AIMS Mathematics, 10(7): 15698–15719.
DOI:10.3934/math.2025703
Received: 08 April 2025
Revised: 26 June 2025
Accepted: 02 July 2025
Published: 09 July 2025

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

Research article

Estimation and diagnostic for a skewed generalized normal partially linear models

Xue Wang[1,2], **Weihu Cheng**[1,*] **and Clécio S. Ferreira**[3]

[1] School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China
[2] College of Science, Qiqihar University, Qiqihar 161006, China
[3] Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, Brazil

* **Correspondence:** Email: chengweihu@bjut.edu.cn.

Abstract: Partially linear models (PLMs) are widely employed in scientific research to analyze hybrid parametric-nonparametric relationships. However, their conventional reliance on symmetric error distributions severely limits applicability to real-world phenomena characterized by pronounced asymmetry and heavy-tailed behavior. To address this gap, we propose a novel PLM framework incorporating skewed generalized normal (SGN) distributed errors, which simultaneously accommodates extreme skewness and heavy-tailed attributes beyond the capabilities of symmetric or skew-normal (SN) specifications. Methodologically, we develop a penalized expectation-maximization (EM) algorithm with provable convergence guarantees and integrated adaptive smoothing selection, effectively resolving optimization instability in high-dimensional settings. Furthermore, we establish a unified diagnostic system that synergizes geometric leverage calculus with local influence analysis to systematically evaluate model robustness against perturbations and outliers. Extensive simulation studies demonstrate the framework's superior estimation accuracy compared to conventional symmetric and SN-based alternatives. Empirical validations based on real-world datasets reveal statistically significant improvements in model fit while maintaining interpretability.

Keywords: skewed generalized normal distribution; partially linear model; EM algorithm; local influence; generalized leverage
Mathematics Subject Classification: 62J02, 62J20

1. Introduction

Partially linear models (PLMs) have emerged as powerful statistical tools due to their unique capability to simultaneously capture linear relationships and nonlinear effects of covariates. This dual

flexibility has driven their widespread adoption across disciplines such as econometrics, biostatistics, and physical sciences. Over the past three decades, substantial theoretical and methodological advancements have been achieved in PLM research (e.g., [1–3]). Nevertheless, a critical limitation persists in conventional PLM frameworks: the predominant error assumption of symmetric error distributions. This assumption proves particularly restrictive when analyzing real-world datasets where response variables frequently exhibit asymmetric patterns. To address this issue, recent methodological developments have focused on integrating skewed error distributions into regression frameworks. A seminal advancement in this direction was made by Cancho et al. [4], who incorporated skewed errors in nonlinear regression models using the skew-normal distribution proposed by Sahu et al. [1]. Subsequent work by Ferreira and Paula [5] extended PLMs to accommodate SN-distributed errors following the parameterization of [6], establishing rigorous methodologies for parameter estimation and diagnostic procedures. Parallel developments by Lyu and Feng [7] demonstrated that generalized linear models with SN-distributed errors exhibit superior adaptability to asymmetric data structures. Further methodological innovations emerged through the work of De Freitas et al. [8], who developed additive PLMs based on a scale mixture of SN distributions. This framework provided enhanced flexibility in modeling datasets with simultaneous skewness and heavy-tailedness. Most recently, Ferreira and Dias [9] proposed a generalized PLM framework incorporating skewed scale mixtures of normal (SSMN) distributions, thereby establishing a unified approach for handling complex error structures with multiple deviation patterns.

Despite these advancements, the SN distribution suffers from inherent parametric constraints. As rigorously demonstrated by Azzalini [10], its theoretical skewness ($-0.995< \gamma_1 <0.995$) and kurtosis ($0< \gamma_2 <0.869$) impose critical limitations on PLMs. Specifically, SN-based models cannot adequately capture responses with substantial asymmetry ($|\gamma_1| > 1$) or heavy-tailedness ($\gamma_2 > 1$). This mismatch necessitates error distributions with extended skewness-kurtosis domains, making the integration of more flexible families into PLMs a crucial research priority.

Diagnostic analysis constitutes a fundamental component of PLM research, providing essential tools for evaluating model robustness and detecting influential observations post-estimation. The methodological foundation was established by Cook [11] through case-deletion diagnostics, later extended by Cook [12] via local influence analysis—a framework assessing parameter sensitivity to data perturbations. Within PLM-specific contexts, Kim et al. [13] systematically developed influence diagnostics for individual parameter estimators, emphasizing the necessity of component-wise evaluation. Subsequent advancements addressed diverse error structures; Galea et al. [14] established diagnostics for nonlinear models with symmetric light/heavy-tailed errors, while Lemus et al. [15] extended these procedures to censored PLMs using scale mixtures of normal (SMN) distributions, demonstrating the persistent effectiveness of case-deletion and local influence methods against outliers. Recent innovations include Ferreira et al.'s [16] development of conditional expectation-based local influence for PLMs with autoregressive SN errors and Ibacache-Pulgar et al.'s [17] adaptation of classical measures (residuals, Cook's distance) for nonparametric regression with symmetric errors. Collectively, these works confirm diagnostic methods' critical role in maintaining PLM reliability across varying error structures.

The paper is organized as follows: Section 2 introduces the PLM with SGN errors and establishes its penalized log-likelihood function. Section 3 develops an EM-type estimation algorithm, including algorithm steps and derivation of effective degrees of freedom for the nonparametric component.

Section 4 proposes a diagnostic framework combining local influence analysis with generalized leverage measures. Section 5 provides simulation studies and empirical applications with systematic comparisons to existing SN-distributed PLMs. Final methodological insights and extensions are discussed in Section 6.

2. Skewed generalized normal PLMs

In this section, we formalize the PLM framework under SGN errors and establish a robust parameter estimation framework via penalized maximum likelihood optimization.

2.1. The skewed generalized normal distribution

Definition 1. *A random variable X follows the skewed generalized normal distribution with location parameter μ, scale parameter σ, shape parameter α, and skew parameter r if the probability density function (pdf) is given by*

$$f(x;\mu,\sigma^2,\alpha,r) = \frac{1}{2\sigma\alpha^{1/\alpha-1}\Gamma(1/\alpha)} \exp\left\{-\frac{|x-\mu|^\alpha}{\alpha\sigma^\alpha[1+r*\text{sign}(x-\mu)]^\alpha}\right\}, \tag{2.1}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > 0$, and $|r| \leq 1$. $\Gamma(\cdot)$ denotes the gamma function, and sign$(\cdot)$ *is symbolic function. We shall write $X \sim \text{SGN}(\mu,\sigma^2,\alpha,r)$.*

The adoption of the SGN distribution is motivated by its superior capacity to characterize extended skewness regimes compared to the classical SN distribution proposed by Azzalini [6]. Crucially, as demonstrated in Table 1, the SGN's shape parameters α and r enable systematic generation of distributional configurations, achieving substantially broader skewness and kurtosis ranges than those attainable under SN specifications. This parametric flexibility is analytically preserved through our proposed estimation procedure, ensuring consistent modeling of tail behaviors and asymmetry patterns.

Table 1. Range of Skewness and Kurtosis of SGN.

α	**Skewness range**	**Kurtosis range**
0.1	$(-5.11\times10^2,\ 5.11\times10^2)$	$(2.82\times10^6,\ 3.55\times10^6)$
1	$(-2.0000,\ 2.0000)$	$(3.0000,\ 6.0000)$
2	$(-0.9953,\ 0.9953)$	$(0.0000,\ 0.8691)$
2.5	$(-0.7808,\ 0.7808)$	$(-0.3688,\ 0.2157)$
4	$(-0.4428,\ 0.4428)$	$(-0.8115,\ -0.5553)$
10	$(-0.0115,\ 0.0115)$	$(-1.1157,\ -1.0576)$
20	$(-0.0350,\ 0.0350)$	$(-1.1755,\ -1.1563)$

Theorem 1. *If $X \sim \text{SGN}(\mu,\sigma^2,\alpha,r)$, $Y \sim \text{Gamma}(1/\alpha,1)$ and W is a variable with the probability mass function as follows:*

$$P\{W=r-1\} = \frac{1-r}{2}, \quad P\{W=r+1\} = \frac{1+r}{2},$$

then

$$X \overset{d}{=} \mu + \alpha^{1/\alpha}\sigma W Y^{1/\alpha}, \tag{2.2}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > 0$, *and* $|r| \leq 1$.

Proposition 2.1. *If* $X \sim \text{SGN}(\mu, \sigma^2, \alpha, r)$, *then*

(1) The mean of X is given by

$$E(X) = \mu + 2\sigma r \alpha^{1/\alpha} \frac{\Gamma(2/\alpha)}{\Gamma(1/\alpha)}. \tag{2.3}$$

(2) The variance of X is given by

$$Var(X) = \frac{\alpha^{2/\alpha}\sigma^2}{\Gamma^2(1/\alpha)}\Big[(1 + 3r^2)\Gamma(1/\alpha)\Gamma(3/\alpha) - 4r^2\Gamma^2(2/\alpha)\Big]. \tag{2.4}$$

2.2. The model

We introduce a novel PLM with SGN errors, defined as

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + g\,(t_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{2.5}$$

where y_i denotes the scalar response, $\mathbf{x}_i \in \mathbb{R}^p$ the covariate vector, $\boldsymbol{\beta} \in \mathbb{R}^p$ the regression coefficients, and $t_i \in \mathbb{R}$ the smoothing variable. The random errors ϵ_i are independent and identically distributed (i.i.d.) as $\text{SGN}\left(0, \sigma^2, \alpha, r\right)$ for $i = 1, \cdots, n$. The nonparametric component $g(\cdot)$ is approximated through B-spline basis expansion following the methodology of He et al. [18]. Let $\mathbf{N}(t) = (N_1(t), \cdots, N_q(t))^\top$ denote a set of normalized B-spline basis functions of order l, where $q = k_n + l + 1$ with k_n interior knots. The B-spline approximation has become a prevalent technique in PLM estimation due to three principal advantages: First, it replaces point-wise estimation with global functional approximation, significantly reducing computational complexity while ensuring consistency. Second, the method inherently avoids boundary effects through its compact support property. Third, polynomial exactness allows precise recovery of underlying trends across the entire domain. These characteristics collectively enhance both theoretical tractability and practical performance in empirical applications. Generally, cubic B-splines ($l = 3$) are the most used, taking into account flexibility and a low number of parameters. Also, we use the Eilers and Marx's method [19], which consists of k_n equally spaced points. The choice of the number of internal nodes is usually empirical, based on analyzing the scatter plot between the response and the explanatory variables. For example, three nodes at each curve of the data are enough to capture all the information of the curve. For a complete construction of B-spline basis, see [16]. Then, $g(t_i)$ can be approximated by

$$g(t_i) \approx \mathbf{n}_i^\top \mathbf{f}, \quad i = 1, \cdots, n,$$

where $\mathbf{n}_i = \mathbf{N}(t_i) = (N_1(t_i), \cdots, N_q(t_i))^\top$ and $\mathbf{f} = (f_1, \cdots, f_q)^\top$ represent the spline coefficients which we need to estimate. Rewriting the PLM-SGN model in (2.5) into matrix form yields

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon}, \tag{2.6}$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ is a $n \times 1$ vector of observed responses, $\mathbf{X}$ is a $n \times p$ design matrix with rows $\mathbf{x}_i^\top$, $\mathbf{f} = \left(f_1, \cdots, f_q\right)^\top$ is a $q \times 1$ spline coefficient vector, $\mathbf{N}$ is an $n \times q$ basis matrix with rows $\mathbf{n}_i^\top$, and $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^\top$ is an $n \times 1$ vector of random errors.

2.3. Penalized log-likelihood function

According to Definition 1, it follows that $Y_i \sim \text{SGN}(\mu_i, \sigma^2, \alpha, r)$ with $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{n}_i^\top \mathbf{f}$. The observed-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \alpha, r, \mathbf{f}^\top)$ is given by

$$\ell(\boldsymbol{\theta}) = -n\Big(\ln 2 + \frac{\ln \sigma^2}{2} + \frac{1-\alpha}{\alpha}\ln\alpha + \ln\Gamma(1/\alpha)\Big) - \frac{1}{\alpha\sigma^\alpha}\sum_{i=1}^{n}\frac{|y_i - \mu_i|^\alpha}{[1 + r * \text{sign}(y_i - \mu_i)]^\alpha}. \tag{2.7}$$

Two primary challenges emerge in this estimation framework. First, direct maximization of the log-likelihood in Eq (2.7) becomes computationally intractable due to the non-differentiable absolute value term, which induces non-convexity in the optimization landscape. Second, unconstrained maximization of Eq (2.7) may result in ill-conditioned matrices during iterations, jeopardizing numerical stability. To resolve these issues, we employ a penalized likelihood approach by introducing a regularization term:

$$\ell_p(\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}) - \frac{\lambda}{2}J(\mathbf{f}), \tag{2.8}$$

where $J(\mathbf{f})$ represents the penalty function enforcing smoothness constraints, and $\lambda > 0$ serves as the smoothing parameter that calibrates the trade-off between model fidelity and functional regularity. This regularization scheme simultaneously mitigates overfitting and guarantees stable parameter identifiability.

3. Parameter estimation

Accurate parameter estimation constitutes a critical prerequisite for ensuring the broad applicability of the proposed model in practical settings. To achieve this, we develop a rigorous estimation framework for the PLM-SGN model by leveraging the global convergence guarantees of the EM algorithm. Furthermore, we derive two key statistical quantities: (i) the effective degrees of freedom for the nonparametric component and (ii) the observed Fisher information matrix for the parametric vector $\boldsymbol{\theta}$. These derivations provide essential theoretical foundations for assessing estimation uncertainty and model complexity, while simultaneously enhancing the framework's operational robustness in empirical applications.

3.1. EM algorithm

In this part, we will apply the EM algorithm to obtain parameter estimation of the PLM-SGN model. The auxiliary distribution was proposed by Devroye [20] and described by Definition 2.

Definition 2. *Let W denote a polynomially tilted α-stable random variable with tilting parameter equal to $1/2$. The probability density function (pdf) of W is given by*

$$f_W(w) = \frac{\sqrt{\pi}}{2\sqrt{w}\Gamma(1 + 1/\alpha)}f_P(w), \tag{3.1}$$

for $w > 0$, where $f_P(\cdot)$ denotes the pdf of $P = S\left(\alpha/2, 1, [\cos(\pi\alpha/4)]^{2/\alpha}, 0\right)$ and the $S(\alpha, \beta, \sigma, \mu)$ denotes an α-stable random variable with parameters α, β, σ and μ.

The characteristic function (chf) of $S(\alpha,\beta,\sigma,\mu)$ is given by Nolan [21]:

$$\varphi(t)=\begin{cases}\exp\left\{-|\sigma t|^{\alpha}\left[1-i\beta\,\mathrm{sign}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right]+it\mu\right\}, & \text{if } \alpha\neq 1,\\ \exp\left\{-|\sigma t|\left[1+i\beta\,\mathrm{sign}(t)\frac{2}{\pi}\ln|t|\right]+it\mu\right\}, & \text{if } \alpha=1,\end{cases}$$

where $i^2=-1$ and sign$(\cdot)$ denotes the sign function. Then, $W\sim PT(\alpha)$ denotes a polynomially tilted α-stable random variable with tail thickness $\alpha\in(0,2]$.

In this context, let us denote the observed data as $\boldsymbol{y}=(y_1,\cdots,y_n)$, while the missing data is represented by $\boldsymbol{w}=(w_1,\cdots,w_n)$. The conditional distribution of Y given by W is directly constructed using Theorem 2.

Theorem 2. *Let Y denote a SGN random variable with pdf given by Eq* (2.1)*. We have*

$$Y\overset{d}{=}\mu+\sigma\alpha^{1/\alpha}\frac{X}{\sqrt{2W}}, \tag{3.2}$$

$$f_{Y|W}(y)=\frac{\sqrt{w}}{\sqrt{\pi}\sigma\alpha^{1/\alpha}}\exp\left\{-\left[\frac{y-\mu}{\alpha^{1/\alpha}\sigma(1+r*\mathrm{sign}(y-\mu))}\right]^2 w\right\}, \tag{3.3}$$

where the pdf of random variable X is given by

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2(1+r*\mathrm{sign}(x))^2}\right\},$$

where the pdf of W is Eq (3.1)*.*

The complete data likelihood function can be expressed as

$$\ell_c(\boldsymbol{\theta})=\prod_{i=1}^{n}f_c\left(y_i,w_i\mid\boldsymbol{\theta}\right)=\prod_{i=1}^{n}f_W\left(w_i\right)f_{Y|W}\left(y_i\right), \tag{3.4}$$

then,

$$\begin{aligned}\ell_c(\boldsymbol{\theta})&=\sum_{i=1}^{n}\ln f_W\left(w_i\right)+\sum_{i=1}^{n}\ln f_{Y|W}\left(y_i\right)\\&=\mathrm{C}+\sum_{i=1}^{n}\ln f_W\left(w_i\right)-\frac{n}{2}\ln\sigma^2-\frac{n}{\alpha}\ln\alpha-\sum_{i=1}^{n}\left\{\frac{y_i-\mu_i}{\alpha^{1/\alpha}\sigma\left[1+r*\mathrm{sign}\left(y_i-\mu_i\right)\right]}\right\}^2 w_i,\end{aligned} \tag{3.5}$$

where C is a constant independent of $\boldsymbol{\theta}$. Given an estimate of $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}^{(t)}$, the current $(t+1)$th iteration of the EM algorithm is as follows:

$$Q\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}^{(t)}\right)=E\left[l_c(\boldsymbol{\theta})\mid\boldsymbol{y},\boldsymbol{\theta}^{(t)}\right]=\mathrm{C}+\sum_{i=1}^{n}E\left[\ln f_W(w_i)\mid y_i,\boldsymbol{\theta}^{(t)}\right]-\frac{n}{2}\ln\sigma^2-\frac{n}{\alpha}\ln\alpha-\sum_{i=1}^{n}\left\{\frac{y_i-\mu_i}{\alpha^{1/\alpha}\sigma[1+r*\mathrm{sign}(y_i-\mu_i)]}\right\}^2 E_i^{(t)},$$

where

$$E_i^{(t)}=E\left[W_i\mid y_i,\boldsymbol{\theta}^{(t)}\right]=\frac{(\alpha^{(t)})^{2/\alpha^{(t)}}}{2}\left\{\frac{|y_i-\mu_i^{(t)}|}{\sigma^{(t)}\left[1+r^{(t)}*\mathrm{sign}\left(y_i-\mu_i^{(t)}\right)\right]}\right\}^{\alpha^{(t)}-2}. \tag{3.6}$$

According to Green [22], maximizing Eq (2.8) is equivalent to optimizing the following expression

$$Q_p(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) = Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) - \frac{\lambda}{2} J(\mathbf{f}). \tag{3.7}$$

Based on Eilers and Marx [19], $J(\mathbf{f}) = \mathbf{f}^\top \mathbf{K} \mathbf{f}$, where $\mathbf{K} \in \mathbb{R}^{q \times q}$ is a non-negative definite matrix that depends only on the second-order difference of the $n \times q$ identity matrix. A complete expression of $\mathbf{K}$ may be found in [19]. Updating $\boldsymbol{\theta}^{(t)}$ by the maximum of $Q_p\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right)$ with respect to $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\beta}}^{(t+1)} = (\mathbf{X}^\top \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{D} (\boldsymbol{y} - \mathbf{N} \hat{\mathbf{f}}^{(t)}), \tag{3.8}$$

$$\hat{\mathbf{f}}^{(t+1)} = (\mathbf{N}^\top \mathbf{D} \mathbf{N} + \lambda / 2 (\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}} \hat{\sigma}^{(t)} \mathbf{K})^{-1} \mathbf{N}^\top \mathbf{D} (\boldsymbol{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(t)}), \tag{3.9}$$

$$\widehat{\sigma^2}^{(t+1)} = \frac{2}{n (\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}}} (\boldsymbol{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(t)} - \mathbf{N} \hat{\mathbf{f}}^{(t)})^\top \mathbf{D} (\boldsymbol{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(t)} - \mathbf{N} \hat{\mathbf{f}}^{(t)}), \tag{3.10}$$

where $\mathbf{D} = \Lambda\left(\dfrac{\boldsymbol{E}^{(t)}}{[1 + r^{(t)} * \operatorname{sign}(\boldsymbol{y} - \boldsymbol{\mu}^{(t)})]^2}\right)$ and $\Lambda(\mathbf{x})$ represents the diagonal matrix of the vector $\mathbf{x}$. A comprehensive computation scheme with detailed information can be found in Appendix A.

Considering r, we apply the *uniroot*$(\cdot)$ in the R environment to solve following the equation:

$$h(\hat{r}) = \sum_{i=1}^{n} \frac{\left(y_i - \mu_i^{(t)}\right)^2 E_i^{(t)}}{\left[1 + r * \operatorname{sign}\left(y_i - \mu_i^{(t)}\right)\right]^3} \operatorname{sign}(y_i - \mu_i^{(t)}) = 0. \tag{3.11}$$

Finally, we update α by maximizing the following log-likelihood function:

$$\hat{\alpha}^{(t+1)} = \arg\max_{\alpha} \sum_{i=1}^{n} \ln f_Y\left(y_i \mid \hat{\boldsymbol{\beta}}^{(t+1)}, \hat{\mathbf{f}}^{(t+1)}, \widehat{\sigma^2}^{(t+1)}, \hat{r}^{(t+1)}\right). \tag{3.12}$$

We set the initial values

$$\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{y}, \tag{3.13}$$

$$\widehat{\sigma^2}^{(0)} = \frac{\|\boldsymbol{y} - \hat{\boldsymbol{\beta}}^{(0)}\|^2}{n - \operatorname{rank}(\mathbf{X})}, \tag{3.14}$$

$$\hat{\mathbf{f}}^{(0)} = (\mathbf{N}^\top \mathbf{N} + \lambda \widehat{\sigma^2}^{(0)} \mathbf{K})^{-1} \mathbf{N}^\top (\boldsymbol{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(0)}), \tag{3.15}$$

$$\hat{r}^{(0)} = \frac{Q(0.8) - 2Q(0.5) + Q(0.2)}{Q(0.8) - Q(0.2)}, \tag{3.16}$$

where $Q(p)$ denotes the p-th sample quantile of $\boldsymbol{y}$ and $\alpha^{(0)}$ is the root of the following equation:

$$h(\alpha) = n \frac{\sum_{i=1}^{n} (y_i - \bar{y})^4}{\left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right]^2} - \frac{\Gamma(5/\alpha) \Gamma^2(1/\alpha)}{\Gamma(3/\alpha)} = 0. \tag{3.17}$$

The algorithm is iterated until a convergence criterion is satisfied, such as when the norm of the difference between successive parameter estimates, $\left\|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\right\|$, falls below a predefined threshold

(e.g., 10^{-5}) or when the number of iterations reaches a maximum limit, such as 5000. The Algorithm 1 is the pseudocode in R for the algorithm.

Algorithm 1 The estimation algorithm for solving the MLE of the model (2.5).

Initialization: Set initial values: $t=0, cont=0$, $(\hat{\boldsymbol{\beta}}^{(0)}, \widehat{\sigma^2}^{(0)}, \hat{\alpha}^{(0)}, \hat{r}^{(0)}, \hat{\mathbf{f}}^{(0)})$ according to Eqs (3.13)–(3.17)
repeat
computing conditional expectation $E^{(t)}$,
updating $\boldsymbol{\theta}^{(t+1)} = (\hat{\boldsymbol{\beta}}^{(t+1)}, \widehat{\sigma^2}^{(t+1)}, \hat{\alpha}^{(t+1)}, \hat{r}^{(t+1)}, \hat{\mathbf{f}}^{(t+1)})$ according to (3.8)–(3.12),
criterio = $\|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\|$,
cont = cont + 1,
until criterio $\leq 10^{-5}$ or cont > 5000.

3.2. Effective degrees of freedom

The effective degrees of freedom (EDF) are closely tied to the penalty parameter λ and play a crucial role in balancing the trade-off between minimizing the fitting errors and avoiding excessive model complexity. In this framework, λ is selected by minimizing the Akaike Information Criterion (AIC). However, since the nonparametric component $g(t)$ represents an infinite-dimensional smooth function, its explicit parameter count remains undefined. We therefore employ EDF to approximate the effective number of free parameters in $g(t)$, enabling rigorous AIC computation for simultaneous model selection and regularization.

$$\begin{aligned} \mathrm{d}g(\lambda) &= \mathrm{tr}\Big[\mathbf{N}\Big(\mathbf{N}^\top\mathbf{D}\mathbf{N} + \lambda/2(\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}}\Big(\widehat{\sigma^2}^{(t)}\Big)^{1/2}\mathbf{K}\Big)^{-1}\mathbf{N}^\top\mathbf{D}\Big] \\ &= \mathrm{tr}\Big[\Big(\mathbf{N}^\top\mathbf{D}\mathbf{N} + \lambda/2(\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}}\Big(\widehat{\sigma^2}^{(t)}\Big)^{1/2}\mathbf{K}\Big)^{-1}\mathbf{N}^\top\mathbf{D}\mathbf{N}\Big], \end{aligned}$$

write $\mathbf{B} = \mathbf{N}^\top\mathbf{D}\mathbf{N}$ and $\mathbf{L} = 1/2(\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}}\Big(\widehat{\sigma^2}^{(t)}\Big)^{1/2}\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{1/2}$. Then the express can be obtained according to [19]

$$\mathrm{d}g(\lambda) = \mathrm{tr}\left[\mathbf{B}^{1/2}\left(\mathbf{B}^{1/2}\mathbf{B}^{1/2} + \lambda/2(\hat{\alpha}^{(t)})^{2/\hat{\alpha}^{(t)}}\Big(\widehat{\sigma^2}^{(t)}\Big)^{1/2}\mathbf{K}\right)^{-1}\mathbf{B}^{1/2}\right] = \mathrm{tr}\left[(\mathbb{I}_q + \lambda\mathbf{L})^{-1}\right] = \sum_{j=1}^{q}\frac{1}{1+\lambda L_j},$$

where L_j is the j-th eigenvalue of $\mathbf{L}$ and $\mathbb{I}_q$ is the identity matrix of order q. Thus, it follows that $\mathrm{d}g(\lambda)$ is inversely proportional to λ. We can determine λ by minimizing the AIC function using the *optimize* function in the R environment.

$$\mathrm{AIC}(\lambda) = -2\ell_p(\hat{\boldsymbol{\theta}}, \lambda) + 2 * (3 + p + \mathrm{d}g(\lambda)).$$

3.3. The observed fisher information matrix

In this part, we establish the observed information matrix by viewing the penalized likelihood as a common likelihood function. Given the PLM model with $\epsilon \sim \mathrm{SGN}(0, \sigma^2, \alpha, r)$, the correspondence penalized log-likelihood function of $\boldsymbol{\theta}$ is of the form $\ell_p(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^{n}\ell_{pi}(\boldsymbol{\theta}, \lambda), i = 1, \cdots, n$ with

$$\ell_{pi}(\boldsymbol{\theta}, \lambda) = -\frac{1}{2}\ln\sigma^2 - \frac{1-\alpha}{\alpha}\ln\alpha - \ln\Gamma(1/\alpha) - \frac{|y_i - \mu_i|^\alpha}{\alpha[\sigma(1 + r * \mathrm{sign}(y_i - \mu_i))]^\alpha} - \frac{\lambda}{2n}\mathbf{f}^\top\mathbf{K}\mathbf{f}, \tag{3.18}$$

the observed information matrix for $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{I}_{\boldsymbol{\theta\theta}} = -\sum_{i=1}^{n}\Big(\frac{\partial^2 \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big), \tag{3.19}$$

$$\begin{aligned}
\frac{\partial \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial\boldsymbol{\beta}} &= \frac{|y_i-\mu_i|^{\alpha-1}*\operatorname{sign}(y_i-\mu_i)}{\sigma^2[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}}\mathbf{x}_i,\\
\frac{\partial \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial\mathbf{f}} &= \frac{|y_i-\mu_i|^{\alpha-1}*\operatorname{sign}(y_i-\mu_i)}{\sigma^2[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}}\mathbf{n}_i-\frac{\lambda}{n}\mathbf{K}\mathbf{f},\\
\frac{\partial \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial\sigma^2} &= -\frac{1}{2\sigma^2}+\frac{|y_i-\mu_i|^{\alpha}}{2\sigma^{\alpha+2}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}},\\
\frac{\partial \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial\alpha} &= \frac{\ln\alpha-1+\psi(1/\alpha)+\alpha}{\alpha^2}+\Big(\frac{1}{\alpha^2}-\frac{1}{\alpha}\ln\frac{|y_i-\mu_i|}{\sigma[1+r*\operatorname{sign}(y_i-\mu_i)]}\Big)*\frac{|y_i-\mu_i|^{\alpha}}{2\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}},\\
\frac{\partial \ell_{pi}(\boldsymbol{\theta},\lambda)}{\partial r} &= \frac{|y_i-\mu_i|^{\alpha}*\operatorname{sign}(y_i-\mu_i)}{\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha+1}},
\end{aligned}$$

then, the second-order derivative of $\ell_p(\boldsymbol{\theta},\lambda)$ is given by

$$\begin{aligned}
\boldsymbol{I}_{\boldsymbol{\beta\beta}^{\top}} &= \frac{\alpha-1}{\sigma^{\alpha}}\mathbf{X}^{\top}\Lambda\Big(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha}}\Big)\mathbf{X},\\
\boldsymbol{I}_{\boldsymbol{\beta}\sigma^2} &= \frac{\alpha}{2\sigma^{\alpha+2}}\mathbf{X}^{\top}\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}\circ\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha}},\\
\boldsymbol{I}_{\boldsymbol{\beta}\alpha} &= -\frac{1}{\sigma^{\alpha}}\mathbf{X}^{\top}\Big(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|}{1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})}\Big)^{\alpha-1}\circ\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})\circ\ln\frac{|\boldsymbol{y}-\boldsymbol{\mu}|}{\sigma[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]},\\
\boldsymbol{I}_{\boldsymbol{\beta}r} &= \frac{\alpha}{\sigma^2}\mathbf{X}^{\top}\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha+1}},\\
\boldsymbol{I}_{\boldsymbol{\beta}\mathbf{f}^{\top}} &= \frac{\alpha-1}{\sigma^{\alpha}}\mathbf{X}^{\top}\Lambda\Big(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha}}\Big)\mathbf{N},\\
\boldsymbol{I}_{\sigma^2\sigma^2} &= -\frac{n}{2\sigma^4}+\frac{\alpha+2}{4\sigma^{\alpha+4}}\sum_{i=1}^{n}\frac{|y_i-\mu_i|^{\alpha}}{[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}},\\
\boldsymbol{I}_{\sigma^2\alpha} &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{|y_i-\mu_i|^{\alpha}}{[\sigma(1+r*\operatorname{sign}(y_i-\mu_i))]^{\alpha}}\circ\ln\frac{|y_i-\mu_i|}{\sigma[1+r*\operatorname{sign}(y_i-\mu_i)]},\\
\boldsymbol{I}_{\sigma^2 r} &= \frac{\alpha}{2\sigma^{\alpha+2}}\sum_{i=1}^{n}\frac{|y_i-\mu_i|^{\alpha}}{[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha+1}}\circ\operatorname{sign}(y_i-\mu_i),\\
\boldsymbol{I}_{\sigma^2\mathbf{f}} &= \frac{\alpha}{2\sigma^{\alpha+2}}\mathbf{N}^{\top}\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha}}\circ\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu}),\\
\boldsymbol{I}_{\alpha\alpha} &= -n\frac{3\alpha-\psi'(1/\alpha)-2\alpha\ln\alpha-2\alpha\psi(1/\alpha)-\alpha^2}{\alpha^4}+\sum_{i=1}^{n}\frac{|y_i-\mu_i|^{\alpha}}{\alpha\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha}}\\
&\quad *\Big(\frac{2}{\alpha^2}-\frac{2}{\alpha}\ln\frac{|y_i-\mu_i|}{\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]}+\ln^2\frac{|y_i-\mu_i|}{\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]}\Big),
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{I}_{\alpha r} &= -\sum_{i=1}^{n} \frac{\operatorname{sign}(y_i-\mu_i)|y_i-\mu_i|^{\alpha}}{\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha+1}} \ln \frac{|y_i-\mu_i|}{\sigma[1+r*\operatorname{sign}(y_i-\mu_i)]},\\
\boldsymbol{I}_{\alpha \mathbf{f}} &= -\frac{1}{\sigma^{\alpha}}\mathbf{N}^{\top}\left(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|}{1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})}\right)^{\alpha-1} \circ \operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu}) \circ \ln \frac{|\boldsymbol{y}-\boldsymbol{\mu}|}{\sigma[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]},\\
\boldsymbol{I}_{rr} &= \sum_{i=1}^{n} \frac{(\alpha+1)|y_i-\mu_i|^{\alpha}}{\sigma^{\alpha}[1+r*\operatorname{sign}(y_i-\mu_i)]^{\alpha+2}},\\
\boldsymbol{I}_{r\mathbf{f}} &= \frac{\alpha}{\sigma^{2}}\mathbf{N}^{\top}\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha+1}},\\
\boldsymbol{I}_{\mathbf{f}\mathbf{f}^{\top}} &= \frac{\alpha-1}{\sigma^{\alpha}}\mathbf{N}^{\top}\Lambda\left(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1+r*\operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha}}\right)\mathbf{N},
\end{aligned}
$$

where $\psi(z) = \dfrac{\mathrm{d}\ln\Gamma(z)}{\mathrm{d}z}$ and the symbol $\circ$ is used to represent the Hadamard product.

4. Model diagnostic

Diagnostic analyses serve as critical tools for enhancing regression models through three key dimensions: accuracy refinement, predictive capability enhancement, and robustness fortification. These analyses systematically identify and resolve latent model deficiencies through two principal methodologies—Local Influence Analysis and Generalized Leverage Assessment—thereby enabling targeted refinements that optimize both predictive performance and interpretability.

4.1. Local influence

The core concept of the local influence methods is to evaluate whether small changes in the model or data lead to significant alterations in parameter estimates. Cook [12] introduced the idea of perturbing the log-likelihood function to examine the impact of parameter estimates. Let $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_n)^{\top}$ denotes an $n \times 1$ perturbation vector within an open subset $\Omega \subseteq \mathcal{R}^n$, and define the perturbed penalized log-likelihood as $\ell_p(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\omega})$. Assume there exists $\boldsymbol{\omega}_0 \in \Omega$ such that $\ell_p(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\omega}_0) = \ell_p(\boldsymbol{\theta}, \lambda)$ holds for all $\boldsymbol{\theta}$ and λ. The likelihood displacement metric is given by

$$
v(\boldsymbol{\omega}) = 2\left[\ell_p(\hat{\boldsymbol{\theta}}, \lambda) - \ell_p\left(\hat{\boldsymbol{\theta}}_{\omega}, \lambda\right)\right],
$$

where $\hat{\boldsymbol{\theta}}_{\omega}$ denotes the maximum penalized likelihood estimate (MPLE) under $\ell_p(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\omega})$, quantifying the divergence between perturbed and unperturbed estimates. As proposed by [12], local influence analysis focuses on the behavior of $v(\boldsymbol{\omega})$ near $\boldsymbol{\omega}_0$. This reduces to examining directional derivatives along unit vectors $\mathbf{h} \in \mathbb{R}^n(\|\mathbf{h}\| = 1)$ through the function $v(\boldsymbol{\omega}_0 + a\mathbf{h})$ for scalar $a \in \mathbb{R}$.

Building on Ibacache-Pulgar et al. [23], the most influential direction $\mathbf{h}_{\max}$ corresponds to the maximal normal curvature

$$
C_{\ell_{\max}}(\boldsymbol{\theta}) = \max_{\|\mathbf{h}\|=1} C_{\ell}(\boldsymbol{\theta}; \mathbf{h}),
$$

where $C_{\ell}(\boldsymbol{\theta}; \mathbf{h})$ represents the curvature at $a = 0$. The index plot of $\mathbf{h}_{\max}$ identifies observations most sensitive to perturbations, thereby revealing influential data points.

Following Cook [12], the normal curvature along a unit direction $\mathbf{h}$ is defined as

$$C_\ell(\boldsymbol{\theta}) = -2\mathbf{h}^\top \boldsymbol{\Delta}_{\boldsymbol{\omega_0}}^\top \mathbf{H}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega_0}} \mathbf{h},$$

where $\mathbf{H}_{\mathrm{p}} = \partial^2 \ell_p(\boldsymbol{\theta}, \lambda)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ denotes the Hessian matrix of the penalized log-likelihood, and $\boldsymbol{\Delta}_{\omega_0} = \partial^2 \ell_p(\boldsymbol{\theta}, \lambda \mid \boldsymbol{\omega})/\, \partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top\big|_{\omega=\omega_0}$ represents the perturbation matrix. Building on Ferreira and Paula [5], we perform spectral decomposition:

$$-2\boldsymbol{\Delta}_{\boldsymbol{\omega_0}}^\top \mathbf{H}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega_0}} = \sum_{k=1}^{n} \xi_k \mathbf{e}_k \mathbf{e}_k^\top,$$

where eigenvalues $\xi_1, \cdots, \xi_n$ are sorted from largest to smallest, and $\mathbf{e}_1, \cdots, \mathbf{e}_n$ are the corresponding eigenvectors to $\xi_1, \cdots, \xi_n$ with $\xi_{q+1} = 0$, $q = 1, 2, \cdots, n$. Let $\tilde{\xi}_k = \xi_k / \left(\xi_1 + \cdots + \xi_q\right)$ and $\mathbf{e}_k^2 = \left(e_{k1}^2, \cdots, e_{kn}^2\right)$. The local influence of individual observations is quantified through the diagnostic measure

$$G_l(0) = \sum_{k=1}^{q} \tilde{\xi}_k \mathrm{e}_{kl}^2, \quad l = 1, \cdots, n.$$

An observation is formally identified as influential if its diagnostic measure exceeds the adaptive threshold:

$$G_l(0) > \frac{1}{n} + c^* \cdot \mathrm{SD}(\{G_l(0)\}_{l=1}^n),$$

where c^* is a tuning constant (typically 2–3) and SD(·) denotes empirical standard deviation.

4.1.1. Case-weight perturbation

Let $\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^\top$ be a vector of dimensions $n \times 1$ with $\boldsymbol{w}_0 = (1, 1, \cdots, 1)^\top = \mathbf{1}_n$. The expectation of the perturbed complete-data log-likelihood function can be formally expressed as

$$\ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} w_i \ell_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) - \frac{\lambda}{2}\mathbf{f}^\top \mathbf{K}\mathbf{f},$$

where $\ell_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \ell_p(y_i, \boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$. In the case the matrix $\boldsymbol{\Delta}_{\boldsymbol{w}0} = \dfrac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{w}^\top}|_{\boldsymbol{w}=\boldsymbol{w_0}} = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_{\sigma^2}^\top, \Delta_\alpha^\top, \Delta_r^\top, \Delta_{\mathbf{f}}^\top)^\top$ has elements given by

$$\Delta_{\boldsymbol{\beta}} = \frac{1}{\sigma^\alpha}\mathbf{X}^\top \Lambda\Big(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1} \circ \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big),$$

$$\Delta_{\sigma^2} = -\frac{1}{2\sigma^2}\mathbf{1}_n^\top + \frac{|(\boldsymbol{y}-\boldsymbol{\mu})^\top|^\alpha}{2\sigma^{\alpha+2}[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})^\top]^\alpha},$$

$$\Delta_\alpha = \frac{\ln\alpha + \psi(1/\alpha) + \alpha - 1}{\alpha^2}\mathbf{1}_n^\top + \frac{|(\boldsymbol{y}-\boldsymbol{\mu})^\top|^\alpha}{\alpha\sigma^\alpha[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})^\top]^\alpha} \circ \Big(\frac{1}{\alpha}\mathbf{1}_n^\top - \ln\frac{|(\boldsymbol{y}-\boldsymbol{\mu})^\top|}{\sigma[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})^\top]}\Big),$$

$$\Delta_r = \frac{|(\boldsymbol{y}-\boldsymbol{\mu})^\top|^\alpha \circ \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})^\top}{\sigma^\alpha[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})^\top]^\alpha},$$

$$\Delta_{\mathbf{f}} = \frac{1}{\sigma^\alpha}\mathbf{N}^\top \Lambda\Big(\frac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1} \circ \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1 + r * \mathrm{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big).$$

For completeness, the important technical derivation supporting these results is rigorously established in Appendix B.

4.1.2. Explanatory variable perturbation

Within this perturbation scheme targeting a specific continuous explanatory variable, we define the perturbed explanatory variable as

$$\mathbf{x}_{r_w} = \mathbf{x}_r + S_r \boldsymbol{w}, \quad r \in \{1, \cdots, p\},$$

where $\mathbf{x}_{r_w}$ represents the r-th column of the perturbed explanatory variable matrix X, while $\mathbf{x}_r$ denotes the r-th column of the original explanatory variable matrix X, S_r denotes the sample standard deviation of explanatory variable $\mathbf{x}_r$, and the baseline perturbation vector satisfies $\boldsymbol{w}_0 = \mathbf{0}_{n\times 1}$. The corresponding perturbed penalized log-likelihood function becomes

$$\ell_p(\boldsymbol{\theta}, \lambda) = -n\Big(\frac{1}{2}\ln\sigma^2 + \frac{1-\alpha}{\alpha}\ln\alpha + \ln\Gamma(1/\alpha)\Big) - \sum_{i=1}^{n} \frac{|y_i - \mu_{iw}|^\alpha}{\alpha[\sigma(1 + r * \text{sign}(y_i - \mu_{iw}))]^\alpha} - \frac{\lambda}{2}\mathbf{f}^\top \mathbf{K}\mathbf{f},$$

where $\mu_{iw} = \mathbf{x}_{iw}^\top \boldsymbol{\beta} - \mathbf{n}_i^\top \mathbf{f}$, with $\mathbf{x}_{iw}$ denoting the i-th row of the perturbed explanatory variable matrix $\mathbf{x}_{r_w}$. It follows that the matrix $\boldsymbol{\Delta}_{\mathbf{w}_0}$ may be expressed in the form:

$$\boldsymbol{\Delta}_{\mathbf{w}_0} = \frac{S_r}{\sigma^\alpha}\begin{pmatrix} (1-\alpha)\beta_r \mathbf{X}^\top \Lambda\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big) + \mathbf{1}_{r_0}\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1} \circ \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big)^\top \\ -\dfrac{\alpha\beta_r}{2\sigma^2}\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1} \circ \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big)^\top \\ \beta_r \ln\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|}{\sigma[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]}\Big)^\top \circ \Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1} \circ \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big)^\top \\ -\alpha\beta_r\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha+1}}\Big)^\top \\ (1-\alpha)\beta_r \mathbf{N}^\top \Lambda\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1 + r * \text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big) \end{pmatrix}, \tag{4.1}$$

where $\mathbf{1}_{r_0}$ denotes a $p \times 1$ vector of zeros with one in the r-th position and β_r represents the r-th element of $\boldsymbol{\beta}$.

4.1.3. Response variable perturbation

A perturbation of the response variables $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ is defined as $\boldsymbol{y}_{\boldsymbol{w}} = \boldsymbol{y} + S_y \boldsymbol{w}$, where S_y is the sample standard deviation of $\boldsymbol{y}$. In this case, $\boldsymbol{w}_0 = \mathbf{0} : n \times 1$ and

$$\ell_p(\boldsymbol{\theta}, \lambda) = -n\Big(\frac{1}{2}\ln\sigma^2 + \frac{1-\alpha}{\alpha}\ln\alpha + \ln\Gamma(1/\alpha)\Big) - \sum_{i=1}^{n} \frac{|y_{iw} - \mu_i|^\alpha}{\alpha[\sigma(1 + r * \text{sign}(y_{iw} - \mu_i))]^\alpha} - \frac{\lambda}{2}\mathbf{f}^\top \mathbf{K}\mathbf{f}.$$

It follows that the matrix $\Delta_{\mathbf{w}_0}$ admits the form:

$$\boldsymbol{\Delta}_{\mathbf{w}_0} = \frac{S_y}{\sigma^\alpha}\begin{pmatrix} (\alpha-1)\mathbf{X}^\top\Lambda\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big) \\ \dfrac{\alpha}{2\sigma^2}\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}\circ\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big)^\top \\ -\ln\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|}{\sigma[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]}\Big)^\top\circ\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}\circ\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})}{[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big)^\top \\ \alpha\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-1}}{[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^{\alpha+1}}\Big)^\top \\ (\alpha-1)\mathbf{N}^\top\Lambda\Big(\dfrac{|\boldsymbol{y}-\boldsymbol{\mu}|^{\alpha-2}}{[1+r*\text{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^\alpha}\Big) \end{pmatrix}. \tag{4.2}$$

4.2. Generalized leverage

Leverage is a fundamental component in the diagnostic assessment of linear regression models. Wei et al. [24] proposed a generalized definition of leverage applicable to a broad class of models, which is defined as follows:

$$\text{GL}(\hat{\boldsymbol{\theta}}) = \mathbf{D}_{\boldsymbol{\theta}}\Big[-\ddot{\ell}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\lambda)\Big]^{-1}\ddot{\ell}_{\boldsymbol{\theta},\boldsymbol{y}}(\hat{\boldsymbol{\theta}},\lambda),$$

where $\mathbf{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}_y/\partial\boldsymbol{\theta}^\top, \ddot{\ell}_{\boldsymbol{\theta},\boldsymbol{y}}(\hat{\boldsymbol{\theta}},\lambda) = \partial^2\ell_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\lambda\mid\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top, \ddot{\ell}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}},\lambda)$ is the negative fisher matrix (3.19). For SGN model, we have $\boldsymbol{\mu}_y = E(\boldsymbol{Y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + 2\sigma r\alpha^{1/\alpha}b\mathbf{1}_n$, where $b = \dfrac{\Gamma(2/\alpha)}{\Gamma(1/\alpha)}$. The $(n\times(p+q+3))$ matrix $D_{\boldsymbol{\theta}}$ is constructed as

$$\mathbf{D}_{\boldsymbol{\theta}} = \left(\mathbf{X}, \frac{r\alpha^{1/\alpha}b}{\sigma}\mathbf{1}_n, 2\sigma r\delta\mathbf{1}_n, 2\sigma b\alpha^{1/\alpha}\mathbf{1}_n, \mathbf{N}\right), \tag{4.3}$$

where $\delta = \alpha^{1/\alpha-2}b[1-\ln\alpha-2\psi(2/\alpha)+\Gamma'(1/\alpha)]$ and $\psi(\cdot)$ denotes the digamma function. The matrix $\ddot{\ell}_{\theta,y}(\hat{\boldsymbol{\theta}})$ retains the structure defined in Eq (4.2), excluding the S_y term. To identify influential cases, we utilize the threshold

$$c*\text{trace(GL)}/n,$$

where c is a selected constant (generally equal to 2 or 3) according to Osorio(2006) [25].

5. Experiment

This section comprises two principal components: a comprehensive Monte Carlo simulation study evaluating the estimation efficiency under varying sample sizes, and an empirical application demonstrating the superior modeling performance of the PLM-SGN through real-world benchmarks.

5.1. Simulation study

We conduct a comprehensive simulation study with two distinct nonparametric functions:
- Model 1: $g(t) = \cos(t), t\in(0,3\pi)$.
- Model 2: $g(t) = \cos(4\pi t)\exp\left(-t^2/2\right), t\in(0.6,1.6)$ (Doppler effect).

The PLM-SGN structure is defined as

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

with true parameters $\beta = 5$ and $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{SGN}\left(\mu = 0, \sigma^2 = 1, \alpha = 1.1, r = 0.8\right)$ as per Theorem 1. Covariates x_i were independently generated from a uniform distribution $\mathcal{U}(0, 1)$. We conducted 500 Monte Carlo replications for each sample size $n = 300, 500, 800$.

Parameter estimation proceeded as follows:

(1) Parameter estimation: Compute MPLE via the EM algorithm.

(2) Uncertainty quantification: Derive standard errors from the observed information matrix inverse.

For each parameter $\boldsymbol{\eta} = \left(\beta, \sigma^2, \alpha, r\right)$, we calculated

$$\text{Mean: } \overline{\hat{\boldsymbol{\eta}}} = \frac{1}{500}\sum_{j=1}^{500}\hat{\boldsymbol{\eta}}_j, \quad \text{SD: } \sqrt{\frac{1}{499}\sum_{j=1}^{500}\left(\hat{\boldsymbol{\eta}}_j^k - \overline{\hat{\boldsymbol{\eta}}}^k\right)^2},$$

$$\text{Empirical SE: } \sqrt{\frac{1}{500}\sum_{j=1}^{500} I_j^{kk}}, \quad k = 1, 2, 3, 4,$$

where I_j^{kk} denotes the k-th diagonal element of the inverse observed information matrix for the j-th replication.

The simulation outcomes presented in Tables 2 and 3 align rigorously with asymptotic theoretical expectations. As sample sizes increase, the MLE bias, quantified by $\overline{\hat{\boldsymbol{\eta}}} - \boldsymbol{\eta}$, demonstrates a systematic decrease across all parameters. Concurrently, empirical standard deviations converge decisively toward their theoretical counterparts derived from the observed information matrix, with both metrics exhibiting the anticipated $O(1/\sqrt{n})$ decay rate. Notably, the diminishing discrepancy between empirical and theoretical standard errors substantiates the statistical efficiency of the observed information matrix for variance-covariance estimation. These findings collectively validate the consistency of our estimators and the precision of the proposed computational framework in large-sample regimes, thereby reinforcing the methodological robustness of PLM-SGN for practical implementations.

Figure 1 presents the nonparametric estimation outcomes from 500 Monte Carlo replications under Models 1 and 2. Three key patterns emerge: First, increasing sample sizes induces progressive shrinkage in the pointwise variability bands of the nonparametric estimates, reflecting enhanced estimation precision. Second, the mean estimated curves exhibit systematic convergence toward the true functional forms $g(t) = \cos(t)$ and $g(t) = \cos(4\pi t)e^{-t^2/2}$, with residual discrepancies diminishing at rates commensurate with $O_p\left(n^{-2/5}\right)$ - the theoretical optimality bound for spline-based estimators. Crucially, the interquartile ranges of the estimates contract symmetrically around the truth as n increases from 300 to 800, providing empirical validation of the estimator's L_2-consistency. These results collectively affirm the asymptotic efficiency of the proposed spline-based estimation framework for capturing complex nonlinear patterns.

Table 2. Mean and standard deviations (SD) for EM estimates and empirical standard errors estimates (SE_{emp}) based on 500 samples from the PLMSN Model 1.

Parameter	True value	$n = 300$			$n = 500$			$n = 800$		
		Mean	SD	SE_{emp}	Mean	SD	SE_{emp}	Mean	SD	SE_{emp}
β	5.0	5.28	0.25	0.54	5.14	0.14	0.32	5.09	0.11	0.23
σ^2	1.0	0.92	0.14	0.16	0.94	0.10	0.11	0.96	0.09	0.09
α	1.1	1.05	0.10	0.17	1.06	0.08	0.11	1.08	0.07	0.08
r	0.8	0.81	0.06	0.04	0.81	0.04	0.03	0.81	0.03	0.03

Table 3. Mean and standard deviations (SD) for EM estimates and empirical standard errors estimates (SE_{emp}) based on 500 samples from the PLMSN Model 2.

Parameter	True value	$n = 300$			$n = 500$			$n = 800$		
		Mean	SD	SE_{emp}	Mean	SD	SE_{emp}	Mean	SD	SE_{emp}
β	5.0	5.2	0.22	0.33	5.13	0.14	0.23	5.08	0.1	0.14
σ^2	1.0	0.92	0.14	0.18	0.94	0.1	0.11	0.96	0.09	0.09
α	1.1	1.06	0.1	0.21	1.07	0.08	0.11	1.08	0.07	0.08
r	0.8	0.81	0.06	0.04	0.81	0.04	0.03	0.81	0.03	0.02

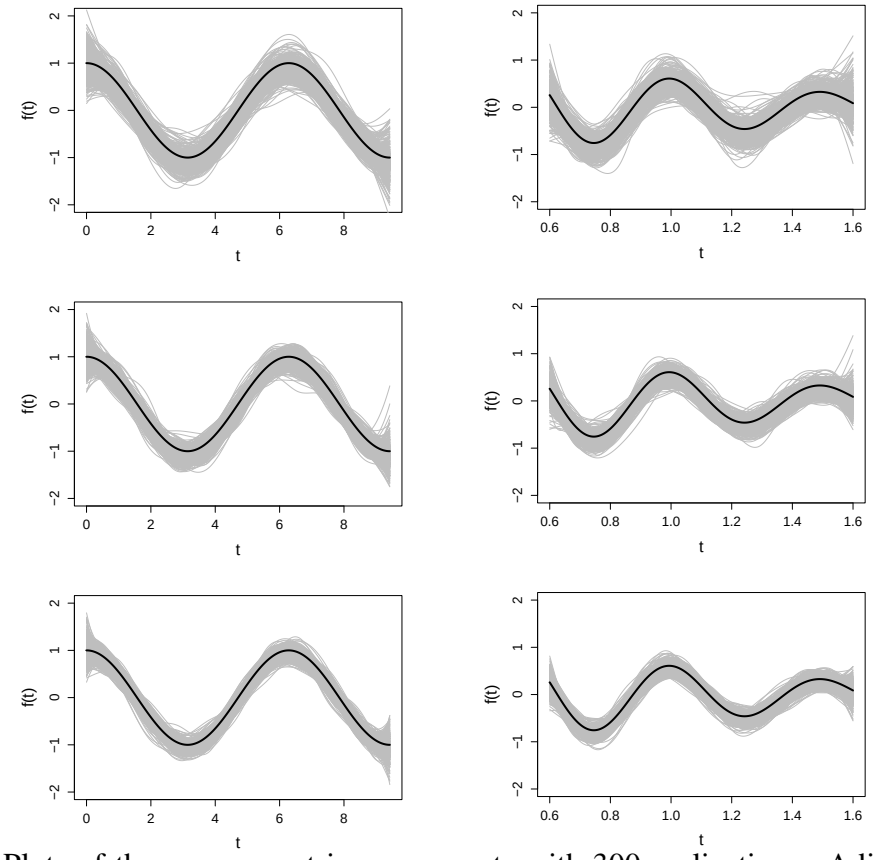


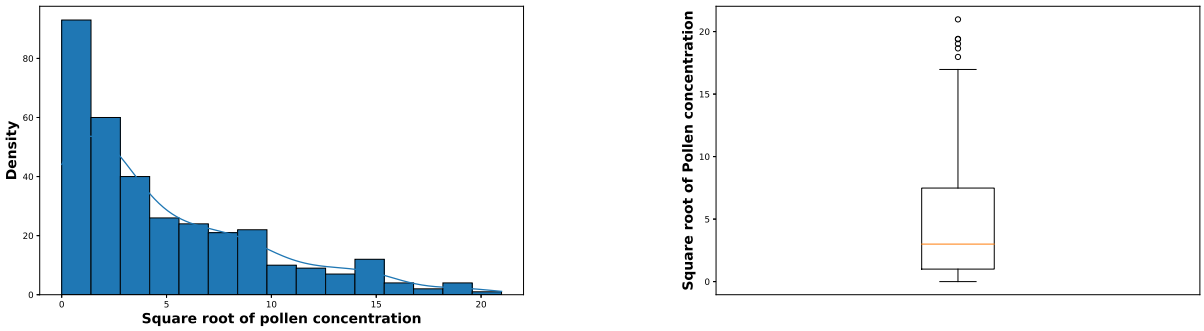
Figure 1. Plots of the nonparametric components with 300 replications. Adjusted curves (gray lines) and true curves (black lines): Model 1 (first column) and Model 2 (second column). The first row for $n = 300$; the second row for $n = 500$ and the last row for $n = 800$.

5.2. Application

This section demonstrates the practical utility of the PLM-SGN framework through analysis of ragweed (Ambrosia spp.) pollen concentration data, previously investigated using the partially linear model with skew-normal errors (PLM-SN) by Ferreira and Paula [5]. Table 4 summarizes key descriptive statistics, while Figure 2 displays the response variable's distribution through a histogram with fitted probability density functions (PDFs) and a boxplot. The visual and numerical evidence reveals pronounced right skewness (skewness = 1.12) exceeding the modeling capacity of conventional SN distributions, thereby justifying our adoption of SGN-distributed errors.

Table 4. The descriptive statistics for the sqrt of ragweed pollen concentration.

n	Min	Max	$\bar{X}$	Median	Upper quarter	Lower quarter	Skewness	Kurtosis
335	0	20.976177	4.713527	3	7.483315	1	1.1201	3.5938



(a) Histogram (b) Boxplot

Figure 2. Square root pollen concentration visualization. (a) Distribution histogram; (b) Variation boxplot.

Following the experimental design of Ferreira and Paula [5], we consider four predictors: precipitation indicator (1, larger than 3 hr steady/intense rain; 0, otherwise), ambient temperature (°F), next-day wind speed forecast (mph), and temporal progression (days since season onset). As described by Ferreira (2016) [5] for the ragweed pollen concentration data: original data presents a very skewed form to the right that may be hard to fit, whereas the square root of the ragweed pollen concentration, which seems to follow some suitable behavior, could be fitted, for instance, by asymmetric distributions such as skew-normal. So, the square root transformed pollen concentration serves as the response variable to mitigate heteroscedasticity.

We implement the PLM-SGN to model the square-root transformed ragweed pollen concentration ($\tilde{y}_i$) as a function of environmental predictors. The proposed semi-parametric structure is formalized as

$$\tilde{y}_i = \beta_1 \,\text{rain}_i + \beta_2 \,\text{temperature}_{\,i} + \beta_3 \,\text{wind}_i + g\,(t_i) + \epsilon_i, \quad i = 1, \cdots, 335,$$

where $\tilde{y}_i = \sqrt{\text{pollen concentration}}$ (response), t_i represents the i-th day within pollen season (smoothing variable), and $\epsilon_i \overset{iid}{\sim} \text{D}$, $\text{D} \in \{\text{N}, \text{SN}, \text{SGN}\}$.

The design matrix $\mathbf{X} \in \mathbb{R}^{335\times4}$ contains four predictors: precipitation indicator, temperature (°F), wind speed (mph), and temporal index. The B-spline basis matrix $\mathbf{N} \in \mathbb{R}^{335\times10}$ is constructed using cubic splines with 6 interior knots and order 4, ensuring temporal coherence through ascending t_i ordering. To establish methodological superiority, we conduct a systematic comparison against normal (PLM-N) and PLM-SN error specifications, evaluating their relative capacities to capture the observed right-skewed error structure.

The comparative analysis in Table 5 establishes the PLM-SGN's superior goodness-of-fit through three key metrics: $\ell_p(\hat{\boldsymbol{\theta}}, \lambda)$, and minimized information criteria values (AIC, BIC). The BIC formulation:

$$\mathrm{BIC}(\lambda) = -2\ell_p(\hat{\boldsymbol{\theta}}, \lambda) + \log(n) * (3 + p + \mathrm{d}g(\lambda)).$$

The PLM-SGN demonstrates remarkable improvements in model fit, with its penalized log-likelihood value exceeding those of the PLM-SN and PLM-N by 26.8% and 27%, respectively. More critically, it achieves substantial gains in parsimony—the AIC and BIC for PLM-SGN are 26.7% and 26.8% lower than PLM-SN, and 26.8% and 26.4% lower than PLM-N, respectively. This dual superiority in both goodness-of-fit and complexity penalization underscores the SGN distribution's enhanced capacity to balance statistical fidelity with parametric efficiency.

Table 5. MPLEs (S.E. represents approximate standard errors) under Normal, SN, and SGN PLMs fitted to the ragweed levels data.

Effect	Normal		Skew-normal		Skewed generalized normal	
	Estimate	SE	Estimate	SE	Estimate	SE
Rain	1.408	0.392	1.456	0.382	0.575	0.303
Temperature	0.100	0.018	0.880	0.018	0.040	0.031
Wind	0.237	0.038	0.228	0.037	0.121	0.037
σ^2	4.504	0.350	9.758	1.362	2.856	0.441
ω	–	–	2.188	0.482	–	–
r	–	–	–	–	0.551	0.006
α	–	–	–	–	1.000	0.147
$l_p(\hat{\boldsymbol{\theta}}, \lambda)$	-733.468	–	-730.804	–	-534.900	–
AIC	1495.154	–	1494.090	–	1093.700	–
BIC	1548.967	–	1556.033	–	1139.000	–

The pointwise confidence bands for the PLMs estimated under the normal, SN, and SGN distributions are also presented in Figure 3. The three models present similar estimated curves, although the confidence bands of the SGN model seem to include more observations than the other models. To supplement the nonparametric curve results presented above, we constructed quantile-quantile plots with simulated confidence bands of 95% [26] based on the Mahalanobis distance $MD_i = (y_i - \mu_i)^\alpha/\sigma^\alpha/W^\alpha) \sim \mathrm{Gamma}(1/\alpha, 1/\alpha)$, for $i = 1, 2, \ldots, n$, where W is the same as in Theorem 1. From Figure 4, the PLMSGN almost always present observations within the confidence bands.

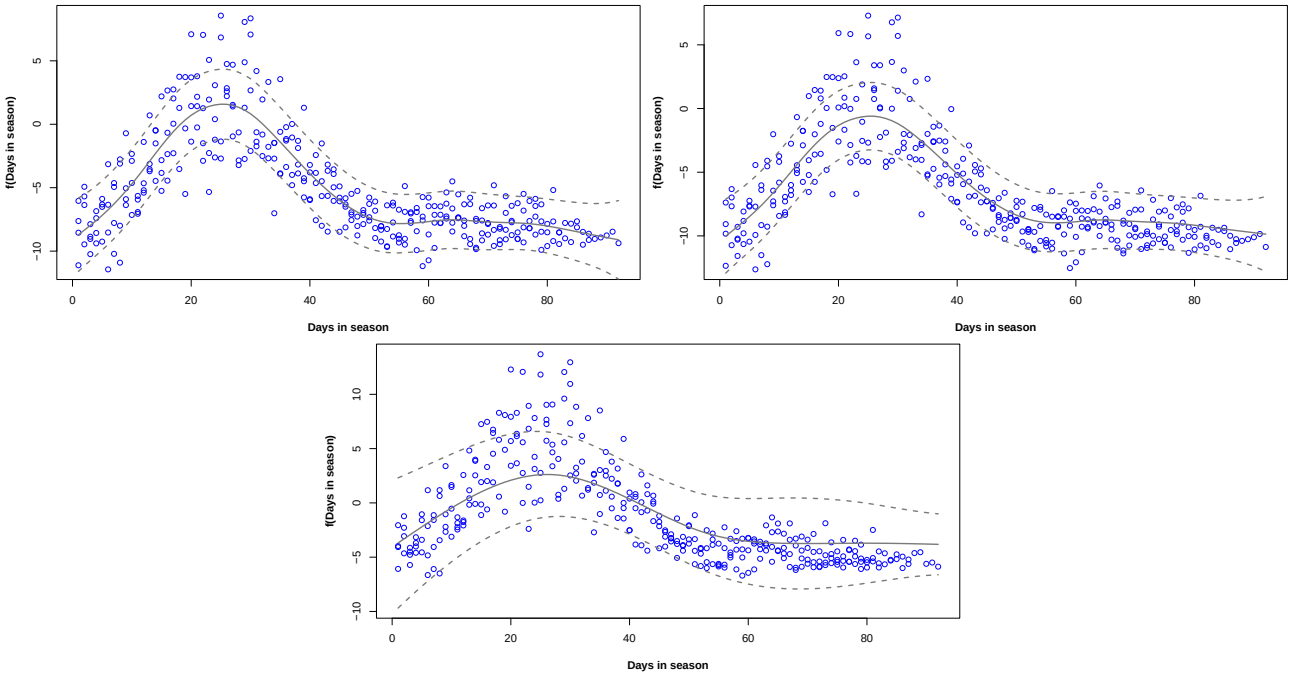


Figure 3. Ragweed levels data: An estimated curve with 95% pointwise confidence bands for g (days in season) from the PLMs fitted to the ragweed levels data under Normal (on the first row), SN (on the second row), and SGN (on the third row).

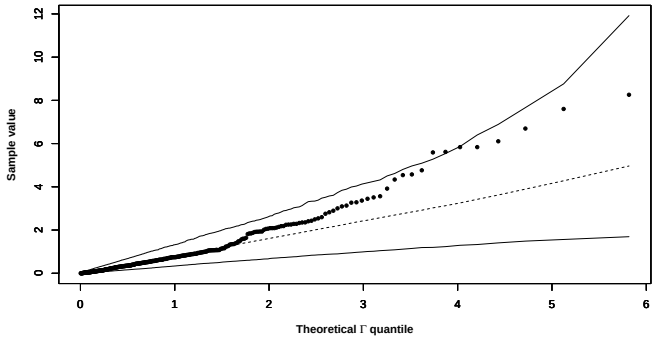


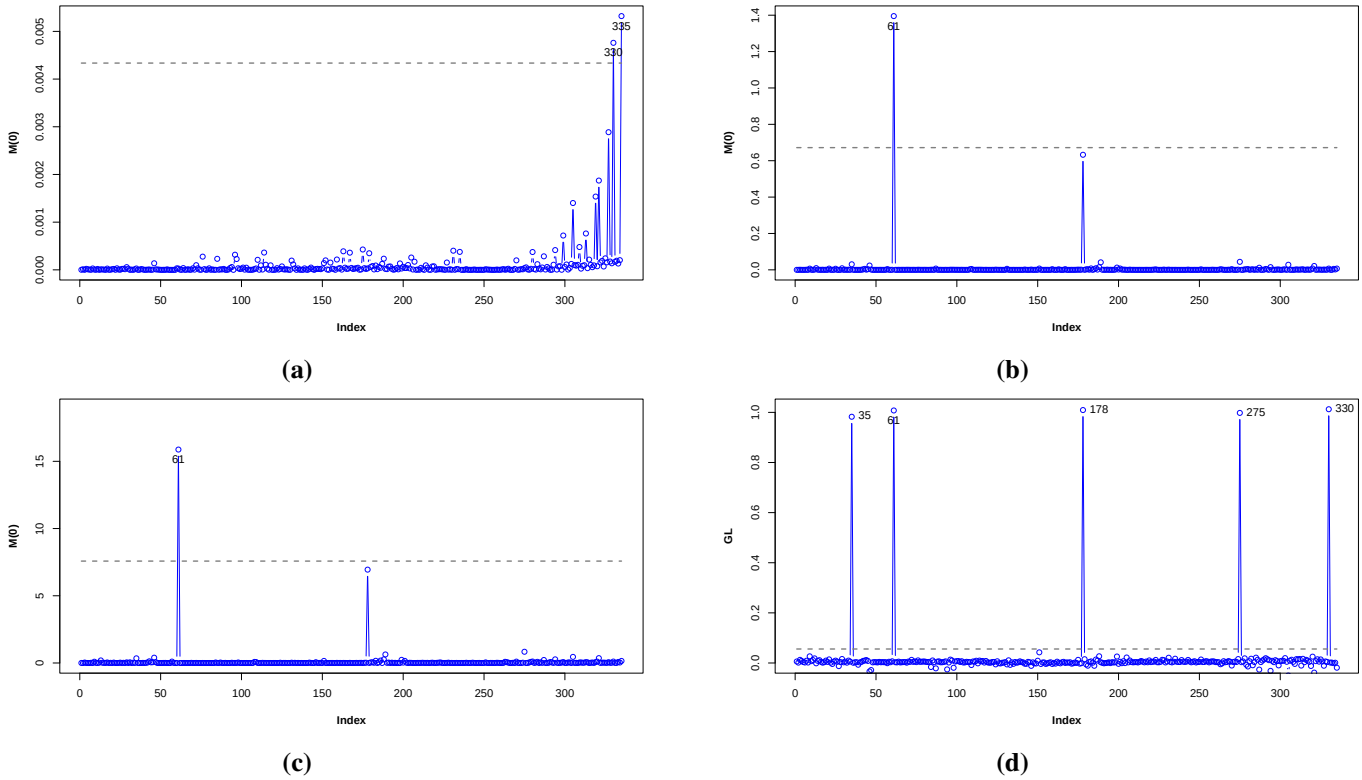
Figure 4. Quantile-quantile plots for the Mahalanobis distance.

We subsequently conduct comprehensive diagnostic analyses using local influence measure $G(0)$ and generalized leverage. The local influence approach systematically evaluates three perturbation schemes: case-weight, response, and explanatory variable (temperature) modifications. Figure 5(a)–(c) displays the $G(0)_l$ index plots for these perturbation scenarios, identifying observations #335, #330, and #61 as potential influential points through their pronounced negative residual magnitudes. Figure 5(d) highlights five high-leverage observations (#35, #61, #178, #275, #330) exhibiting anomalously low ragweed pollen concentrations. Notably:

- Observations #178, #330, and #335 correspond to minimal response values.
- Observation #61 records the maximum wind speed measurement.
- Observations #35 and #275 fall within the highest temperature quartile.

This multi-method diagnostic framework demonstrates the PLM-SGN's capacity to detect influential observations across diverse data anomalies, enhancing model reliability for ecological

forecasting.



(a) **(b)** **(c)** **(d)**

Figure 5. Diagnostic graphs from the PLM-SGN fitted to the ragweed levels data with: (a) case-weight perturbation; (b) response (square root of ragweed) perturbation; (c) explanatory (wind speed) perturbation; (d) generalized leverage.

6. Conclusions

This study proposes a novel PLM-SGN to address error structures exhibiting extreme skewness and heavy-tailedness beyond conventional symmetric or SN frameworks. We develop a penalized EM algorithm integrated with adaptive smoothing parameter selection, which ensures stable high-dimensional optimization through theoretically guaranteed convergence. The estimation framework incorporates effective degrees of freedom calculation for regularizing nonparametric components, maintaining estimation stability across diverse sample sizes.

Under large-sample conditions, the proposed estimators demonstrate strong agreement with theoretical benchmarks, rigorously verified through explicit computation of the observed information matrix—a critical component for evaluating estimation precision. A comprehensive diagnostic system is established, combining local influence analysis under case-weight, response, and covariate perturbations with generalized leverage measures.

Through two systematic simulation studies, we validate the consistency of both parametric and nonparametric estimators. Empirical analysis of ragweed pollen concentration data demonstrates the PLM-SGN's superiority over conventional PLM-N and PLM-SN errors. To broaden the applicability of this model, we will further investigate its efficiency in prediction scenarios.

Author contributions

Xue Wang: contributed the creative ideals and computation techniques for this paper; Weihu Cheng: consulted the relevant background of the paper and revised the article; Ferreira: made important contributions to the nonparametric fitting method and the search for smooth parameters. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declared that they have no conflicts of interest.

References

1. Y. Fan, Q. Li, Root-n-consistent estimation of partially linear time series models, *J. Nonparametr. Stat.*, **11** (1999), 251–269. https://doi.org/10.1080/10485259908832783
2. G. Tabakan, Performance of the difference-based estimators in partially linear models, *Statistics*, **47** (2013), 329–347. https://doi.org/10.1080/02331888.2011.592189
3. M. Han, D. Han, L. Sun, A class of partially linear transformation models for recurrent gap times, *Commun. Stat.-Theor. M.*, **47** (2018), 739–766. https://doi.org/10.1080/03610926.2017.1313986
4. V. Cancho, V. Lachos, E. Ortega, A nonlinear regression model with skew-normal errors, *Stat. Papers*, **51** (2010), 547–558. https://doi.org/10.1007/s00362-008-0139-y
5. C. Ferreira, G. Paula, Estimation and diagnostic for skew-normal partially linear models, *J. Appl. Stat.*, **44** (2017), 3033–3053. https://doi.org/10.1080/02664763.2016.1267124
6. A. Azzalini, A class of distributions which includes the normal ones, *Scand. J. Stat.*, **12** (1985), 171–178.
7. W. Lyu, Z. Feng, Estimation and application of skew-normal data for generalized linear regression, *Proceedings of the 2018 International Conference on Computer Modeling, Simulation and Algorithm (CMSA 2018)*, 2018, 208–211. https://doi.org/10.2991/cmsa-18.2018.48
8. J. de Freitas, C. Azevedo, J. Nobre, A stochastic approximation ECME algorithm to semi-parametric scale mixtures of centred skew normal regression models, *Stat. Comput.*, **33** (2023), 51. https://doi.org/10.1007/s11222-023-10223-5
9. C. S. Ferreira, R. Dias, Semiparametric regression models under skew scale mixtures of normal distributions, *Commun. Stat.-Simul. C.*, in press. https://doi.org/10.1080/03610918.2024.2372667
10. A. Azzalini, The skew-normal distribution and related multivariate families, *Scand. J. Stat.*, **32** (2005), 159–188. https://doi.org/10.1111/j.1467-9469.2005.00426.x
11. R. Cook, Detection of influential observation in linear regression, *Technometrics*, **19** (1977), 15–18. https://doi.org/10.1080/00401706.1977.10489493

12. R. Cook, Assessment of local influence, *J. R. Stat. Soc. B*, **48** (1986), 133–155. https://doi.org/10.1111/j.2517-6161.1986.tb01398.x

13. C. Kim, B. Park, W. Kim, Influence diagnostics in semiparametric regression models, *Stat. Probabil. Lett.*, **60** (2002), 49–58. https://doi.org/10.1016/S0167-7152(02)00268-7

14. M. Galea, G. Paula, F. Cysneiros, On diagnostics in symmetrical nonlinear models, *Stat. Probabil. Lett.*, **73** (2005), 459–467. https://doi.org/10.1016/j.spl.2005.04.033

15. M. Lemus, V. Lachos, C. Galarza, L. Matos, Estimation and diagnostics for partially linear censored regression models based on heavy-tailed distributions, *Stat. Interface*, **14** (2021), 165–182.

16. C. Ferreira, G. Paula, G. Lana, Estimation and diagnostic for partially linear models with first-order autoregressive skew-normal errors, *Comput. Stat.*, **37** (2022), 445–468. https://doi.org/10.1007/s00180-021-01130-2

17. G. Ibacache-Pulgar, C. Villegas, J. Lopez-Gonzales, M. Moraga, Influence measures in nonparametric regression model with symmetric random errors, *Stat. Methods Appl.*, **32** (2023), 1–25. https://doi.org/10.1007/s10260-022-00648-z

18. X. He, Z. Zhu, W. Fung, Estimation in a semiparametric model for longitudinal data with unspecified dependence structure, *Biometrika*, **89** (2002), 579–590. https://doi.org/10.1093/biomet/89.3.579

19. P. Eilers, B. Marx, Flexible smoothing with B-splines and penalties, *Stat. Sci.*, **11** (1996), 89–121. https://doi.org/10.1214/ss/1038425655

20. L. Devroye, Random variate generation for exponentially and polynomially tilted stable distributions, *ACM Trans. Model. Comput. S.*, **19** (2009), 18. https://doi.org/10.1145/1596519.1596523

21. J. Nolan, Parameterizations and modes of stable distributions, *Stat. Probabil. Lett.*, **38** (1998), 187–195. https://doi.org/10.1016/S0167-7152(98)00010-8

22. P. Green, On use of the EM algorithm for penalized likelihood estimation, *J. R. Stat. Soc. B*, **52** (1990), 443–452. https://doi.org/10.1111/j.2517-6161.1990.tb01798.x

23. G. Ibacache-Pulgar, G. Paula, F. Cysneiros, Semiparametric additive models under symmetric distributions, *TEST*, **22** (2013), 103–121. https://doi.org/10.1007/s11749-012-0309-z

24. B. Wei, Y. Hu, W. Fung, Generalized leverage and its applications, *Scand. J. Stat.*, **25** (1998), 25–37. https://doi.org/10.1111/1467-9469.00086

25. F. Salgado, Diagnóstico de influência em modelos elıpticos com efeitos mistos, Ph. D Thesis, Universidade de Sao Paulo, 2006. https://doi.org/10.11606/T.45.2006.tde-20210729-151424

26. A. Atkinson, Two graphical displays for outlying and influential observations in regression, *Biometrika*, **68** (1981), 13–20. https://doi.org/10.1093/biomet/68.1.13

Appendix A

The penalized Q function is shown by

$$Q_p\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right) = \mathrm{C} + \sum_{i=1}^{n} E\left(\ln f_W\left(w_i\right) \mid y_i, \boldsymbol{\theta}^{(t)}\right) - n\ln\sigma - \frac{n}{\alpha}\ln\alpha - (\boldsymbol{y}-\boldsymbol{\mu})^{\top}\mathbf{D}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{\lambda}{2}\mathbf{f}\mathbf{K}^{\top}\mathbf{f}, \quad \text{(A.1)}$$

where $\mathbf{D}$ is diagonal matrix of the vector $\dfrac{\boldsymbol{Z}}{[1 + r * \operatorname{sign}(\boldsymbol{y}-\boldsymbol{\mu})]^2}$, and $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f}$,

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^{\top}\mathbf{D}\boldsymbol{y} + 2\mathbf{X}^{\top}\mathbf{D}\mathbf{N}\mathbf{f} + 2\mathbf{X}^{\top}\mathbf{D}\mathbf{X}\boldsymbol{\beta}, \quad \text{(A.2)}$$

$$\frac{\partial Q}{\partial \mathbf{f}} = -2\mathbf{N}^{\top}\mathbf{D}\boldsymbol{y} + 2\mathbf{N}^{\top}\mathbf{D}\mathbf{X}\boldsymbol{\beta} + 2\mathbf{N}^{\top}\mathbf{D}\mathbf{N}\mathbf{f} - \lambda\mathbf{K}\mathbf{f}. \quad \text{(A.3)}$$

The other parameters can be directly differentiated to obtain the result.

Appendix B

$$\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^{\alpha}}\sum_{i=i}^{n} w_i \frac{|y_i-\mu_i|^{\alpha-1}\operatorname{sign}(y_i-\mu_i)}{[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha}}\mathbf{x}_i,$$

$$\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} w_i + \frac{1}{2\sigma^{\alpha+2}}\sum_{i=1}^{n} w_i \frac{|y_i-\mu_i|^{\alpha-1}}{[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha}},$$

$$\begin{aligned}\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial \alpha} &= \sum_{i=1}^{n} w_i \frac{\ln\alpha + \psi(1/\alpha) + \alpha - 1}{\alpha^2} + \frac{1}{\alpha^2}\sum_{i=1}^{n} w_i \frac{|y_i-\mu_i|^{\alpha}}{\sigma^{\alpha}[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha}} \\ &\quad - \frac{1}{\alpha}\sum_{i=1}^{n} w_i \frac{|y_i-\mu_i|^{\alpha}}{\sigma^{\alpha}[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha}} \ln\frac{|y_i-\mu_i|}{\sigma[1 + r * \operatorname{sign}(y_i-\mu_i)]},\end{aligned}$$

$$\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial r} = \frac{1}{\sigma^{\alpha}}\sum_{i=1}^{n} w_i \frac{|y_i-\mu_i|^{\alpha}\operatorname{sign}(y_i-\mu_i)}{[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha+1}},$$

$$\frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{w}|\hat{\boldsymbol{\theta}})}{\partial \mathbf{f}} = \frac{1}{\sigma^{\alpha}}\sum_{i=i}^{n} w_i \frac{|y_i-\mu_i|^{\alpha-1}\operatorname{sign}(y_i-\mu_i)}{[1 + r * \operatorname{sign}(y_i-\mu_i)]^{\alpha}}\mathbf{N}_i.$$