standard normal random variables.

This representation facilitates EM-type inference procedures due to its hierarchical latent structure.

Definition 3 (Hierarchical Representation)**.** *The SLN distribution may be expressed through a three-layer hierarchy as follows:*

$$\begin{aligned} Y|\gamma,v &\sim N\left(\mu + \frac{\sigma\lambda\gamma}{v^2+\lambda^2}, \frac{\sigma^2}{v^2+\lambda^2}\right), \\ \gamma|v &\sim TN\left(\left(0, \frac{v^2+\lambda^2}{v^2}\right); (0,\infty)\right), \\ V &\sim f_V(v) = v^{-3}\exp(-(2v^2)^{-1}), \end{aligned} \tag{2}$$

where $\gamma = \sqrt{\frac{V^2+\lambda^2}{V^2}}|Z_1|$, and $TN(a,b;(0,\infty))$ refers to a truncated normal distribution with mean a, variance b, restricted to the positive real line.

Proposition 1. *Given the hierarchical structure above, conditional expectations for latent variables can be derived, which are instrumental for ML estimation. Specifically,*

$$\begin{aligned} E(V^2|y) &= \frac{\sigma}{|y-\mu|}, \\ E(\gamma|y) &= \lambda\frac{u}{\sigma} + \frac{\phi(\lambda\frac{u}{\sigma})}{\Phi(\lambda\frac{u}{\sigma})}, \end{aligned}$$

where $u = y-\mu$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively.

These results are foundational for constructing the ECM algorithm used in the subsequent estimation procedure.

3 PLMs with SLN Errors

We formulate a PLM framework that incorporates the SLN distribution for the error structure, accommodating skewness and heavy tails in a unified setting. Consider the following regression structure:

$$y_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta} + f(t_i) + \epsilon_i, i = 1, 2, \cdots, n$$

where y_i is the response, $\boldsymbol{x}_i$ is a $p \times 1$ covariate vector, $\boldsymbol{\beta}$ is the corresponding regression coefficient vector, $f(\cdot)$ is a smooth, nonparametric function of a continuous predictor t_i (e.g., time), and ϵ_i are mutually independent error terms.

In this setup, we assume the error terms follow SLN distribution, that is, $\epsilon_i \sim SLN(0, \sigma^2, \lambda)$ as described in (1), for all $i = 1, \ldots, n$. The model can be reformulated in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}\boldsymbol{f} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ is the $n \times 1$ response vector, $\boldsymbol{X}$ is the $n \times p$ design matrix with rows $\boldsymbol{x}_i^\top$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ contains the SLN-distributed errors. The vector $\boldsymbol{f} = \left(f(t_1^0), \ldots, f(t_q^0)\right)^\top$ consists of the smooth function evaluated at q unique time points $t_1^0, \ldots, t_q^0$. The incidence matrix $\boldsymbol{N}$ of dimension $n \times q$ is defined elementwise as $N_{ij} = \mathbb{I}(t_i = t_j^0)$.

3.1 Penalized Log-likelihood

Given that $\epsilon_i \sim SLN(0, \sigma^2, \lambda)$, the conditional distribution of y_i inherits the SLN structure with location parameter $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \boldsymbol{n}_i^\top\boldsymbol{f}$, where $\boldsymbol{n}_i^\top$ is the i-th row of matrix $\boldsymbol{N}$. The corresponding log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \lambda, \boldsymbol{f}^\top)^\top$ is expressed as:

$$\ell(\boldsymbol{\theta}) \propto -\frac{n}{2}\log\sigma^2 - \frac{1}{\sigma}\sum_{i=1}^{n}|y_i - \mu_i| + \sum_{i=1}^{n}\log\Phi\left(\frac{\lambda(y_i - \mu_i)}{\sigma}\right). \tag{3}$$

Maximizing this log-likelihood directly is difficult due to the non-linear form of the cdf. Additionally, without constraining the smooth component $f(\cdot)$, the model may suffer from over-fitting or identifiability issues between the linear and non-parametric parts (see [15]). To overcome these challenges, we adopt a penalized likelihood approach:

$$\ell_p(\boldsymbol{\theta}, \alpha) = \ell(\boldsymbol{\theta}) - \frac{\alpha}{2}J(\boldsymbol{f}),$$

where $J(\boldsymbol{f})$ penalizes roughness in the nonparametric component, and $\alpha > 0$ is a tuning parameter balancing model fit and smoothness. Specifically, higher values of $\ell(\boldsymbol{\theta})$ reflect better model fit, while lower values of $J(\boldsymbol{f})$ encourage smoother estimates of $\boldsymbol{f}$. The function $\boldsymbol{f}$ is represented using a natural cubic spline basis, as described in [11, 12, 16, 18, 19], and the penalty term is based on its second derivative:

$$J(\boldsymbol{f}) = \int_a^b [f^{(2)}(t)]^2 dt.$$

Here, $f^{(2)}(t)$ represents the second derivative of $f(t)$ and $[a, b]$ covers range of $t_1^0, \cdots, t_q^0$.

By Theorem 2.1 of [16], the integral penalty simplifies to a quadratic form:

$$\int_a^b [f^{(2)}(t)]^2 dt = \boldsymbol{f}^\top \boldsymbol{K} \boldsymbol{f},$$

where $\boldsymbol{K} \in \mathbb{R}^{q \times q}$ is a symmetric, positive semi-definite matrix determined by the spline knot positions. A detailed construction of $\boldsymbol{K}$ can be found in [16], and we follow the implementation details outlined in [11].

4 Parameter Estimation

This section outlines the ML estimation process for the PLM-SLN model. Due to the intractability of directly maximizing the log-likelihood function in Equation (3), we employ the ECM algorithm introduced by [21].

4.1 ECM Algorithm

To apply the ECM algorithm, we utilize the hierarchical representation of the SLN distribution, derived from Equation (2). This hierarchical form facilitates the construction of the complete-data log-likelihood required for the algorithm. Then, let $\boldsymbol{y} = (y_1, \ldots, y_n)$ denote the observed data and $\boldsymbol{\gamma}, \boldsymbol{v}$ be the unobserved variables. The complete-data $(\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{v})$ log-likelihood is:

$$\ell_c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{v}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_i-\mu_i)^2}{\sigma^2}{v_i}^2 - 2\eta(y_i-\mu_i)\gamma_i + \eta^2(y_i-\mu_i)^2\right),$$

where $\eta = \lambda/\sigma$. Because of the latent variables, direct maximization does not yield the MLE of $\boldsymbol{\theta}$. Instead, we calculate the expected complete-data log-likelihood conditional on the observed responses:

$$\begin{aligned} E\left(\ell_c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{v})|y_i\right) &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_i-\mu_i)^2}{\sigma^2}E(V_i^2|y_i)\right) + \sum_{i=1}^{n}\eta(y_i-\mu_i)E(\gamma_i|y_i) \\ &- \frac{\eta^2}{2}\sum_{i=1}^{n}(y_i-\mu_i)^2. \end{aligned}$$

This expression requires computing $E(V_i^2|y_i)$ and $E(\gamma_i|y_i)$, which are derived using Proposition (1). The ECM steps are outlined as follows.

ECM Algorithm Steps:

1. **Initialization:** Choose an initial parameter vector $\boldsymbol{\theta}^{(0)}$ and a small convergence tolerance Δ.

2. **E-step:** For each iteration $k = 0, 1, \ldots$, compute:

$$\hat{v}_i^{(k)} = E(V^2|y_i, \hat{\boldsymbol{\theta}}^{(k)}) = \frac{\hat{\sigma}^{(k)}}{\left|\hat{u}_i^{(k)}\right|},$$

$$\hat{\gamma}_{1i}^{(k)} = E(\gamma_i|y_i, \hat{\boldsymbol{\theta}}^{(k)}) = \hat{\lambda}^{(k)} \frac{\hat{u}_i^{(k)}}{\hat{\sigma}^{(k)}} + \frac{\phi\left(\hat{\lambda}^{(k)} \frac{\hat{u}_i^{(k)}}{\hat{\sigma}^{(k)}}\right)}{\Phi\left(\hat{\lambda}^{(k)} \frac{\hat{u}_i^{(k)}}{\hat{\sigma}^{(k)}}\right)}$$

where $\hat{u}_i^{(k)} = y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{n}_i^\top \hat{\boldsymbol{f}}^{(k)}$ and $i = 1, \cdots, n$.

Construct the penalized objective function:

$$Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = Q_c(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) - \frac{\alpha}{2} J(\boldsymbol{f})$$

where

$$Q_c(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} \hat{v}_i^{(k)}(y_i - \mu_i)^2 + \eta \sum_{i=1}^{n} \hat{\gamma}_{1i}^{(k)}(y_i - \mu_i) - \frac{\eta^2}{2}\sum_{i=1}^{n}(y_i - \mu_i)^2.$$

3. **CM-step:** Maximize Q_p with respect to $\boldsymbol{\theta}$ to obtain updated parameters:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(k+1)} &= (\boldsymbol{X}^\top \hat{\boldsymbol{W}}^{(k)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \left(\hat{\boldsymbol{W}}^{(k)}(\boldsymbol{y} - \boldsymbol{N}\hat{\boldsymbol{f}}^{(k)}) - \hat{\lambda}^{(k)}\hat{\sigma}^{(k)}\hat{\boldsymbol{\gamma}}_1^{(k)}\right), \\
\hat{\boldsymbol{f}}^{(k+1)} &= (\boldsymbol{N}^\top \hat{\boldsymbol{W}}^{(k)} \boldsymbol{N} + \alpha\hat{\sigma}^{2(k)}\boldsymbol{K})^{-1} \boldsymbol{N}^\top (\hat{\boldsymbol{W}}^{(k)}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{(k+1)}) \\
&- \hat{\lambda}^{(k)}\hat{\sigma}^{(k)}\hat{\boldsymbol{\gamma}}_1^{(k)}), \\
\hat{\sigma}^{2(k+1)} &= \frac{1}{n}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})^\top \hat{\boldsymbol{V}}^{(k)} (\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)}), \\
\hat{\eta}^{(k+1)} &= \frac{\hat{\boldsymbol{\gamma}}_1^{(k)^\top}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})}{(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})^\top (\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})}, \\
\hat{\lambda}^{(k+1)} &= \frac{\hat{\boldsymbol{\gamma}}_1^{(k)^\top}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})\hat{\sigma}^{(k+1)}}{(\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})^\top (\boldsymbol{y} - \hat{\boldsymbol{\mu}}^{(k+1)})},
\end{aligned}$$

where $\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{N}\hat{\boldsymbol{f}}$, $\hat{\boldsymbol{W}}$ is a diagonal matrix with elements $(\hat{v}_i + \hat{\lambda}^2)$ of dimension $n \times n$, $\hat{\boldsymbol{V}}$ is a diagonal matrix with elements $\hat{v}_i$ of dimension $n \times n$, and $\hat{\boldsymbol{\gamma}}_1 = (\hat{\gamma}_{11}, \cdots, \hat{\gamma}_{1n})^\top$ is a $n \times 1$ vector.

4. **Update Smoothing Parameter** α: For each iteration, $\hat{\alpha}^{(k+1)}$ is determined by minimizing the AIC:

$$\hat{\alpha}^{(k+1)} = \underset{\alpha}{\text{argmin}}\ AIC(\alpha|\hat{\boldsymbol{\theta}}^{(k+1)}).$$

5. **Convergence:** Repeat Steps 2–4 until changes in parameters fall below the threshold Δ.

4.2 Model Selection and Degrees of Freedom

The smoothing parameter α in the penalized log-likelihood influences the trade-off between data fidelity and smoothness of the nonparametric component. To determine an optimal α, we use the Akaike Information

Criterion (AIC) by [1], defined as:

$$AIC(\alpha) = -2\ell_p(\hat{\boldsymbol{\theta}}, \alpha) + 2(p + m + df(\alpha)), \tag{4}$$

where $\ell_p(\cdot)$ denotes the penalized log-likelihood evaluated at the estimated parameters $\hat{\boldsymbol{\theta}}$ for a given α, and m counts the estimated parameters. The term $df(\alpha)$ represents the effective degrees of freedom associated with the smooth term. The AIC is a very effective tool for selecting the best model and finds the best model by minimizing the function given in (4), see [1] and [5] for more details. In this function, the α can be estimated by optimizing using a software program in searching α between 0.001 and 10^3. In the context of semiparametric modeling, the effective degrees of freedom can be interpreted as the equivalent number of parameters used to model the nonlinear structure [17, 19]. For the PLM-SLN model, $df(\alpha)$ is computed as:

$$\begin{aligned} df(\alpha) &= tr(\boldsymbol{N}(\boldsymbol{N}^\top\boldsymbol{N} + \alpha\hat{\sigma}^2\boldsymbol{K})^{-1}\boldsymbol{N}^\top), \\ &= tr(\boldsymbol{N}^\top\boldsymbol{N}(\boldsymbol{N}^\top\boldsymbol{N} + \alpha\hat{\sigma}^2\boldsymbol{K})^{-1}). \end{aligned}$$

Further, by following [12], let $\boldsymbol{B} = \boldsymbol{N}^\top\boldsymbol{N}$ and $\boldsymbol{L} = \hat{\sigma}^2\boldsymbol{B}^{-1/2}\boldsymbol{K}\boldsymbol{B}^{-1/2}$. Then, we form

$$\begin{aligned} df(\alpha) &= tr(\boldsymbol{B}^{1/2}(\boldsymbol{B}^{1/2}\boldsymbol{B}^{1/2} + \alpha\hat{\sigma}^2\boldsymbol{K})^{-1}\boldsymbol{B}^{1/2}), \\ &= tr((I_q + \alpha\boldsymbol{L})^{-1}) = \sum_{j=1}^{q}\frac{1}{1+\alpha l_j}, \end{aligned}$$

where I_q is the q-dimensional identity matrix and l_j are the eigenvalues of matrix $\boldsymbol{L}$.

4.3 The Hessian Matrix

The Hessian matrix plays a fundamental role in assessing local curvature of the penalized objective and is especially useful in local influence diagnostics. For the PLM-SLN model, the Hessian is derived by computing the second derivatives of the penalized complete-data objective function $Q_p(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}})$ with respect to the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \lambda, \boldsymbol{f}^\top)^\top$:

$$\ddot{Q}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

The following expressions outline the components of the Hessian matrix:

$$\begin{aligned} \frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} &= -\frac{1}{\sigma^2}\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}, \\ \frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2} &= -\frac{1}{\sigma^4}\boldsymbol{X}^\top\left(\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{\sigma\lambda\hat{\boldsymbol{\gamma}}_1}{2}\right), \\ \frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\lambda} &= \frac{1}{\sigma^2}\boldsymbol{X}^\top\left(2\lambda(\boldsymbol{y}-\boldsymbol{\mu}) - \sigma\hat{\boldsymbol{\gamma}}_1\right), \\ \frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{f}\partial\boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2}\boldsymbol{N}^\top\boldsymbol{W}\boldsymbol{X}, \end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{f} \partial \boldsymbol{f}^\top} &= -\frac{1}{\sigma^2}\boldsymbol{N}^\top \boldsymbol{W}\boldsymbol{N} - \alpha \boldsymbol{K}, \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{f} \partial \sigma^2} &= -\frac{1}{\sigma^4}\boldsymbol{N}^\top \left(\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{\sigma\lambda\hat{\boldsymbol{\gamma}}_1}{2}\right), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{f} \partial \lambda} &= \frac{1}{\sigma^2}\boldsymbol{N}^\top \left(2\lambda(\boldsymbol{y}-\boldsymbol{\mu}) - \sigma\hat{\boldsymbol{\gamma}}_1\right), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \lambda^2} &= -\frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu})^\top(\boldsymbol{y}-\boldsymbol{\mu}), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \lambda} &= \frac{1}{\sigma^4}\left(\lambda(\boldsymbol{y}-\boldsymbol{\mu})^\top(\boldsymbol{y}-\boldsymbol{\mu}) - (\boldsymbol{y}-\boldsymbol{\mu})^\top\frac{\hat{\boldsymbol{\gamma}}_1\sigma}{2}\right), \\
\frac{\partial^2 Q_p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \sigma^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(\boldsymbol{y}-\boldsymbol{\mu})^\top\hat{\boldsymbol{V}}(\boldsymbol{y}-\boldsymbol{\mu}).
\end{aligned}
$$

5 Applications

This section illustrates the practical performance of the proposed PLM-SLN model and ECM estimation algorithm through both simulation studies and a real-world dataset. The simulations aim to evaluate the accuracy and consistency of the parameter estimates, while the real data analysis highlights the method's applicability.

Computational Settings:

(i) All numerical experiments and data analyses were conducted using MATLAB R2017b.

(ii) Convergence for the ECM algorithm was achieved when parameter updates fell below a threshold of $\Delta = 10^{-6}$.

(iii) Three different sample sizes were considered: $n = 200$, 400, and 600.

(iv) Each simulation design was repeated independently $N = 500$ times.

(v) Estimation accuracy was evaluated using two metrics:

$$
\begin{aligned}
\widehat{\text{bias}}(\hat{\theta}) &= \overline{\theta} - \theta, \\
\widehat{\text{MSE}}(\hat{\theta}) &= \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2,
\end{aligned}
$$

where θ presents the true parameter, $\hat{\theta}_i$ is the estimate from the i-th run, and $\overline{\theta} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i$ is the average estimate across all replications.

To assess the statistical properties of the ECM-based estimation procedure, we consider the following simulation settings:

Scenario 1: The non-parametric function is defined as $f(t) = \cos(t)$ over the interval $t \in (0, 3\pi)$. The PLM-SLN model is expressed as:

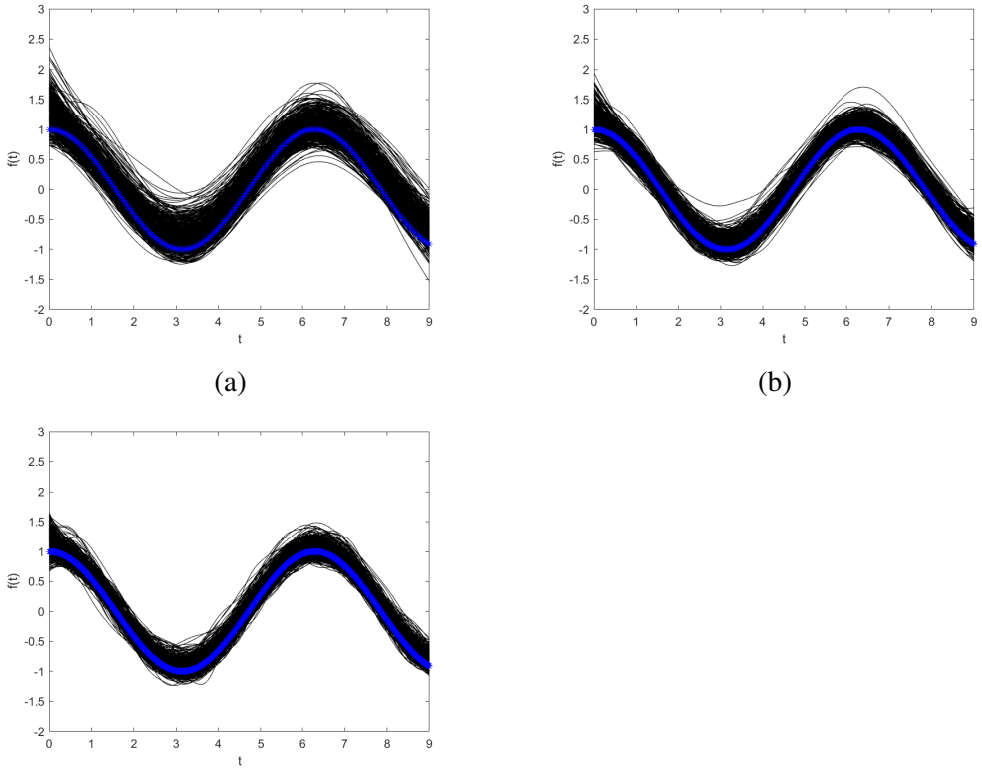
$$y_i = \beta x_i + f(t_i) + \epsilon_i,\ i = 1, \cdots, n,$$

with the true parameters set to $\beta = 5$, $\sigma^2 = 1$, and $\lambda = 3$. Covariates x_i are generated from the uniform distribution $U(0, 1)$.

Estimation results for Scenario 1 are presented in Table 1. As expected, the MSE values decrease as the sample size increases, indicating improved estimation accuracy. The proposed ECM-based estimation procedure demonstrates reliable performance across all parameters of the PLM-SLN model. Furthermore, Figure 1 displays the non-parametric function estimates based on 500 simulations. The plots show that with larger n, the average fitted curves converge more closely to the true underlying function, and variability among the fitted lines is reduced. This behavior indicates that the spline-based nonparametric estimator becomes more stable and accurate with more data.

Table 1: Bias and MSE of the parameter estimates for the PLM-SLN model (Scenario 1)

		$n = 200$		$n = 400$		$n = 600$	
Parameter	True Value	Bias	MSE	Bias	MSE	Bias	MSE
β	5	−0.1389	0.0621	−0.0852	0.0226	−0.0919	0.0185
σ^2	1	−0.0505	0.0113	−0.0257	0.0055	−0.0246	0.0038
λ	3	−0.2940	1.4945	−0.0198	0.7673	−0.0429	0.4516



(a) (b)

(c)

Figure 1: Estimated nonparametric functions from 500 simulations. The blue curve is the true function $f(t) = \cos(t)$, while the black curves represent individual estimates for (a) $n = 200$, (b) $n = 400$, and (c) $n = 600$

Scenario 2: In this setting, we adopt a Doppler-type nonparametric function defined by $f(t) = \cos(4\pi t)\exp(-t^2/2)$ over the domain $t \in (0.6, 1.6)$. The overall model structure remains identical to that in Scenario 1, with $\beta = 5$, $\sigma^2 = 1$, and $\lambda = 3$ as the true parameter values. The covariates x_i are again sampled independently from the uniform distribution $U(0, 1)$.

Table 2 reports the bias and MSE values obtained for each parameter across different sample sizes. The estimation accuracy improves as the sample size increases, consistent with the observations from the previous scenario. This further confirms the stability and efficiency of the ECM algorithm. Figure 2 presents the estimated non-parametric functions, where the reduction in variability with larger sample sizes is again evident, supporting the consistency of the estimator.

Table 2: Bias and MSE for PLM-SLN model under Scenario 2

		$n = 200$		$n = 400$		$n = 600$	
Parameter	True Value	Bias	MSE	Bias	MSE	Bias	MSE
β	5	−0.0866	0.0340	−0.0762	0.0206	−0.0666	0.0124
σ^2	1	−0.0148	0.0074	−0.0147	0.0059	−0.0144	0.0033
λ	3	0.1277	1.0455	0.1852	0.8811	0.1553	0.4660

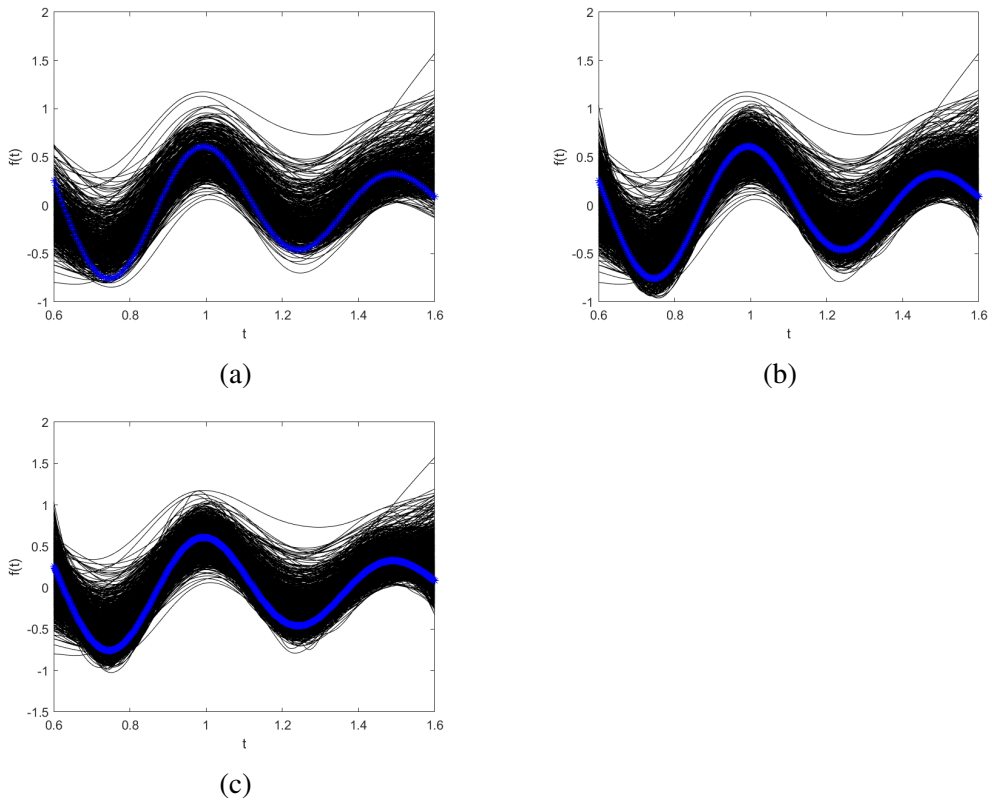


Figure 2: Estimated non-parametric functions for 500 simulations under Scenario 2. The blue line shows the true function $f(t)$, and the black lines are the fitted curves for (a) $n = 200$, (b) $n = 400$, and (c) $n = 600$

Scenario 3: To evaluate the performance of the proposed estimation method in a multivariate setting, we extend the model to include two explanatory variables. Specifically, we consider the following PLM-SLN:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{2i} + f(t_i) + \epsilon_i,\ i = 1, \cdots, n,$$

where covariates x_{i1} and x_{i2} are independently generated from $U(0, 1)$. The true regression coefficients are set to $\beta_1 = 1$ and $\beta_2 = -1$, and the nonparametric component is defined as $f(t) = \sin(2\pi t)$ over the interval $t \in (0, 1)$. The error term has the SLN distribution with parameters $\mu = 0, \sigma^2 = 1, \lambda = 3$.

Results averaged over 500 replications are presented in Table 3, and the corresponding nonparametric function estimates are illustrated in Figure 3 for increasing sample sizes. As expected, MSE values decrease as the sample size increases, indicating improved estimation accuracy and confirming the robustness of the proposed ECM-based estimator in a multivariate framework. Although the estimation of the skewness parameter λ exhibits slightly higher variability, the overall performance remains reliable. Moreover, the estimated nonparametric functions approach the true curve more closely and display reduced variability with larger samples, supporting the consistency and stability of the method.

Table 3: Bias and MSE of parameter estimates for the PLM-SLN model under Scenario 3

		$n = 200$		$n = 400$		$n = 600$	
Parameter	True Value	Bias	MSE	Bias	MSE	Bias	MSE
β_1	1	0.0038	0.0023	0.0021	0.0010	−0.0013	0.0006
β_2	−1	0.0021	0.0024	−0.0012	0.0011	0.0002	0.0006
σ^2	1	−0.0808	0.0175	−0.0224	0.0053	−0.0306	0.0042
λ	3	−0.5657	1.6077	0.0104	0.5976	−0.0290	0.4230

5.1 Real Data Example

We analyze a dataset consisting of 335 daily observations of ragweed pollen concentrations (Ambrosia spp.), recorded during the 1993 pollen season in Kalamazoo, Michigan. The data were originally published by [24]. Figure 4 provides visual summaries of the untransformed and square-root-transformed pollen counts, including histograms and a boxplot for the transformed series.

Based on these visualizations and descriptive statistics for the square-root-transformed values (including mean, median, standard deviation, skewness, and kurtosis), it is evident that the data exhibit skewness, heavy tails, and possible outliers. These features indicate that a distribution accounting for both skewness and heavy tails may be better suited for modeling this dataset.

Given the environmental nature of the study, we also examine the influence of several meteorological variables: rainfall (binary; 1 if ≥ 3 hours of steady or intense rainfall occurred, 0 otherwise), temperature (in degrees Fahrenheit), forecasted wind speed for the subsequent day (in mph), and time (expressed as the number of days since the start of the season).

Figure 5 presents scatter plots between the square-root-transformed pollen concentration and three predictors: temperature, wind speed, and time (seasonal day index). A pronounced nonlinear relationship between pollen levels and time is visible (Figure 5c), while the relationship with temperature is approximately linear (Figure 5a). Wind speed appears to exhibit little to no clear pattern (Figure 5b).

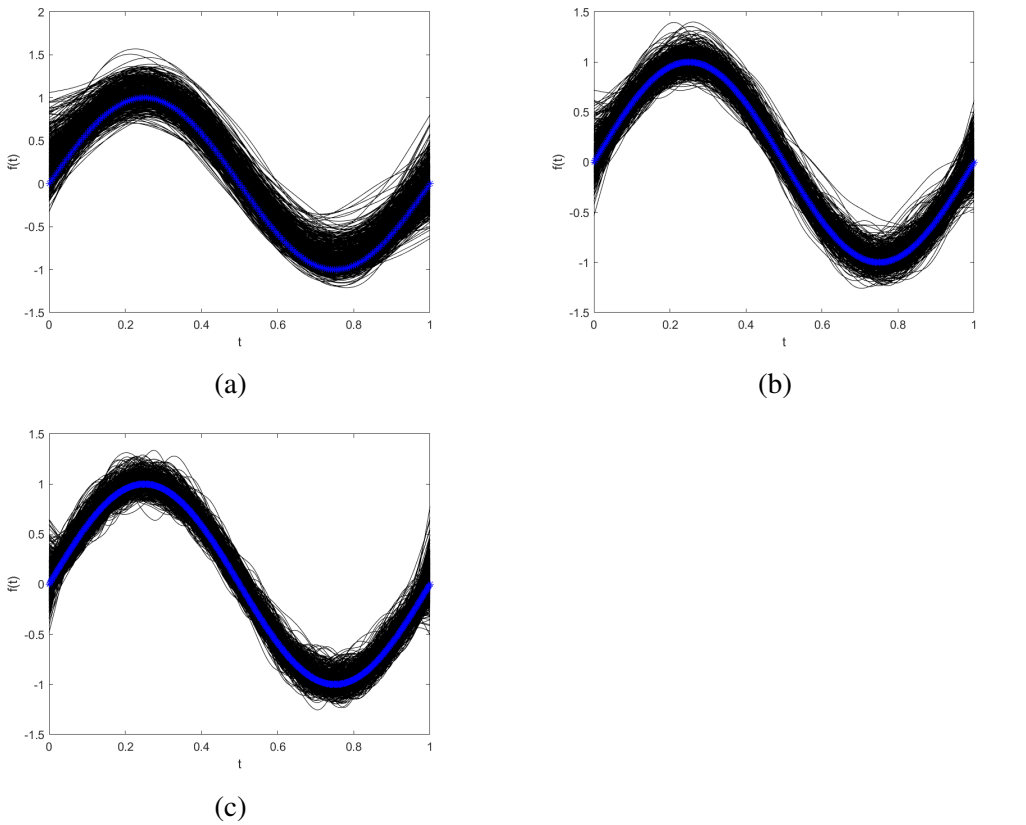


Figure 3: Estimated non-parametric functions for 500 simulations under Scenario 3. The blue line shows the true function $f(t)$, and the black lines are the fitted curves for (a) $n = 200$, (b) $n = 400$, and (c) $n = 600$

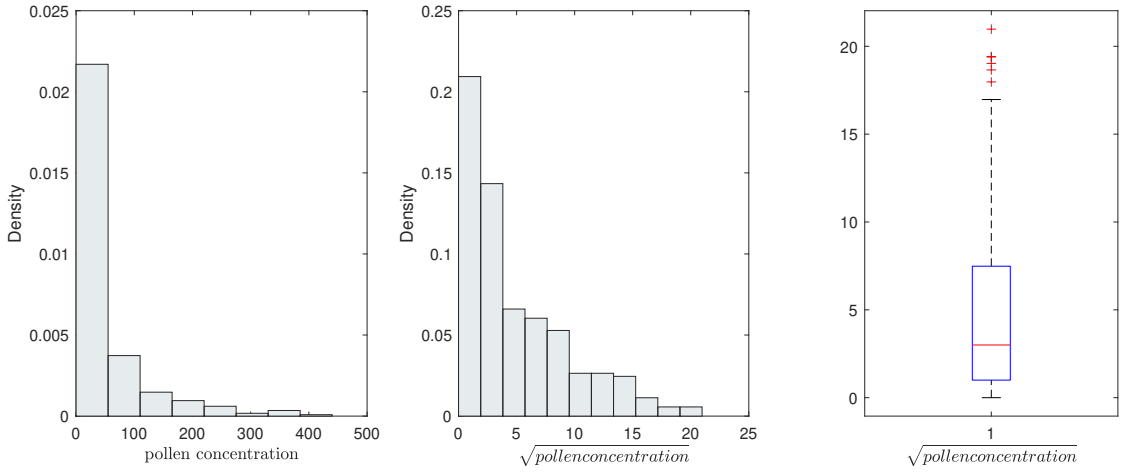


Figure 4: Histograms and boxplot for original and square-root-transformed pollen concentration values

This dataset has been previously explored in several studies, including [4, 11, 12, 23]. Notably, [12] analyzed this ragweed dataset under the PLM-SN model, while [11] applied a heteroscedastic PLM-SN framework. Both studies found that incorporating skewness significantly improved model fit compared to the

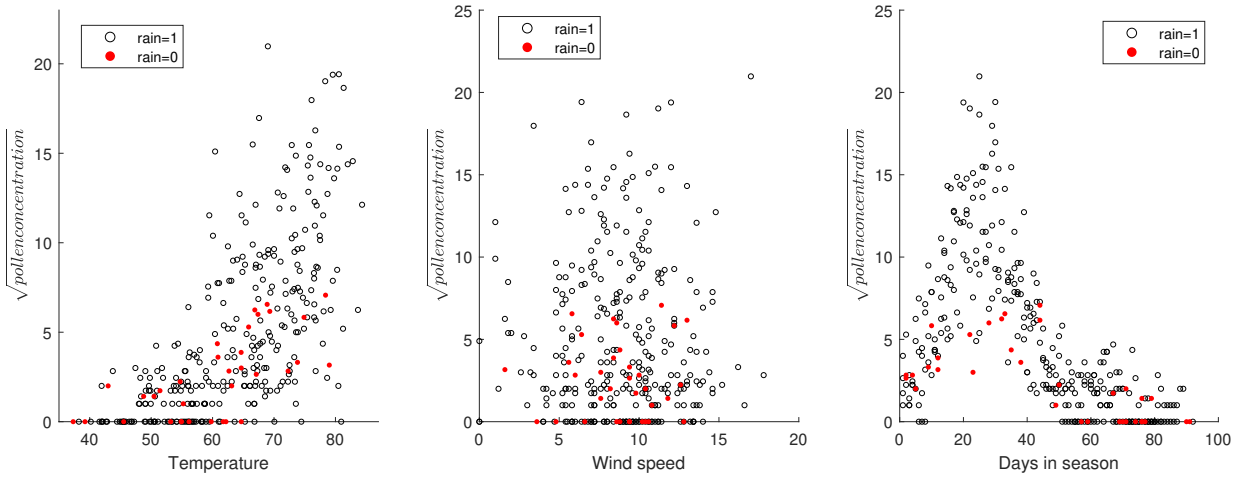


Figure 5: Scatter plots showing the square-root of pollen concentration vs. (a) temperature, (b) wind speed, and (c) time (days in season)

normal assumption. Extending this line of work, we propose the PLM-SLN model to simultaneously account for both skewness and heavy-tailedness. For benchmarking purposes, we also compare its performance against the standard PLM-N and PLM-SN models.

It has been proposed that the association between ragweed pollen levels and environmental factors such as rain, temperature, and wind speed can be modeled by incorporating a smooth function for the effect of time (days in the season) into the linear predictor. Based on preliminary findings, the following PLM structure is adopted:

$$\sqrt{y_i} = \beta_1 rain_i + \beta_2 temperature_i + \beta_3 wind_i + f(t_i) + \epsilon_i,\ i = 1, \cdots, 335,$$

where y_i denotes the pollen count for day i, t_i is the seasonal index, and ϵ_i follows an SLN distribution. A natural cubic spline is used to estimate $f(\cdot)$ flexibly.

Table 4 presents the parameter estimates, standard errors, log-likelihoods, and model comparison criteria such as AIC, Mean Integrated Squared Error (MISE), and the Kolmogorov–Smirnov (KS) test with corresponding p-values for the PLM-N, PLM-SN, and PLM-SLN models applied to the ragweed pollen dataset. The MISE criterion evaluates the accuracy of the estimated nonparametric function $\hat{f}(t)$ relative to the true function $f(t)$ and is defined as

$$\text{MISE} = \mathbb{E}\left[\int (\hat{f}(t) - f(t))^2\, dt\right] \approx \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(t_i) - f(t_i)\right)^2.$$

The KS-test assesses the goodness-of-fit of the residual distribution by measuring the maximum deviation between the empirical distribution function of the standardized residuals and the assumed cdf:

$$D = \sup_x |F_n(x) - F(x)|.$$

Lower MISE and KS values, along with higher p-values, indicate better model performance.

Among the three models, the PLM-SLN model achieves the highest log-likelihood and the lowest AIC,

indicating better overall fit. Additionally, it yields the lowest MISE value, reflecting improved estimation accuracy for the nonparametric component. To assess the goodness-of-fit of the error distribution, the KS-test is applied to the standardized residuals. The PLM-SLN model produces the smallest KS statistic and the largest p-value, confirming the adequacy of the SLN error assumption. These results suggest that PLM-SLN is the most appropriate model among those considered.

Table 4: Parameter estimates and model selection criteria for PLM-N, PLM-SN, and PLM-SLN applied to the ragweed dataset

	Normal		SN		SLN	
Effect	Estimate	SE	Estimate	SE	Estimate	SE
Rain	1.411	0.392	1.456	0.382	1.059	0.117
Temperature	0.101	0.018	0.088	0.018	0.079	0.004
Wind	0.237	0.038	0.228	0.037	0.188	0.011
σ^2	4.511	0.351	9.760	1.362	2.822	0.249
λ	-	-	2.189	0.448	0.451	0.043
α	141.622		211.486		216.89	
$df(\alpha)$	9.838		11.234		17.670	
Comparison Criteria						
$\ell_p(\hat{\boldsymbol{\theta}}, \alpha)$	−734.212		−730.822		**−719.266**	
AIC	1497.100		1494.100		**1485.874**	
MISE	2.747		2.057		**1.868**	
KS-test	0.206		0.050		**0.026**	
p-value	1.013e-06		0.771		**0.999**	

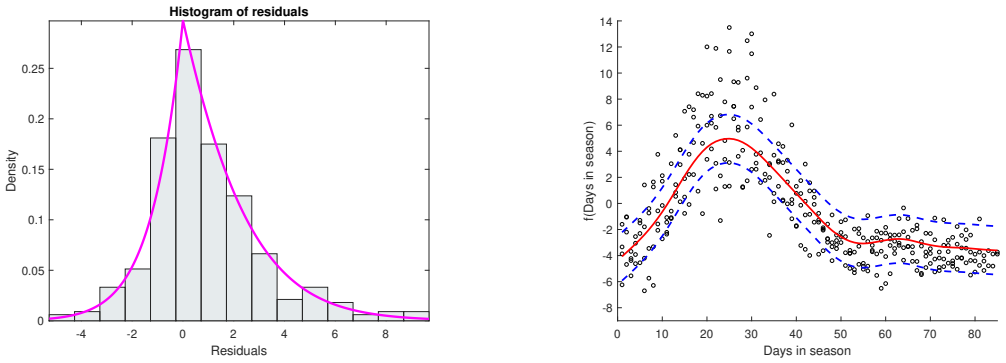
Figure 6 illustrates the model's adequacy: Panel (a) displays a histogram of residuals from the PLM-SLN model overlaid with the fitted SLN density, while Panel (b) shows the estimated smooth function $f(t)$ with corresponding 95% pointwise confidence bands. The results confirm that the PLM-SLN model effectively captures seasonal dynamics, particularly toward the later part of the season. Additionally, the model yields higher estimated degrees of freedom than PLM-N and PLM-SN, reinforcing its greater flexibility in modeling complex data structures.

6 Conclusion

In this study, we investigated parameter estimation in PLMs under the assumption that the error terms follow an SLN distribution, a flexible alternative to the SN distribution that accommodates both skewness and heavy tails, while maintaining the same number of parameters.

We introduced an ECM algorithm to compute the ML estimates and demonstrated its effectiveness through a simulation study. The proposed method was then applied to real-world data on ragweed pollen concentration. The results indicated that the PLM-SLN model provided a superior fit compared to both PLM-N and PLM-SN models, as evaluated by the AIC criterion, MISE, and KS-test statistic.

These findings highlight the usefulness of the SLN distribution in capturing the complex features, such as asymmetry, heavy tails, and non-linear trends in real data, and demonstrate the applicability and flexibility of the proposed PLM-SLN approach in practical scenarios.



(a) (b)

Figure 6: (a) Residual histogram from PLM-SLN with fitted SLN density curve; (b) Estimated function f (Days in season) with 95% pointwise confidence intervals

A promising direction for future research is the development of local influence diagnostics under the SLN error assumption, which would allow for the assessment of robustness against perturbations in the data or model structure. Furthermore, the current model can be extended to account for heteroscedasticity by incorporating a scale function, resulting in a heteroscedastic PLM with SLN errors. Such an extension would enhance the model's flexibility in handling non-constant error variances and further improve estimation accuracy.

Author Contribution Statements

F. Z. D. and O. A. contributed to the conceptualization, methodology, writing, and supervision. F. Z. D. was responsible for investigation and software development. Both authors contributed to the visualization, review, and editing.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data Availability Statement

This manuscript has no associated data.

Funding Source Declaration

No funding was received for conducting this study.

Ethical Standard

This article does not contain any studies with human participants or animals performed by the author.

Declaration of Generative Artificial Intelligence Use

The author declares that no generative Artificial Intelligence tools were used in the preparation of this manuscript.

Acknowledgements

The authors would like to thank the associate editor and the anonymous reviewers for their valuable comments and suggestions, which have significantly improved the quality of the manuscript.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Caski, F., editors, *Proceeding of the Second International Symposium on Information Theory*, pages 267–281. Akademiai Kiado.

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

[3] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistics*, 46(2):199–208.

[4] Bazrafkan, M., Zare, K., Maleki, M., and Khodadi, Z. (2022). Partially linear models based on heavy-tailed and asymmetrical distributions. *Stochastic Environmental Research and Risk Assessment*, 36(5):1243–1253.

[5] Burnham, K. and Anderson, D. (2002). *Model Selection and Multi-Model Inference*. Springer, New York.

[6] DiCiccio, T. J. and Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association*, 99(466):439–450.

[7] Doğru, F. Z. and Arslan, O. (2017). Parameter estimation for mixtures of skew Laplace normal distributions and application in mixture regression modeling. *Communications in Statistics – Theory and Methods*, 46(21):10879–10896.

[8] Doğru, F. Z. and Arslan, O. (2019). Joint modelling of the location, scale and skewness parameters of the skew Laplace normal distribution. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(3):1249–1257.

[9] Doğru, F. Z. and Arslan, O. (2021). Finite mixtures of skew Laplace normal distributions with random skewness. *Computational Statistics*, 36(1):423–447.

[10] Doğru, F. Z., Yu, K., and Arslan, O. (2019). Heteroscedastic and heavy-tailed regression with mixtures of skew Laplace normal distributions. *Journal of Statistical Computation and Simulation*, 89(17):3213–3240.

[11] Ferreira, C. S., Borelli Zeller, C., and de Oliveira Garcia, R. R. (2023). Heteroscedastic partially linear model under skew-normal distribution with application in ragweed pollen concentration. *Journal of Applied Statistics*, 50(6):1255–1282.

[12] Ferreira, C. S. and Paula, G. A. (2017). Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics*, 44(16):3033–3053.

[13] Ferreira, C. S., Paula, G. A., and Lana, G. C. (2022). Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors. *Computational Statistics*, 37(1):445–468.

[14] Gómez, H. W., Venegas, O., and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, 18(4):395–407.

[15] Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, pages 245–259.

[16] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, Boca Raton.

[17] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Monographs on Statistics & Applied Probability. Chapman and Hall/CRC.

[18] Ibacache-Pulgar, G. and Paula, G. A. (2011). Local influence for student t partially linear models. *Computational Statistics & Data Analysis*, 55(3):1462–1478.

[19] Ibacache-Pulgar, G., Paula, G. A., and Cysneiros, F. J. A. (2013). Semiparametric additive models under symmetric distributions. *TEST*, 22(1):103–121.

[20] Maleki, M. and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics – Theory and Methods*, 46(24):12356–12369.

[21] Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278.

[22] Relvas, C. E. M. and Paula, G. A. (2016). Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers*, 57(3):795–825.

[23] Ruppert, D., Wand, M. P., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

[24] Stark, P. C., Ryan, L. M., McDonald, J. L., and Burge, H. A. (1997). Using meteorologic data to predict daily ragweed pollen levels. *Aerobiologia*, 13(3):177–184.

[25] Vanegas, L. H. and Paula, G. A. (2015). A semiparametric approach for joint modeling of median and skewness. *TEST*, 24:110–135.

[26] Zhu, H.-T. and Lee, S.-Y. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):111–126.