axioms

MDPI

Article

Maximum Penalized Likelihood Estimation of the Skew–*t* Link Model for Binomial Response Data

Omar Chocotea-Poca [1,2,*,†], Orietta Nicolis [3,*,†] and Germán Ibacache-Pulgar [2,4,*,†]

1 Department of Mathematics and Computer Science, Universidad de Santiago de Chile, Avenida Libertador Bernardo O'Higgins 3363, Santiago 9170022, Chile
2 Statistical Institute, Universidad de Valparaíso, Avenida Gran Bretaña 1111, Valparaíso 2360102, Chile
3 Engineering Faculty, Universidad Andres Bello, Calle Quillota 980, Viña del Mar 2520000, Chile
4 Centro de Estudios Atmosféricos y Cambio Climático, Universidad de Valparaíso, Avenida Gran Bretaña 644, Valparaíso 2360102, Chile
* Correspondence: omar.chocotea@usach.cl (O.C.-P.); orietta.nicolis@unab.cl (O.N.); german.ibacache@uv.cl (G.I.-P.)
† These authors contributed equally to this work.

Abstract: A critical aspect of modeling binomial response data is selecting an appropriate link function, as an improper choice can significantly affect model precision. This paper introduces the skew–*t* link model, an extension of the skew–probit model, offering increased flexibility by incorporating both asymmetry and heavy tails, making it suitable for asymmetric and complex data structures. A penalized likelihood-based estimation method is proposed to stabilize parameter estimation, particularly for the asymmetry parameter. Extensive simulation studies demonstrate the model's superior performance in terms of lower bias, root mean squared error (RMSE), and robustness compared to traditional symmetric models like probit and logit. Furthermore, the model is applied to two real-world datasets: one concerning women's labor participation and another related to cardiovascular disease outcomes, both showing superior fitting capabilities compared to more traditional models (with probit and the skew–probit links). These findings highlight the model's applicability to socioeconomic and medical research, characterized by skew and asymmetric data. Moreover, the proposed model could be applied in various domains where data exhibit asymmetry and complex structures.

Keywords: binomial response data; skew–probit link model; model flexibility; penalized likelihood

MSC: 62P12; 62J20; 62J12

Citation: Chocotea-Poca, O.; Nicolis, O.; Ibacache-Pulgar, G. Maximum Penalized Likelihood Estimation of the Skew–*t* Link Model for Binomial Response Data. *Axioms* **2024**, *13*, 749. https://doi.org/10.3390/axioms13110749

Academic Editors: Francisco Novoa-Muñoz and Bernardo M. Lagos-Álvarez

Received: 17 September 2024
Revised: 23 October 2024
Accepted: 28 October 2024
Published: 30 October 2024

1. Introduction

When analyzing binomial response data, models with probit, logit, and cloglog links are commonly considered. Probit and logit links are symmetric, while the cloglog link is asymmetric. Ref. [1] explores these common links. The researchers conduct various simulation studies to investigate whether the choice of the link influences the model's fit or predictive ability. Additionally, they perform analyses on three sets of real-world data. Both simulation studies and real data analysis highlight that an incorrect choice of the link can lead to underfitting or overfitting. They conclude that making a wrong decision can occur if the link function is specified incorrectly. Therefore, one option for a specified link is an asymmetric link that has a symmetric link as a particular case. One of the links that has gained importance in various contexts is the skew–probit link, which incorporates a parameter to regulate the asymmetry of the link function and considers the probit link as a specific case. This link was proposed by [2] in the field of item response theory from a Bayesian perspective. Furthermore, the skew–probit link has been examined in the context of binary response data, both from a Bayesian perspective in [3] and from a frequentist perspective in [4]. The skew–probit link is the inverse of the cumulative distribution function of the skew–normal distribution proposed by [5]. The literature, from the work of [5,6], highlights the difficulties associated with estimating the skewness parameter in these distributions. Therefore, Refs. [7,8] suggest the application of a penalty function.

Recognizing the need for a more flexible and adaptable link function in the analysis of binomial response data, we introduce the skew–*t* link model in this work. Our proposed model significantly extends the capabilities of existing methods by offering greater flexibility compared to traditional links, including the skew–probit link, which it encompasses as a special case. A key innovation of our approach lies in the use of penalized maximum likelihood estimation, which enhances the stability and accuracy of parameter estimation, particularly in handling the skewness parameter. This combination of flexibility and robust estimation makes our skew–*t* link model a powerful tool for addressing complex data structures in binomial response models.

The structure of this article is organized systematically as follows. Section 2 presents the proposed skew–*t* link model for analyzing binomial response data, detailing its formulation and the penalized log-likelihood function. In Section 3, we focus on parameter estimation, including the derivation of the score function and relevant asymptotic results. Section 4 evaluates the performance of the proposed model through two comprehensive simulation studies, each considering different data generation configurations. The application of the model to real-world datasets is presented in Section 5, where we analyze both binomial and binary response data. Finally, Sections 6 and 7 offer a detailed discussion of the findings and conclude the paper.

2. Skew–*t* Link Model

2.1. Formulation

As a starting point, we establish an essential part of the general notation. Let us assume that $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ is a vector of covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ is a vector of regression coefficients, both of dimension $k \times 1$, where $i = 1, \ldots, m$. Additionally, $x_{i1} = 1$ due to the intercept term. In this context, the skew–*t* link model is specified by

$$\begin{aligned} y_i &\overset{ind.}{\sim} \text{Binomial}(n_i, p_i) \\ p_i &= \Psi_\alpha(\mathbf{x}_i^\top \boldsymbol{\beta}; \nu), \end{aligned} \tag{1}$$

where $\Psi_\alpha(\cdot;\nu)$ is the cumulative distribution function of the skew–*t* distribution introduced by [9,10] with skewness parameter $-\infty < \alpha < \infty$ and degrees of freedom $\nu > 0$. In the context of the Generalized Linear Models theory, $\Psi_\alpha(\cdot;\nu)$ is the inverse link function [11],

$$\Psi_\alpha^{-1}(p_i;\nu) = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

The model (1) will be represented as $\mathcal{M}_{\alpha,\nu}$.

By definition, the cumulative distribution function for the skew–*t* distribution is given by

$$\Psi_\alpha(\eta;\nu) = 2\int_{-\infty}^{\eta} \psi_0(t;\nu)\Psi_0\left(\alpha t\sqrt{\frac{\nu+1}{\nu+t^2}};\nu+1\right)\mathrm{d}t, \tag{2}$$

where $\psi_0(\cdot;\nu^*)$ and $\Psi_0(\cdot;\nu^*)$ denote the probability density function and cumulative distribution function of the standard Student's *t*-distribution with degrees of freedom $\nu^* > 0$, respectively,

$$\psi_0(\eta;\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left(1+\frac{\eta^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

$\Psi_\alpha(\cdot;\nu)$ is continuous, smooth, and infinitely differentiable. By the fundamental theorem of calculus, the first derivative of (2) with respect to η is the probability density function of the skew–t distribution,

$$\psi_\alpha(\eta;\nu) = 2\psi_0(\eta;\nu)\Psi_0\left(\alpha\eta\sqrt{\frac{\nu+1}{\nu+\eta^2}};\nu+1\right). \quad (3)$$

The second derivative of (2) with respect to η is

$$\frac{\partial}{\partial\eta}\psi_\alpha(\eta;\nu) = -\frac{\nu+1}{\nu}\eta\left(1+\frac{\eta^2}{\nu}\right)^{-1}\psi_\alpha(\eta;\nu) + \frac{\alpha}{2}\psi_0(\eta;\nu)\psi_0\left(\alpha\eta\sqrt{\frac{\nu+1}{\nu+\eta^2}};\nu+1\right)\left\{\sqrt{\frac{\nu+1}{\nu+\eta^2}} - \frac{\sqrt{\nu+1}}{(\nu+\eta^2)^{3/2}}\eta^2\right\}.$$

Refs. [12,13] derived explicit expressions for the cumulative distribution function when $\nu = 1,2,3,4$. Ref. [12] for $\nu = 1$,

$$\Psi_\alpha(\eta;1) = \frac{1}{\pi}\left[\tan^{-1}(\eta) + \cos^{-1}\left\{\frac{\alpha}{\sqrt{(1+\alpha^2)(1+\eta^2)}}\right\}\right]$$

and Ref. [13] for $\nu = 2,3,4$, for example,

$$\Psi_\alpha(\eta;2) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}(\alpha) + \frac{\eta}{\sqrt{2+\eta^2}}\left\{\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{\alpha\eta}{\sqrt{2+\eta^2}}\right)\right\}.$$

The model $\mathcal{M}_{\alpha,\nu}$ is connected to several other important models. When $\nu = 1$, it reduces to the new skew–Cauchy link model. As $\nu \longrightarrow \infty$, it converges to the skew–probit link model, represented by $\mathcal{M}_\alpha$. When $\alpha = 0$, the model simplifies to the t-link model, represented by $\mathcal{M}_\nu$. When $\alpha = 0$ and $\nu = 1$, it becomes the Cauchy link model. Finally, as $\alpha = 0$ and $\nu \longrightarrow \infty$, the model converges to the probit link model, represented by $\mathcal{M}_0$. These relationships are illustrated in Figure 1.

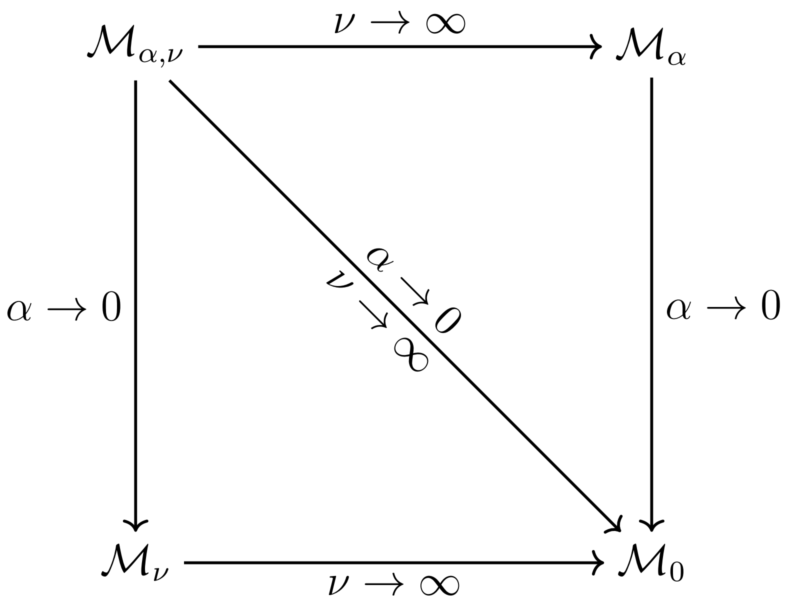

Figure 1. Some of the relationships between models derived from the $\mathcal{M}_{\alpha,\nu}$ model.

Figure 2 displays the probability as a function of η for selected values of the parameters α and ν. This figure is highly informative. When $\alpha = 0$, the probability approaches both zero and one at the same rate. However, when $\alpha > 0$, the probability approaches zero more rapidly than it does one. The inverse behavior is observed when $\alpha < 0$. Additionally, as the degrees of freedom decrease, the tails of the distribution become heavier. These characteristics highlight the advantage of this model compared to those that use symmetric links, such as $\mathcal{M}_0$ and $\mathcal{M}_\nu$, as well as models that employ asymmetric links, such as $\mathcal{M}_\alpha$. Furthermore, this model can adapt to the aforementioned models, making it a flexible link option.

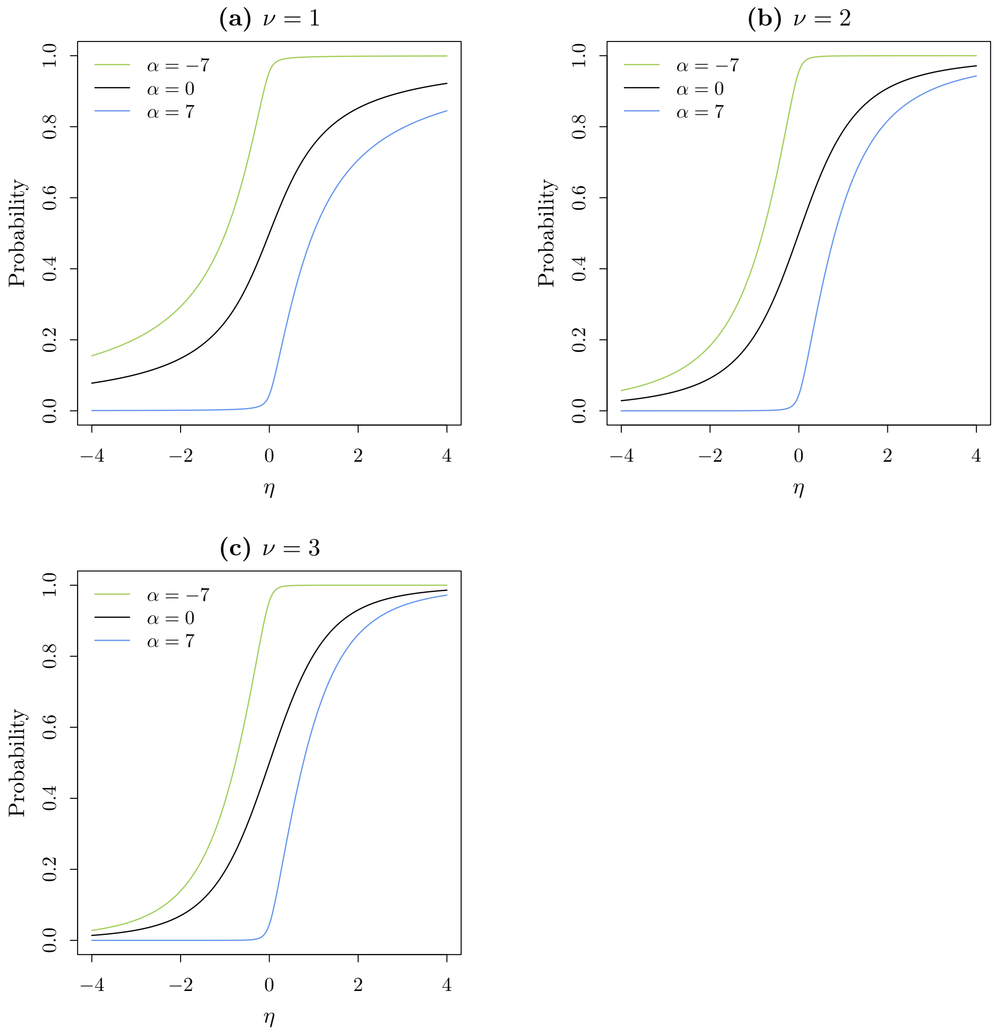

Figure 2. Probability curves of $\mathcal{M}_{\alpha,\nu}$ for some values of α and ν.

2.2. Penalized Log-Likelihood Function

In the Student's t-distribution, estimating the degrees of freedom parameter presents considerable challenges, which also impact the proposed model. According to [14], in a general regression model, two estimation scenarios are considered. In the first, all parameters are estimated simultaneously, while in the second, a subset of parameters is estimated, with the remaining parameters fixed at their true values. In this work, the parameters $\boldsymbol{\beta}$ and α will be estimated, while the parameter ν is fixed at its true value, ν_0.

Consider the complete dataset, $\mathcal{D} = \{(y_i, n_i, \mathbf{x}_i)\}$ and $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$. The log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \ell_i(\boldsymbol{\theta}), \tag{4}$$

where

$$\ell_i(\boldsymbol{\theta}) \propto y_i \log(p_i) + (n_i - y_i) \log(1 - p_i)$$

with parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ of dimension $(k+1) \times 1$.

A second challenge arises in the stability of estimating the skewness parameter in the skew–t distribution, a problem that also impacts the proposed model, as it relies on the cumulative distribution function of this distribution. According to [8], available methods to address this issue, both frequentist and Bayesian, are limited to simple scenarios. This study adopts the penalized log-likelihood approach, which necessitates an appropriate penalty function. This function, also known as shrinkage estimation, semi-Bayes, or partial-Bayes, is implemented without requiring a strictly Bayesian justification [15].

Given $\mathcal{D}$, the penalized log-likelihood function is given by

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - Q(\alpha), \tag{5}$$

where $Q(\cdot)$ is the penalty function.

Ref. [8] proposes that the penalization function should meet the following properties:

1. $Q(\alpha) \geq 0$
2. $Q(\alpha)\big|_{\alpha=0} = 0$ and
3. $\lim_{\alpha \to \pm\infty}(\alpha) = \infty$.

The penalization function that we can derive from the proposal of [8] is

$$Q(\alpha) = c_1 \log(1 + c_2\alpha^2), \tag{6}$$

where c_1 and c_2 are positive constants. In [3,4], this penalty function was incorporated into the skew–probit link model for binary response data without prior knowledge of its formulation. In [3], from a Bayesian perspective, the values were set as $c_1 = 3/2$ and $c_2 = 1$. Meanwhile, in [4], a frequentist approach was taken, setting $c_1 = 1$ and $c_2 = 4/25$. These values were derived using the kernel of the density of the Student's t distribution, based on the following relationship:

$$\exp\{-Q(\alpha)\} = \left(1 + \frac{\alpha^2}{\sigma^2\nu}\right)^{-(\nu+1)/2}.$$

Ref. [3] used the parameters $\sigma^2 = 1/2$ and $\nu = 2$, while [4] employed $\sigma^2 = 25/4$ and $\nu = 1$. In the estimation scenario, where ν is assumed to be known, we adopt the approach of [8], selecting the following values:

$$c_1 = \frac{1}{4\nu_2} \qquad \text{and} \qquad c_2 = \frac{\nu_2}{\nu_1},$$

where

$$\nu_1 = \frac{(\nu+2)(\nu+3)}{3(\nu+1)^2} \qquad \text{and} \qquad \nu_2 = 0.285 \times \left(1 + \frac{4}{\nu + 0.577}\right).$$

Figure 3 shows the behavior of the penalty functions for both types of links.

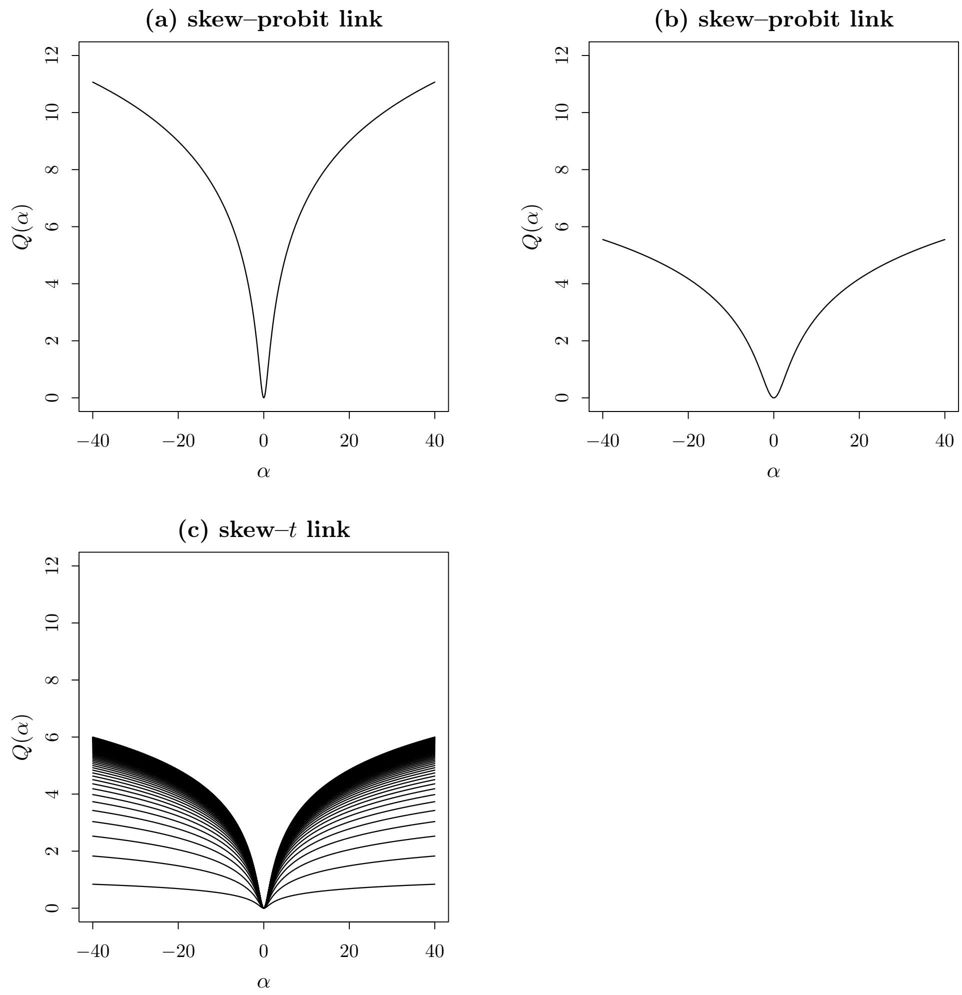

Figure 3. Penalty functions for link models: (**a**) skew–probit by Bazán et al. (2010) [3]; (**b**) skew–probit by Lee and Sinha (2019) [4]; and (**c**) skew–*t* by Azzalini and Arellano-Valle (2013) [8].

3. Parameter Estimation

3.1. Score Function

The estimation of the model parameters with skew–*t* link is carried out using the maximum penalized likelihood method. The penalized score function, obtained as the partial derivative of the penalized log-likelihood function with respect to the parameters, provides the estimation equations that must be solved to find the penalized maximum likelihood estimators. The strategy of fitting a model based on the penalized likelihood function has been widely used in various fields. Moreover, the asymptotic properties of these estimators have been extensively studied and documented in the literature for different models; see, for example, [16–18]. A key reference on this method is the book by [19].

Given the function (5) and $\mathcal{D}$, the penalized score function of $\boldsymbol{\theta}$ is given by

$$\mathbf{s}_{\mathrm{p}}(\boldsymbol{\theta}) = \frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{s}_{\mathrm{p}}^{\beta}(\boldsymbol{\theta}) \\ \mathrm{s}_{\mathrm{p}}^{\alpha}(\boldsymbol{\theta}) \end{bmatrix}, \quad (7)$$

where

$$\begin{aligned}\mathbf{s}_{\mathrm{p}}^{\beta}(\boldsymbol{\theta}) &= \mathbf{X}^{\top}\mathrm{diag}(\mathbf{u}_1)(\mathbf{y}-\boldsymbol{\mu}) \\ \mathrm{s}_{\mathrm{p}}^{\alpha}(\boldsymbol{\theta}) &= \mathbf{u}_2^{\top}(\mathbf{y}-\boldsymbol{\mu}) - \frac{2c_1c_2\alpha}{1+c_2\alpha^2}.\end{aligned}$$

In (7), $\mathbf{u}_1 = (u_{11}, \ldots, u_{1m})^{\top}$ and $\mathbf{u}_2 = (u_{21}, \ldots, u_{2m})^{\top}$, where

$$u_{1i} = \frac{v_{1i}}{p_i(1-p_i)} \qquad \text{and} \qquad u_{2i} = \frac{v_{2i}}{p_i(1-p_i)}$$

with

$$v_{1i} = \psi_{\alpha}(\eta_i;\nu) \qquad \text{and} \qquad v_{2i} = 2\int_{-\infty}^{\eta_i} t\sqrt{\frac{\nu+1}{\nu+t^2}}\,\psi_0(t;\nu)\psi_0\left(\alpha t\sqrt{\frac{\nu+1}{\nu+t^2}};\nu+1\right)\mathrm{d}t.$$

Also, $\boldsymbol{\mu} = (n_1p_1, \ldots, n_mp_m)^{\top}$, $\mathbf{y} = (y_1, \ldots, y_m)^{\top}$ and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^{\top}$.

The maximum penalized likelihood estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{p}}$ of $\boldsymbol{\theta}$ is obtained by solving the score equation,

$$\mathbf{s}_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}_{\mathrm{p}}) = \mathbf{0},$$

with $\mathbf{0}$ being the zero vector of dimension $(k+1)\times 1$.

3.2. Asymptotic Results

In asymptotic theory, maximum penalized likelihood estimators are shown to be consistent and have an approximate multivariate normal distribution in large samples. This approach allows us to construct confidence intervals and perform hypothesis tests on the model parameters, using the expected penalized information matrix and its inverse.

Given the function (5) and $\mathcal{D}$, the expectation of the penalized information matrix is given by

$$\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{s}_{\mathrm{p}}(\boldsymbol{\theta})\mathbf{s}_{\mathrm{p}}(\boldsymbol{\theta})^{\top}] = \begin{bmatrix}\boldsymbol{\mathcal{I}}_{11} & \mathbf{i}_{12} \\ \mathbf{i}_{12}^{\top} & i_{22}\end{bmatrix}, \tag{8}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{I}}_{11} &= \mathbf{X}^{\top}\mathrm{diag}(\mathbf{u}_1\circ\mathbf{v}_1)\mathbf{X} \\ \mathbf{i}_{12} &= \mathbf{X}^{\top}\mathrm{diag}(\mathbf{u}_1)\mathbf{v}_2 \\ i_{22} &= \mathrm{trace}\{\mathrm{diag}(\mathbf{u}_2\circ\mathbf{v}_2)\} + \frac{2a_1a_2(1-a_2)\alpha^2}{(1+a_2\alpha^2)^2},\end{aligned}$$

with $\circ$ denoting the Hadamard product. In (8), $\mathbf{v}_1 = (v_{11}, \ldots, v_{1m})^{\top}$ and $\mathbf{v}_2 = (v_{21}, \ldots, v_{2m})^{\top}$.

Consequently, the inverse from $\boldsymbol{\mathcal{I}}_{\mathrm{p}}(\boldsymbol{\theta})$ is

$$\boldsymbol{\Omega}_{\mathrm{p}}(\boldsymbol{\theta}) = \begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{12}^{\top} & \omega_{22}\end{bmatrix}, \tag{9}$$

where

$$\begin{aligned}\boldsymbol{\Omega}_{11} &= \boldsymbol{\mathcal{I}}_{11}^{-1}\{\mathbf{I}_k + \mathbf{i}_{12}\omega_{22}\mathbf{i}_{12}^{\top}\boldsymbol{\mathcal{I}}_{11}^{-1}\} \\ \boldsymbol{\omega}_{12} &= -\boldsymbol{\mathcal{I}}_{11}^{-1}\mathbf{i}_{12}\omega_{22} \\ \omega_{22} &= \{i_{22} - \mathbf{i}_{12}^{\top}\boldsymbol{\mathcal{I}}_{11}^{-1}\mathbf{i}_{12}\}^{-1},\end{aligned}$$

with $\mathbf{I}_k$ being the identity matrix of dimension $k\times k$. The convenient formula for the inverse of a partitioned matrix can be seen, for example, in [20].

Assuming that $\widehat{\boldsymbol{\theta}}_{\text{p}}$ is consistent, we can replace (9) with $\boldsymbol{\Omega}_{\text{p}}(\widehat{\boldsymbol{\theta}}_{\text{p}})$. Consequently,

$$\widehat{\boldsymbol{\theta}}_{\text{p}} \dot{\sim} N_{k+1}(\boldsymbol{\theta}, \boldsymbol{\Omega}_{\text{p}}(\widehat{\boldsymbol{\theta}}_{\text{p}})). \tag{10}$$

The result (10) is essential for constructing confidence intervals for both β_j and α.

In this article, we utilize the open-source software R-4.4.1 [21], specifically the `maxLik` package [22]. This package provides a range of optimization routines that facilitate the implementation of advanced features, such as fixing specific parameter values, which is essential for the proposed model in this study. For the estimation of the parameters of the model $\mathcal{M}_{\alpha,\nu}$, the following procedure is adopted for selecting the initial values. The regression coefficients are initialized based on the estimates from the baseline model $\mathcal{M}_0$, represented as $\widehat{\boldsymbol{\beta}}^{\text{probit}}_{\text{initial}} = (\widehat{\beta}_{01}, \ldots, \widehat{\beta}_{0k})^\top$, where $\widehat{\beta}_{0j}$ corresponds to the estimates obtained under the probit model. Additionally, the asymmetry parameter α is initialized at zero, $\widehat{\alpha}_{\text{initial}} = 0$. These initial values are chosen to ensure a stable foundation for the iterative optimization process.

4. Simulation Study

This section aims to verify the accuracy of the parameter estimates for the $\mathcal{M}_{\alpha,\nu}$ model under the proposed methodology. It also assesses the adaptability of the proposed model to two asymmetric links and describes the process for determining the parameter ν.

4.1. Study 1

This study will address 2 scenarios for generating datasets, creating a total of 500 datasets for each scenario. The construction of the complete datasets follows the outlined procedure: in Step (1), the values for $\boldsymbol{\beta}$ and α are chosen, where $\nu_0 = 1, 2, 3$. For scenario 1, the values are set as $\theta_1 = \beta_1 = 0.5$, $\theta_2 = \beta_2 = 1$, $\theta_3 = \beta_3 = 1.5$ and $\theta_4 = \alpha = 0.2$. For scenario 2, the values are set as $\theta_1 = \beta_1 = -0.7$, $\theta_2 = \beta_2 = 1.3$, $\theta_3 = \beta_3 = -2.5$ and $\theta_4 = \alpha = 2$. In Step (2), the values for m and n_i are chosen, where $n_i = 50$. For scenario 1, the values are chosen as $m = 20, 40, \ldots, 200$. For scenario 2, the values are chosen as $m = 200, 400, \ldots, 2000$. In Step (3), x_{i2} and x_{i3} are generated using the normal distribution. For scenario 1, the values $x_{i2} \sim N(0, 2)$ and $x_{i3} \sim N(0, 3)$ are generated. For scenario 2, the values $x_{i2} \sim N(-3, 7)$ and $x_{i3} \sim N(3, 7)$ are generated. In Step (4), p_i is structured, $p_i = \Psi_\alpha(\beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}; \nu_0)$. In Step (5), y_i is generated using the binomial distribution, $y_i \overset{ind.}{\sim} \text{Binomial}(n_i, p_i)$.

Assuming that $\widehat{\theta}^{(j)}$ is the estimator of θ in the j-th dataset crafted, the Monte Carlo mean of these estimates is consequently

$$\bar{\widehat{\theta}} = \frac{1}{500} \sum_{j=1}^{500} \widehat{\theta}^{(j)}. \tag{11}$$

The absolute bias (AB) is defined as

$$\text{AB}(\widehat{\theta}) = |\bar{\widehat{\theta}} - \theta|. \tag{12}$$

The Monte Carlo standard deviation (SD) and the root mean square error (RMSE) are expressed as

$$\text{SD}(\widehat{\theta}) = \sqrt{\frac{1}{499} \sum_{j=1}^{500} (\widehat{\theta}^{(j)} - \bar{\widehat{\theta}})^2} \tag{13}$$

and

$$\text{RMSE}(\widehat{\theta}) = \sqrt{\text{SD}^2(\widehat{\theta}) + \text{AB}^2(\widehat{\theta})}, \tag{14}$$

respectively.

The results of the estimations for scenarios 1 and 2 are detailed in Tables 1 and 2, respectively. Overall, the absolute bias (12) values are close to zero. Additionally, both the SD (13) and RMSE (14) tend to decrease as the simulation size m increases, indicating greater precision in parameter estimation. These results are illustrated in Figures 4 and 5, corresponding to scenarios 1 and 2, respectively. In both figures, a diagonal line highlights the proximity between the SD and RMSE values, confirming that absolute biases are practically negligible.

Table 1. Simulation study results for the first scenario: AB, SD, and RMSE in the parameter estimation of $\mathcal{M}_{\alpha,\nu}$ under three values of ν and five values of m.

Parameter	m	$\nu = 1$			$\nu = 2$			$\nu = 3$		
		AB	**SD**	**RMSE**	**AB**	**SD**	**RMSE**	**AB**	**SD**	**RMSE**
β_1	20	0.001	0.304	0.304	0.024	0.330	0.331	0.033	0.376	0.377
	60	0.008	0.165	0.165	0.005	0.190	0.189	0.008	0.214	0.214
	100	0.003	0.121	0.120	0.000	0.140	0.140	0.003	0.166	0.166
	140	0.002	0.112	0.112	0.001	0.130	0.129	0.004	0.154	0.154
	180	0.000	0.092	0.092	0.005	0.105	0.105	0.006	0.125	0.125
β_2	20	0.002	0.148	0.148	0.014	0.128	0.129	0.029	0.123	0.126
	60	0.001	0.074	0.074	0.005	0.064	0.064	0.010	0.063	0.063
	100	0.004	0.056	0.056	0.000	0.047	0.047	0.004	0.046	0.046
	140	0.000	0.044	0.044	0.003	0.037	0.037	0.006	0.036	0.036
	180	0.001	0.043	0.043	0.002	0.036	0.036	0.005	0.035	0.035
β_3	20	0.002	0.173	0.173	0.021	0.149	0.150	0.046	0.147	0.154
	60	0.000	0.091	0.091	0.009	0.079	0.080	0.017	0.080	0.082
	100	0.007	0.071	0.071	0.000	0.059	0.059	0.006	0.060	0.060
	140	0.000	0.060	0.060	0.004	0.050	0.050	0.009	0.050	0.051
	180	0.002	0.054	0.054	0.003	0.046	0.046	0.007	0.045	0.045
α	20	0.018	0.254	0.255	0.042	0.836	0.836	0.123	1.925	1.927
	60	0.010	0.127	0.127	0.016	0.194	0.194	0.029	0.251	0.253
	100	0.004	0.092	0.092	0.007	0.141	0.141	0.008	0.188	0.188
	140	0.004	0.083	0.083	0.004	0.125	0.125	0.005	0.167	0.167
	180	0.002	0.068	0.068	0.001	0.101	0.101	0.002	0.137	0.137

Table 2. Simulation study results for the second scenario: AB, SD, and RMSE in the parameter estimation of $\mathcal{M}_{\alpha,\nu}$ under three values of ν and five values of m.

Parameter	m	$\nu = 1$			$\nu = 2$			$\nu = 3$		
		AB	**SD**	**RMSE**	**AB**	**SD**	**RMSE**	**AB**	**SD**	**RMSE**
β_1	200	0.054	0.301	0.306	0.061	0.327	0.332	0.074	0.378	0.384
	600	0.008	0.139	0.139	0.009	0.153	0.153	0.015	0.180	0.180
	1000	0.012	0.113	0.114	0.013	0.124	0.125	0.021	0.148	0.149
	1400	0.001	0.092	0.092	0.003	0.099	0.099	0.005	0.119	0.119
	1800	0.006	0.081	0.081	0.005	0.092	0.092	0.001	0.104	0.104
β_2	200	0.035	0.210	0.213	0.035	0.203	0.206	0.038	0.219	0.222
	600	0.004	0.097	0.097	0.003	0.095	0.095	0.006	0.104	0.104
	1000	0.006	0.080	0.081	0.007	0.078	0.078	0.010	0.087	0.087
	1400	0.002	0.067	0.067	0.002	0.064	0.064	0.003	0.070	0.070
	1800	0.003	0.060	0.060	0.003	0.060	0.060	0.001	0.063	0.063
β_3	200	0.069	0.385	0.391	0.069	0.378	0.384	0.075	0.412	0.419
	600	0.009	0.185	0.185	0.007	0.180	0.180	0.012	0.197	0.197
	1000	0.011	0.154	0.154	0.012	0.150	0.151	0.019	0.166	0.167
	1400	0.002	0.128	0.128	0.004	0.122	0.122	0.005	0.135	0.135
	1800	0.006	0.111	0.111	0.005	0.112	0.113	0.002	0.119	0.119
α	200	0.012	0.181	0.181	0.069	0.311	0.311	0.075	1.095	1.097
	600	0.001	0.088	0.087	0.007	0.149	0.149	0.014	0.218	0.218
	1000	0.002	0.075	0.075	0.012	0.126	0.126	0.003	0.179	0.179
	1400	0.003	0.061	0.061	0.004	0.099	0.099	0.008	0.140	0.140
	1800	0.000	0.052	0.052	0.005	0.092	0.092	0.011	0.126	0.126

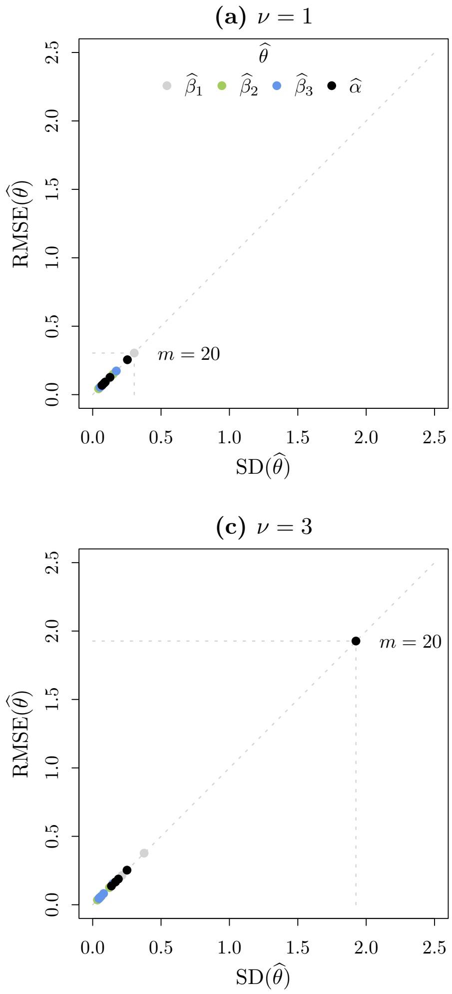

Figure 4. Simulation study results for the first scenario: SD versus RMSE in the parameter estimation of $\mathcal{M}_{\alpha,\nu}$ under three values of ν and ten values of m.

However, it is important to note that in the scenarios with less available information, the variability of the estimators can increase significantly. This effect is observed in Figure 4, where for $m = 20$ and $\nu = 3$, the SD of the skewness parameter estimator reaches a value of 1.925, while the RMSE is 1.927. This behavior explains why the SD and RMSE values for the skewness estimator deviate from the main group of estimates. Additionally, this dispersion may be influenced by the inherent variability in data generation, which emphasizes the need for a more detailed analysis in scenarios with smaller sample sizes.

4.2. Study 2

This subsection presents a second simulation study aimed at evaluating the adaptability of the proposed model in comparison with other asymmetric link models, specifically the cloglog and loglog links. Additionally, the process for setting the parameter ν is detailed. The dataset construction for each link follows the procedure outlined below. In Step (1), the value of $\beta = 1$ is set. In Step (2), the value $m = 70$ is selected. In Step (3), the values $n_i \sim \text{Poisson}(\lambda = 15)$ and $x_{i2} \sim N(0, 3)$ are generated. In Step (4), the probability p_i is defined as

$$p_i = \begin{cases} 1 - \exp(-\exp(x_i)), & \text{for the cloglog link (Scenario 1)} \\ \exp(-\exp(-x_i)), & \text{for the loglog link (Scenario 2).} \end{cases}$$

In Step (5), finally, the responses y_i are generated using a binomial distribution, i.e., $y_i \overset{ind.}{\sim} \text{Binomial}(n_i, p_i)$.

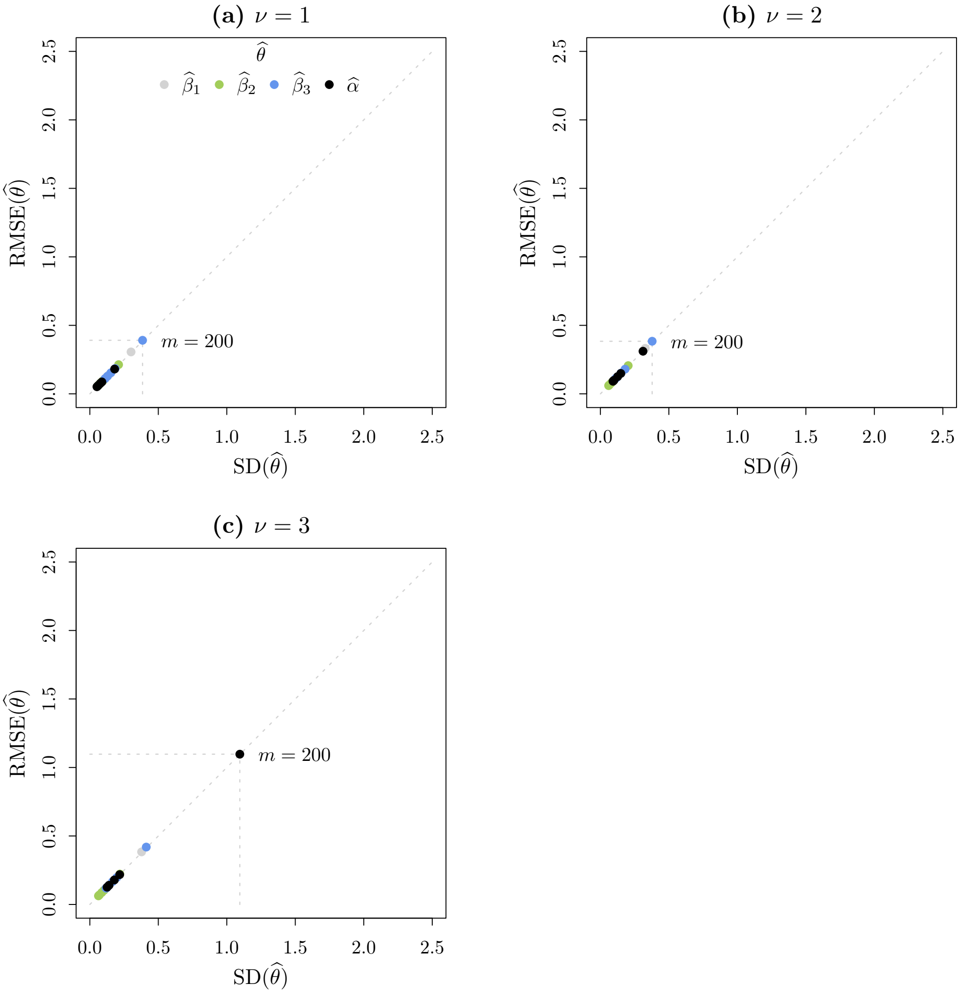

Figure 5. Simulation study results for the second scenario: SD versus RMSE in the parameter estimation of $\mathcal{M}_{\alpha,\nu}$ under three values of ν and ten values of m.

In both scenarios, the cloglog, loglog, and $\mathcal{M}_{\alpha,\nu}$ models are fitted. The determination of the degrees of freedom is performed using the profile likelihood procedure described by [23]. Figure 6 displays the profile likelihood curves as a function of ν. The sequence used for ν is $0.1, 0.2, \ldots, 30$ (with an increment of 0.1). For data generated using the cloglog link, the appropriate degrees of freedom, where the profile-penalized log-likelihood reaches its maximum, is $\nu_0 = 2.8$. For data generated using the loglog link, the appropriate degrees of freedom, corresponding to the maximum of the profile-penalized log-likelihood, is $\nu_0 = 1.9$.

Table 3 focuses on the parameter estimates for the $\mathcal{M}_{\alpha,\nu}$ model in both scenarios. The estimates of β are very close to the true values, and the 95% confidence interval includes the true value. The 95% confidence interval for α does not include zero, which suggests the continued need for a model such as the one proposed, as it is adapting to datasets generated from asymmetric links.

Reinforcing the flexibility of the proposed model, the formula for calculating the distance between two cumulative distribution functions of interest is presented below,

$$\mathrm{D}(x) = \begin{cases} |\Psi_{-0.431}(1.085x; 2.8) - \{1 - \exp(-\exp(x))\}|, & \text{for the cloglog link} \\ |\Psi_{0.523}(1.121x; 1.9) - \exp(-\exp(-x))|, & \text{for the loglog link.} \end{cases}$$

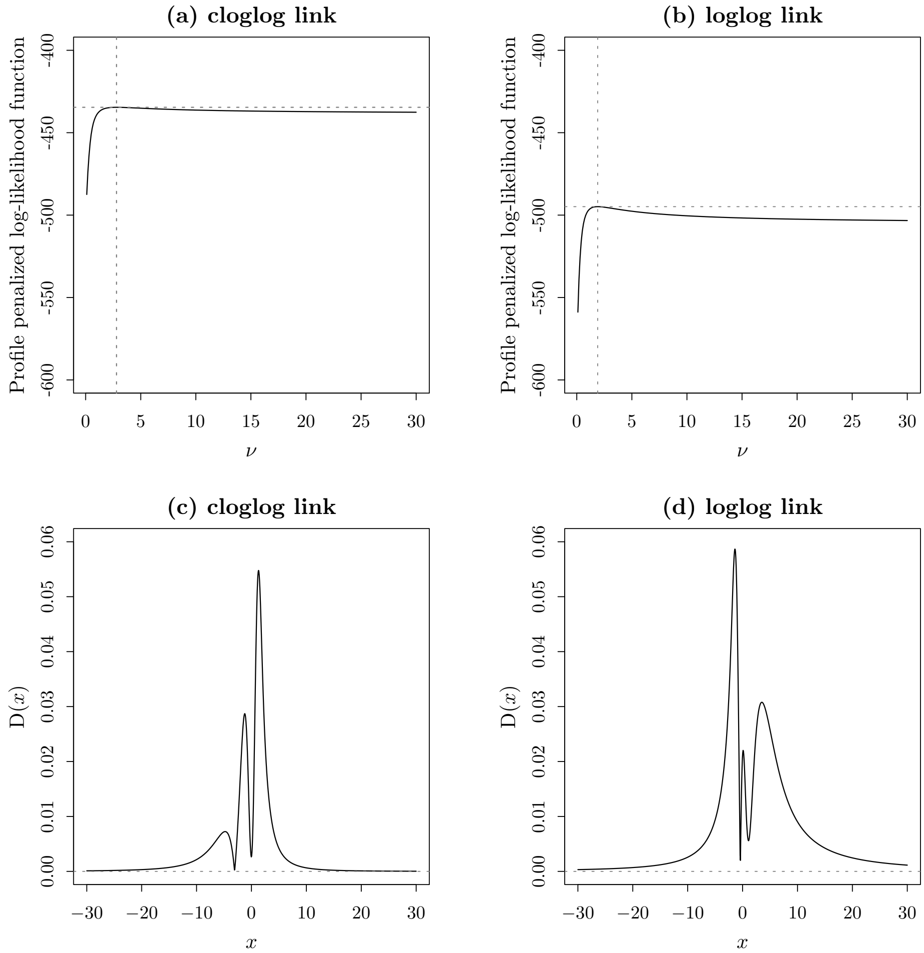

Figure 6. Profile log-likelihood function and distance between the $\mathcal{M}_{\alpha,\nu}$ model and the cloglog and loglog models: (**a**) profile log-likelihood, scenario 1; (**b**) profile log-likelihood, scenario 2; (**c**) distance, scenario 1; (**d**) distance, scenario 2.

Table 3. Summary of parameter estimation for the cloglog, loglog and $\mathcal{M}_{\alpha,\nu}$ models in simulated data.

Scenario	Model	Parameter	Estimation	SD	2.5%	97.5%
1	cloglog	β	0.952	0.056	0.844	1.061
	$\mathcal{M}_{\alpha,\nu}$	β	1.085	0.077	0.932	1.237
		α	−0.431	0.066	−0.559	−0.302
		$\nu_0 = 2.8$				
2	loglog	β	0.951	0.055	0.843	1.059
	$\mathcal{M}_{\alpha,\nu}$	β	1.121	0.080	0.964	1.278
		α	0.523	0.074	0.377	0.668
		$\nu_0 = 1.9$				

Figure 6 illustrates how this distance behaves, with the horizontal axis at zero.

A notable implication of Table 3 and Figure 6 is that the inverse of the skew–t link function approximates the inverse of the cloglog and loglog links,

$$\Psi_{-0.431}(1.085\eta; 2.8) \approx 1 - \exp(-\exp(\eta)) \tag{15}$$

and

$$\Psi_{0.523}(1.121\eta; 1.9) \approx \exp(-\exp(-\eta)), \tag{16}$$

respectively.

5. Applications

In this section, we motivate the proposed link by comparing it with the probit and skew–probit links by fitting them to two real datasets. The first dataset has binomial response data, while the second dataset has binary response data.

5.1. Women's Labor

The National Institute of Statistics (Instituto Nacional de Estadística: INE) of Bolivia conducts the Household Survey (Encuesta de Hogares: EH) semi-annually with the aim of assessing income poverty and collecting essential socioeconomic data to monitor the Sustainable Development Goals. The EH 2021 covers topics such as health, education, employment, non-labor income, and housing, and is publicly available on the INE website at https://www.ine.gob.bo, accessed on 1 January 2024 .

For this illustration, specific data from the EH 2021 were processed. The following variables were considered: the total number of women in a specific age group (n_i), the number of women who worked at least one hour in the previous week (y_i), and the age of the women in years (x_i). Additionally, this process was carried out by differentiating between urban and rural areas, considering women aged 10 to 55, in accordance with current legislation, such as the "Law 321/2013-2014 of the Code of Girls, Boys, and Adolescents" and the "Pension Law 065". The dataset is presented in Table 4, which also includes the proportion of women who worked at least one hour in the previous week (p_i^*).

It is important to note that women participating in labor activities demonstrated a wide diversity of productive and commercial occupations.

Table 5 presents the parameter estimates, standard deviations and 95% confidence intervals for models $\mathcal{M}_0$ (with probit link), $\mathcal{M}_\alpha$ (with skew–probit link) and $\mathcal{M}_{\alpha,\nu}$ (with skew–t link). It also includes model comparison criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan–Quinn Information Criterion (HQIC). The information criterion values unanimously indicate that the model to consider for analysis is $\mathcal{M}_{\alpha,\nu}$.

Figure 7 shows the proportion of women who participated in work activities according to their area of residence, with the lines representing the fitted model. This figure demonstrates that, when comparing models $\mathcal{M}_0$ and $\mathcal{M}_\alpha$ with $\mathcal{M}_{\alpha,\nu}$, the first two do not fit the data in any of the three scenarios. This underscores the need for a model with a flexible link function, which $\mathcal{M}_{\alpha,\nu}$ successfully addresses.

Table 4. Women's labor data.

Urban					Rural				
i	x_i	n_i	y_i	p_i^*	i	x_i	n_i	y_i	p_i^*
1	10	368	1	0.003	1	10	136	9	0.066
2	11	296	5	0.017	2	11	95	14	0.147
3	12	296	4	0.014	3	12	108	15	0.139
4	13	327	12	0.037	4	13	108	18	0.167
5	14	308	12	0.039	5	14	98	33	0.337
6	15	325	16	0.049	6	15	99	27	0.273
7	16	286	27	0.094	7	16	82	25	0.305
8	17	301	34	0.113	8	17	88	29	0.330
9	18	313	58	0.185	9	18	61	20	0.328
10	19	275	61	0.222	10	19	58	19	0.328
11	20	306	94	0.307	11	20	59	22	0.373
12	21	338	101	0.299	12	21	72	29	0.403
13	22	295	104	0.353	13	22	47	17	0.362
14	23	299	116	0.388	14	23	53	22	0.415
15	24	292	119	0.408	15	24	49	20	0.408
16	25	324	144	0.444	16	25	64	32	0.500
17	26	310	151	0.487	17	26	65	33	0.508
18	27	291	145	0.498	18	27	63	38	0.603
19	28	307	156	0.508	19	28	65	38	0.585
20	29	255	130	0.510	20	29	48	26	0.542

Table 4. *Cont.*

Urban					Rural				
i	x_i	n_i	y_i	p_i^*	i	x_i	n_i	y_i	p_i^*
21	30	319	186	0.583	21	30	47	26	0.553
22	31	206	111	0.539	22	31	45	23	0.511
23	32	263	148	0.563	23	32	57	29	0.509
24	33	240	135	0.562	24	33	58	35	0.603
25	34	226	140	0.619	25	34	52	30	0.577
26	35	319	181	0.567	26	35	65	42	0.646
27	36	253	162	0.640	27	36	66	40	0.606
28	37	233	151	0.648	28	37	50	32	0.640
29	38	291	176	0.605	29	38	62	43	0.694
30	39	221	147	0.665	30	39	50	34	0.680
31	40	272	177	0.651	31	40	61	43	0.705
32	41	209	144	0.689	32	41	62	40	0.645
33	42	245	158	0.645	33	42	58	41	0.707
34	43	203	129	0.635	34	43	55	36	0.655
35	44	171	106	0.620	35	44	49	41	0.837
36	45	235	160	0.681	36	45	68	49	0.721
37	46	153	103	0.673	37	46	36	21	0.583
38	47	159	100	0.629	38	47	44	29	0.659
39	48	195	124	0.636	39	48	58	39	0.672
40	49	174	117	0.672	40	49	46	28	0.609
41	50	197	121	0.614	41	50	71	46	0.648
42	51	151	99	0.656	42	51	44	33	0.750
43	52	157	101	0.643	43	52	55	37	0.673
44	53	143	101	0.706	44	53	41	24	0.585
45	54	132	76	0.576	45	54	30	18	0.600
46	55	154	95	0.617	46	55	45	32	0.711

Table 5. Summary of parameter estimation for $\mathcal{M}_0$, $\mathcal{M}_\alpha$ and $\mathcal{M}_{\alpha,\nu}$, along with model selection performance evaluation using AIC, BIC, and HQIC on data related to women's labor.

Area	Model	Parameter	Estimation	SD	2.5%	97.5%	AIC	BIC	HQIC
Urban	$\mathcal{M}_0$	β_1	−1.653	0.035	−1.721	−1.585	13,544.280	13,547.940	13,545.650
		β_2	0.048	0.001	0.046	0.050			
	$\mathcal{M}_\alpha$	β_1	−0.346	0.020	−0.385	−0.306	13,442.100	13,447.590	13,444.160
		β_2	0.031	0.001	0.030	0.032			
		α	4.152	0.401	3.367	4.937			
	$\mathcal{M}_{\alpha,\nu}$	β_1	−2.531	0.123	−2.772	−2.290	12,821.050	12,826.540	12,823.110
		β_2	0.157	0.006	0.145	0.168			
		α	4.111	0.418	3.292	4.931			
		$\nu_0 = 0.40$							
Rural	$\mathcal{M}_0$	β_1	−1.134	0.060	−1.251	−1.017	3595.020	3598.677	3596.390
		β_2	0.036	0.002	0.033	0.040			
	$\mathcal{M}_\alpha$	β_1	−1.134	0.180	−1.487	−0.780	3597.020	3602.506	3599.075
		β_2	0.036	0.002	0.033	0.040			
		α	0.000	0.213	−0.417	0.417			
	$\mathcal{M}_{\alpha,\nu}$	β_1	−1.699	0.222	−2.134	−1.265	3546.702	3552.188	3548.757
		β_2	0.110	0.010	0.092	0.129			
		α	1.003	0.261	0.493	1.514			
		$\nu_0 = 0.40$							
Total	$\mathcal{M}_0$	β_1	−1.528	0.030	−1.587	−1.469	17,201.110	17,204.760	17,202.480
		β_2	0.045	0.001	0.043	0.047			
	$\mathcal{M}_\alpha$	β_1	−0.270	0.018	−0.306	−0.234	17,106.840	17,112.330	17,108.900
		β_2	0.029	0.001	0.028	0.030			
		α	3.784	0.339	3.118	4.449			
	$\mathcal{M}_{\alpha,\nu}$	β_1	−2.341	0.117	−2.571	−2.111	16,545.460	16,550.950	16,547.520
		β_2	0.144	0.005	0.134	0.154			
		α	2.338	0.196	1.953	2.722			
		$\nu_0 = 0.40$							

Table 6 presents the results of the Kolmogorov–Smirnov (K-S) test for assessing the normality of the residuals. Following the fitting of the three models, $\mathcal{M}_0$, $\mathcal{M}_\alpha$ and $\mathcal{M}_{\alpha,\nu}$, we calculated the estimated probabilities and Pearson residuals. The K-S test was then applied to compare the distribution of the residuals with a theoretical normal distribution. In this table, the D value represents the K-S statistic, which measures the maximum difference between the empirical and theoretical distributions. At a significance level of τ, the null hypothesis is rejected if the p-value falls below this threshold, indicating a departure from normality. Conversely, if the p-value exceeds this level, it suggests that the residuals can be considered normally distributed, thereby supporting the model's adequacy. Specifically, with $\tau = 0.05$, the p-values for both the urban area and the total area indicate that the proposed model is adequate. Furthermore, at $\tau = 0.1$, all p-values suggest that the proposed model is appropriate.

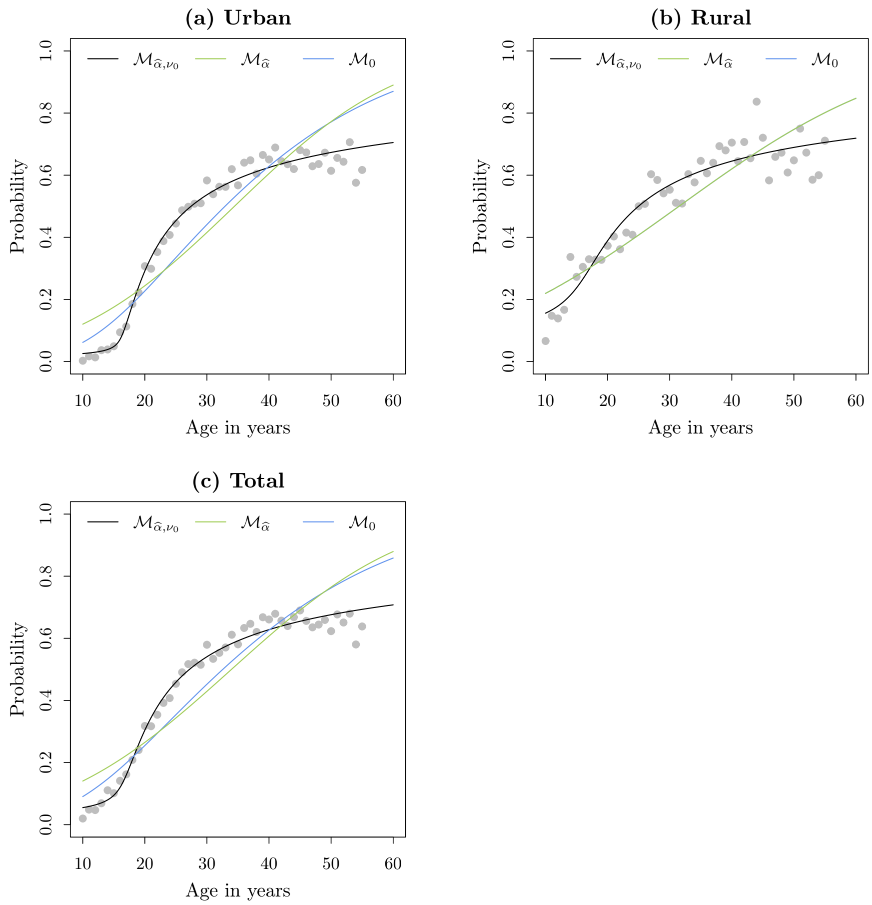

Figure 7. Proportion of women who participated in work activities for at least one hour during the previous week, according to the area of residence in Bolivia. The lines represent the fitted model.

Table 6. Kolmogorov–Smirnov test results for assessing the normality of residuals by area and model on data related to women's labor.

Area	Model	D	p-Value
Urban	$\mathcal{M}_0$	0.394	0.000
	$\mathcal{M}_\alpha$	0.370	0.000
	$\mathcal{M}_{\alpha,\nu}$	0.103	0.672
Rural	$\mathcal{M}_0$	0.180	0.089
	$\mathcal{M}_\alpha$	0.180	0.089
	$\mathcal{M}_{\alpha,\nu}$	0.134	0.351
Total	$\mathcal{M}_0$	0.431	0.000
	$\mathcal{M}_\alpha$	0.364	0.000
	$\mathcal{M}_{\alpha,\nu}$	0.099	0.724

5.2. Cardiovascular Disease

One of the leading causes of death in the world is cardiovascular disease (CVD). Ref. [24] proposes a methodology to predict cardiovascular diseases. The study [24] is from the Sleep Heart Health Study (SHHS), a multicenter cohort study implemented by the National Heart, Lung, and Blood Institute to determine the cardiovascular and other consequences of sleep-disordered breathing (the full dataset can be downloaded from https://sleepdata.org, accessed on 1 January 2024). In the original study, the SHHS was designed to determine whether sleep-disordered breathing and obstructive sleep apnea were a risk factor for cardiovascular and cerebrovascular disease [25]. The SHHS dataset is a prospective cohort study that recruited participant data from nine cohort studies. It contains the data of 5804 adults (40+ years), divided into 2 datasets: SHHS1 as a baseline for measuring the health status of the subjects, based on the polysomnography test and medical indices, collected between 1995 and 1998; and SHHS2, which contains the results of the selected cohort follow-up from January 2001 to June 2003, classifying the subjects who did or did not present cardiovascular disease (CVD) or related health problems during that period of time. Given that the dataset includes more than 1200 variables, the main challenge in CVD prediction is to select a subset of variables that might be most useful in explaining CVD. Although many alternatives have been proposed, we use the variables selected by [24] as its use in a machine learning model seems to outperform other options in terms of prediction ability.

The variables considered are the following: y, indicating cardiovascular disease (CVD) events (TRUE = a CVD event occurred, FALSE, otherwise) since Sleep Heart Health Study Visit One (Sleep Heart Health Study Visit One: SHHS1); x_2, indicating age at time of study in years, based on start date of Sleep Heart Health Study Visit One (SHHS1) Polysomnography (PSG) recording; x_3, indicating history of diabetes (Sleep Heart Health Study Visit One (SHHS1)); x_4, indicating Doctor of Medicine (MD)-reported coronary artery bypass graft (CABG) since Sleep Heart Health Study Visit One (SHHS1); x_5, indicating cigarette pack-years since Sleep Heart Health Study Visit One (SHHS1).

The minimum age in years at the time of the study is 39, the maximum age is 90, and the mean age is approximately 65 years. The proportion of patients with a history of diabetes is 0.067. The proportion of patients with coronary artery graft report is 0.035. The mean number of packs of cigarettes consumed per year by patients is approximately 13. The variable y is labeled with one and zero, which represent the presence or not, respectively, of CVD events from SHHS2. Out of a total of 2939 patients (patients belonging to the first and second studies), 749 had almost 1 of the CVD events, which shows an imbalance in the dataset.

The skew–t link model for binary response data is specified by

$$\begin{aligned} y_i &\overset{ind.}{\sim} \text{Bernoulli}(p_i) \\ p_i &= \Psi_\alpha(\mathbf{x}_i^\top \boldsymbol{\beta}; \nu), \end{aligned} \tag{17}$$

where $i = 1, 2, \ldots, 2939$.

Table 7 provides a summary of the parameter estimates and model comparison criteria applied to the dataset. The information criteria values indicate that the most suitable model for the analysis is $\mathcal{M}_{\alpha,\nu}$, which underscores once again the need for a model with a flexible link function, a need that $\mathcal{M}_{\alpha,\nu}$ effectively meets. It is worth noting that the estimated skewness parameter is 2.943. As shown in Figure 2, when $\alpha > 0$, the probability p_i tends to approach zero more rapidly than one, which explains the imbalance observed in the binary response data.

The assessment of residual normality using the K-S test showed that the K-S statistics for models $\mathcal{M}_0$, $\mathcal{M}_\alpha$ and $\mathcal{M}_{\alpha,\nu}$ were $D_{\mathcal{M}_0} = 0.306$, $D_{\mathcal{M}_\alpha} = 0.285$ and $D_{\mathcal{M}_{\alpha,\nu}} = 0.312$, respectively, with p-values less than 2.2×10^{-16}. These results suggest a significant deviation from normality in the residuals of all three models. However, the K-S test is known for its sensitivity to large sample sizes, which may result in detecting minor deviations from normality that do not necessarily affect the overall model fit.

Table 7. Summary of parameter estimation for $\mathcal{M}_0$, $\mathcal{M}_\alpha$ and $\mathcal{M}_{\alpha,\nu}$, along with model selection performance evaluation using AIC, BIC, and HQIC on data related to cardiovascular disease.

Model	Parameter	Estimation	SD	2.5%	97.5%	AIC	BIC	HQIC
$\mathcal{M}_0$	β_1	−3.826	0.203	−4.228	−3.429	2944.676	2974.605	2955.453
	β_2	0.045	0.003	0.039	0.051			
	β_3	0.532	0.097	0.344	0.720			
	β_4	0.831	0.134	0.573	1.092			
	β_5	0.007	0.001	0.004	0.009			
$\mathcal{M}_\alpha$	β_1	−1.156	0.187	−1.522	−0.790	2935.866	2971.781	2948.798
	β_2	0.021	0.002	0.016	0.026			
	β_3	0.303	0.063	0.179	0.426			
	β_4	0.504	0.096	0.315	0.692			
	β_5	0.003	0.001	0.002	0.005			
	α	4.129	1.218	1.742	6.515			
$\mathcal{M}_{\alpha,\nu}$	β_1	−2.095	0.312	−2.706	−1.484	2930.889	2966.803	2943.821
	β_2	0.035	0.004	0.027	0.043			
	β_3	0.590	0.133	0.329	0.851			
	β_4	1.177	0.315	0.560	1.794			
	β_5	0.007	0.001	0.004	0.009			
	α	2.943	0.723	1.527	4.360			
	$\nu_0 = 0.80$							

6. Discussion

The skew–t link model proposed in this paper provides a significant contribution to the field of binomial and binary response data analysis by addressing the limitations of traditional symmetric link functions like probit and logit, as well as the less flexible skew–probit link. This model introduces an additional level of flexibility through the incorporation of both an asymmetry parameter and degrees of freedom, enabling it to handle more complex data patterns. The effectiveness of the skew–t link model is demonstrated in both simulation studies and real-world applications, highlighting its potential as a valuable tool for researchers in various fields.

Another main contribution of the paper is the use of a penalized likelihood-based estimation method to stabilize the estimation process, particularly for the asymmetry parameter. Estimating this parameter can be challenging due to potential divergence, but the penalization strategy mitigates this issue, leading to more robust and reliable parameter estimates. This approach is essential in cases where traditional maximum likelihood estimation methods may fail, making the model more practical for real-world applications.

The comprehensive simulation studies conducted in this paper demonstrate the accuracy and reliability of the skew–t link model. In both scenarios tested, the model consistently exhibited low absolute bias and decreasing Monte Carlo standard deviations and root mean squared errors as the sample size increased. These findings underscore the model's ability to provide consistent and reliable parameter estimates across different data configurations. The second study, which highlights the model's adaptability to asymmetric link functions such as cloglog and loglog, further emphasizes its flexibility in handling a variety of data structures.

Additionally, an alternative approach that could complement the findings of this study is the implementation of a Bayesian perspective. Specifically, one could consider a hierarchical representation where the response y_i follows a binomial distribution given by $y_i|\boldsymbol{\beta}, \alpha, \nu \overset{ind.}{\sim} \text{Binomial}(n_i, p_i)$ with $p_i = \Psi_\alpha(\mathbf{x}_i^\top \boldsymbol{\beta}; \nu)$, where $\boldsymbol{\beta} \sim \pi_1(\boldsymbol{\beta})$, $\alpha \sim \pi_2(\alpha)$ and $\nu \sim \pi_3(\nu)$. Here, $\pi_1(\cdot)$, $\pi_2(\cdot)$ and $\pi_3(\cdot)$ represent suitable prior distributions. Notably, the estimation of the degrees of freedom parameter can be complex; the authors of [26] suggest employing the Jeffreys prior distribution proposed by [27] for the skew–t distribution, which can be expressed as

$$\pi_3(\nu) \propto \sqrt{\frac{\nu}{\nu+3}\left\{\psi_3\left(\frac{\nu}{2}\right) - \psi_3\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}},$$

where $\psi_3(\cdot)$ is the trigamma function. Alternatively, one might consider fixing the degrees of freedom in the estimation scenario presented in this article.

The application of the skew–t link model to two real-world datasets, women's labor and cardiovascular disease data, further validates its practical utility. In both cases, the model outperformed traditional symmetric link functions and the skew–probit link in terms of model fit, as indicated by lower AIC, BIC, and HQIC values. These results indicate that the skew–t link model is particularly well suited for datasets with inherent asymmetry, which is commonly found in socioeconomic and medical research as well as in other fields like environmental science and finance. For instance, in environmental and atmospheric sciences, the skew–t link model can be applied to analyze data related to events such as pollution levels or extreme weather occurrences, where outcome distributions often exhibit non-symmetry due to natural variability. Additionally, in finance and economics, this model can enhance the analysis of binary outcomes such as credit defaults, investment decisions, or market participation, where irregularities in data distributions frequently challenge traditional models. We believe the skew–t model provides a flexible alternative that can more effectively capture the complexities of asymmetric and heavy-tailed data across these diverse domains.

Despite the clear advantages of the skew–t link model, there are several limitations that warrant further exploration. First, while the penalized likelihood estimation method effectively addresses the challenges in estimating the asymmetry parameter, the choice of penalty function could potentially influence the results. Future research could explore alternative penalty functions and assess their impact on model performance. Second, the skew–t link model, as presented in this paper, is limited to binomial and binary response data. Extending this model to more complex data structures, such as multinomial or mixed response models, could significantly broaden its applicability. Additionally, exploring the use of the skew–t link in hierarchical or multilevel models could provide further insights into its potential for analyzing nested data structures, which are common in fields like education and healthcare. Another promising direction for future research lies in the development of computational tools and algorithms that can efficiently estimate the skew–t link model's parameters for larger datasets. Although the penalized likelihood estimation method is effective, its computational cost may increase with larger sample sizes or more complex data configurations. Developing optimized algorithms for parameter estimation could make the skew–t link model more accessible for big data applications.

7. Conclusions

The skew–t link model introduced in this work represents a significant advance in the analysis of binomial and binary response data. Its flexibility, particularly in handling asymmetric distributions and heavy tails, makes it a valuable tool for researchers in a range of fields. The model's ability to outperform traditional symmetric links and the skew–probit link in both simulated and real-world data underscores its potential to address longstanding challenges in statistical modeling. Future research should focus on extending the model to more complex data structures, exploring alternative estimation methods, and optimizing computational tools to improve its applicability in diverse research settings.

Author Contributions: Conceptualization, O.C.-P., G.I.-P. and O.N.; methodology, O.C.-P., O.N. and G.I.-P.; software, O.C.-P.; validation, O.C.-P., O.N. and G.I.-P.; formal analysis, O.C.-P.; investigation, O.C.-P.; data curation, O.N. and O.C.-P.; writing—original draft preparation, O.C.-P.; writing—review and editing, O.N., G.I.-P. and O.C.-P.; visualization, O.C.-P.; supervision, O.N. and G.I.-P.; project administration, O.N.; funding acquisition, O.N. All authors have read and agreed to the published version of the manuscript.

Funding: O.C.-P.'s research was partially supported by the PhD scholarship of the Universidad de Valparaíso (Chile), and O.N. was partially funded by the Chilean national ANID-Fondecyt grant 1241881-E2024.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors thank the editor and the three anonymous reviewers for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, M.-H.; Dey, D.K.; Shao, Q.-M. On robustness of choice of links in binomial regression. *Calcutta Stat. Assoc. Bull.* **2022**, *53*, 145–164. [CrossRef]
2. Bazán, J.L.; Bolfarine, H.; Branco, M.D. A skew item response model. *Bayesian Anal.* **2006**, *1*, 861–892. [CrossRef]
3. Bazán, J.L.; Bolfarine, H.; Branco, M.D. A framework for skew-probit links in binary regression. *Commun. Stat.-Theory Methods* **2010**, *4*, 678–697. [CrossRef]
4. Lee, D.; Sinha, S. Identifiability and bias reduction in the skew-probit model for a binary response. *J. Stat. Comput. Simul.* **2019**, *89*, 1621–1648. [CrossRef]
5. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
6. Liseo, B.; Loperfido, N. A note on reference priors for the scalar skew-normal distribution. *J. Stat. Plan. Inference* **2006**, *136*, 373–389. [CrossRef]
7. Sartori, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew *t* distributions. *J. Stat. Plan. Inference* **2006**, *136*, 4259–4275. [CrossRef]
8. Azzalini, A.; Arellano-Valle, R.B. Maximum penalized likelihood estimation for skew-normal and skew-*t* distributions. *J. Stat. Plan. Inference* **2013**, *143*, 419–433. [CrossRef]
9. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [CrossRef]
10. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc. B Stat. Methodol.* **2003**, *65*, 367–389. [CrossRef]
11. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [CrossRef]
12. Behboodian, J.; Jamalizadeh, A.; Balakrishnan, N. A new class of skew-cauchy distributions. *Stat. Probab. Lett.* **2006**, *76*, 1488–1493. [CrossRef]
13. Jamalizadeh, A.; Khosravi, M.; Balakrishnan, N. Recurrence relations for distributions of a skew-*t* and a linear combination of order statistics from a bivariate-*t*. *Comput. Stat. Data Anal.* **2009**, *53*, 847–852. [CrossRef]
14. Taylor, J.M.G.; Siqueira, T.A.; Weiss, R.E. The cost of adding parameters to a model. *J. R. Stat. Soc. B Stat. Methodol.* **1996**, *58*, 593–607. [CrossRef]
15. Cole, S.R.; Chu, H.; Greenland, S. Maximum likelihood, profile likelihood, and penalized likelihood: A primer. *Am. J. Epidemiol.* **2013**, *179*, 252–260. [CrossRef]
16. Ciuperca, G.; Ridolfi, A.; Idier, J. Penalized maximum likelihood estimator for normal mixtures. *Scand. J. Stat.* **2003**, *30*, 45–59. [CrossRef]
17. Otsu, T. Penalized empirical likelihood estimation of semiparametric models. *J. Multivar. Anal.* **2007**, *98*, 1923–1954. [CrossRef]
18. Pötscher, B.M.; Leeb, H. On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *J. Multivar. Anal.* **2009**, *100*, 2065–2082. [CrossRef]
19. Eggermont, P.P.; LaRiccia, V.N. *Maximum Penalized Likelihood Estimation: Volume II: Regression*, 1st ed.; Springer: New York, NY, USA, 2009; p. 7.
20. Abadir, K.M.; Magnus, J.R. *Matrix Algebra*, 1st ed.; Cambridge University Press: New York, NY, USA, 2005; p. 106.
21. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024.
22. Henningsen, A.; Toomet, O. maxLik: A package for maximum likelihood estimation in R. *Comput. Stat.* **2010**, *26*, 443–458. [CrossRef]
23. Murphy, S.A.; Van Der Vaart, A.W. On profile likelihood. *J. Am. Stat. Assoc.* **2000**, *95*, 449–465. [CrossRef]
24. Rodriguez-Segura, M.; Nicolis, O.; Peralta-Marquez, B.; Carrillo-Azocar, J. Predicting cardiovascular disease by combining optimal feature selection methods with machine learning. In Proceedings of the 2020 39th International Conference of the Chilean Computer Science Society (SCCC), Coquimbo, Chile, 16–20 November 2020; pp. 1–8.
25. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'Connor G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The sleep heart health study: Design, rationale, and methods. *Sleep* **1997**, *20*, 1077–1085. [PubMed]
26. Branco, M.D.; Genton, M.G.; Liseo, B. Objective bayesian analysis of skew-*t* distributions. *Scand. J. Stat.* **2012**, *40*, 63–85. [CrossRef]
27. Fonseca, T.C.O.; Ferreira, M.A.R.; Migon, H.S. Objective bayesian analysis for the Student-*t* regression model. *Biometrika* **2008**, *95*, 325–333. [CrossRef]