mathematics

MDPI

Article

New Statistical Residuals for Regression Models in the Exponential Family: Characterization, Simulation, Computation, and Applications

Raydonal Ospina [1,2], Patrícia L. Espinheira [1,2], Leilo A. Arias [1], Cleber M. Xavier [3], Víctor Leiva [4,*] and Cecilia Castro [5]

[1] Departamento de Estatística, CASTLab, Universidade Federal de Pernambuco, Recife 50670-901, Brazil; parespipa@de.ufpe.br (P.L.E.); aariasv@dane.gov.co (L.A.A.)
[2] Departamento de Estatística, LInCa, Universidade Federal da Bahia, Salvador 40170-110, Brazil
[3] Departamento de Estatística e Ciências Atuariais, Universidade Federal de Sergipe, São Cristóvão 49107-230, Brazil
[4] School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile
[5] Centre of Mathematics, Universidade do Minho, 4710-057 Braga, Portugal; cecilia@math.uminho.pt
* Correspondence: victor.leiva@pucv.cl or victorleivasanchez@gmail.com

Abstract: Residuals are essential in regression analysis for evaluating model adequacy, validating assumptions, and detecting outliers or influential data. While traditional residuals perform well in linear regression, they face limitations in exponential family models, such as those based on the binomial and Poisson distributions, due to heteroscedasticity and dependence among observations. This article introduces a novel standardized combined residual for linear and nonlinear regression models within the exponential family. By integrating information from both the mean and dispersion sub-models, the new residual provides a unified diagnostic tool that enhances computational efficiency and eliminates the need for projection matrices. Simulation studies and real-world applications demonstrate its advantages in efficiency and interpretability over traditional residuals.

Keywords: advanced residual analysis; computational efficiency; exponential family models; Fisher scoring; mean and dispersion integration; model adequacy; regression diagnostics

MSC: 62J05; 62E15

Citation: Ospina, R. ; Espinheira, P.L.; Arias, L.A.; Xavier, C.M.; Leiva, V.; Castro, C. New Statistical Residuals for Regression Models in the Exponential Family: Characterization, Simulation, Computation, and Applications. *Mathematics* **2024**, *12*, 3196. https://doi.org/10.3390/math12203196

Academic Editor: Niansheng Tang

Received: 27 August 2024
Revised: 4 October 2024
Accepted: 6 October 2024
Published: 12 October 2024

1. Introduction

Evaluating model adequacy in regression analysis is essential as it directly impacts the validity of the inferences made. Residuals, traditionally stated as the difference between observed and predicted values, serve as key diagnostic tools, offering insights into model assumptions, outliers, and overall specification [1–6]. Residual analysis can uncover model flaws that might provide misleading conclusions.

Linear regression models, which assume normally distributed errors, are commonly used due to their simplicity. However, many real phenomena do not meet this assumption. Generalized linear models (GLMs) extend linear models by accommodating a wider class of probability laws from the exponential family, including binomial, gamma, and Poisson distributions [7–9]. Nevertheless, this extension complicates the residual analysis. The assumptions of constant variance and independence of errors are more nuanced in GLM, requiring more sophisticated residuals tailored to the specific distribution [10–14]. For example, in binary or count models, the variance depends on the mean, complicating interpretation of the residuals [15–17]. Recent studies have applied GLM to diverse contexts, such as economic data [18], quantile regression [19], and environmental data through Birnbaum-Saunders regression models [20–22].

Residuals are crucial in various analytical techniques, including variable selection, screening, and machine learning [23]. They help to identify predictor contributions by highlighting patterns, guiding feature selection in high-dimensional data, and improving model performance in methods like gradient boosting [24–26]. Moreover, diagnostic methods, such as those discussed in [27–29], expand on how residuals can be used for goodness-of-fit tests and model adequacy checks, further demonstrating their versatility. The statistic named predicted residual error sum of squares, based on residuals, is also used to assess predictive performance through cross-validation [30].

Despite their importance, traditional residuals, like studentized or deviance, have limitations when applied to GLM. Advanced methods are needed to address these limitations, especially in handling heteroscedasticity and dependencies [31–33]. Recent studies have explored alternative approaches, such as skew-normal inflated models [34] and semi-parametric additive modeling with Birnbaum-Saunders distributions [35]. These approaches contribute to the development of more accurate diagnostic tools in complex settings.

Standardizing residuals in GLM often involve projection matrices, derived from the likelihood maximization process [1,5]. Nonetheless, these matrices can be computationally demanding, particularly for large datasets, decreasing their practical use [16,17,36]. Moreover, projection matrices may not fully capture data variability when changes in the dispersion exist, complicating diagnostics [17,34,37]. Recent research has addressed the above-mentioned limitations and proposed methods to improve residual analysis in GLM. Alternative residuals tailored to specific distributions within the exponential family were investigated in [37,38]. However, these proposals often involve complex calculations or are restricted to specific data types, reducing their broader applicability [39].

Modern approaches, presented in studied for semi-parametric additive models with complex data structures [35] as well as continuous Bernoulli [40] and quasi-Cauchy [41] quantile regression models, underscore the need for innovative diagnostic methods and demonstrate the challenges in dealing with data supported within specific intervals, emphasizing the restrictions of traditional approaches. While these studies represent advancements, important gaps in residual analysis for GLMs remain. Proposed methods, such as skew-normal inflated [34] and semi-parametric additive [35] models, focus on either mean or dispersion models without considering both simultaneously. Moreover, several methods rely on computationally intensive procedures, such as projection matrices, which reduces their applicability in large datasets or models with varying dispersion.

In response to the earlier-mentioned limitations, the standardized combined residual integrates information from both mean and dispersion models. Unlike methods that focus on a single model, this integration offers a more comprehensive approach to residual analysis, improving both computational efficiency and interpretability. Importantly, the standardized combined residuals avoid the need for projection matrices, so reducing computational cost and complexity, particularly for large datasets. Thus, one can propose a methodology that enhances the detection of model inadequacies, particularly in scenarios involving heteroscedasticity or interdependence between observations. Therefore, the present study proposes a novel standardized combined residual that provides both theoretical advancements and practical improvements over existing residuals, including a more efficient computational behavior and enhanced diagnostic capabilities.

The main objectives of this study are: (i) to develop a unified framework for residual analysis in GLMs that integrates both mean and dispersion sub-models, (ii) to assess the computational efficiency of the proposed residual in comparison with traditional methods, and (iii) to evaluate the performance of the proposed residual through extensive simulations and applications to real-world datasets. To achieve these objectives, we employ the Fisher scoring iterative method for parameter estimation in GLMs, which allows the simultaneous modeling of both mean and dispersion effects. The new residual is specifically designed to accommodate this without the need for projection matrices, as mentioned, offering both computational efficiency and clearer interpretation.

The structure of the article is as follows. Section 2 covers the framework for modeling and parameter estimation in the exponential family, emphasizing the relationship between mean and variance. In Section 3, nonlinear models within the exponential family are discussed, exploring both general structure and specific cases. In Section 4, we introduce standardized combined residuals, which integrate information from both mean and dispersion sub-models. The proposed residual effectiveness is demonstrated through simulations and two empirical applications, highlighting its utility in practical contexts. Specifically, Section 5 provides methods for evaluating model fit using simulated envelopes, offering a robust approach to residual analysis. In Section 6, empirical applications of the proposed residual to real datasets are presented, demonstrating its practical utility, while Section 7 provides concluding remarks and future research directions.

Additionally, this article includes appendices that provide detailed mathematical derivations (Appendix A); a description of the Fisher scoring iterative method (Appendix B); a complete example in `R` code [42] to show the implementation of the proposed methods (Appendixes C and D) containing the results of extensive Monte Carlo simulations, further elucidating the performance of the proposed methods.

2. Modeling and Estimation in Exponential Family Distributions

This section presents the framework for modeling and parameter estimation within the exponential family of distributions, emphasizing the critical relationship between the mean and variance. Understanding this relationship is essential for constructing regression models that accurately capture the variability in real-world data, especially when the variability is not constant but depends on the magnitude of the mean.

2.1. Simultaneous Modeling of the Mean and Dispersion in the Exponential Family

The mean and variance are fundamental characteristics that describe the behavior of a random variable within a distribution, assuming that these moments exist. In many distributions belonging to the exponential family, a specific functional relationship exists between the mean and variance. Accurately modeling this relationship is crucial when the variance changes with the mean, which is often the case in practical applications. Such modeling ensures that the variability observed in the data is properly accounted for, particularly in contexts where the dispersion (or spread) of the data is not constant but varies proportionally with the mean. Let $Y_1, \ldots, Y_n$ be independent random variables, where each Y_i, for $i \in \{1, \ldots, n\}$, has a probability density function that can be expressed in the canonical form of the exponential family [5,43,44] as

$$f(y_i; \theta_i, \phi_i) = \exp(\phi_i(y_i\theta_i - b(\theta_i)) + c(y_i; \phi_i)), \quad i \in \{1, \ldots, n\}, \tag{1}$$

with θ_i being the canonical parameter, which uniquely determines the distribution of Y_i within the exponential family; $\phi_i^{-1} > 0$ being the dispersion parameter, which scales the variance independently of the mean, capturing additional variability in the data; $b(\theta_i)$ being the cumulant function, crucial for determining the mean and variance of the distribution, as well as for deriving higher-order cumulants and moments; and $c(y_i; \phi_i)$ being the normalizing function, ensuring that the probability density function integrates to one.

The mean and variance of the distribution with density described in (1) are given by

$$\mathrm{E}(Y_i) = \mu_i = b'(\theta_i), \quad \mathrm{Var}(Y_i) = \phi_i^{-1} b''(\theta_i), \quad i \in \{1, \ldots, n\}, \tag{2}$$

where $b''(\theta_i)$ is the variance function, denoted as $\mathrm{V}(\mu_i)$, which depends only on the mean μ_i. The function $\mathrm{V}(\mu_i)$ describes the theoretical variance of Y_i under the assumption of constant dispersion. However, the true variance observed in the data, $\mathrm{Var}(Y_i)$ namely, is influenced by the dispersion parameter ϕ_i. This parameter ϕ_i allows the model to adjust for overdispersion (where the observed variance is greater than the theoretical variance) or underdispersion (where the observed variance is less than the theoretical variance), thereby providing a more accurate reflection of the data variability compared to the standard

exponential family model. The flexibility of the exponential family of distributions in modeling is crucial in practical applications, as it enables the model to more accurately reflect the true variability present in the data, particularly in situations where the variability cannot be fully explained by the mean alone. Such flexibility is evident across various distributions within the exponential family, each with its own distinct relationship between the mean, variance, and dispersion parameters.

Table 1 summarizes some common distributions within the exponential family. In this table, CV denotes the coefficient of variation, defined as $\mathrm{CV} = \sqrt{\mathrm{Var}(Y_i)}/\mathrm{E}(Y_i)$, and Γ represents the Euler gamma function, which generalizes the factorial function to real and complex numbers [45,46]. Understanding these parameters is crucial for accurately modeling the dispersion and shape characteristics of different distributions, which play an important role in the analysis of real-world data.

Table 1. Examples of distributions belonging to the exponential family with their indicated parameter.

Distribution	$\boldsymbol{\theta}$	$\boldsymbol{b(\theta)}$	$\boldsymbol{\phi}$	$\mathbf{V}(\boldsymbol{\mu})$	$\boldsymbol{c(y;\phi)}$
Normal	μ	$\theta^2/2$	$1/\sigma^2$	1	$-(1/2)\log(2\pi/\phi) - \phi y^2/2$
Gamma	$-1/\mu$	$-\log(-\theta)$	$1/\mathrm{CV}^2$	μ^2	$(\phi-1)\log(y) + \phi\log(\phi) - \log(\Gamma(\phi))$
Inverse Gaussian	$-1/2\mu^2$	$-\sqrt{-2\theta}$	ϕ	μ^3	$\log(\phi/(2\pi y^3))/2 - \phi/(2y)$
Poisson	$\log(\mu)$	$\exp(\theta)$	1	μ	$-\log(y!)$
Binomial	$\log(\mu/(1-\mu))$	$\log(1+\exp(\theta))$	n	$\mu(1-\mu)$	$\log(\binom{\phi}{\phi y/n})$

CV is the variation coefficient of the gamma distribution and Γ is the standard gamma function.

The theoretical framework outlined above provides a foundation for understanding how the mean and variance interact within the exponential family of distributions. However, applying this framework in practice requires defining a regression structure that incorporates both the mean and dispersion parameters effectively.

2.2. Regression Structure and Parameter Estimation in Exponential Family Models

Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ represent values of the set of k covariates associated with the mean response μ_i for observation i, and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})^\top$ be the values of the set of q covariates related to the dispersion parameter ϕ_i for the same observation, where $i \in \{1, \ldots, n\}$. These values of covariates capture the effect of different predictors on the mean and dispersion, respectively. Importantly, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ can, but do not have to, be the same, depending on the model specification. This modeling allows for different sets of covariates to influence the mean and the dispersion independently, which is an advantage in modeling with the exponential family.

We assume that the mean μ_i and the dispersion ϕ_i, as defined in (2), satisfy the relationships stated as

$$g(\mu_i) = \eta_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} = \sum_{j=1}^{k} x_{ij}\beta_j, \tag{3}$$

$$h(\phi_i) = \eta_{2i} = \boldsymbol{z}_i^\top \boldsymbol{\gamma} = \sum_{j=1}^{q} z_{ij}\gamma_j, \quad i \in \{1, \ldots, n\}, \tag{4}$$

where $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\boldsymbol{\gamma} \in \mathbb{R}^q$ are unknown parameter vectors.

The functions g presented in (3) and h in (4) are known as link functions, which are assumed to be monotonic, twice differentiable, and chosen based on the nature of the response variable as well as in the desired relationship between the predictors and the mean or dispersion. The linear predictors established as η_{1i} in (3) and as η_{2i} in (4), formed by linear combinations of the covariates with their respective coefficients, encapsulate the effects of the covariates on the mean and dispersion, respectively.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ represent the full parameter vector of the model to be estimated, combining both the coefficients related to the mean and dispersion. The log-likelihood function for the model, which forms the basis for parameter estimation, is expressed as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\mu_i; \phi_i), \quad i \in \{1, \ldots, n\}, \tag{5}$$

where ℓ_i is the log-likelihood contribution of observation i formulated as

$$\ell_i(\mu_i; \phi_i) = \phi_i(y_i\theta_i - b(\theta_i)) + c(y_i; \phi_i), \quad i \in \{1, \ldots, n\},$$

with θ_i, $b(\theta_i)$, and $c(y_i; \phi_i)$ being defined as in (1). Parameter estimation in GLM is typically performed using the method of maximum likelihood (ML), which involves solving the score equations derived by taking the partial derivatives of the log-likelihood function presented in (5) with respect to the parameters. The ML estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ are generated by solving the system of score equations given by

$$\boldsymbol{U}_{\boldsymbol{\beta}} = 0, \quad \boldsymbol{U}_{\boldsymbol{\gamma}} = 0.$$

If the dispersion parameter ϕ_i is known and constant, the probability density function obtained in (1) defines the class of exponential family models with the canonical parameter θ_i. However, when ϕ_i is unknown, the expression established in (1) defines a valid exponential family model only if the function $c(y_i; \phi_i)$ takes the specific form stated as

$$c(y_i; \phi_i) = d(\phi_i) + \phi_i a(y_i) + u(y_i), \quad i \in \{1, \ldots, n\},$$

where $d(\phi_i)$, $a(y_i)$, and $u(y_i)$ are arbitrary functions that ensure the model adheres to the exponential family structure. The specific form of $c(y_i; \phi_i)$ is crucial because it maintains the tractability of the likelihood function and the validity of the model within the exponential family. For further details, see [38].

In many cases, a closed-form solution is not feasible, necessitating the use of numerical optimization methods. A commonly employed approach to obtain the mentioned solution is the Fisher scoring iterative method, which is presented generically as

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \left(\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(m)}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\beta}}^{(m)}, \tag{6}$$

$$\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)} + \left(\boldsymbol{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{(m)}\right)^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^{(m)}, \tag{7}$$

where m denotes the iteration number; $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\gamma}^{(m)}$ represent the parameter estimates at the mth iteration; $\boldsymbol{U}_{\boldsymbol{\beta}}^{(m)}$ and $\boldsymbol{U}_{\boldsymbol{\gamma}}^{(m)}$ are the score function components corresponding to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively; whereas $\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(m)}$ and $\boldsymbol{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{(m)}$ are the Fisher information matrices at the mth iteration, which represent the curvature of the log-likelihood function stated in (5) with respect to the parameters. This curvature refers to the negative expected value of the second derivative (Hessian matrix) of the log-likelihood function with respect to the parameters. It captures how sensitive the likelihood function is to changes in the parameters. The Fisher information matrices measure the curvature by quantifying the rate at which the slope of the log-likelihood function changes with respect to the parameters. A higher curvature indicates that the log-likelihood function has a sharper peak around the parameter estimates, leading to more precise estimates.

By substituting the necessary quantities (see Appendix B), the iterative formulas established in (6) and (7) can be rewritten as a weighted least squares process presented as

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \left(\boldsymbol{X}^\top \boldsymbol{\Phi}^{(m)} \boldsymbol{W}^{(m)} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Phi}^{(m)} \boldsymbol{W}^{(m)} \boldsymbol{u}_1^{(m)}, \\ \boldsymbol{\gamma}^{(m+1)} &= \left(\boldsymbol{Z}^\top \boldsymbol{P}^{(m)} \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{P}^{(m)} \boldsymbol{u}_2^{(m)}, \end{aligned} \tag{8}$$

where

- $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the design matrices corresponding to the values of covariates associated with the mean μ_i and dispersion ϕ_i, respectively;
- $\boldsymbol{\Phi}^{(m)}$ and $\boldsymbol{P}^{(m)}$ are diagonal matrices, where $\boldsymbol{\Phi}^{(m)} = \text{diag}\{\phi_i^{(m)}\}$ and $\boldsymbol{P}^{(m)}$ depends on the variance function related to the dispersion model, capturing the structure of the data variability at mth iteration;
- $\boldsymbol{W}^{(m)}$ is a weight matrix that typically depends on the variance function $\text{V}(\mu^{(m)})$ at mth iteration, adjusting for the variability of the response variable;
- $\boldsymbol{u}_1^{(m)} = \boldsymbol{\eta}_1^{(m)} + \boldsymbol{W}^{1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)})$, where $\boldsymbol{\eta}_1^{(m)}$ is the linear predictor for the mean model, $\boldsymbol{V}_\gamma = \text{diag}\{-d''(\phi_1), \ldots, -d''(\phi_n)\}$, $\boldsymbol{y}$ is the vector of observed responses, and $\boldsymbol{\mu}^{(m)}$ is the vector of fitted values at mth iteration;
- $\boldsymbol{u}_2^{(m)} = \boldsymbol{\eta}_2^{(m)} + \boldsymbol{V}_\gamma^{-1(m)}\boldsymbol{H}_\gamma^{(m)}(\boldsymbol{t} - \boldsymbol{\mu}^{*(m)})$, where $\boldsymbol{\eta}_2^{(m)}$ is the linear predictor for the dispersion model, $\boldsymbol{H}_\gamma = \text{diag}\{h'(\phi_1), \ldots, h'(\phi_n)\}$, with $h'(\phi_i) = \partial\boldsymbol{\eta}_{2i}/\partial\phi_i\boldsymbol{t}$ being the response variable associated with the dispersion, and $\boldsymbol{\mu}^{*(m)}$ is the fitted value for the dispersion at mth iteration.

Upon convergence of the iterative process stated in (8), we obtain the estimates given by

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}}\boldsymbol{u}_1, \tag{9}$$

$$\widehat{\boldsymbol{\gamma}} = \left(\boldsymbol{Z}^\top\widehat{\boldsymbol{P}}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\widehat{\boldsymbol{P}}\boldsymbol{u}_2, \tag{10}$$

where $\boldsymbol{u}_1 = \widehat{\boldsymbol{\eta}}_1 + \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{V}}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}})$ and $\boldsymbol{u}_2 = \widehat{\boldsymbol{\eta}}_2 + \widehat{\boldsymbol{V}}_\gamma^{-1}\widehat{\boldsymbol{H}}_\gamma(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^*)$. The expressions defined in (9) and (10) can be interpreted as ordinary least squares (OLS) estimators after applying appropriate weights and transformations. Specifically, to reach $\widehat{\boldsymbol{\beta}}$, we regress the auxiliary variable $\widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{u}_1$ on the covariate matrix $\widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}$. Similarly, to get $\widehat{\boldsymbol{\gamma}}$, we perform a regression of the auxiliary variable $\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{u}_2$ on the covariate matrix $\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{Z}$.

2.3. Residuals

The corresponding residuals from the above-mentioned weighted regressions are given by

$$r^\beta = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}(\boldsymbol{u}_1 - \widehat{\boldsymbol{\eta}}_1), \quad r^\gamma = \widehat{\boldsymbol{P}}^{1/2}(\boldsymbol{u}_2 - \widehat{\boldsymbol{\eta}}_2), \quad r_i^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\phi}_i^{-1}\widehat{V}_i}}, \quad r_i^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d''(\widehat{\phi}_i)}}, \quad i \in \{1, \ldots, n\},$$

where $\widehat{V}_i = \partial\widehat{\mu}_i/\partial\widehat{\theta}_i$. These residuals can be standardized using the projection matrix derived from the OLS solution, which normalizes them for better comparability across observations. Standardization is particularly important when dealing with heteroscedasticity, as it ensures that the residuals have consistent variability, making them more effective for model diagnostics. For each component, the projection matrices are formulated as

$$\widehat{\boldsymbol{S}}^\beta = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{X}\left(\boldsymbol{X}^\top\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{\Phi}}^{1/2}, \tag{11}$$

$$\widehat{\boldsymbol{S}}^\gamma = \widehat{\boldsymbol{P}}^{1/2}\boldsymbol{Z}\left(\boldsymbol{Z}^\top\widehat{\boldsymbol{P}}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\widehat{\boldsymbol{P}}^{1/2}. \tag{12}$$

Then, the standardized ordinary residuals are defined as

$$r_{\text{s}_i}^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\phi}_i^{-1}\widehat{V}_i(1 - \widehat{s}_{ii}^\beta)}}, \quad r_{\text{s}_i}^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d''(\widehat{\phi}_i)(1 - \widehat{s}_{ii}^\gamma)}}, \quad i \in \{1, \ldots, n\},$$

where $\widehat{s}_{ii}^\beta$ and $\widehat{s}_{ii}^\gamma$ are the diagonal elements of the projection matrices $\widehat{\boldsymbol{S}}^\beta$ and $\widehat{\boldsymbol{S}}^\gamma$, respectively. The calculation of $r_{\text{s}_i}^\beta$ depends on obtaining the projection matrix $\widehat{\boldsymbol{S}}^\beta$. After assessing the residuals, another important statistic to consider in evaluating the model fit within the exponential family is the deviance, which is a key metric used for assessing how well the model fits the data. The deviance was first translated as deviation in [47].

The deviance function for the mean is established as the difference between the log-likelihood function of the saturated model (with n parameters) and the model of interest (with k parameters, where $k < n$), evaluated at the ML estimate $\widehat{\boldsymbol{\beta}}$, and is given by

$$D^*(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} \phi_i\big(y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - b(\widetilde{\theta}_i) + b(\widehat{\theta}_i)\big),$$

where $\widetilde{\theta}_i = \theta(y_i)$ and $\widehat{\theta}_i = \theta(\widehat{\mu}_i)$ represent the ML estimates of θ for the saturated model (with n parameters) and the model under study (with $k < n$ parameters), respectively. Similarly, the deviance function for the dispersion can be defined as

$$D^*(\boldsymbol{y}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} \big(t_i(\widetilde{\phi}_i - \widehat{\phi}_i) + (d(\widetilde{\phi}_i) - d(\widehat{\phi}_i))\big),$$

where $\widetilde{\phi}_i$ is the solution to $d'(\widetilde{\phi}_i) = -t_i$. For further details on the deviance, see [1,5].

The deviance residuals are commonly used in GLM to assess the contribution of each observation to the overall model fit. The standardized deviance residuals for the mean and dispersion are given, respectively, by

$$d_\mu(\widehat{\mu}_i, \widehat{\phi}_i) = \frac{d^*_\mu(\widehat{\mu}_i, \widehat{\phi}_i)}{\sqrt{1 - \widehat{s}^{\beta}_{ii}}}, \quad d_\phi(\widehat{\mu}_i, \widehat{\phi}_i) = \frac{d^*_\phi(\widehat{\mu}_i, \widehat{\phi}_i)}{\sqrt{1 - \widehat{s}^{\gamma}_{ii}}}, \quad i \in \{1, \ldots, n\}, \tag{13}$$

where $\widehat{s}^{\beta}_{ii}$ and $\widehat{s}^{\gamma}_{ii}$ are the ith diagonal elements of the projection matrices defined in (11) and (12), respectively. The terms $d^*_\mu(\widehat{\mu}_i, \widehat{\phi}_i)$ and $d^*_\phi(\widehat{\mu}_i, \widehat{\phi}_i)$ represent the signed square roots of the deviance components for the mean and dispersion, respectively, and are established, for $i \in \{1, \ldots, n\}$, as

$$\begin{aligned} d^*_\mu(\widehat{\mu}_i, \widehat{\phi}_i) &= \operatorname{sign}(y_i - \widehat{\mu}_i)(2y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - 2(b(\widetilde{\theta}_i) - b(\widehat{\theta}_i)))^{1/2}, \\ d^*_\phi(\widehat{\mu}_i, \widehat{\phi}_i) &= \operatorname{sign}(\widehat{t}_i + d'(\widehat{\phi}_i))(2t_i(\widetilde{\phi}_i - \widehat{\phi}_i) + 2(d(\widetilde{\phi}_i) - d(\widehat{\phi}_i)))^{1/2}. \end{aligned}$$

These terms capture the contribution of each observation to the deviance, with the sign indicating whether the observed value is above or below the fitted value.

3. Nonlinear Models in the Exponential Family

As the modeling of exponential family distributions extends beyond linear relationships, it becomes necessary to explore nonlinear formulations that offer greater flexibility in capturing complex data patterns. In this section, we transition to nonlinear models within the exponential family framework, focusing on the general structure and specific cases such as the nonlinear negative binomial model.

3.1. General Structure of Nonlinear Models

Let $Y_1, \ldots, Y_n$ be independent random variables, each following the distribution with density as defined in (1), where μ_i and $\phi_i > 0$ satisfy the nonlinear structures given by

$$g(\mu_i) = \eta_{1i} = f_1(\boldsymbol{x}_i^\top \boldsymbol{\beta}), \tag{14}$$

$$h(\phi_i) = \eta_{2i} = f_2(\boldsymbol{z}_i^\top \boldsymbol{\gamma}), \quad i \in \{1, \ldots, n\}, \tag{15}$$

where $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\boldsymbol{\gamma} \in \mathbb{R}^q$ are unknown parameter vectors with $k + q < n$. Here, η_{1i} and η_{2i} are the nonlinear predictors, $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ik})$ and $\boldsymbol{z}_i^\top = (z_{i1}, \ldots, z_{iq})$ are observations of k and q known and fixed covariates, respectively, for each $i \in \{1, \ldots, n\}$.

The link functions g and h are strictly monotonic and twice differentiable. The functions f_1 and f_2 are continuous and differentiable, ensuring that the Jacobian matrices, $\boldsymbol{J}_1 = \partial\boldsymbol{\eta}_1/\partial\boldsymbol{\beta}$ and $\boldsymbol{J}_2 = \partial\boldsymbol{\eta}_2/\partial\boldsymbol{\gamma}$ namely, have full ranks k and q, respectively.

The log-likelihood function for the structures presented in (14) and (15) is stated as in (2), with the jth component of the score function with respect to $\boldsymbol{\beta}$ defined as

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i(\mu_i;\phi_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_j} = \sum_{i=1}^{n} \phi_i (Y_i - \mu_i) \sqrt{\frac{\omega_i}{V_i}} \frac{\partial \eta_{1i}}{\partial \beta_j}, \quad j \in \{1,\ldots,k\},$$

where $V_i = \partial \mu_i / \partial \theta_i$ and $\omega_i = (\partial \mu_i / \partial \eta_{1i})^2 / V_i$. The score function can be expressed in matrix form as $\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{J}_1^\top \boldsymbol{\Phi} \boldsymbol{W}^{1/2} \boldsymbol{V}^{-1/2} (\boldsymbol{Y} - \boldsymbol{\mu})$, where $\boldsymbol{J}_1 = \partial \boldsymbol{\eta}_1 / \partial \boldsymbol{\beta}$ is an $n \times k$ matrix, $\boldsymbol{\Phi} = \operatorname{diag}\{\phi_1,\ldots,\phi_n\}$, $\boldsymbol{W} = \operatorname{diag}\{\omega_1,\ldots,\omega_n\}$, $\boldsymbol{V} = \operatorname{diag}\{V_1,\ldots,V_n\}$, $\boldsymbol{Y} = (Y_1,\ldots,Y_n)^\top$, and $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^\top$. Assuming that $c(Y_i;\phi_i) = d(\phi_i) + \phi_i a(Y_i) + u(Y_i)$, the score function with respect to $\boldsymbol{\gamma}$ is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \ell_i(\mu_i;\phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j} = \sum_{i=1}^{n} (t_i + d'(\phi_i)) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j \in \{1,\ldots,q\},$$

where $t_i = Y_i \theta_i - b(\theta_i) + a(Y_i)$. Similarly, the score function for $\boldsymbol{\gamma}$ can be expressed as $\boldsymbol{U}_{\boldsymbol{\gamma}} = \boldsymbol{J}_2^\top \boldsymbol{H}_{\gamma}^{-1} (\boldsymbol{t} - \boldsymbol{\mu}^*)$, where $\boldsymbol{J}_2 = \partial \boldsymbol{\eta}_2 / \partial \boldsymbol{\gamma}$ is an $n \times q$ matrix, $\boldsymbol{H}_{\gamma} = \operatorname{diag}\{h'(\phi_1),\ldots,h'(\phi_n)\}$, $h'(\phi_i) = \partial \boldsymbol{\eta}_{2i} / \partial \phi_i$, $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$, and $\boldsymbol{\mu}^* = (\mu_1^*,\ldots,\mu_n^*)^\top$.

The Fisher information matrix is derived from the second-order derivatives of the log-likelihood function with respect to the parameter vector. For $j, l \in \{1,\ldots,k\}$, we have

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_\ell} = \sum_{i=1}^{n} \phi_i \frac{\partial}{\partial \beta_\ell} \left(\sqrt{\frac{\omega_i}{V_i}} \right) (Y_i - \mu_i) \frac{\partial \eta_{1i}}{\partial \beta_j} - \sum_{i=1}^{n} \phi_i \sqrt{\frac{\omega_i}{V_i}} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_l},$$

and for $j, l \in \{1,\ldots,q\}$, we reach

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_l} = \sum_{i=1}^{n} (t_i + d'(\phi_i)) \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial}{\partial \gamma_l} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right) + \sum_{i=1}^{n} d''(\phi_i) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_l}.$$

For $j \in \{1,\ldots,k\}$ and $l \in \{1,\ldots,q\}$, we obtain

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_l} = \sum_{i=1}^{n} (Y_i - \mu_i) \sqrt{\frac{\omega_i}{V_i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{2i}}{\partial \gamma_l}.$$

Since $\mathrm{E}(Y_i - \mu_i) = 0$ and $\mathrm{E}(t_i - \mu_{T_i}) = 0$ (under regularity conditions), the expected values of the above-presented derivatives are established as

$$\mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} \right) = -\sum_{i=1}^{n} \phi_i \omega_i \frac{\partial \eta_{1i}}{\partial \beta_j} \frac{\partial \eta_{1i}}{\partial \beta_l}, \quad \mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_l} \right) = -\sum_{i=1}^{n} \frac{d''(\phi_i)}{(h'(\phi_i))^2} \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_l}, \quad \mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \gamma_l} \right) = 0.$$

Thus, the Fisher information matrix for $\boldsymbol{\theta}$ is a block diagonal matrix given as $\boldsymbol{K}_{\boldsymbol{\theta\theta}} = \operatorname{diag}\{\boldsymbol{K}_{\boldsymbol{\beta\beta}}, \boldsymbol{K}_{\boldsymbol{\gamma\gamma}}\}$, where $\boldsymbol{K}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{K}_{\boldsymbol{\gamma\gamma}}$ are the Fisher information matrices for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. These expected values can be expressed in matrix form as

$$\boldsymbol{K}_{\boldsymbol{\beta\beta}} = \mathrm{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\beta};\boldsymbol{\gamma})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right) = \boldsymbol{J}_1^\top \boldsymbol{\Phi} \boldsymbol{W} \boldsymbol{J}_1, \quad \boldsymbol{K}_{\boldsymbol{\gamma\gamma}} = \mathrm{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\beta};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} \right) = \boldsymbol{J}_2^\top \boldsymbol{P} \boldsymbol{J}_2,$$

where $\boldsymbol{P} = \boldsymbol{V}_{\gamma} \boldsymbol{H}_{\gamma}^{-2}$ and $\boldsymbol{V}_{\gamma} = \operatorname{diag}\{-d''(\phi_1),\ldots,-d''(\phi_n)\}$. The Poisson distribution is commonly employed for modeling count data. However, it may present issues, particularly in cases of overdispersion, where the observed variance exceeds the mean. In such situations, the Poisson model becomes inadequate because it inherently assumes that the mean and variance are equal.

A similar issue arises with the binomial model, where the dispersion parameter is fixed as the value n; see Table 1. However, in a regression context, n is typically known and does not vary, which can limit the model flexibility in accommodating overdispersion.

3.2. Nonlinear Negative Binomial Model

In recent years, the negative binomial distribution has been increasingly utilized as an alternative in situations where overdispersion is present. Overdispersion occurs when the observed variance exceeds the mean, a scenario in which the Poisson regression model becomes inadequate because it assumes that the mean and variance are equal. Ignoring overdispersion can lead to relevant consequences, such as incorrect or underestimated standard errors in parameter estimates.

Let Y be a random variable following a negative binomial distribution with parameters $\mu > 0$ and $\phi > 0$, with its probability function given by

$$f(y;\mu,\phi) = \frac{\Gamma(y+\phi)}{\Gamma(y+1)\Gamma(\phi)}\left(\frac{\phi}{\phi+\mu}\right)^{\phi}\left(\frac{\mu}{\phi+\mu}\right)^{y}, \quad y \in \{0,1,\dots\}$$

where, as mentioned, Γ is the gamma function. We use the notation $Y \sim \mathrm{NB}(\mu,\phi)$ for the negative binomial distribution, and its expected value and variance are given by $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu + \mu^2/\phi$, respectively. When the parameter ϕ is known, the negative binomial distribution belongs to the exponential family, with the canonical parameter being $\theta = \log(\mu/(\mu+\phi))$, the cumulant function $b(\theta) = -\phi\log(\mu/(\mu+\phi))$, the normalizing function $c(y;\phi) = \Gamma(y+\phi)/(\Gamma(y+1)\Gamma(\phi))$, and the variance function $\mathrm{V}(\mu) = \mu + \mu^2/\phi$.

Consider $Y_1,\dots,Y_n$ to be independent random variables such that $Y_i \sim \mathrm{NB}(\mu_i,\phi_i)$, for $i \in \{1,\dots,n\}$, where the dispersion parameter ϕ_i is unknown. We assume a regression structure in which μ_i and ϕ_i satisfy the nonlinear relationships defined in (14) and (15), respectively. Let the vector of model parameters be $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ and the corresponding log-likelihood function be expressed as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i\log(\mu_i) + \phi_i\log(\phi_i) - (y_i+\phi_i)\log(\mu_i+\phi_i) + \log(\Gamma(\phi_i+y_i)) - \log(\Gamma(\phi_i)) - \log(\Gamma(y_i+1))),$$

where μ_i is the mean and ϕ_i the dispersion parameter, for $i \in \{1,\dots,n\}$. The score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be represented in matrix form as $\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{J}_1^\top \boldsymbol{W}(\boldsymbol{y}-\boldsymbol{\mu})$ and $\boldsymbol{U}_{\boldsymbol{\gamma}} = \boldsymbol{J}_2^\top \boldsymbol{W}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{c}$, where $\boldsymbol{J}_1 = \partial\boldsymbol{\eta}_1/\partial\boldsymbol{\beta}$ is an $n \times k$ matrix, $\boldsymbol{J}_2 = \partial\boldsymbol{\eta}_2/\partial\boldsymbol{\gamma}$ is an $n \times q$ matrix, $\boldsymbol{W} = \mathrm{diag}\{\omega_1,\dots,\omega_n\}$ with $\omega_i = (\partial\mu_i/\partial\eta_{1i})^2\mathrm{V}(\mu_i)^{-1}$, $\boldsymbol{W}_{\boldsymbol{\gamma}} = \mathrm{diag}\{h'(\phi_1),\dots,h'(\phi_n)\}$ with $h'(\phi_i) = \partial\eta_{2i}/\partial\phi_i$, $\boldsymbol{y} = (y_1,\dots,y_n)^\top$, and $\boldsymbol{\mu} = (\mu_1,\dots,\mu_n)^\top$.

According to [48], the Fisher information matrix for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ is block-diagonal and given by $\boldsymbol{K}_{\boldsymbol{\theta\theta}} = \mathrm{diag}\{\boldsymbol{K}_{\boldsymbol{\beta\beta}}, \boldsymbol{K}_{\boldsymbol{\gamma\gamma}}\}$, where $\boldsymbol{K}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{K}_{\boldsymbol{\gamma\gamma}}$ are the Fisher information matrices for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Thus, we can represent $\boldsymbol{K}_{\boldsymbol{\beta\beta}}$ and $\boldsymbol{K}_{\boldsymbol{\gamma\gamma}}$ in matrix form as

$$\boldsymbol{K}_{\boldsymbol{\beta\beta}} = \mathrm{E}\left(-\frac{\partial^2\ell(\boldsymbol{\beta};\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right) = \boldsymbol{J}_1^\top\boldsymbol{W}\boldsymbol{J}_1, \quad \boldsymbol{K}_{\boldsymbol{\gamma\gamma}} = \mathrm{E}\left(-\frac{\partial^2\ell(\boldsymbol{\beta};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right) = \boldsymbol{J}_2^\top\boldsymbol{P}\boldsymbol{J}_2,$$

where $\boldsymbol{P} = \boldsymbol{T}\boldsymbol{W}_{\boldsymbol{\gamma}}^{-2}$ and $\boldsymbol{T} = \mathrm{diag}\{\mathrm{E}(c_1'),\dots,\mathrm{E}(c_n')\}$, with $\boldsymbol{c} = (c_1,\dots,c_n)^\top$ being the residual vector for the dispersion model. The deviance function from the mean, assuming ϕ_i is known, for $i \in \{1,\dots,n\}$, is given by

$$D^*(\boldsymbol{y},\widehat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left(\phi_i\log\left(\frac{\widehat{\mu}_i+\phi_i}{y_i+\phi_i}\right) + y_i\log\left(\frac{y_i(\widehat{\mu}_i+\phi_i)}{\widehat{\mu}_i(y_i+\phi_i)}\right)\right).$$

In practice, ϕ_i is replaced by its ML estimate $\widehat{\phi}_i$. When $y_i = 0$ for some i, the ith component of the deviance is stated as

$$2\phi_i\log\left(\frac{y_i+\phi_i}{\widehat{\mu}_i+\phi_i}\right).$$

The construction of ordinary residuals depends on the Fisher scoring method used to obtain the ML estimate; see Appendix B.

By replacing the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ with $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$, respectively, we can express the Fisher scoring method as an iterative weighted least squares process presented as

$$\boldsymbol{\beta}^{(m+1)} = \left(\boldsymbol{J}_1^\top \boldsymbol{\Phi}^{(m)} \boldsymbol{W}^{(m)} \boldsymbol{J}_1\right)^{-1} \boldsymbol{J}_1^\top \boldsymbol{\Phi}^{(m)} \boldsymbol{W}^{(m)} \boldsymbol{u}_1^{(m)}, \tag{16}$$

$$\boldsymbol{\gamma}^{(m+1)} = \left(\boldsymbol{J}_2^\top \boldsymbol{P}^{(m)} \boldsymbol{J}_2\right)^{-1} \boldsymbol{J}_2^\top \boldsymbol{P}^{(m)} \boldsymbol{u}_2^{(m)}, \tag{17}$$

where $\boldsymbol{u}_1^{(m)} = \boldsymbol{J}_1\boldsymbol{\beta}^{(m)} + \boldsymbol{W}^{1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)})$ and $\boldsymbol{u}_2^{(m)} = \boldsymbol{J}_2\boldsymbol{\gamma}^{(m)} + \boldsymbol{V}_\gamma^{-1(m)}\boldsymbol{H}_\gamma^{(m)}(\boldsymbol{t} - \boldsymbol{\mu}^{*(m)})$. Upon convergence, we have

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{J}_1^\top \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}} \boldsymbol{J}_1\right)^{-1} \boldsymbol{J}_1^\top \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}} \boldsymbol{u}_1, \tag{18}$$

$$\widehat{\boldsymbol{\gamma}} = \left(\boldsymbol{J}_2^\top \widehat{\boldsymbol{P}} \boldsymbol{J}_2\right)^{-1} \boldsymbol{J}_2^\top \widehat{\boldsymbol{P}} \boldsymbol{u}_2, \tag{19}$$

with

$$\boldsymbol{u}_1 = \widehat{\boldsymbol{\eta}}_1 + \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{V}}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}), \quad \boldsymbol{u}_2 = \widehat{\boldsymbol{\eta}}_2 + \widehat{\boldsymbol{V}}_\gamma^{-1}\widehat{\boldsymbol{H}}_\gamma(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^*).$$

The expressions stated in (18) and (19) can be interpreted as the OLS estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. In other words, to obtain $\widehat{\boldsymbol{\beta}}$, consider the regression of the auxiliary variable $\widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{u}_1$ on the covariate $\widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{J}_1$, and to reach $\widehat{\boldsymbol{\gamma}}$, consider the regression of the auxiliary variable $\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{u}_2$ on the covariate $\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{J}_2$. Consequently, the ordinary residuals (see Appendix B) from these regressions are given by

$$r^\beta = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}(\boldsymbol{u}_1 - \widehat{\boldsymbol{\eta}}_1), \quad r^\gamma = \widehat{\boldsymbol{P}}^{1/2}(\boldsymbol{u}_2 - \widehat{\boldsymbol{\eta}}_2).$$

Thus, considering the residuals based on the sub-models for the mean and dispersion, we have

$$r_i^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\phi}_i^{-1}\widehat{V}_i}}, \quad r_i^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d''(\widehat{\phi}_i)}}, \quad i \in \{1, \ldots, n\}.$$

These residuals can be standardized using the projection matrices from the OLS solutions stated in (18) and (19). Specifically, we obtain

$$\widehat{\boldsymbol{S}}^\beta = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\boldsymbol{J}_1\left(\boldsymbol{J}_1^\top \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{W}}\boldsymbol{J}_1\right)^{-1}\boldsymbol{J}_1^\top \widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{\Phi}}^{1/2}, \tag{20}$$

$$\widehat{\boldsymbol{S}}^\gamma = \widehat{\boldsymbol{P}}^{1/2}\boldsymbol{J}_2\left(\boldsymbol{J}_2^\top \widehat{\boldsymbol{P}}\boldsymbol{J}_2\right)^{-1}\boldsymbol{J}_2^\top \widehat{\boldsymbol{P}}^{1/2}. \tag{21}$$

Therefore, the corresponding standardized residuals are formulated as

$$r_{s_i}^\beta = \frac{y_i - \widehat{\mu}_i}{\sqrt{\widehat{\phi}_i^{-1}\widehat{V}_i(1 - \widehat{s}_{ii}^\beta)}}, \quad r_{s_i}^\gamma = \frac{t_i - \widehat{\mu}_i^*}{\sqrt{-d''(\widehat{\phi}_i)(1 - \widehat{s}_{ii}^\gamma)}}, \quad i \in \{1, \ldots, n\},$$

where $\widehat{s}_{ii}^\beta$ and $\widehat{s}_{ii}^\gamma$ are the diagonal elements of the matrices presented in (20) and (21), respectively.

We have analyzed various types of residuals within the context of GLM, focusing on their roles in model diagnostics and goodness-of-fit evaluation. However, in scenarios where both the mean and dispersion are modeled simultaneously, it is beneficial to consider a combined residual that encapsulates the information from both aspects.

4. Standardized Combined Residuals

In this section, we introduce the concept of standardized combined residuals, designed to integrate information from both the mean and dispersion sub-models into a single diagnostic tool. This unified approach facilitates a more comprehensive assessment of model adequacy. We begin by presenting a new standardized combined residual and then extend the concept to accommodate broader model structures.

4.1. New Standardized Combined Residual

The combined residual is calculated by summing the residuals from the mean and dispersion components. Using the estimates from the formulas given in (16) and (17), the combined residual for each observation i is defined as

$$r_i^{\beta\gamma} = (y_i - \widehat{\mu}_i) + (t_i - \widehat{\mu}_i^*), \quad i \in \{1, \ldots, n\}, \tag{22}$$

where y_i is the observed response, $\widehat{\mu}_i$ is the estimated mean from the mean sub-model, t_i is a function related to the dispersion component, and $\widehat{\mu}_i^*$ is the estimated value related to t_i. By combining these residuals, we capture discrepancies in both models.

To standardize the combined residual stated in (22), we calculate its variance. Using the variances and covariances from the equations presented in (A2), (A5), and (A6) of Appendix A, the variance of $r_i^{\beta\gamma}$, denoted by ζ_i, is given as

$$\zeta_i = \phi_i^{-1} b''(\theta_i) - d''(\phi_i), \quad i \in \{1, \ldots, n\}, \tag{23}$$

where, as mentioned, $b''(\theta_i)$ is the second derivative of the cumulant generating function $b(\theta_i)$ and $d''(\phi_i)$ is the second derivative of the dispersion function $d(\phi_i)$. The term $\phi_i^{-1} b''(\theta_i)$ established in (23) represents the variance from the mean sub-model, while $d''(\phi_i)$ accounts for variability in the dispersion sub-model.

With ζ_i computed, we define the standardized combined residual as

$$r_{S_i}^{\beta\gamma} = \frac{(y_i - \widehat{\mu}_i) + (t_i - \widehat{\mu}_i^*)}{\sqrt{\widehat{\zeta}_i}}, \quad i \in \{1, \ldots, n\},$$

where $\widehat{\zeta}_i$ is the variance ζ_i evaluated at the estimated parameters $\widehat{\mu}_i$ and $\widehat{\phi}_i$. This standardization allows for comparison across observations and aids in identifying outliers or model inadequacies by scaling the residuals.

Furthermore, we can estimate the covariance matrices of the parameter estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ to assess the variability in the model parameters. The approximate covariance matrices are presented as

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) \approx \left(\boldsymbol{J}_1^\top \boldsymbol{\Phi} \boldsymbol{W} \boldsymbol{J}_1\right)^{-1}, \quad \mathrm{Cov}(\widehat{\boldsymbol{\gamma}}) \approx \left(\boldsymbol{J}_2^\top \boldsymbol{P} \boldsymbol{J}_2\right)^{-1},$$

where $\boldsymbol{J}_1$ and $\boldsymbol{J}_2$ are the Jacobian matrices of the mean and dispersion models, respectively, and $\boldsymbol{\Phi}$ and $\boldsymbol{P}$ are diagonal weight matrices. These covariance matrices help to quantify the uncertainty in the estimated parameters due to the data variability. Using these covariance estimates, the normal equations for parameter estimation, originally expressed in (18) and (19), can be approximated as

$$\left(\boldsymbol{J}_1^\top \boldsymbol{\Phi} \boldsymbol{W} \boldsymbol{J}_1\right) \widehat{\boldsymbol{\beta}} \approx \boldsymbol{J}_1^\top \boldsymbol{\Phi} \boldsymbol{W} u_1, \quad \left(\boldsymbol{J}_2^\top \boldsymbol{P} \boldsymbol{J}_2\right) \widehat{\boldsymbol{\gamma}} \approx \boldsymbol{J}_2^\top \boldsymbol{P} u_2,$$

where u_1 and u_2 are the working responses for the mean and dispersion models. The left-hand sides represent the information matrices, adjusted by projection matrices, projecting residuals onto the parameter space. This simplifies the parameter estimation process by solving weighted least squares problems.

Assuming that the estimated matrices $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{P}}$ are good approximations of their true counterparts $\boldsymbol{W}$, $\boldsymbol{\Phi}$, and $\boldsymbol{P}$, respectively, we can simplify the covariance for the working responses as

$$\widehat{\mathrm{Cov}}(u_1) \approx \widehat{\boldsymbol{W}}^{-1} \widehat{\boldsymbol{\Phi}}^{-1}, \quad \widehat{\mathrm{Cov}}(u_2) \approx \widehat{\boldsymbol{P}}^{-1},$$

where $\widehat{\boldsymbol{W}}$, $\widehat{\boldsymbol{\Phi}}$, and $\widehat{\boldsymbol{P}}$ are the diagonal matrices of weights and adjustments.

With these approximations, the covariance matrices for $r^{\boldsymbol{\beta}}$ and $r^{\boldsymbol{\gamma}}$ can be derived. In regression, the covariance of residuals is expressed using projection matrices given by

$$\widehat{\mathrm{Cov}}(r^{\boldsymbol{\beta}}) \approx (\boldsymbol{I} - \widehat{\boldsymbol{S}}^{\boldsymbol{\beta}}), \quad \widehat{\mathrm{Cov}}(r^{\boldsymbol{\gamma}}) \approx (\boldsymbol{I} - \widehat{\boldsymbol{S}}^{\boldsymbol{\gamma}}),$$

where $\boldsymbol{I}$ is the identity matrix, and $\widehat{\boldsymbol{S}}^{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{S}}^{\boldsymbol{\gamma}}$ are the projection matrices for the mean and dispersion models, respectively. These matrices reflect the variability explained by the models, while $\boldsymbol{I} - \widehat{\boldsymbol{S}}$ captures the unexplained portion.

To account for the covariances of u_1 and u_2, we use the standardized combined residual by summing the individual standardized residuals. For observation i, it is defined as

$$r_{\mathrm{S}_i}^{\boldsymbol{\beta}\boldsymbol{\gamma}} = \frac{r_i^{\boldsymbol{\beta}} + r_i^{\boldsymbol{\gamma}}}{\sqrt{(1 - \widehat{s}_{ii}^{\boldsymbol{\beta}}) + (1 - \widehat{s}_{ii}^{\boldsymbol{\gamma}}) + 2\widehat{m}_{ii}}}, \quad i \in \{1, \ldots, n\}, \tag{24}$$

where $r_i^{\boldsymbol{\beta}}$ and $r_i^{\boldsymbol{\gamma}}$ are the standardized residuals from the mean and dispersion models, respectively; $\widehat{s}_{ii}^{\boldsymbol{\beta}}$ and $\widehat{s}_{ii}^{\boldsymbol{\gamma}}$ are the ith diagonal elements of $\widehat{\boldsymbol{S}}^{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{S}}^{\boldsymbol{\gamma}}$, representing the leverage of observation i in each model; and $\widehat{m}_{ii}$ is the ith diagonal element of the covariance matrix $M = \mathrm{Cov}(r^{\boldsymbol{\beta}}, r^{\boldsymbol{\gamma}})$, capturing the covariance between the residuals of the two sub-models.

The denominator stated in (24) represents the combined variance of the residuals, incorporating both individual variances and their covariance. This standardization ensures that $r_{\mathrm{S}_i}^{\boldsymbol{\beta}\boldsymbol{\gamma}}$ has a unit variance, allowing for consistent comparison across observations and aiding in the detection of outliers or influential data points. However, evaluating the standardized combined residual given in (24) requires calculating the projection matrices $\widehat{\boldsymbol{S}}^{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{S}}^{\boldsymbol{\gamma}}$, as well as the covariance matrix. These calculations can be computationally expensive, especially for large datasets where matrix operations become complex. The difficulty arises from the need for matrix inversions and multiplications involving the design matrices, which scale poorly with the number of observations. Due to this, the practical implementation of the standardized combined residual presented in (24), while it theoretically sounds good, may be limited in applications with large datasets.

4.2. Background for the Generalized Standardized Combined Residual

Next, we develop a generalized form of the combined residual by relaxing certain assumptions on the functional form of $c(y_i; \phi_i)$. These conditions were previously used to derive variance expressions, but by removing them, we focus on a broader estimation of the dispersion parameter ϕ_i, which allows greater flexibility in model specification.

Let $Y_1, \ldots, Y_n$ be independent random variables, where each Y_i, for $i \in \{1, \ldots, n\}$, follows the density stated in (1). Then, the probability density function of Y_i is given by

$$f(y_i; \theta_i, \phi_i) = \exp\left(\frac{\theta_i y_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i) \right), \quad i \in \{1, \ldots, n\}, \tag{25}$$

where θ_i is the canonical parameter, $\phi_i > 0$ is the dispersion parameter, b is a known function related to the cumulant generating function, and $c(y_i; \phi_i)$ is a normalizing function.

Building on the nonlinear relationships defined in (14) and (15), we now provide a more detailed explanation of the link functions and covariate effects.

The mean μ_i and dispersion parameter ϕ_i are linked to the covariates through nonlinear functional relationships established such as in (14) and (15), that is, $g(\mu_i) = \eta_{1i} = f_1(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ and $h(\phi_i) = \eta_{2i} = f_2(\boldsymbol{z}_i^\top \boldsymbol{\gamma})$, for $i \in \{1, \ldots, n\}$, where, as mentioned, g and h are known link functions, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are vectors of covariates values for the mean and dispersion submodels, respectively, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the parameter vectors to be estimated. The functions f_1 and f_2 represent the nonlinear effects of the covariates on the mean and dispersion components, allowing the model to capture more complex relationships between covariates and the response.

The log-likelihood function for the model based on all observations is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\mu_i; \phi_i) = \sum_{i=1}^{n} \left\{ \frac{\theta_i y_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i) \right\}, \tag{26}$$

where $\boldsymbol{\theta}$ represents all the parameters in the model.

To derive the score function for the parameter vector $\boldsymbol{\beta}$, we calculate the gradient of the log-likelihood function with respect to $\boldsymbol{\beta}$. Applying the chain rule, we obtain

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}}. \tag{27}$$

Recognizing that $\partial \ell_i / \partial \theta_i = (y_i - b'(\theta_i))/\phi_i$ and noting that $b'(\theta_i) = \mu_i$, the formulation presented in (27) simplifies to

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}}. \tag{28}$$

To express the equation stated in (28) in matrix form, we define:

- $\boldsymbol{J}_1 = \partial \boldsymbol{\eta}_1 / \partial \boldsymbol{\beta}$ as the $n \times k$ Jacobian matrix for the mean model;
- $\boldsymbol{\Phi} = \text{diag}\{\phi_1, \ldots, \phi_n\}$ as the diagonal matrix of dispersion parameters;
- $\boldsymbol{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$, where $\omega_i = (\partial \mu_i / \partial \eta_{1i})^2 / V_i$, for $i \in \{1, \ldots, n\}$;
- $\boldsymbol{V} = \text{diag}\{V_1, \ldots, V_n\}$, where $V_i = \text{Var}(Y_i) = \phi_i b''(\theta_i)$, for $i \in \{1, \ldots, n\}$;
- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$.

Then, we have that the score function is formulated as $\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{J}_1^\top \boldsymbol{\Phi}^{-1} \boldsymbol{W} (\boldsymbol{y} - \boldsymbol{\mu})$. This formulation shows that the score function is a weighted sum of the residuals $(\boldsymbol{y} - \boldsymbol{\mu})$, where the weights take into account the variance structure and the link function. The information matrix for $\boldsymbol{\beta}$, which quantifies the information of the data about the parameters, is given by the expected value of the negative second derivative of the log-likelihood function stated as

$$\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \text{E}[-\partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top] = \boldsymbol{J}_1^\top \boldsymbol{\Phi}^{-1} \boldsymbol{W} \boldsymbol{J}_1.$$

This matrix plays a crucial role in parameter estimation and inference, as it is used to compute standard errors and conduct hypothesis tests.

Next, we derive the score function for $\boldsymbol{\gamma}$, the parameter vector associated with the dispersion model. Starting with the partial derivative of the log-likelihood function with respect to the jth element of $\boldsymbol{\gamma}$, denoted by γ_j, and applying the chain rule, we obtain

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j \in \{1, \ldots, q\},$$

where the derivative $\partial \ell_i / \partial \phi_i$ represents how the log-likelihood function changes with respect to the dispersion parameter ϕ_i and is given by

$$\frac{\partial \ell_i}{\partial \phi_i} = -\frac{\theta_i y_i - b(\theta_i)}{\phi_i^2} + \frac{\partial c(y_i; \phi_i)}{\partial \phi_i}, \quad i \in \{1, \ldots, n\}.$$

To simplify the notation in the model, we define the term t_i, which captures the change in the log-likelihood function with respect to ϕ_i. Specifically, we express t_i as

$$t_i = \frac{\partial \ell_i}{\partial \phi_i} = -\frac{\theta_i y_i - b(\theta_i)}{\phi_i^2} + c'(y_i; \phi_i), \quad i \in \{1, \ldots, n\}, \tag{29}$$

where $c'(y_i; \phi_i)$ is the derivative of $c(y_i; \phi_i)$ with respect to ϕ_i. This quantity t_i can be interpreted as a residual that reflects the variability associated with the dispersion parameter.

Next, we use t_i to derive the score function for the dispersion parameters γ. Recognizing that $\partial\phi_i/\partial\eta_{2i} = h'(\phi_i)^{-1}$, where $h'(\phi_i)$ is the derivative of the dispersion link function with respect to ϕ_i, we can express the score function for γ as

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^{n} t_i h'(\phi_i)^{-1} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j \in \{1, \ldots, q\}, \tag{30}$$

We can further express the formulation stated in (30) in matrix form, leading to the score function for γ given by

$$\boldsymbol{U}_\gamma = \boldsymbol{J}_2^\top \boldsymbol{H}_\gamma^{-1} \boldsymbol{t}, \tag{31}$$

where $\boldsymbol{J}_2 = \partial\eta_2/\partial\boldsymbol{\gamma}$ is the Jacobian of the dispersion model, $\boldsymbol{H}_\gamma = \text{diag}\{h'(\phi_1), \ldots, h'(\phi_n)\}$ is the diagonal matrix of link function derivatives, and $\boldsymbol{t} = (t_1, \ldots, t_n)^\top$ are the dispersion residuals. The expression stated in (31) shows that the score function for γ is a weighted sum of the components of $\boldsymbol{t}$, incorporating both the link function and covariate effects, and reflecting the role of the dispersion parameter ϕ_i in the model. To derive the second derivative of the log-likelihood function presented in (26) with respect to γ_j and γ_l, we differentiate the score function stated in (30) with respect to γ_l to obtain

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_l} = \sum_{i=1}^{n} \left\{ \frac{\partial t_i}{\partial \phi_i} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 + t_i \frac{\partial^2 \phi_i}{\partial \eta_{2i}^2} \right\} \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_l}, \quad j, l \in \{1, \ldots, q\}, \tag{32}$$

where $t_i' = \partial t_i/\partial\phi_i$ describes how the dispersion residual changes with respect to ϕ_i, and $\partial^2\phi_i/\partial\eta_{2i}^2$ represents the second derivative of ϕ_i with respect to η_{2i}, capturing the curvature of the link function. The terms $\partial\eta_{2i}/\partial\gamma_j$ and $\partial\eta_{2i}/\partial\gamma_l$ are the derivatives of the linear predictor with respect to the parameters. The curvature of the link function indicates how changes in ϕ_i affect the linear predictor η_{2i} through the link function.

The second derivative stated in (32) is crucial for constructing the Fisher information matrix for γ. Under regularity conditions, the expected value of the score function is zero, that is, $\text{E}(\boldsymbol{U}_\gamma) = \boldsymbol{0}$. This fundamental property ensures that ML estimators are unbiased, meaning that the average score across repeated samples equals zero. Substituting the expression for the score function of γ from the formula given in (31), we get

$$\text{E}\left(\sum_{i=1}^{n} t_i \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j} \right) = 0, \quad j \in \{1, \ldots, q\}.$$

Since $\partial\phi_i/\partial\eta_{2i}$ and $\partial\eta_{2i}/\partial\gamma_j$ are deterministic functions (given the data and parameter values), it follows that $\text{E}(t_i) = 0$, for $i \in \{1, \ldots, n\}$. This result shows that the expected value of the dispersion residuals t_i is zero, similar to the property that the mean residuals $(y_i - \mu_i)$ have zero expectation. Next, we consider the expected value of the second derivative of the log-likelihood function expressed in (26) with respect to γ_j and γ_l, as shown in the formula given in (32). Taking expectation, we obtain

$$\text{E}\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_l} \right) = \sum_{i=1}^{n} \text{E}(t_i') \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \frac{\partial \eta_{2i}}{\partial \gamma_j} \frac{\partial \eta_{2i}}{\partial \gamma_l}, \quad j, l \in \{1, \ldots, q\},$$

where $t_i' = \partial t_i/\partial\phi_i$ represents the sensitivity of the dispersion residuals to changes in the dispersion parameter ϕ_i. The term $\text{E}(t_i')$ captures the expected curvature of the log-likelihood function established in (26) with respect to ϕ_i.

Defining $\boldsymbol{T} = \text{diag}\{-\text{E}(t_1'), \ldots, -\text{E}(t_n')\}$, we can express the Fisher information matrix for γ in matrix form as $\boldsymbol{K}_{\gamma\gamma} = \boldsymbol{J}_2^\top \boldsymbol{T} \boldsymbol{H}_\gamma^{-2} \boldsymbol{J}_2$, where $\boldsymbol{J}_2 = \partial\eta_2/\partial\gamma$ is the Jacobian of the dispersion model, containing the derivatives of the linear predictors η_{2i} with respect to γ, and $\boldsymbol{H}_\gamma = \text{diag}\{h'(\phi_1), \ldots, h'(\phi_n)\}$ is a diagonal matrix of the derivatives of the link function h with respect to ϕ_i. The matrix $\boldsymbol{T}$ encapsulates the expected information provided by each observation about the dispersion parameters.

This Fisher information matrix quantifies the amount of information the data provide about the dispersion parameters γ, and it is crucial for parameter estimation and inference, such as computing standard errors and confidence intervals. Under the assumption that the mean and dispersion parameters are orthogonal (that is, the cross-partial derivatives of the log-likelihood with respect to $\boldsymbol{\beta}$ and γ are zero), the Fisher information matrix for the full parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ is block-diagonal stated as

$$\boldsymbol{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \operatorname{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \boldsymbol{K}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\}. \tag{33}$$

The block-diagonal structure presented in (33) simplifies the estimation process, as it allows us to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ separately without loss of efficiency.

4.3. Generalized Standardized Combined Residual

Our goal is to develop a residual that incorporates the Fisher scoring iterative process, an optimization algorithm used to obtain ML estimates. This process is represented as an iterative weighted least squares procedure, leveraging the quadratic approximation of the log-likelihood function.

For the mean parameters $\boldsymbol{\beta}$, this iterative procedure is expressed as in (16) by

$$\boldsymbol{\beta}^{(m+1)} = \left(\boldsymbol{J}_1^\top \boldsymbol{W}^{(m)} \boldsymbol{J}_1\right)^{-1} \boldsymbol{J}_1^\top \boldsymbol{W}^{(m)} \boldsymbol{u}^{(m)},$$

where $\boldsymbol{W}^{(m)}$ is the weight matrix at mth iteration, reflecting the variance of the observations and the link function, and $\boldsymbol{u}^{(m)}$ is the working response vector, which incorporates the current estimates and residuals.

Similarly, for the dispersion parameters γ, the Fisher scoring update is given as

$$\boldsymbol{\gamma}^{(m+1)} = \left(\boldsymbol{J}_2^\top \boldsymbol{P}^{(m)} \boldsymbol{J}_2\right)^{-1} \boldsymbol{J}_2^\top \boldsymbol{P}^{(m)} \boldsymbol{u}^{*(m)}, \tag{34}$$

where $\boldsymbol{P}^{(m)} = \boldsymbol{T}^{(m)} (\boldsymbol{H}_\gamma^{(m)})^{-2}$ is the weight matrix for the dispersion model at mth iteration. This matrix combines the expected information from $\boldsymbol{T}^{(m)}$ with adjustments from the link function derivatives. The working response vector for the dispersion model, $\boldsymbol{u}^{*(m)} = \boldsymbol{J}_2 \boldsymbol{\gamma}^{(m)} + (\boldsymbol{T}^{(m)})^{-1} \boldsymbol{H}_\gamma^{(m)} \boldsymbol{t}$ namely, accounts for the current parameter estimates and the dispersion residuals $\boldsymbol{t}$.

In the expression stated in (34), $\boldsymbol{H}_\gamma^{(m)}$ and $\boldsymbol{T}^{(m)}$ are evaluated using the current parameter estimates $\boldsymbol{\gamma}^{(m)}$. This iterative weighted least squares procedure refines the parameter estimates by solving a sequence of linear approximations to the score equations. At each iteration, the weights and working responses are updated based on the current estimates, progressively improving the parameter estimates until convergence.

By developing a residual based on this iterative process, we can create diagnostic measures that reflect the influence of individual observations on the estimation of both mean and dispersion parameters. These residuals help to identify outliers or influential data points that may disproportionately affect the model fit.

Upon convergence of the iterative weighted least squares procedure, we obtain the estimate of the dispersion parameters $\widehat{\boldsymbol{\gamma}}$. Specifically, the converged estimate is given by

$$\widehat{\boldsymbol{\gamma}} = \left(\boldsymbol{J}_2^\top \widehat{\boldsymbol{P}} \boldsymbol{J}_2\right)^{-1} \boldsymbol{J}_2^\top \widehat{\boldsymbol{P}} \boldsymbol{u}^*, \tag{35}$$

where $\widehat{\boldsymbol{P}}$ is the weight matrix evaluated at the converged estimates, $\boldsymbol{J}_2$ is the Jacobian matrix of the dispersion model, and $\boldsymbol{u}^*$ is the working response vector for the dispersion sub-model.

The expression presented in (35) shows that $\widehat{\boldsymbol{\gamma}}$ can be interpreted as the weighted least squares estimator of the transformed response $\boldsymbol{u}^*$ on the covariates represented by $\boldsymbol{J}_2$, with weights given by $\widehat{\boldsymbol{P}}$.

By pre-multiplying both the response and the covariate matrix by $\widehat{\boldsymbol{P}}^{1/2}$, we can view the estimation process as an OLS problem formulated as

$$\widehat{\boldsymbol{\gamma}} = \left((\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{J}_2)^\top(\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{J}_2)\right)^{-1}(\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{J}_2)^\top(\widehat{\boldsymbol{P}}^{1/2}\boldsymbol{u}^*).$$

This formulation provides an intuitive understanding of the estimation process, emphasizing the role of the weights in adjusting for heteroscedasticity in the dispersion model.

The residuals for the dispersion parameters $\boldsymbol{\gamma}$ can be defined based on the least squares interpretation. The ordinary residual vector for $\boldsymbol{\gamma}$ is expressed as

$$\boldsymbol{r}^{\gamma} = \widehat{\boldsymbol{P}}^{1/2}(\boldsymbol{u}^* - \boldsymbol{J}_2\widehat{\boldsymbol{\gamma}}). \tag{36}$$

This expression interprets the residuals as the difference between the observed working responses $\boldsymbol{u}^*$ and the fitted values $\boldsymbol{J}_2\widehat{\boldsymbol{\gamma}}$, scaled by the square root of the weight matrix $\widehat{\boldsymbol{P}}$.

By substituting $\boldsymbol{u}^*$ given in (34) into the definition of the residual vector $\boldsymbol{r}^{\gamma}$ from the expression presented in (36), we obtain

$$\boldsymbol{r}^{\gamma} = \widehat{\boldsymbol{P}}^{1/2}\left(\boldsymbol{J}_2\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{T}}^{-1}\widehat{\boldsymbol{H}}_{\gamma}\boldsymbol{t} - \boldsymbol{J}_2\widehat{\boldsymbol{\gamma}}\right) = \widehat{\boldsymbol{P}}^{1/2}\widehat{\boldsymbol{T}}^{-1}\widehat{\boldsymbol{H}}_{\gamma}\boldsymbol{t}.$$

Recall that the weight matrix is defined as $\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{T}}\widehat{\boldsymbol{H}}_{\gamma}^{-2}$. Then, its square root is $\widehat{\boldsymbol{P}}^{1/2} = \widehat{\boldsymbol{T}}^{1/2}\widehat{\boldsymbol{H}}_{\gamma}^{-1}$. Substituting this into the expression for the residuals, we get

$$\boldsymbol{r}^{\gamma} = \left(\widehat{\boldsymbol{T}}^{1/2}\widehat{\boldsymbol{H}}_{\gamma}^{-1}\right)\widehat{\boldsymbol{T}}^{-1}\widehat{\boldsymbol{H}}_{\gamma}\boldsymbol{t} = \widehat{\boldsymbol{T}}^{-1/2}\boldsymbol{t}.$$

This simplification shows that the residuals $\boldsymbol{r}^{\gamma}$ are scaled versions of the dispersion residuals $\boldsymbol{t}$, adjusted by the inverse square root of the expected information from the Fisher information matrix. Therefore, for each observation, the individual residual is expressed as

$$r_i^{\gamma} = t_i/(-\mathrm{E}(t_i'))^{1/2}, \quad i \in \{1, \ldots, n\},$$

where $-\mathrm{E}(t_i')$ represents the negative expected value of the derivative of t_i with respect to ϕ_i, effectively capturing the curvature of the log-likelihood function with respect to the dispersion parameter.

Building on the residuals from both the mean and dispersion models, we can define a combined residual that incorporates information from both components. Specifically, the combined residual for observation i is defined as

$$r_i^{\beta\gamma} = (y_i - \widehat{\mu}_i) + t_i, \quad i \in \{1, \ldots, n\}, \tag{37}$$

reflecting the total deviation of the observed response from its expected value, accounting for both the mean and dispersion effects.

To standardize the combined residual presented in (37) and facilitate comparison across observations, we compute its variance given by

$$\zeta_i = \mathrm{Var}(r_i^{\beta\gamma}) = \mathrm{Var}((y_i - \widehat{\mu}_i) + t_i) = \mathrm{Var}(y_i) + \mathrm{Var}(t_i) + 2\,\mathrm{Cov}(y_i, t_i), \tag{38}$$

where $\mathrm{Var}(y_i) = \phi_i b''(\theta_i)$ represents the variance of the response variable from the mean model, and $\mathrm{Var}(t_i)$ captures the variance of the dispersion residuals. The covariance term $\mathrm{Cov}(y_i, t_i)$ accounts for any dependency between the response variable and the dispersion residuals. Calculating $\mathrm{Var}(t_i)$ and $\mathrm{Cov}(y_i, t_i)$ in the expression defined in (38) requires careful consideration of the statistical properties of t_i. As defined previously in (29), t_i represents the change in the log-likelihood function with respect to ϕ_i. The term $c'(y_i; \phi_i) = \partial c(y_i; \phi_i)/\partial\phi_i$ depends on the specific form of the function $c(y_i; \phi_i)$, which is model-dependent. Without specific assumptions about $c(y_i; \phi_i)$, the direct computation of $\mathrm{Var}(t_i)$ and $\mathrm{Cov}(y_i, t_i)$ can be challenging. However, these terms are crucial for accurately quantifying the combined residual variance and ensuring robust model diagnostics.

To address this challenge, we can use asymptotic approximations or empirical estimates. Under regularity conditions and assuming a large sample size, $\mathrm{Var}(t_i)$ can be approximated using the observed information as $\mathrm{Var}(t_i) \approx -\mathrm{E}(t_i') = -\mathrm{E}(\partial t_i/\partial \phi_i)$. Similarly, $\mathrm{Cov}(y_i, t_i)$ may be approximated by considering the covariance between the score functions for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. If the mean and dispersion parameters are orthogonal (that is, their cross-derivatives are zero), the covariance term may be negligible. Substituting these approximations, the variance of the combined residual simplifies to $\zeta_i = \mathrm{Var}(y_i) + \mathrm{Var}(t_i)$, for $i \in \{1, \ldots, n\}$. Thus, the standardized combined residual is expressed as

$$r_{\mathrm{S}_i}^{\boldsymbol{\beta}\boldsymbol{\gamma}} = \frac{(y_i - \widehat{\mu}_i) + t_i}{\sqrt{\phi_i b''(\theta_i) - \mathrm{E}(t_i')}}, \quad i \in \{1, \ldots, n\}.$$

In practice, the unknown quantities are replaced with their estimates evaluated at the fitted values as $\widehat{\zeta}_i = \widehat{\phi}_i b''(\widehat{\theta}_i) - \widehat{\mathrm{E}}(t_i')$, for $i \in \{1, \ldots, n\}$, yielding the estimated standardized combined residual formulated as

$$r_{\mathrm{S}_i}^{\boldsymbol{\beta}\boldsymbol{\gamma}} = \frac{(y_i - \widehat{\mu}_i) + \widehat{t}_i}{\sqrt{\widehat{\zeta}_i}}, \quad i \in \{1, \ldots, n\}. \tag{39}$$

This standardized residual presented in (39) combines the deviations from both the mean and dispersion models into a single metric, making it easier to identify observations that may be outliers or exert undue influence on the model fit. By appropriately scaling the combined residual, we ensure consistent interpretation across observations, facilitating effective diagnostic assessment.

In the case where the response variable follows a negative binomial distribution, calculating $\mathrm{Var}(t_i)$ and $\mathrm{Cov}(y_i, t_i)$ for the standardized combined residual may not be straightforward due to the complexity of the function $c'(y_i; \phi_i)$. This function arises from the derivative of the normalization term $c(y_i; \phi_i)$ in the exponential family representation of the negative binomial distribution. To estimate $\widehat{\mathrm{Var}}(c'(y_i; \widehat{\phi}_i))$ and $\widehat{\mathrm{Cov}}(y_i, c'(y_i; \widehat{\phi}_i))$, we can employ numerical methods. Specifically, these estimates are computed by evaluating the expectations over the distribution of Y_i using the fitted values of the model parameters. The estimated variance of $c'(y_i; \widehat{\phi}_i)$ is given by

$$\widehat{\mathrm{Var}}(c'(y_i; \widehat{\phi}_i)) = \widehat{\mathrm{E}}(c'(y_i; \widehat{\phi}_i)^2) - (\widehat{\mathrm{E}}(c'(y_i; \widehat{\phi}_i)))^2, \quad i \in \{1, \ldots, n\},$$

and the estimated covariance between y_i and $c'(y_i; \widehat{\phi}_i)$ is represented as

$$\widehat{\mathrm{Cov}}(y_i, c'(y_i; \widehat{\phi}_i)) = \widehat{\mathrm{E}}(y_i\, c'(y_i; \widehat{\phi}_i)) - \widehat{\mathrm{E}}(y_i)\, \widehat{\mathrm{E}}(c'(y_i; \widehat{\phi}_i)), \quad i \in \{1, \ldots, n\}.$$

To compute these above-mentioned expectations numerically, we use the probability function of the negative binomial distribution evaluated at the estimated parameters. Hence, the expectations are approximated as

$$\widehat{\mathrm{E}}(c'(y_i; \widehat{\phi}_i)) = \sum_{j=0}^{\infty} c'(j; \widehat{\phi}_i)\, \widehat{P}(Y_i = j), \quad \widehat{\mathrm{E}}(c'(y_i; \widehat{\phi}_i)^2) = \sum_{j=0}^{\infty} c'(j; \widehat{\phi}_i)^2\, \widehat{P}(Y_i = j), \quad \widehat{\mathrm{E}}(y_i\, c'(y_i; \widehat{\phi}_i)) = \sum_{j=0}^{\infty} j\, c'(j; \widehat{\phi}_i)\, \widehat{P}(Y_i = j),$$

where $\widehat{P}(Y_i = j)$ is the estimated probability that Y_i takes the value j, given by the negative binomial probability function evaluated at the estimated parameters $\widehat{\mu}_i$ and $\widehat{\phi}_i$, given by

$$\widehat{P}(Y_i = j) = \binom{j + r_i - 1}{j} \left(\frac{\widehat{\mu}_i}{\widehat{\mu}_i + r_i}\right)^j \left(\frac{r_i}{\widehat{\mu}_i + r_i}\right)^{r_i},$$

with $r_i = 1/\widehat{\phi}_i$. Since the sums over j from 0 to infinity are infinite, in practice, we truncate the sums at a sufficiently large value $\boldsymbol{J}_{\max}$, where the probabilities become negligible. This truncation allows for numerical computation while maintaining accuracy.

For example, we can choose $J_{\max}$ such that $\sum_{j=0}^{J_{\max}} \widehat{P}(Y_i = j) \geq 1 - \varepsilon$, where ε is a small number (for instance, 10^{-6}). This ensures that we capture almost all the probability mass in our calculations. By computing these sums numerically for each observation i, we obtain the estimates $\widehat{\mathrm{Var}}\big(c'(y_i;\widehat{\phi}_i)\big)$ and $\widehat{\mathrm{Cov}}\big(y_i, c'(y_i;\widehat{\phi}_i)\big)$ needed for the variance $\widehat{\zeta}_i$ of the combined residual. Substituting these estimates into the expression for the standardized combined residual, we have

$$r_{\mathrm{S}_i}^{\beta\gamma} = \frac{(y_i - \widehat{\mu}_i) + \widehat{t}_i}{\sqrt{\widehat{\phi}_i b''(\widehat{\theta}_i) + \widehat{\mathrm{Var}}\big(c'(y_i;\widehat{\phi}_i)\big) + 2\,\widehat{\mathrm{Cov}}\big(y_i, c'(y_i;\widehat{\phi}_i)\big)}}, \quad i \in \{1, \ldots, n\}. \tag{40}$$

Our approach allows us to compute the standardized combined residuals defined in (40) numerically, accommodating the complexities inherent in the negative binomial distribution. Thus, we maintain the diagnostic capabilities of the residuals while ensuring that the calculations remain feasible in practice.

5. Evaluation of Model Fit Using Simulated Envelopes

This section explores methods for assessing model fit through simulated envelopes, focusing on the analysis of residuals. The proposed techniques provide a robust approach to identifying model misspecifications and aberrant observations, even when traditional residual plots may be inadequate.

5.1. Simulated Envelopes for Residual Analysis

The residuals proposed in this study do not limit to normality. Consequently, the conventional thresholds of $[-2,\ 2]$ commonly used in residual plots for evaluating model fit against observation indices, covariate values, or predicted values may not be appropriate. To address this limitation, we use a method based on the empirical quantiles derived from simulated envelopes for residual analysis. This method serves as a helpful diagnostic tool, accurately capturing the underlying probabilistic structure of the residuals, even in situations where key assumptions regarding the data or model are not satisfied.

Simulated envelope-based probability plots are highly effective for evaluating the goodness of fit of statistical models and detecting potential outliers. Residuals that fall outside the bounds of the simulated envelope are flagged as outliers [49]. The procedure for constructing these plots is rigorously outlined in Algorithm 1. A visual representation of this algorithm is provided in the accompanying flowchart of Figure 1. To facilitate understanding of how this procedure can be implemented in practice, we provide a complete example in `R` code; see Appendix C. This example demonstrates the simulation and residual analysis steps discussed in this section, helping readers bridge the gap between theory and application.

The adequacy of the model fit is confirmed if, in the QQ plots with simulated envelopes, the residuals lie inside the bounds of the simulated confidence bands. This indicates that the distribution of the residuals from the model closely approximates the expected empirical distribution under the assumed model, thereby validating the model assumptions.

To evaluate the performance of the proposed residuals under different conditions for the mean and dispersion sub-models, we analyze the following scenarios:

(i) [Model 1] $Y_i \sim \mathrm{Gamma}(\mu_i;\phi_1)$ with $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, where $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1.5$, x_i are values from $X \sim \mathrm{U}(0.1, 1.1)$, z_i are values from $Z \sim \mathrm{U}(0.4, 1.4)$, and $i \in \{1, \ldots, 80\}$.

(ii) [Model 2] $Y_i \sim \mathrm{BN}(\mu_i;\phi_1)$ with $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, where $\beta_1 = 1.6$, $\beta_2 = 0.8$, $\gamma_1 = -0.3$, $\gamma_2 = 0.5$, x_i are values from $X \sim \mathrm{U}(0.1, 1.1)$, z_i are values from $Z \sim \mathrm{U}(0.4, 1.4)$, and $i \in \{1, \ldots, 80\}$.

The choice of these parameters was made to reflect realistic scenarios where both the mean and dispersion sub-models vary according to different covariates, allowing for a comprehensive evaluation of the residuals performance under varying conditions.

Figure 2 shows plots with envelopes for Models 1 and 2, generated under the correct model and $n = 80$. In these plots, the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, whereas the dashed blue line is the median. The empirical distributions of the residuals match the estimated distributions, indicating a good fit. Thus, we conclude that the standardized combined residual proves satisfactory performance across all analyzed scenarios, as all observed quantiles lie inside the simulated envelopes, suggesting a suitable fit.

Algorithm 1 Construction of simulated envelopes for residual analysis

Input: Fitted model, number of simulations m, sample size n
Output: Simulated envelopes for residuals, diagnostic plots
Step 1: Model fitting and initial residual calculation
Fit the initial model to the observed data to obtain the estimated residuals. The residuals can be ordinary residuals, deviance residuals, or any other residual type relevant to the model being analyzed.
Compute the residuals from the fitted model.
Step 2: Simulation procedure
for $j = 1$ to m **do**
    Generate a simulated sample of n independent observations based on the fitted model parameters.
    Refit the model to the simulated sample.
    Calculate and store the absolute values of the residuals from the refitted model.
end for
Step 3: Construction of empirical quantiles
for $i = 1$ to n **do**
    Obtain the empirical quantiles: 2.5th percentile, mean, and 97.5th percentile, for each ordered set of residuals. These quantiles represent the interval where 95% of the residuals are expected to fall, assuming the model is correctly specified.
end for
Step 4: Diagnostic plotting
Construct a quantile versus quantile (QQ) plot by plotting the ordered residuals from the original data against the theoretical quantiles of the standard normal distribution.
Overlay the simulated envelope, constructed from the empirical quantiles, onto the plot.
Step 5: Model fit assessment
Assess the model fit by evaluating whether the residuals lie inside the simulated envelope.
if the residuals fall within the simulated envelope **then**
    Conclude that the model provides an adequate fit.
else
    Identify potential model inadequacies.
    Optional step: Model check
    If important deviations are observed, consider checking the model. Possible check includes re-specifying the model structure, adjusting for heteroscedasticity, or considering alternative distributions.
    Refit the checked model and repeat Steps 1-5.
end if
Step 6: Application of model-specific diagnostics (if necessary)
if applicable (for example, in specific distributions such as binomial or Poisson models) **then**
    Use additional diagnostic checks relevant to the specific type of model (for example, checks for overdispersion, zero-inflation).
end if

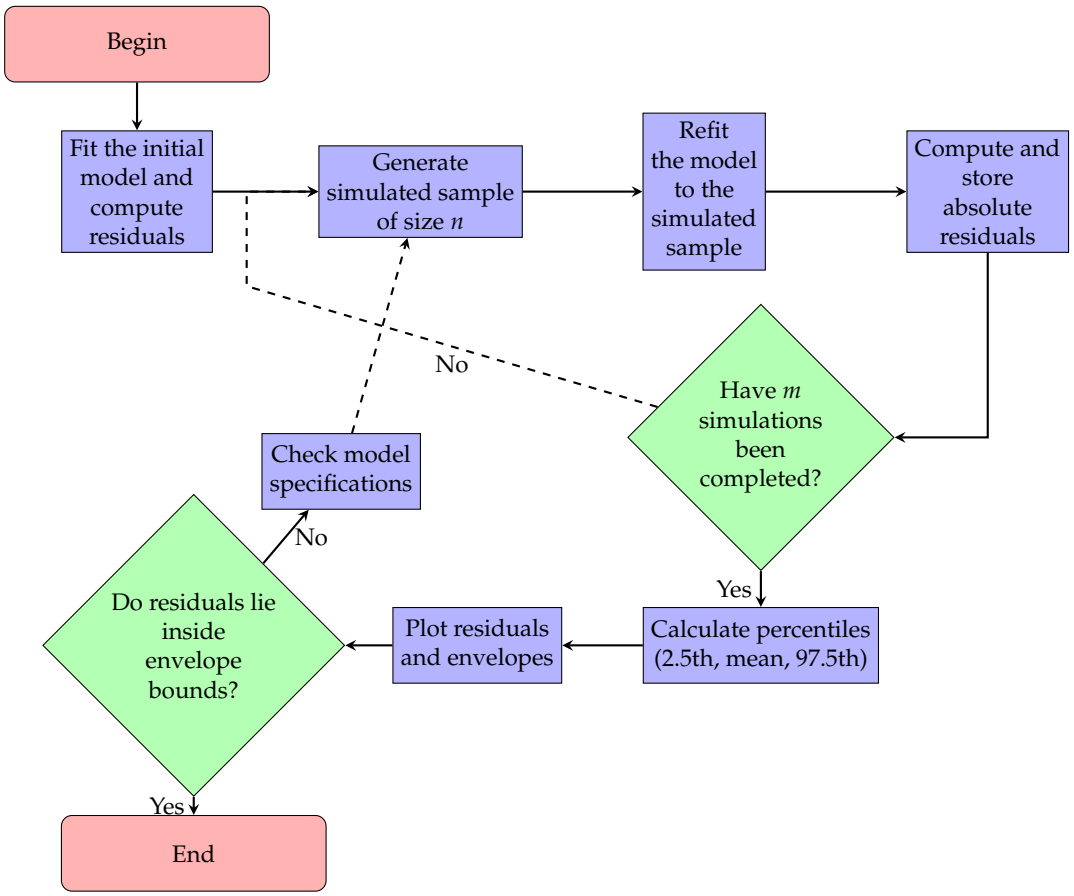
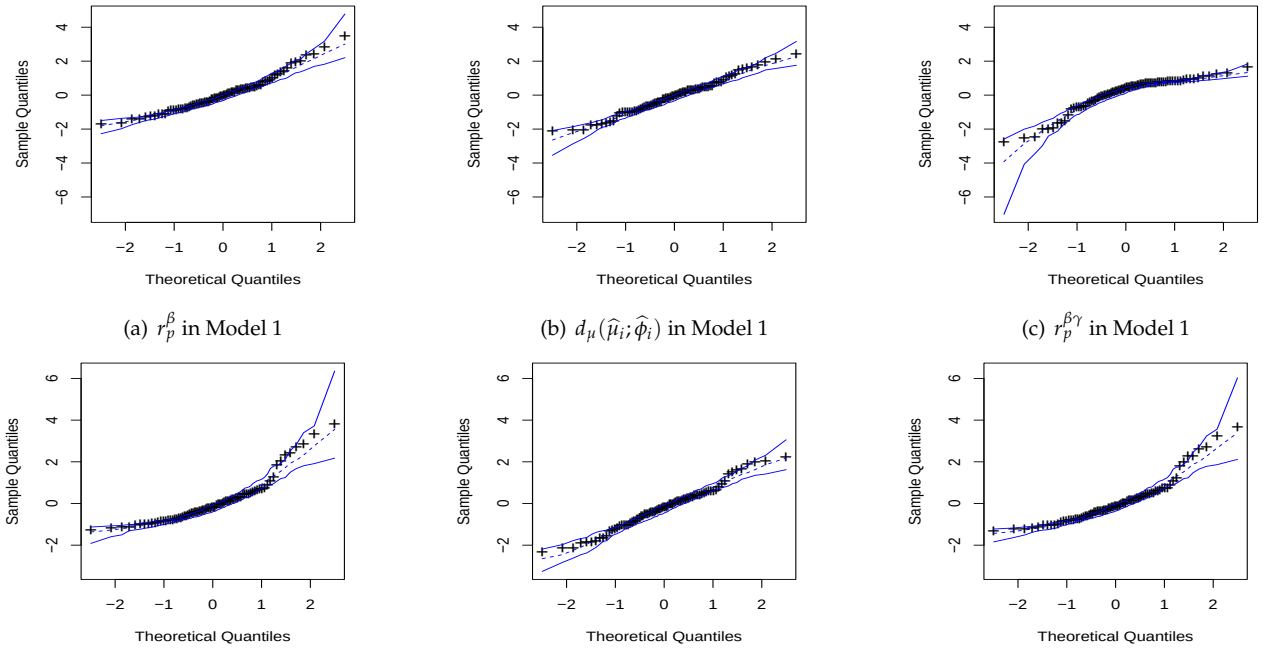


Figure 1. Flowchart illustrating the construction of simulated envelopes for residual analysis.



(a) r_p^{β} in Model 1 (b) $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$ in Model 1 (c) $r_p^{\beta\gamma}$ in Model 1

(d) r_p^{β} in Model 2 (e) $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$ in Model 2 (f) $r_p^{\beta\gamma}$ in Model 2

Figure 2. Probability plots with simulated envelope for the indicated model, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values.

5.2. Residual Analysis under Incorrect Model Specification

One of the common errors in modeling a dataset is assuming constant dispersion, while the real data-generating process may involve varying dispersion. Such misspecification can importantly impact residual analysis, as the empirical distribution of the residuals may deviate substantially from expectations based on correct model assumptions.

To detect this type of misspecification, we evaluate the empirical distribution of model residuals using probability plots. Typically, these plots are employed to assess model fit under correct specifications. However, when model assumptions are violated, the residual distributions can exhibit important deviations, such as skewness or kurtosis, that differ from those of a standard normal distribution. Even if the residuals appear approximately normally distributed, caution is necessary. Relying on conventional thresholds of the standard normal distribution—for example, $[-2,\ 2]$—to identify outliers may lead to misleading conclusions, as the true residual distribution may not adhere to this standard. To address this, we propose using thresholds derived from simulated envelopes. Specifically, we employ the 2.5th and 97.5th percentiles of the empirical distribution of simulated residuals as cut-off bands in residual plots. This employment is computationally efficient because it leverages the residuals already simulated.

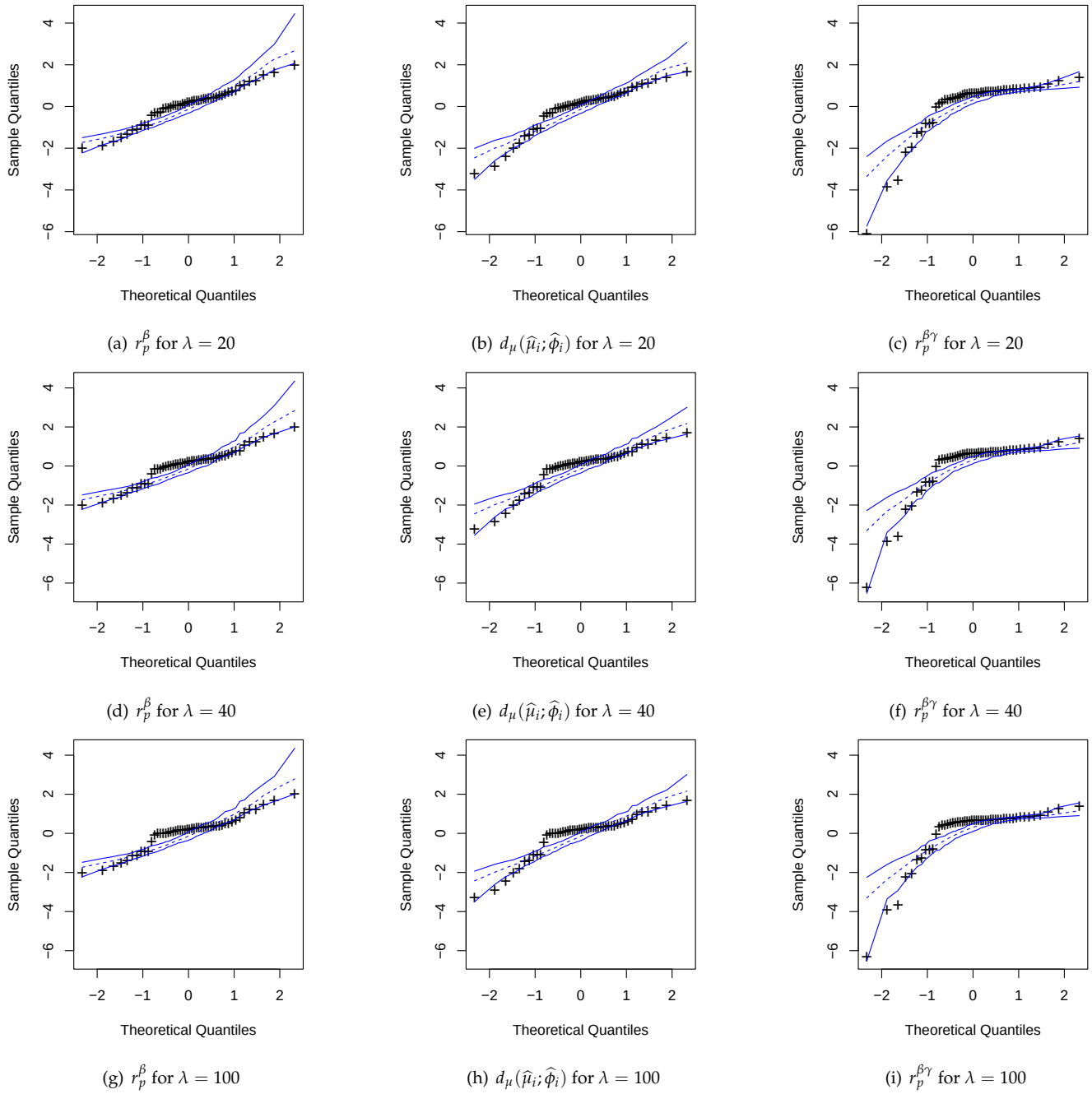
To illustrate our diagnostic method, we present examples where dispersion modeling is ignored, meaning that a model with constant dispersion is assumed. In these simulations, the true dispersion sub-model is given by $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, where the covariate value $z_i = 1$, for $i \in \{1, \ldots, 25\}$, and $z_i = 1.5$, for $i \in \{26, \ldots, 50\}$. The degree of heterogeneity is defined as $\lambda = \max\{\phi_1, \ldots, \phi_n\}/\min\{\phi_1, \ldots, \phi_n\}$. Note that λ depends on the true values of the parameters γ_1 and γ_2. For models where the response follows a continuous distribution (for example, normal, gamma, or inverse Gaussian —IG—), we consider $\gamma_1 = 1$ and $\gamma_2 = 2$, implying $\lambda \approx 20$; if $\gamma_2 = 2.46$, then $\lambda \approx 40$; and if $\gamma_2 = 3.07$, then $\lambda \approx 100$. For a response following a negative binomial distribution, we consider $\gamma_1 = 1$ and $\gamma_2 = 3.07$, implying $\lambda \approx 100$; if $\gamma_2 = 3.54$, then $\lambda \approx 200$; and if $\gamma_2 = 4$, then $\lambda \approx 400$. However, during parameter estimation, varying dispersion is disregarded, and only the mean of the response variable is modeled. Thus, we analyze the following scenarios:

(i) [Model 1] $Y_i \sim \text{Gamma}(\mu_i; \phi_i)$ with $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$, where $\beta_1 = 1$, $\beta_2 = 0.8$, and x_i are values from $X \sim \text{U}(0.1, 1.1)$, for $i \in \{1, \ldots, 50\}$.

(ii) [Model 2] $Y_i \sim \text{BN}(\mu_i; \phi_i)$ with $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$, where $\beta_1 = 1.6$, $\beta_2 = 0.8$, and x_i are values from $X \sim \text{U}(0.1, 1.1)$, for $i \in \{1, \ldots, 50\}$.

Figures 3 and 4 illustrate the behavior of residuals when the response variable follows gamma and negative binomial distributions, respectively. In both cases, the residuals disagree with the model specification, as the residuals fall outside the simulated envelope. This disagreement becomes more pronounced with increasing heterogeneity, denoted by λ. However, for the negative binomial response model, the detection of incorrect specification is more subtle. As λ increases, the probability plots show only slight differences, likely due to the robustness of the parameter estimation process across varying dispersion parameters.

The residual plots for detecting atypical observations are presented in Figures 5 and 6, which show the distribution of residuals versus observation indices for the gamma and negative binomial models, respectively. A consistent pattern is evident across all plots, initially indicating an incorrect model specification. Typically, a good model fit is characterized by randomly distributed residuals. Notably, the combined residuals more prominently highlight aberrant observations, as these points deviate further from the established cut-off bands. The proposed method using simulated envelopes offers an improvement over traditional methods of residual analysis, which often rely on fixed thresholds, such as $[-2, 2]$, that assume normality of residuals. This assumption is frequently violated in complex models, especially in GLM with varying dispersion. By constructing envelopes based on empirical quantiles from simulations, our diagnostic method provides a more accurate reflection of the underlying distribution of residuals, even when traditional residual plots may fail to capture deviations from model assumptions.



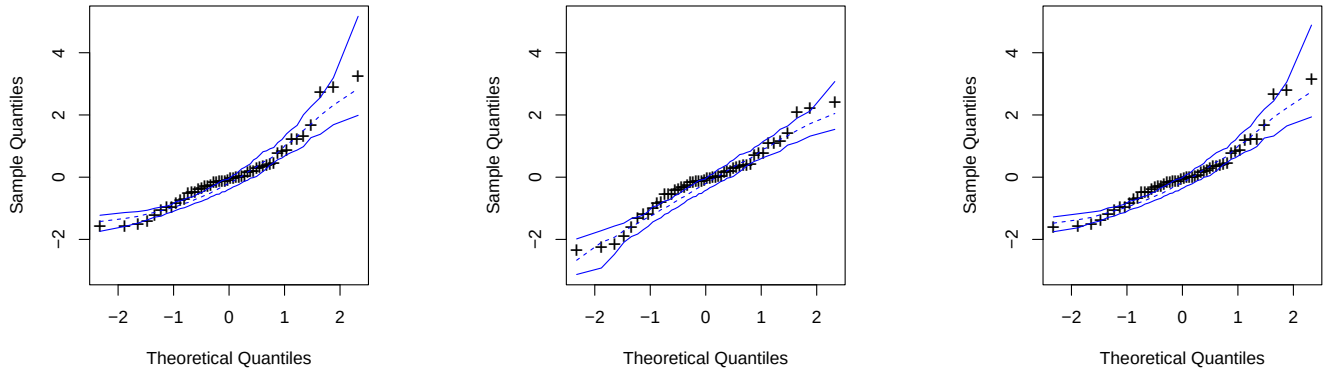
(a) r_p^{β} for $\lambda = 20$ (b) $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$ for $\lambda = 20$ (c) $r_p^{\beta\gamma}$ for $\lambda = 20$

(d) r_p^{β} for $\lambda = 40$ (e) $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$ for $\lambda = 40$ (f) $r_p^{\beta\gamma}$ for $\lambda = 40$

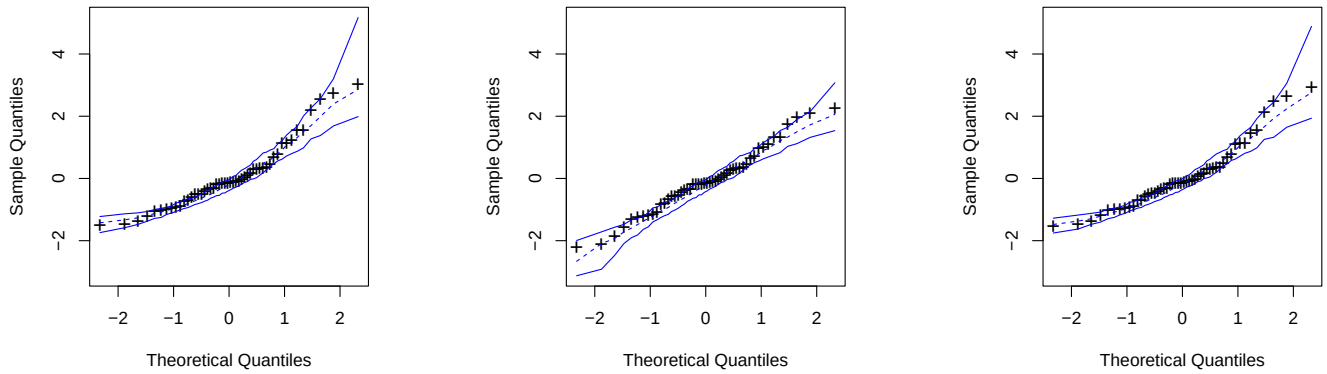
(g) r_p^{β} for $\lambda = 100$ (h) $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$ for $\lambda = 100$ (i) $r_p^{\beta\gamma}$ for $\lambda = 100$

Figure 3. Probability plots of standardized ordinary residuals, deviance residuals, and combined residuals under incorrect specification for Model 1 across different levels of heterogeneity λ, with simulated envelope for the indicated model, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values.

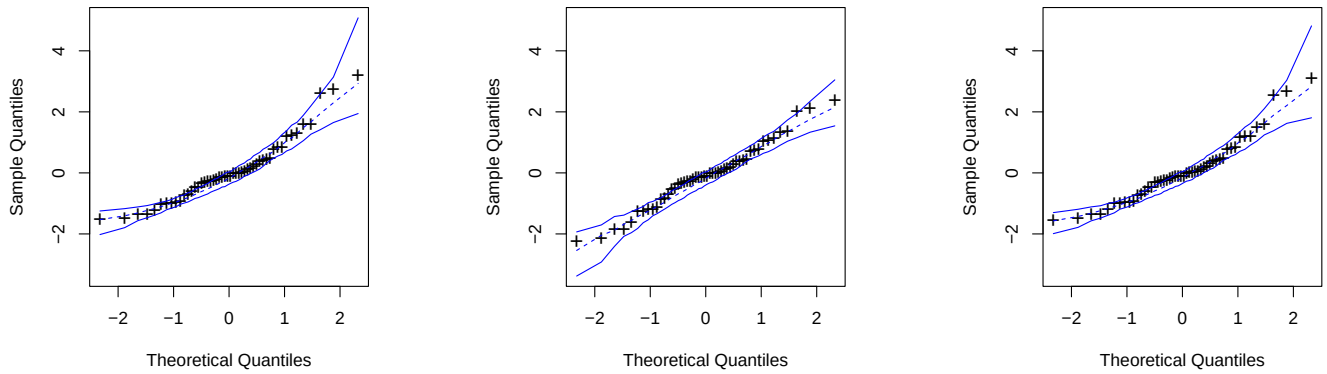
The above-mentioned construction makes our method more robust and reliable for detecting model misspecifications and outliers, as demonstrated in the performance evaluation across different simulated scenarios. It is also noteworthy that the new limits for detecting outliers differ from the conventional $[-2, 2]$ range in several situations. For instance, in a model with a negative binomial distribution, the upper limit is less than 2; see Figure 6.



(a) r_p^{β} for $\lambda = 100$ (b) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 100$ (c) $r_p^{\beta\gamma}$ for $\lambda = 100$



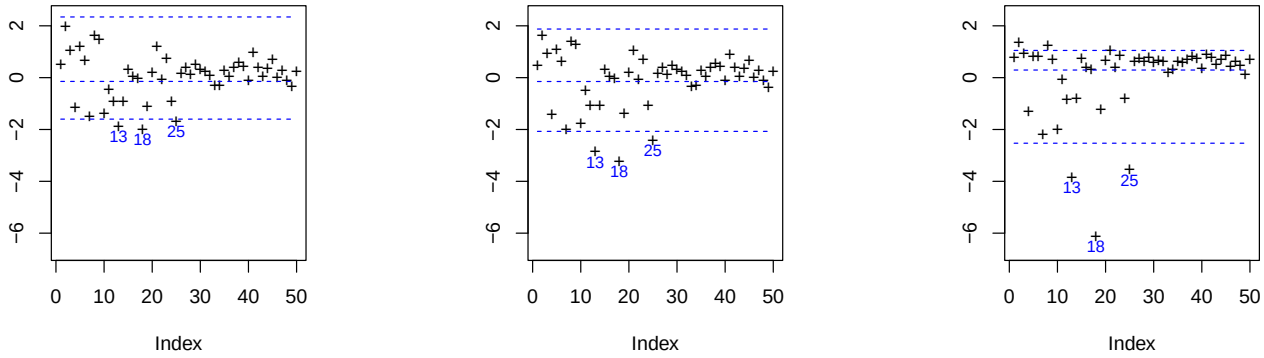
(d) r_p^{β} for $\lambda = 200$ (e) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 100$ (f) $r_p^{\beta\gamma}$ for $\lambda = 200$



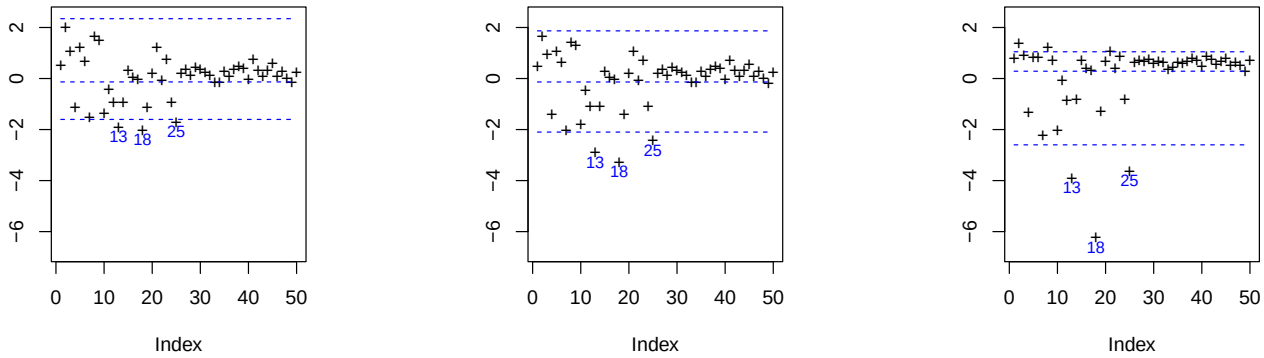
(g) r_p^{β} for $\lambda = 400$ (h) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 400$ (i) $r_p^{\beta\gamma}$ for $\lambda = 400$

Figure 4. Probability plots of standardized ordinary residuals (r_p^{β}), deviance residuals ($d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$), and combined residuals ($r_p^{\beta\gamma}$) under incorrect specification for Model 2, evaluated at different levels of heterogeneity λ, where the residuals are plotted for $\lambda = 100$, $\lambda = 200$, and $\lambda = 400$, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values.

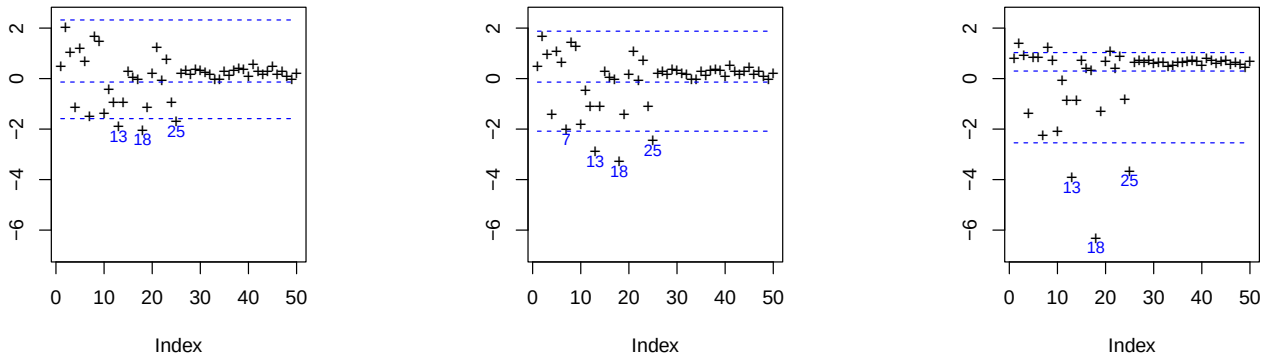
In the previous sections, we developed and evaluated the effectiveness of various residuals for diagnosing the goodness of fit in models with both mean and dispersion components. We now transition from theoretical considerations to practical illustrations of these concepts by means of empirical applications.



(a) r_p^{β} for $\lambda = 20$ (b) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 20$ (c) $r_p^{\beta\gamma}$ for $\lambda = 20$

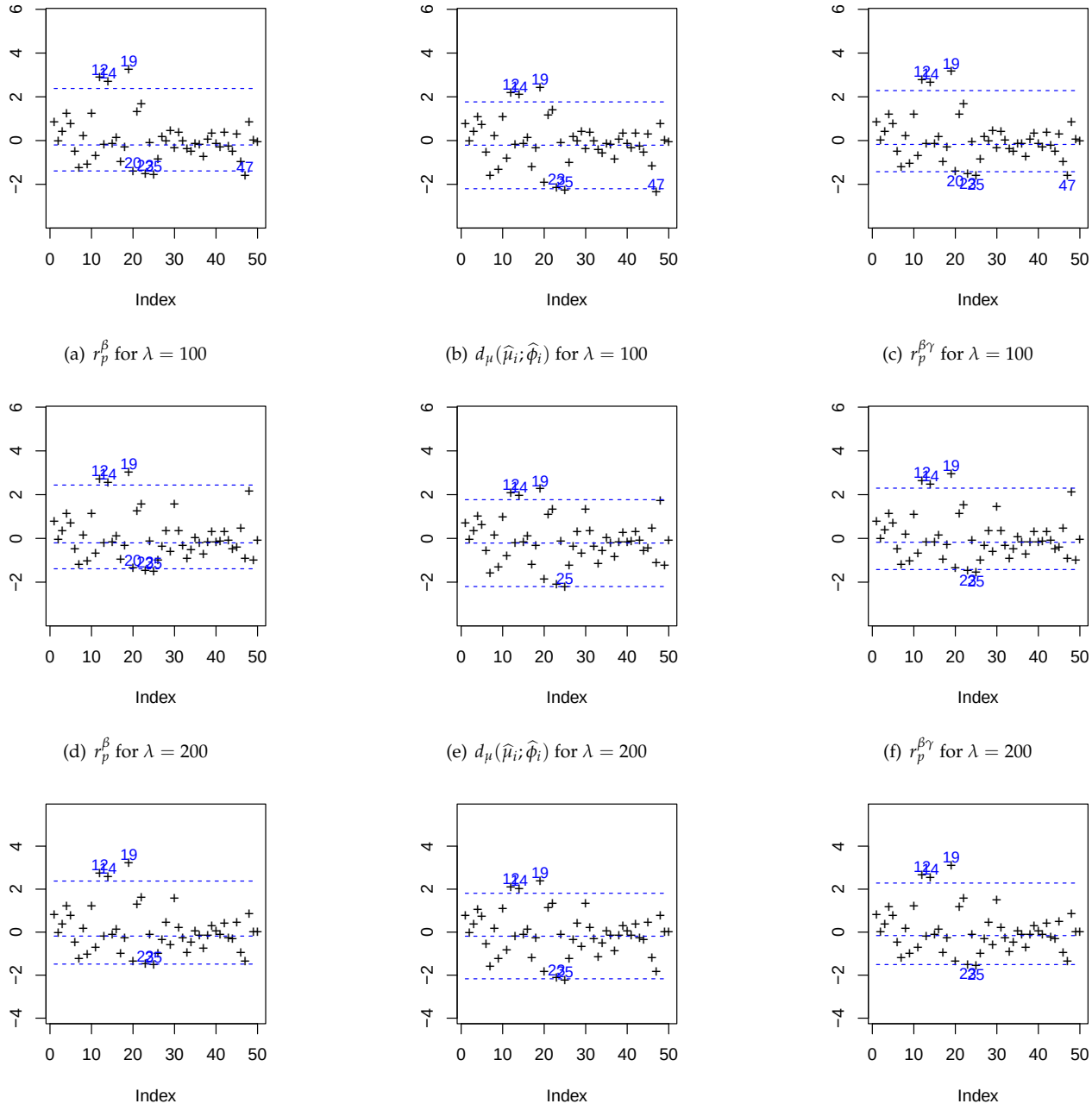


(d) r_p^{β} for $\lambda = 40$ (e) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 40$ (f) $r_p^{\beta\gamma}$ for $\lambda = 40$



(g) $r_p^{\beta\gamma}$ for $\lambda = 100$ (h) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 100$ (i) $r_p^{\beta\gamma}$ for $\lambda = 100$

Figure 5. Plots of residuals against observation indices under incorrect specification for Model 1, evaluated at different levels of heterogeneity λ, where the plots compare standardized ordinary residuals (r_p^{β}), deviance residuals ($d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$), and combined residuals ($r_p^{\beta\gamma}$) for $\lambda = 20$, $\lambda = 40$, and $\lambda = 100$, where the dashed blue lines are the thresholds $[-2, 2]$ of the standard normal distribution, the symbol "+" indicates the empirical values, and the blue numbers are outliers.



(a) r_p^{β} for $\lambda = 100$ (b) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 100$ (c) $r_p^{\beta\gamma}$ for $\lambda = 100$

(d) r_p^{β} for $\lambda = 200$ (e) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 200$ (f) $r_p^{\beta\gamma}$ for $\lambda = 200$

(g) r_p^{β} for $\lambda = 400$ (h) $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$ for $\lambda = 400$ (i) $r_p^{\beta\gamma}$ for $\lambda = 400$

Figure 6. Plots of standardized ordinary residuals (r_p^{β}), deviance residuals ($d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$), and combined residuals ($r_p^{\beta\gamma}$) against observation indices under incorrect specification for Model 2, evaluated at different levels of heterogeneity λ, where the plots correspond to $\lambda = 100$, $\lambda = 200$, and $\lambda = 400$, with the dashed blue lines being the thresholds $[-2, 2]$ of the standard normal distribution, the symbol "+" indicating the empirical values, and the blue numbers being outliers.

6. Empirical Applications

This section presents two empirical applications where the proposed standardized combined residual is used to analyze real datasets, highlighting its utility in detecting model inadequacies and identifying potential outliers.

6.1. Application 1: Shear Force in Snack Products

The dataset analyzed in this application, as presented in [1], originates from an experiment conducted at the Nutrition Department of the Faculty of Public Health, University of São Paulo, Brazil. This experiment was conducted to evaluate a new type of snack product, specifically focusing on the impact of replacing hydrogenated vegetable fat—traditionally used to fix the aroma in snacks—with canola oil. Five different formulations of the snack were tested over a period of 20 weeks, varying in the proportions of fat and canola oil, as detailed below:

- Characteristic A —22% fat, 0% canola oil.
- Characteristic B —0% fat, 22% canola oil.
- Characteristic C —17% fat, 5% canola oil.
- Characteristic D —11% fat, 11% canola oil.
- Characteristic E —5% fat, 17% canola oil.

The experimental design involved collecting 15 samples for each formulation in even-numbered weeks, resulting in a total of 150 observations per formulation across the study period. The primary response variable in this study was the force required to shear the snack, which serves as a measure of the product texture. The covariates include the group (A, B, C, D, or E) representing the snack formulation, and the week of collection.

Step 1: Model fitting and initial residual calculation

We denote y_{ijk} as the shear force for the kth replicate of the ith group in the jth week, where $k \in \{1, \ldots, 15\}$ represents the replicates within each group-week combination; $j \in \{2, 4, 6, \ldots, 20\}$ denotes the even-numbered weeks during which measurements were taken, and i corresponds to the different snack formulations, specifically groups A, B, C, D, and E. To analyze the data, we apply a joint modeling approach for both the mean and dispersion of the gamma-distributed response variable. The systematic components of the model are expressed as $\mu_{ij} = \beta_0 + \beta_i + \beta_6 \text{week}_j + \beta_7 \text{week}_j^2$ and $\log(\phi_{ij}) = \gamma_0 + \gamma_i + \gamma_6 \text{week}_j + \gamma_7 \text{week}_j^2$, where the baseline parameters β_0 and γ_0 represent the effects on the mean and dispersion, respectively, for Group A (22% fat, 0% canola oil); the coefficients β_i and γ_i ($i \in \{B, C, D, E\}$) account for the deviations in the mean and dispersion associated with groups B, C, D, and E relative to Group A; the terms β_6 and γ_6 capture the linear effect of the week on the mean and dispersion; while β_7 and γ_7 capture potential nonlinear trends over time through the quadratic term for the week. This model structure allows us to assess the impact of different formulations and the passage of time on the texture (shear force) of the snack products, while also accounting for variability within and between different formulations.

Step 2: Simulation procedure

To assess the adequacy of the model, we generate simulated samples and refit the model for each simulated sample. For each simulation, the absolute values of the residuals are computed and stored. This step follows the approach detailed in Appendix D.2, where the behavior of various residuals is empirically evaluated through Monte Carlo simulations.

Step 3: Construction of empirical quantiles

After generating and refitting the model m times, we calculate the empirical quantiles (2.5th percentile, mean, and 97.5th percentile) for each ordered set of residuals, as described in Appendix D, particularly in the context of gamma-distributed response variables as shown in Table A2.

Step 4: Diagnostic plotting

Then, we plot the ordered residuals from the original data against the theoretical quantiles of the standard normal distribution. The simulated envelopes are overlaid on the plot, as shown in Figure 7. This aligns with the simulation-based diagnostics discussed in Appendix D, which confirm the effectiveness of using simulated envelopes for assessing model adequacy.

Step 5: Model fit assessment

The parameter estimates provided in Table 2 reveal that Group A exhibits the highest shear force, indicating a firmer texture compared to the other groups, particularly groups D and E, which show the lowest shear forces. Additionally, Group A shows the highest variability in terms of the coefficient of variation. The standardized combined residual introduced in this study integrates information from both the mean and dispersion models, providing a more comprehensive diagnostic tool. Figure 7 presents the probability plot with simulated envelopes for these residuals, finding no evidence of model misfit, as the residuals fall within the expected bounds. This finding is consistent with the results from the Monte Carlo simulations presented in Appendix D.2, which validate the reliability of the standardized combined residual for diagnosing model adequacy in similar scenarios.

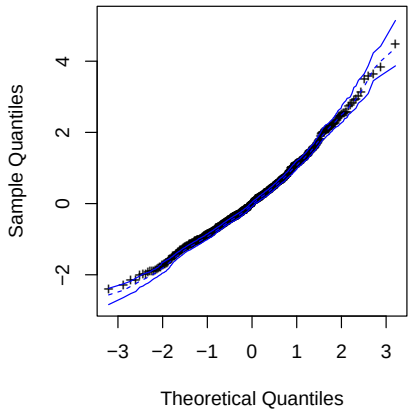


Figure 7. Probability plot of the standardized combined residual for the joint modeling of the mean and dispersion, assuming a gamma response for the snack data, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, with simulated envelopes, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values.

Table 2. ML estimates for the parameters from the joint modeling of mean and dispersion with a gamma response fitted to the snack data.

	Mean			Dispersion		
Effect	**Estimate**	**Standard Error**	***p*-Value**	**Estimate**	**Standard Error**	***p*-Value**
Constant	36.990	3.208	0.00	−1.560	0.214	0.00
Group B	−10.783	1.684	0.00	−0.477	0.161	0.00
Group C	−3.478	1.763	0.00	−0.050	0.160	0.07
Group D	−14.829	1.614	0.00	−0.815	0.161	0.00
Group E	−15.198	1.622	0.00	−0.817	0.161	0.00
week	5.198	0.525	0.00	−0.155	0.039	0.00
week^2	−0.189	0.212	0.00	0.005	0.001	0.00

6.2. Application 2: Root Production in Apple Shoots

In this application, we examine a dataset provided in [50], which explores the standardized combined residual when the model response variable is discrete. The data come from an experiment on apple crops, focusing on the number of roots produced by 270 micropropagated shoots. These shoots were grown under two different photoperiod conditions (8 or 16 h of light exposure) and five varying concentrations of benzylaminopurine (BAP) cytokinin. The variables used in the model are defined as follows:

- Y_i—Number of roots for the ith shoot;
- X_{i1}—Photoperiod (0 = 8 h, 1 = 16 h);
- Group 1—BAP cytokinin concentration of 2.2;
- Group 2—BAP cytokinin concentration of 4.4;
- Group 3—BAP cytokinin concentration of 8.8;
- Group 4—BAP cytokinin concentration of 17.6;

- Group 5—BAP cytokinin concentration of 18.6.

Step 1: Model fitting and initial residual calculation

We perform joint modeling of the mean and dispersion for the response variable, assuming a negative binomial distribution. The systematic component of the model is specified as $\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_j$ and $\log(\phi_{ij}) = \gamma_0 + \gamma_1 x_{i1} + \gamma_j$, where the index i represents the individual micropropagated shoot, for $i \in \{1, \ldots, 270\}$. Each shoot is subject to one of two photoperiod conditions (8 or 16 h of light exposure) and one of five BAP cytokinin concentration groups, whereas the index j corresponds to the BAP cytokinin concentration group, for $j \in \{1, \ldots, 5\}$, representing the five levels of cytokinin concentration tested in the experiment. The parameter β_0 represents the baseline effect on the mean number of roots (response variable) for Group 1 (BAP concentration of 2.2) under the 8-hour photoperiod condition. Similarly, γ_0 represents the baseline effect on the dispersion for the same group and condition. The coefficients β_1 and γ_1 capture the effect of increasing the photoperiod from 8 to 16 h on the mean and dispersion, respectively. The terms β_j and γ_j account for deviations in the mean and dispersion associated with each BAP concentration group relative to Group 1. The choice of setting $\beta_2 = 0$ and $\gamma_2 = 0$ implies that Group 1 serves as the reference group, allowing us to interpret the other groups' effects as deviations from this baseline. This model structure enables us to quantify both the direct effect of photoperiod on root production and how this effect varies across different cytokinin concentrations, providing a comprehensive understanding of the factors influencing root production in the experiment.

Step 2: Simulation procedure

To evaluate the model fit, we generate simulated samples and refit the model for each simulated dataset. For each simulation, the absolute values of the residuals are computed and stored. This procedure is aligned with the approach detailed in Appendix D, where similar simulations were conducted to assess the behavior of residuals in negative binomial models; see Table A4.

Step 3: Construction of empirical quantiles

After generating and refitting the model m times, we calculate the empirical quantiles (2.5th percentile, mean, and 97.5th percentile) for each ordered set of residuals, consistent with the methodology applied in Appendix D.

Step 4: Diagnostic plotting

Then, we plot the ordered residuals from the original data against the theoretical quantiles of the standard normal distribution. The simulated envelopes are overlaid on the plot, as shown in Figure 8. This technique of using simulated envelopes is validated in Appendix D, where the reliability of the standardized combined residuals in the context of negative binomial models is confirmed; see Table A4.

Step 5: Model fit assessment

The parameter estimates reported in Table 3 indicate that Group 1 exhibits the highest root production among the shoots, while groups 4 and 5 have the lowest root counts. Additionally, Group 1 shows the greatest variability in terms of the coefficient of variation. The probability plot of the standardized combined residual, shown in Figure 8, strongly suggests that the proposed model does not adequately fit the data.

Given the inadequacy of the initial model, we propose a simplified systematic component that only accounts for the photoperiod effect formulated as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1}, \quad \log(\phi_{ij}) = \gamma_0 + \gamma_1 x_{i1}. \tag{41}$$

Table 4 presents the ML estimates for the simplified model specified in (41).

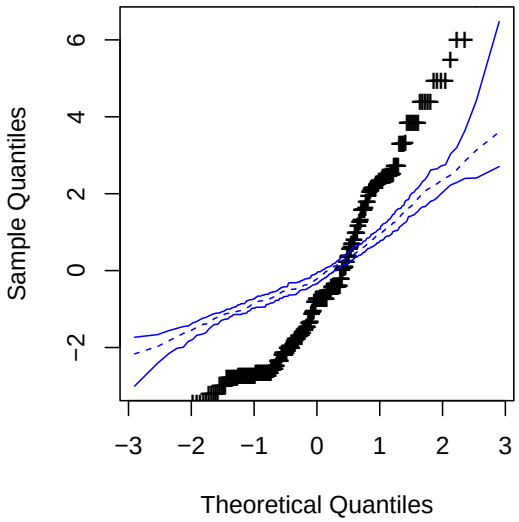


Figure 8. Probability plot of the standardized combined residual considering the joint modeling of the mean and dispersion, assuming a negative binomial response fitted to the apple dataset, with simulated envelopes, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, with simulated envelopes, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values.

Table 3. ML estimates for the parameters in the joint modeling of the mean and dispersion with a negative binomial response fitted to the apple dataset.

	Mean			Dispersion		
Effect	**Estimate**	**Standard Error**	***p*-Value**	**Estimate**	**Standard Error**	***p*-Value**
Constant	1.78	0.091	0.00	2.34	0.527	0.00
Photoperiod	−0.92	0.151	0.00	−3.79	0.463	0.00
Group 2	0.26	0.119	0.03	0.46	0.485	0.34
Group 3	0.23	0.110	0.03	0.92	0.491	0.06
Group 4	0.17	0.112	0.12	0.69	0.484	0.15
Group 5	−0.87	1.013	0.16	10.08	105.6	0.92

Table 4. ML estimates for the parameters in the simplified joint modeling of the mean and dispersion with a negative binomial response fitted to the apple dataset.

	Mean			Dispersion		
Effect	**Estimate**	**Standard Error**	***p*-Value**	**Estimate**	**Standard Error**	***p*-Value**
Constant	1.96	0.419	0.00	2.77	0.038	0.00
Photoperiod	−0.91	0.458	0.00	−3.68	0.152	0.00

The probability plot for the standardized combined residual from the checked model is presented in Figure 9. Unlike Figure 8, where most points are outside the simulated envelopes, the checked model probability plot shows that only a few points, particularly in the tails, are outside the envelopes. This suggests that the checked model provides a much better fit to the data, consistent with findings from the simulations of the negative binomial model presented in Appendix D; see Table A4.

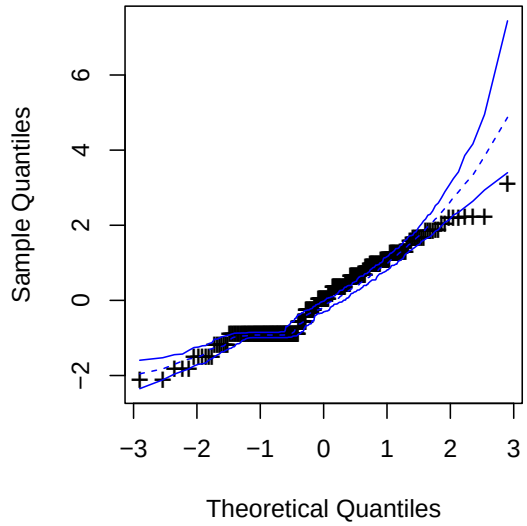


Figure 9. Probability plot of the standardized combined residual considering the joint modeling of the mean and dispersion, assuming a negative binomial response fitted to the apple dataset after model selection, with simulated envelopes, where the solid blue lines are the 2.5th and 97.5th percentiles of 100 simulated points for each observation, with simulated envelopes, whereas the dashed blue line is the median, with the symbol "+" indicating the empirical values..

7. Conclusions and Future Work

In this work, we introduced a new residual—the standardized combined residual—designed for classes of linear and nonlinear models within the exponential family, aimed at evaluating the goodness of fit in various scenarios. The need for this new residual arises from the limitations observed in traditional residuals, particularly when dealing with models that incorporate both mean and dispersion sub-models. Traditional residuals often fall short in adequately capturing the complexities of such models, especially in the presence of heteroscedasticity or nonconstant variance.

The standardized combined residual offers several advantages over existing methods. An important advantage is its construction, which integrates the Fisher iterative scoring process for both the mean and dispersion models. This allows for a unified diagnostic analysis, as it considers a single group of residuals that encapsulate information from both sub-models, rather than requiring separate analyses for each component. Our approach simplifies the evaluation process, particularly in generalized linear models and other models where variance is not constant. Moreover, the proposed residual does not require the computation of projection matrices, which are often computationally intensive and challenging to interpret. This results in reduced computational costs and enhances the practicality of the diagnostic process. The standardized combined residual also behaves similarly to the deviance residual, which is widely used in practical applications, making it an accessible tool for practitioners. A notable feature of the standardized combined residual is its independence from the method of maximization used to estimate model parameters. Whether using traditional optimization techniques or more advanced methods like simulated annealing [51], the residual relies solely on the parameter estimates, making it versatile and robust across different scenarios.

The empirical applications presented in this article, including the analysis of shear force in snack products and root production in apple shoots, showed the utility of the standardized combined residual in practical settings. In both cases, the new residual provided clear insights into model fit and highlighted inadequacies that were not as evident with traditional residuals. Specifically, the application to the apple shoot data underscored the importance of considering both the mean and dispersion in the model, as the initial model was found to be inadequate when evaluated with the new residual.

However, the standardized combined residual has certain limitations. Its effectiveness depends on the correct specification of both the mean and dispersion models. In scenarios where either sub-model is misspecified, the residual may not fully capture the inadequacies. Additionally, while it is computationally less intensive than methods requiring projection matrices, it still relies on accurate parameter estimation, which can be challenging in complex models or with limited data.

While the detailed derivation of certain asymptotic properties of the estimators is not the primary focus of this work, the developed framework provides clear insights into the robustness of the methods employed. The estimators, obtained through maximum likelihood methods and the Fisher scoring procedure, are expected to exhibit desirable properties, such as consistency and asymptotic normality, under regular conditions presented in Appendix B. The regular conditions applied, particularly in the context of exponential family distributions, support the notion that the proposed methodology provides accurate estimates for large samples, as further elaborated in the appendix. This reinforces the practical utility of the estimators, even in the absence of formal proofs of all properties in the present work.

Although the standardized combined residual offers several advantages, particularly in simplifying the diagnostic process and reducing computational costs, some trade-offs are present. By combining the information from both the mean and dispersion sub-models into a single residual, some level of detail may be sacrificed compared to using separate residuals for each component. However, this trade-off was a deliberate decision to balance simplicity with practicality, aiming to create an efficient and unified diagnostic tool. For most practical applications, the benefits of a straightforward interpretation and reduced computational burden outweigh the potential loss of specificity from separate residual analysis.

Future work should explore the extension of this residual to other classes of models, particularly those outside the exponential family. There is also potential for further refinement of the residual to enhance its robustness against model misspecifications. Additionally, the development of software tools to automate the computation and interpretation of the standardized combined residual would greatly benefit practitioners, making advanced diagnostic methods more accessible in applied settings.

In summary, while the standardized combined residual offers a unified and computationally efficient method for evaluating models that incorporate both mean and dispersion components, it also balances the trade-off between simplicity and the potential loss of granularity. Nonetheless, its development represents a step forward in statistical diagnostics, with the potential for broader applications and continued refinement in future research.

Author Contributions: Conceptualization: R.O., P.L.E., V.L. and C.C.; data curation: R.O. and L.A.A.; formal analysis: R.O., P.L.E., L.A.A., C.M.X., V.L. and C.C.; investigation: R.O., P.L.E., L.A.A., C.M.X., V.L. and C.C.; methodology: R.O., P.L.E. L.A.A., C.M.X., V.L. and C.C.; writing—original draft: R.O., P.L.E., L.A.A. and C.M.X.; writing—review and editing: V.L. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), under Grant Numbers 303192/2022-4 and 402519/2023-0 and Fundação de Amparo a Ciência e Tecnologia do Estado da Bahia —FAPESB—, No. APP0021/2023 (R.O.); the Vice-rectorate for Research, Creation, and Innovation (VINCI) of the Pontificia Universidad Católica de Valparaíso —PUCV—, Chile, under grants VINCI 039.470/2024 (regular research), VINCI 039.493/2024 (interdisciplinary associative research), VINCI 039.309/2024 —PUCV centenary—, and FONDECYT 1200525 (V.L.) by rom the National Agency for Research and Development —ANID— of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation; and by Portuguese funds through the Research Centre of Mathematics —CMAT— of University of Minho, Portugal, within projects UIDB/00013/2020 —https://doi.org/10.54499/UIDB/00013/2020, accessed on 5 October 2024— and UIDP/00013/2020 —https://doi.org/10.54499/UIDP/00013/2020, accessed on 5 October 2024— (C.C.).

Data Availability Statement: The data and codes used in this study are available in the present article and also can be requested from the authors.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their valuable comments and suggestions, which helped us to improve the quality of this article.

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

Appendix A. Canonical Form of the Exponential Family

Consider the function $c(y_i;\phi_i)$ defined as $c(y_i;\phi_i) = d(\phi_i) + \phi_i a(y_i) + u(y_i)$. Using this definition, the expression stated in (25) can be reformulated as $f(y_i,\theta_i,\phi_i) = \exp(\phi_i\theta_i y_i + \phi_i a(y_i) - (\phi_i b(\theta_i) - d(\phi_i)) + u(y_i))$. From this formulation, the following components can be identified: $\tau_1(y_i) = y_i$, $\tau_2(y_i) = a(y_i)$, $c_1(\theta_i;\phi_i) = \theta_i\phi_i$, $c_2(\theta_i;\phi_i) = \phi_i$, and $\delta(\theta_i;\phi_i) = \phi_i b(\theta_i) - d(\phi_i)$. The canonical parameters are, therefore, given by $\nu_1 = \theta_i\phi_i \Rightarrow \quad \theta_i = \nu_1/\nu_2$, and $\nu_2 = \phi_i$. Consequently, the expression for $\delta_0(\nu_1;\nu_2)$ becomes $\delta_0(\nu_1;\nu_2) = \nu_2 b(\nu_1/\nu_2) - d(\nu_2)$. The expectations and variances of the sufficient statistics τ_1 and τ_2 are then derived as

$$\mathrm{E}(\tau_1) = \frac{\partial\delta_0(\nu_1;\nu_2)}{\partial\nu_1} = b'\left(\frac{\nu_1}{\nu_2}\right) = b'(\theta_i), \tag{A1}$$

$$\mathrm{Var}(\tau_1) = \frac{\partial^2\delta_0(\nu_1;\nu_2)}{\partial\nu_1^2} = \frac{1}{\nu_2}b''\left(\frac{\nu_1}{\nu_2}\right) = \frac{1}{\phi_i}b''(\theta_i), \tag{A2}$$

$$\mathrm{E}(\tau_2) = \frac{\partial\delta_0(\nu_1;\nu_2)}{\partial\nu_2} = b\left(\frac{\nu_1}{\nu_2}\right) - \frac{\nu_1}{\nu_2}b'\left(\frac{\nu_1}{\nu_2}\right) - d'(\nu_2), \tag{A3}$$

$$= b(\theta_i) - \theta_i b'(\theta_i) - d'(\phi_i), \tag{A4}$$

$$\mathrm{Var}(\tau_2) = \frac{\partial^2\delta_0(\nu_1;\nu_2)}{\partial\nu_2^2} = \frac{\nu_1^2}{\nu_2^3}b''\left(\frac{\nu_1}{\nu_2}\right) - d''(\nu_2) = \frac{\theta_i^2}{\phi_i}b''(\theta_i) - d''(\phi_i), \tag{A5}$$

$$\mathrm{Cov}(\tau_1,\tau_2) = \frac{\partial\delta_0(\nu_1;\nu_2)}{\partial\nu_1\partial\nu_2} = -\frac{\nu_1}{\nu_2}b''\left(\frac{\nu_1}{\nu_2}\right) = -\frac{\theta_i}{\phi_i}b''(\theta_i). \tag{A6}$$

For further details on these derivations and the broader context of the exponential family, refer to [46,52].

Appendix B. Fisher Scoring Iterative Method

The score function derived from (3), (4), and (25) is given by $\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = U_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(U_\beta^\top, U_\gamma^\top\right)^\top$, where $U_\beta = \partial\ell(\boldsymbol{\theta})/\partial\beta$ and $U_\gamma = \partial\ell(\boldsymbol{\theta})/\partial\gamma$. The Fisher scoring iterative method for β is expressed as

$$\begin{aligned}
\beta^{(m+1)} &= \boldsymbol{\beta}^{(m)} + (\boldsymbol{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(m)})^{-1}\boldsymbol{U}_{\boldsymbol{\beta}^{(m)}} = \boldsymbol{\beta}^{(m)} + \left(\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y}-\boldsymbol{\mu}^{(m)}) \\
&= \left(\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\boldsymbol{\beta}^{(m)} + \left(\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y}-\boldsymbol{\mu}^{(m)}) \\
&= \left(\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\left(\boldsymbol{\eta}_1^{(m)} + \boldsymbol{W}^{-1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y}-\boldsymbol{\mu}^{(m)})\right) \\
&= \left(\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{\Phi}^{(m)}\boldsymbol{W}^{(m)}\boldsymbol{u}_1,
\end{aligned}$$

where $\boldsymbol{u}_1 = \boldsymbol{\eta}_1^{(m)} + \boldsymbol{W}^{-1/2(m)}\boldsymbol{V}^{-1/2(m)}(\boldsymbol{y}-\boldsymbol{\mu}^{(m)})$. Similarly, the Fisher scoring iterative method for γ is given by

$$\begin{aligned}
\gamma^{(m+1)} &= \gamma^{(m)} + \{\boldsymbol{K}_{\gamma\gamma}^{(m)}\}^{-1}\boldsymbol{U}_{\gamma^{(m)}} = \gamma^{(m)} + \left(\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{H}_\gamma^{-1(m)}(\boldsymbol{t}-\boldsymbol{\mu}^{(m)}) \\
&= \left(\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\gamma^{(m)} + \left(\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{H}_{\gamma^{-1(m)}}(\boldsymbol{t}-\boldsymbol{\mu}^{(m)}) \\
&= \left(\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\left(\boldsymbol{\eta}_2 + \boldsymbol{P}^{-1(m)}\boldsymbol{H}_{\gamma^{-1(m)}}(\boldsymbol{t}-\boldsymbol{\mu}^{(m)})\right) = \left(\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{P}^{(m)}\boldsymbol{u}_2,
\end{aligned}$$

where $\boldsymbol{u}_2 = \boldsymbol{\eta}_2 + \boldsymbol{P}^{-1(m)}\boldsymbol{H}_{\gamma^{-1(m)}}(\boldsymbol{t} - \boldsymbol{\mu}^{(m)})$. The ordinary residual for β is defined as $r^{\beta} = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}(\boldsymbol{u}_1 - \widehat{\boldsymbol{\eta}}_1) = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\Big(\widehat{\boldsymbol{\eta}}_1 + \widehat{\boldsymbol{W}}^{-1/2}\widehat{\boldsymbol{V}}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}) - \widehat{\boldsymbol{\eta}}_1\Big) = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{W}}^{1/2}\widehat{\boldsymbol{W}}^{-1/2}\widehat{\boldsymbol{V}}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}}) = \widehat{\boldsymbol{\Phi}}^{1/2}\widehat{\boldsymbol{V}}^{-1/2}(\boldsymbol{y} - \widehat{\boldsymbol{\mu}})$. Thus, the residual for observation i is stated as $r_i^{\beta} = (y_i - \widehat{\mu}_i)/(\phi_i^{-1}V_i)^{1/2}$, for $i \in \{1, \ldots, n\}$. Similarly, the ordinary residual for γ is given by $r^{\gamma} = \widehat{\boldsymbol{P}}^{1/2}(\boldsymbol{u}_2 - \widehat{\boldsymbol{\eta}}_2) = \widehat{\boldsymbol{P}}^{1/2}(\widehat{\boldsymbol{\eta}}_2 + \widehat{\boldsymbol{P}}^{-1}\widehat{\boldsymbol{H}}_{\gamma}^{-1}(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^{*}) - \widehat{\boldsymbol{\eta}}_2) = \widehat{\boldsymbol{P}}^{1/2}\widehat{\boldsymbol{P}}^{-1}\widehat{\boldsymbol{H}}_{\gamma^{-1}}(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^{*}) = \widehat{\boldsymbol{P}}^{-1/2}\widehat{\boldsymbol{H}}_{\gamma^{-1}}(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^{*})$. For $\boldsymbol{P} = \boldsymbol{V}_{\gamma}\boldsymbol{H}_{\gamma^{-2}}$, we have $r^{\gamma} = \boldsymbol{V}_{\gamma^{-1/2}}(\boldsymbol{t} - \widehat{\boldsymbol{\mu}}^{*})$. Therefore, the residual for observation i is defined as $r_i^{\gamma} = (t_i - \widehat{\mu}_i^{*})/(-d''(\phi_i)))^{1/2}$, for $\in \{1, \ldots, n\}$.

Appendix C. R Code for Simulation and Residual Analysis

This appendix provides an example of how the theoretical models discussed in the article can be implemented in practice using R. The code simulates data, fits a GLM with varying dispersion, calculates residuals, and performs residual analysis using simulated envelopes.

Listing A1. R code for simulation and residual analysis

```
# Load necessary libraries
library(MASS)        # For negative binomial GLM fitting
library(gamlss.dist) # For distribution functions like~rDPO

# Set seed for reproducibility
set.seed(12345)

# Define parameters
N = 1         # Number of simulations (set to 1 for illustration)
n = 50        # Sample size
b.1 = 1       # True parameter beta_1 for the mean sub-model
b.2 = 0.8     # True parameter beta_2 for the mean sub-model
g.1 = -0.4    # True parameter gamma_1 for the dispersion
    sub-model
g.2 = -0.2    # True parameter gamma_2 for the
    dispersion~sub-model

# Generate covariates x and z
x = runif(n, 0.1, 1.1)    # Covariate for the mean sub-model
z = runif(n, 0.2, 1.2)    # Covariate for the
    dispersion~sub-model

# Calculate linear predictors for mu and phi
eta.mu = b.1 + x^b.2
eta.phi = g.1 + z^g.2

# Transform linear predictors to obtain mu and phi
mu = exp(eta.mu)     # Mean parameter
phi = exp(eta.phi)   # Dispersion~parameter

# Define functions for the simulation and residual~analysis

# Log-likelihood function for the negative binomial model
logbn <- function(theta, y, x, z) {
  n = length(y)
  eta.mu = theta[1] + x^theta[2]
  eta.phi = theta[3] + z^theta[4]
  mu = exp(eta.mu)
  phi = exp(eta.phi)
  log.lik = numeric(n)
  for (i in 1:n) {
```

```
    # Negative binomial log-likelihood for each observation
    log.lik[i] = phi[i] * log(phi[i] / (phi[i] + mu[i])) +
    y[i] * log(mu[i] / (mu[i] + phi[i])) +
    lgamma(y[i] + phi[i]) - lgamma(phi[i]) - lgamma(y[i] + 1)
  }
  return(-sum(log.lik))  # Return negative log-likelihood for
      minimization
}

# Function to calculate the adjustment term cc for the combined
    residual
cc <- function(y, mu, phi) {
  temp = phi^(-1) - mu + y * (log(mu * phi) + 1) - digamma(phi
      * y + 1)
  return(temp)
}

# Function to calculate the variance components for ksi
    estimation
e_psi2 <- function(func, mu, phi) {
  n = length(mu)
  temp1 = numeric(n)
  for (j in 1:n) {
    i = 1
    temp0 = 0
    repeat {
      term = func(k = phi[j], v = i) * dDPO(i, mu = mu[j],
          sigma = phi[j])
      temp0 = temp0 + term
      i = i + 1
      if (abs(term) < 1e-8) break
    }
    temp1[j] = temp0 + func(phi[j], 0) * dDPO(0, mu = mu[j],
        sigma = phi[j])
  }
  return(temp1)
}

# Functions needed for variance calculations
e1 <- function(k, v) { (-digamma(k * v + 1))^2 }
e2 <- function(k, v) { -digamma(k * v + 1) }
e3 <- function(k, v) { -v * digamma(k * v + 1) }

# Function to estimate ksi for the combined residual variance
ksi_est <- function(mu, phi, var_psi, cov_psi) {
  temp = ((log(mu * phi) + 2)^2) * mu / phi + var_psi +
  2 * (log(mu * phi) + 2) * cov_psi
  return(temp)
}

# Function to calculate deviance residuals
devr <- function(y, mu, phi) {
  temp = numeric(length(y))
  for (i in 1:length(y)) {
    if (y[i] == 0) {
      temp[i] = sqrt(2 * mu[i] * phi[i])
    } else {
      temp[i] = sqrt(2 * phi[i] * (mu[i] - y[i] * (1 +
          log(mu[i] / y[i]))))
```

```
    }
    temp[i] = sign(y[i] - mu[i]) * temp[i]
  }
  return(temp)
}

# Main simulation loop
k = 1  # Simulation~counter

while (k <= N) {

  # Generate responses from the negative binomial distribution
  y = rDPO(n, mu = mu, sigma = phi)  # Simulated response~variable

  # Fit the initial model using glm.nb to obtain starting values
  fit = glm.nb(y ~ x, link = log)
  beta0 = coef(fit)

  # Estimate initial phi values using residuals from the mean model
  mu.est_1 = exp(beta0[1] + x * beta0[2])
  e_1 = y - mu.est_1
  var_1 = sum(e_1^2) / (n - 2)
  phi_1 = mu.est_1 / var_1

  # Fit a preliminary dispersion model using gamma regression
  fit2 = glm(log(phi_1) ~ z, family = Gamma(link = "log"))
  gamma0 = coef(fit2)

  # Combine initial estimates
  val_ini = c(beta0, gamma0)

  # Estimate model parameters using maximum likelihood estimation
  lp = optim(val_ini, logbn, method = "BFGS", hessian = TRUE,
  y = y, x = x, z = z)

  # Extract estimated parameters
  eta.mu.est = lp$par[1] + x^lp$par[2]
  eta.phi.est = lp$par[3] + z^lp$par[4]
  mu.est = exp(eta.mu.est)
  phi.est = exp(eta.phi.est)

  # Calculate the adjustment term cc
  ci = cc(y, mu.est, phi.est)

  # Compute the numerator of the combined residual
  r = y - mu.est + ci

  # Calculate variance components for ksi
  var_psi = e_psi2(func = e1, mu.est, phi.est) -
  (e_psi2(func = e2, mu.est, phi.est))^2
  cov_psi = e_psi2(func = e3, mu.est, phi.est) -
  mu.est * e_psi2(func = e2, mu.est, phi.est)

  # Estimate ksi for the combined residual variance
  ksi = ksi_est(mu.est, phi.est, var_psi, cov_psi)

```

```
    # Compute the standardized combined residual
    r_p = r / sqrt(ksi)

    # (Optional) Plot the QQ plot of the standardized combined
        residual
    qqnorm(r_p, main = "QQ plot of standardized combined
        residual")
    qqline(r_p, col = "red")

    # Increment simulation counter
    k = k + 1
  }
```

Appendix D. Further Results of Simulations

This appendix provides additional details and results that complement the main findings presented in the article. It includes the numerical evaluations performed through simulations to assess the behavior of various residuals within the context of the exponential family with varying dispersion. The material is organized into subsections that present the setup and results of Monte Carlo simulations designed to evaluate the empirical distribution of these residuals under different model specifications.

Appendix D.1. Numerical Evaluation

Next, we present the numerical evaluation of the empirical distribution of standardized ordinary residuals, deviance residuals, and combined residuals. The analysis is performed using models from the exponential family with varying dispersion parameters. The evaluation includes simulations under correctly specified models, as well as scenarios involving model misspecification and the presence of outliers.

Appendix D.2. Residual Evaluation through Monte Carlo Simulations

To characterize the empirical distribution of standardized ordinary residuals, standardized deviance, and standardized combined residuals, we conducted Monte Carlo simulations with 5000 replications under various scenarios using a generalized nonlinear model with varying dispersion. Note that, in this study, we focus exclusively on the deviance residual for the mean, as defined in (13). The empirical distribution of these residuals was evaluated using several key statistical measures as follows: mean, standard error, skewness, and kurtosis. For the standard normal distribution, these measures are expected to be 0, 1, 0, and 0, respectively.

In the first scenario, we assume that the response variable follows a normal distribution with mean μ_i and variance ϕ_i^{-1}, which is a commonly used distribution within the exponential family. The sample size is set to $n = 30$. Thus, the variables $y_1, \ldots, y_{30}$ are independent random variables, each following a normal distribution with mean μ_i and variance ϕ_i^{-1}, denoted as $Y_i \sim \mathrm{N}(\mu_i; \phi_i^{-1})$, for $i \in \{1, \ldots, 30\}$. In this scenario, the mean of the response variable and the dispersion parameter are modeled as functions of the covariates, with the relationships stated as

$$\mu_i = \beta_1 + x_i^{\beta_2}, \quad \log(\phi_i) = \gamma_1 + z_i^{\gamma_2}, \quad i \in \{1, \ldots, 30\}, \tag{A7}$$

where the values x_i and z_i of covariates are independently generated from uniform distributions. Specifically, x_i are values from $X \sim \mathrm{U}(0.1, 1.1)$ and z_i are values from $Z \sim \mathrm{U}(0.4, 1.4)$. These covariates values are held constant across all scenarios analyzed. The parameter values used in this scenario are $\beta_1 = -1.4$, $\beta_2 = 0.8$, $\gamma_1 = 0.9$, and $\gamma_2 = -1$.

In Table A1, we observe that the means of the evaluated residuals are close to zero, and, similarly, the standard errors are close to one, indicating that the residuals are centered and scaled appropriately. However, the skewness and, more notably, the kurtosis values

indicate a departure from the standard normal distribution, suggesting that the residuals do not follow a perfect normal distribution. It is important to note that achieving a residual distribution that is approximately normal is not always the primary goal. Instead, the focus should be on accurately capturing the empirical distribution of the residuals, which can be effectively estimated through simulations [49]. The combined standardized residual offers several advantages over traditional residuals, particularly in its ability to account for the joint modeling of both the mean and dispersion sub-models. Additionally, it reduces computational complexity by avoiding the need for projection matrix calculations. It is also worth noting that the values of the mean, standard error, skewness, and kurtosis for the ordinary residuals and the deviance components are identical in Table A1, which is due to the fact that when the response variable follows a normal distribution and the identity link function is used, these residuals share the same functional form.

Table A1. Empirical means, standard errors, skewness, and kurtosis of standardized ordinary residuals r_p^{β}, deviance $d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$, and combined $r_p^{\beta\gamma}$ for the model: $Y_i \sim \mathrm{N}(\mu_i;\phi_1^{-1})$ with $\mu_i = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_1 + z_i^{\gamma_2}$, $\beta_1 = -1.4$, $\beta_2 = 0.8$, $\gamma_1 = 0.9$, $\gamma_2 = -1$, x_i are values from $X \sim \mathrm{U}(0.1, 1.1)$, z_i are values from $Z \sim \mathrm{U}(0.4, 1.4)$, and $i \in \{1, \ldots, 30\}$.

	Mean			**Standard Error**			**Skewness**			**Kurtosis**		
i	r_p^{β}	$d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i;\widehat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0529	0.0529	0.0501	1.0362	1.0362	0.9573	0.0019	0.0019	−0.9437	−0.2490	−0.2490	0.8401
2	−0.1111	−0.1111	−0.0974	1.1502	1.1502	1.0590	−0.0058	−0.0058	−0.8852	−0.3649	−0.3649	0.5246
3	0.0635	0.0635	0.0592	1.0394	1.0394	0.9667	−0.0389	−0.0389	−0.9625	−0.2414	−0.2414	0.8919
4	0.0313	0.0313	0.0254	1.0583	1.0583	1.0049	−0.0578	−0.0578	−1.0292	−0.3378	−0.3378	1.0147
5	0.0013	0.0013	0.0005	1.0915	1.0915	1.0096	−0.0015	−0.0015	−0.7594	−0.3119	−0.3119	0.4026
6	0.0214	0.0214	0.0165	1.0540	1.0540	0.9989	−0.0304	−0.0304	−0.9679	−0.3951	−0.3951	0.7662
7	−0.0923	−0.0923	−0.0863	1.1760	1.1760	1.0462	0.0809	0.0809	−0.5716	−0.4337	−0.4337	−0.1301
8	−0.0854	−0.0854	−0.0751	1.0852	1.0852	1.0171	0.0140	0.0140	−0.8861	−0.2477	−0.2477	0.7573
9	−0.0742	−0.0742	−0.0497	0.9771	0.9771	0.8851	0.0639	0.0639	−0.5967	−0.3986	−0.3986	0.0725
10	−0.0161	−0.0161	−0.0163	1.0562	1.0562	1.0027	0.0037	0.0037	−0.9609	−0.0796	−0.0796	1.1221
11	0.0200	0.0200	0.0195	1.0335	1.0335	0.9795	0.0081	0.0081	−0.9239	−0.1361	−0.1361	1.0070
12	0.0029	0.0029	0.0076	1.0231	1.0231	0.9762	−0.0230	−0.0230	−0.9498	−0.4250	−0.4250	0.7264
13	−0.0619	−0.0619	−0.0539	1.0405	1.0405	1.0025	0.0003	0.0003	−0.9311	−0.1951	−0.1951	0.9274
14	0.0795	0.0795	0.0664	1.0798	1.0798	1.0051	−0.0481	−0.0481	−0.9750	−0.1878	−0.1878	0.9593
15	0.0031	0.0031	0.0062	1.0848	1.0848	0.9493	0.0045	0.0045	−0.5384	−0.6729	−0.6729	−0.3277
16	−0.0058	−0.0058	−0.0100	1.1008	1.1008	1.0191	0.0464	0.0464	−0.7126	−0.3134	−0.3134	0.2808
17	−0.0310	−0.0310	−0.0266	1.1289	1.1289	0.9966	−0.0003	−0.0003	−0.6177	−0.5437	−0.5437	−0.0570
18	0.0017	0.0017	0.0018	1.0455	1.0455	0.9959	−0.0378	−0.0378	−0.9825	−0.3026	−0.3026	0.9239
19	0.0218	0.0218	0.0200	1.0805	1.0805	0.9983	0.0028	0.0028	−0.7909	−0.2823	−0.2823	0.4899
20	−0.0164	−0.0164	−0.0070	1.0136	1.0136	0.9509	0.0280	0.0280	−0.9129	−0.3559	−0.3559	0.6219
21	−0.1167	−0.1167	−0.0963	1.2312	1.2312	1.0411	−0.0140	−0.0140	−0.9626	−0.1495	−0.1495	0.9737
22	0.0592	0.0592	0.0540	1.0718	1.0718	0.9827	−0.0275	−0.0275	−0.6926	−0.4931	−0.4931	0.1835
23	0.0689	0.0689	0.0649	1.0327	1.0327	0.9681	−0.0809	−0.0809	−1.0778	−0.3255	−0.3255	1.1394
24	0.0498	0.0498	0.0486	1.0264	1.0264	0.9447	−0.0106	−0.0106	−0.9274	−0.4568	−0.4568	0.6166
25	0.0531	0.0531	0.0473	1.0527	1.0527	0.9813	−0.0466	−0.0466	−1.0164	−0.2997	−0.2997	1.0107
26	0.0298	0.0298	0.0283	1.0374	1.0374	0.9528	0.0317	0.0317	−0.8929	−0.4148	−0.4148	0.5735
27	0.0651	0.0651	0.0570	1.1328	1.1328	0.9910	−0.0783	−0.0783	−0.5823	−0.7951	−0.7951	−0.3369
28	−0.0468	−0.0468	−0.0326	1.0019	1.0019	0.9511	0.0320	0.0320	−0.8650	−0.1997	−0.1997	0.7188
29	0.0811	0.0811	0.0726	1.0818	1.0818	0.9745	0.0105	0.0105	−0.7302	−0.3677	−0.3677	0.2215
30	−0.0857	−0.0857	−0.0780	1.1094	1.1094	1.0294	0.0496	0.0496	−0.8707	−0.3277	−0.3277	0.5302

Next, we conducted another simulation where the response variable follows a gamma distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, denoted as $y_i \sim \mathrm{Gamma}(\mu_i;\phi_i)$. This distribution is particularly useful for modeling scenarios where the response variable consists of positive real numbers ($\mathbb{R}^+$). In this simulation, we consider a sample of size $n = 30$, where $y_1, \ldots, y_{30}$ are independent random variables, each following a gamma distribution with mean μ_i and dispersion parameter ϕ_i. The gamma distribution for each

observation is expressed as $Y_i \sim \text{Gamma}(\mu_i; \phi_i)$, for $i \in \{1, \ldots, 30\}$. In this scenario, the mean of the response variable and the dispersion parameter are assumed to satisfy the functional relationships stated as

$$\mu_i^{-1} = \beta_1 + x_i^{\beta_2}, \quad \log(\phi_i) = \gamma_i + z_i^{\gamma_2}, \quad i \in \{1, \ldots, n\}, \tag{A8}$$

where the parameters are set as $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, and $\gamma_2 = 1.5$. The values x_i and z_i of covariates are generated independently from uniform distributions, specifically, from $X \sim \text{U}(0.1, 1.1)$ and $Z \sim \text{U}(0.4, 1.4)$, respectively.

Table A2. Empirical means, standard errors, skewness, and kurtosis of standardized ordinary residuals r_p^{β}, deviance $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$, and combined residuals $r_p^{\beta\gamma}$ for the model: $Y_i \sim \text{Gamma}(\mu_i; \phi_i)$ with $\mu_i^{-1} = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = 1$, $\gamma_2 = 1.5$, x_i are values from $X \sim \text{U}(0.1; 1.1)$, z_i are values from $Z \sim \text{U}(0.4; 1.4)$, and $i \in \{1, \ldots, 30\}$.

	Mean			**Standard Error**			**Skewness**			**Kurtosis**		
i	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$
1	−0.0681	−0.1901	−0.0633	1.0069	1.0473	1.0426	0.6518	0.0229	−1.6062	0.3642	−0.2360	2.9941
2	0.1376	0.0352	0.1295	1.2066	1.1792	0.9105	0.3711	−0.0892	−0.6799	−0.1334	−0.3099	0.2738
3	−0.0911	−0.2193	−0.0826	0.9845	1.0308	1.0341	0.6961	0.0606	−1.4265	0.3802	−0.2877	2.4229
4	−0.0654	−0.1620	−0.0551	1.0403	1.0691	1.0100	0.4759	0.0433	−1.2266	−0.1053	−0.4049	1.3695
5	0.0408	−0.1035	0.0218	1.0884	1.0788	0.8775	0.7183	0.0388	−1.8128	0.3847	−0.3716	3.5617
6	−0.0752	−0.1766	−0.0710	1.0448	1.0744	1.0416	0.5371	0.0464	−1.2865	0.1252	−0.3143	1.6389
7	0.0568	−0.0897	0.0466	1.1075	1.0969	0.9690	0.6480	−0.0024	−0.9617	0.1305	−0.4083	0.8442
8	0.0802	−0.0409	0.0756	1.1169	1.0936	0.9398	0.6013	−0.0229	−0.8558	0.3787	−0.2179	0.7340
9	0.1150	−0.0191	0.1128	1.0787	1.0251	0.6807	0.6784	0.0249	−2.0631	0.3947	−0.3186	5.3635
10	0.0056	−0.1215	0.0204	1.0252	1.0285	0.8790	0.6721	0.0293	−1.9264	0.3551	−0.2699	4.7902
11	−0.0916	−0.2250	−0.0857	0.9613	1.0251	1.0574	0.6735	−0.0072	−1.2530	0.3850	−0.2520	1.6858
12	−0.0650	−0.1516	−0.0471	1.0229	1.0411	0.9241	0.4629	0.0932	−0.5624	−0.2520	−0.5480	−0.1751
13	0.0638	−0.0612	0.0579	1.0820	1.0748	0.9571	0.5854	−0.0347	−0.9877	0.1551	−0.2563	1.1088
14	−0.0942	−0.2283	−0.0873	0.9743	1.0290	1.0520	0.7399	0.0479	−1.3609	0.5323	−0.2109	2.1490
15	0.0059	−0.1462	−0.0058	1.0640	1.0737	0.9096	0.7334	0.0395	−1.7826	0.3235	−0.4162	3.4976
16	−0.0241	−0.1645	−0.0194	1.0013	1.0359	0.9927	0.6848	−0.0055	−1.1316	0.3598	−0.3201	1.2526
17	0.0183	−0.1264	0.0210	1.0286	1.0448	0.9516	0.6639	−0.0099	−1.0117	0.1783	−0.3976	1.0217
18	0.0409	−0.0710	0.0315	1.0911	1.0818	0.9075	0.5661	0.0108	−1.9305	0.1144	−0.2480	5.1510
19	0.0227	−0.1132	0.0254	1.0315	1.0427	0.9057	0.6252	−0.0231	−1.8073	0.2184	−0.3521	3.6561
20	0.0523	−0.0366	0.0622	1.1244	1.1179	0.8100	0.3460	−0.0199	−1.4416	−0.3247	−0.4876	2.1293
21	0.2360	0.1138	0.2038	1.2859	1.2057	0.8969	0.4666	−0.0862	−0.7408	0.0220	−0.2888	0.4351
22	−0.0534	−0.2005	−0.0506	1.0163	1.0415	0.9697	0.7783	0.0948	−1.3372	0.4787	−0.3966	1.7910
23	−0.1021	−0.1970	−0.0892	1.0148	1.0521	1.0181	0.4988	0.0796	−1.0082	−0.0953	−0.4596	0.7226
24	−0.0968	−0.1919	−0.0987	1.0644	1.1025	1.0552	0.4335	0.0694	−0.8332	−0.3531	−0.6255	0.2011
25	−0.1021	−0.2108	−0.1049	1.0456	1.0892	1.0939	0.5640	0.0527	−1.3735	0.0698	−0.3166	2.0412
26	−0.0757	−0.1747	−0.0750	1.0639	1.1027	1.0617	0.4483	0.0358	−1.1779	−0.2370	−0.4699	1.1893
27	−0.0344	−0.1896	−0.0438	1.0327	1.0671	0.9761	0.7311	0.0517	−1.4981	0.2856	−0.4978	2.2192
28	0.1027	−0.0232	0.0702	1.1020	1.0902	0.8793	0.5139	−0.1023	−2.1642	0.0722	−0.2594	5.7557
29	−0.0879	−0.2298	−0.0737	0.9649	1.0198	1.0101	0.7244	0.0576	−1.2519	0.3863	−0.3727	1.4555
30	0.0895	−0.0173	0.0767	1.2022	1.1978	0.9942	0.3846	−0.0833	−0.7137	−0.1861	−0.3712	0.2669

As observed in Table A2, the mean of the residuals is close to zero across all three types of residuals analyzed, with the deviance residuals showing the largest deviation from zero. The standard errors are approximately one for all residuals, but notable differences appear in the skewness and kurtosis values. Specifically, the empirical distribution of the deviance residuals exhibits a roughly symmetric shape but with negative kurtosis, indicating a deviation from the standard normal distribution.

Following this, we considered an alternative model for positive real-valued data, where the response variable follows an IG with mean μ_i and dispersion parameter ϕ_i^{-1}, denoted as $Y_i \sim \text{IG}(\mu_i; \phi_i)$. In this scenario, $Y_1, \ldots, Y_{30}$ are independent random variables, each following an IG distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, for $i \in \{1, \ldots, 30\}$. In this case, the mean of the response variable and the dispersion parameter are assumed to follow the functional relationships stated as

$$\log(\mu_i) = \beta_1 + x_i^{\beta_2}, \quad \log(\phi_i) = \gamma_i + z_i^{\gamma_2}, i \in \{1, \ldots, n\}, \tag{A9}$$

with parameter values set as $\beta_1 = -0.45$, $\beta_2 = 0.9$, $\gamma_1 = -0.2$, and $\gamma_2 = -0.4$.

Table A3. Empirical means, standard errors, skewness, and kurtosis of standardized ordinary residuals r_p^{β}, deviance residuals $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$, and combined residuals $r_p^{\beta\gamma}$ for the model: $Y_i \sim \text{IG}(\mu_i; \phi_1)$ with $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = -0.45$, $\beta_2 = 0.9$, $\gamma_1 = -0.2$, $\gamma_2 = -0.4$, x_i are values from $X \sim \text{U}(0.1, 1.1)$, z_i are values from $Z \sim \text{U}(0.4, 1.4)$, for $i \in \{1, \ldots, 30\}$.

	Mean			**Standard Error**			**Skewness**			**Kurtosis**		
i	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.1093	−0.3234	0.1032	1.0611	1.4007	1.0224	1.6000	−0.0137	0.4466	3.3368	−0.3908	0.9456
2	−0.1667	−0.6437	−0.1358	0.8971	1.4081	0.9823	1.2245	−0.0270	−0.8427	1.8915	−0.3926	0.9439
3	0.0996	−0.3231	0.0841	1.0926	1.4206	1.0276	1.7995	0.0397	0.2460	4.7126	−0.2011	1.0853
4	0.1128	−0.3366	0.1004	1.0890	1.4225	1.0355	1.6730	0.0360	0.4411	3.6695	−0.3967	0.8769
5	−0.0101	−0.4639	−0.0013	0.9499	1.3508	0.9471	1.5946	−0.0067	0.7523	3.4482	−0.3329	1.3493
6	0.1084	−0.3394	0.1006	1.1031	1.4101	1.0332	1.7601	0.0940	0.5602	4.1436	−0.3957	0.9542
7	−0.1175	−0.5480	−0.1074	0.9410	1.4120	0.9854	1.2048	0.0426	−0.5765	1.7194	−0.4286	0.5012
8	−0.1185	−0.5685	−0.1025	0.9185	1.3916	0.9855	1.3471	0.0120	−0.8529	2.3364	−0.3026	1.4265
9	−0.1864	−0.7934	−0.1752	0.7176	1.2638	0.7471	1.5659	0.0023	0.8128	3.3358	−0.3195	1.3763
10	0.0238	−0.4481	0.0290	0.9855	1.3582	0.9735	1.7326	0.0008	0.8750	3.8052	−0.2624	1.6053
11	0.0674	−0.3200	0.0620	1.0323	1.4047	0.9971	1.5250	−0.0360	−0.2924	3.4876	−0.2037	0.8526
12	−0.0076	−0.4208	0.0030	0.9706	1.3722	0.9581	1.3744	0.0570	−0.4015	2.4670	−0.4719	0.3383
13	−0.0959	−0.5233	−0.0749	0.9079	1.3590	0.9713	1.3549	−0.0174	−0.7717	2.3474	−0.2642	1.4153
14	0.0506	−0.3670	0.0347	1.0512	1.4193	1.0163	1.6107	0.0125	−0.0660	3.5881	−0.2619	0.7162
15	0.0479	−0.3829	0.0484	1.0291	1.4034	0.9840	1.4119	0.1017	0.7719	2.2877	−0.5801	0.8494
16	−0.0155	−0.4031	−0.0071	0.9877	1.3955	0.9814	1.3742	0.0212	−0.4087	2.6827	−0.2890	0.6487
17	−0.0921	−0.5093	−0.0839	0.9608	1.4159	0.9756	1.2148	0.0758	−0.4736	1.7728	−0.4597	0.2306
18	0.0230	−0.4610	0.0284	0.9835	1.3532	0.9680	1.8352	0.0311	0.9112	4.7223	−0.3412	1.9152
19	0.0684	−0.3762	0.0675	1.0502	1.3978	1.0221	1.7285	−0.0117	0.8049	4.2068	−0.2745	1.7230
20	−0.0324	−0.5466	−0.0228	0.9283	1.3268	0.9174	1.8440	0.0693	1.0149	4.7498	−0.3704	2.0863
21	−0.1724	−0.6437	−0.1068	0.9212	1.4511	0.9582	1.1546	−0.0950	−0.9650	1.7367	−0.2621	1.3915
22	0.0628	−0.3381	0.0516	1.0718	1.4345	1.0157	1.4677	0.1032	0.2225	2.8493	−0.4636	0.3202
23	0.1068	−0.3255	0.0980	1.0955	1.4015	1.0130	1.7240	0.1198	0.3676	3.8983	−0.3703	0.9237
24	0.0939	−0.3418	0.0764	1.0765	1.4313	1.0281	1.5342	0.0510	0.1184	3.0591	−0.4712	0.4936
25	0.0974	−0.3486	0.0944	1.0784	1.3995	1.0182	1.7553	0.0799	0.5558	4.4933	−0.3461	1.1892
26	0.1018	−0.3511	0.0896	1.1014	1.4122	1.0335	1.7790	0.0991	0.5271	4.6299	−0.4332	1.0196
27	0.1016	−0.2805	0.0958	1.0744	1.4277	1.0023	1.2800	0.1152	0.4578	1.8635	−0.6681	0.1886
28	−0.1422	−0.7401	−0.1180	0.8133	1.3977	0.8243	1.8185	−0.6816	1.0622	5.1946	3.3745	2.6832
29	0.1012	−0.2774	0.0906	1.0600	1.4336	1.0148	1.3300	−0.0321	−0.0864	2.1824	−0.3847	0.3986
30	−0.1209	−0.5643	−0.0835	0.9070	1.3702	0.9568	1.3087	0.0295	−0.7676	2.1934	−0.3329	1.1623

Table A3 illustrates that the residuals do not follow an approximately $\text{N}(0, 1)$ distribution. This is evident from the empirical means and variances of the deviance residuals, as well as the skewness and kurtosis of the other types of residuals. Among them, the deviance residuals show the most important deviation, with their mean and standard error differing markedly from those of a standard normal distribution.

To analyze count data, the negative binomial distribution is often employed as an alternative to the Poisson distribution in situations of overdispersion. In this scenario, we assume that the response variable follows a negative binomial distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, denoted as $Y_i \sim \text{NB}(\mu_i, \phi_i)$. Here, $y_1, \ldots, y_{30}$ are independent random variables, each following a negative binomial distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, for $i \in \{1, \ldots, 30\}$.

In this context, the mean of the response variable and the dispersion parameter are assumed to follow the functional relationships provided in (A9), with parameters set as $\beta_1 = 1.6$, $\beta_2 = 0.8$, $\gamma_1 = -0.3$, and $\gamma_2 = 0.5$.

Table A4. Empirical means, standard errors, skewness, and kurtosis of standardized ordinary residuals r_p^{β}, deviance residuals $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$, and combined residuals $r_p^{\beta\gamma}$ for the model: $Y_i \sim \text{NB}(\mu_i; \phi_1)$ with $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1.6$, $\beta_2 = 0.8$, $\gamma_1 = -0.3$, $\gamma_2 = 0.5$, x_i are values from $X \sim \text{U}(0.1; 1.1)$, z_i are values from $Z \sim \text{U}(0.4; 1.4)$, and $i \in \{1, \ldots, 30\}$.

	Mean			**Standard Error**			**Skewness**			**Kurtosis**		
i	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0800	−0.1576	0.0784	1.0543	1.0440	1.0303	1.1830	−0.0145	1.0567	1.6297	−0.2099	1.2593
2	−0.1285	−0.3632	−0.1148	0.9790	1.0834	0.9020	0.8964	−0.1011	0.7846	0.6302	−0.3266	0.3963
3	0.0521	−0.1940	0.0507	1.0299	1.0495	1.0129	1.1202	−0.0737	0.9810	1.4170	−0.2828	1.0352
4	0.0664	−0.1639	0.0637	1.0782	1.0744	1.0498	1.0030	0.0108	0.9080	0.8986	−0.4082	0.6631
5	−0.0020	−0.2557	−0.0025	1.0275	1.0544	0.9994	1.1614	0.0250	1.0509	1.3966	−0.3572	1.0897
6	0.0681	−0.1624	0.0659	1.0745	1.0688	1.0474	1.0042	0.0051	0.9082	0.8893	−0.3812	0.6530
7	−0.0718	−0.3220	−0.0684	0.9586	1.0343	0.9302	1.0829	−0.0360	0.9278	1.2114	−0.3675	0.7840
8	−0.1105	−0.3527	−0.1037	0.9465	1.0401	0.9102	1.0490	−0.0747	0.8936	1.0519	−0.3181	0.6783
9	−0.1257	−0.3551	−0.1152	0.8751	0.9842	0.8423	0.9571	−0.1475	0.8891	0.9414	−0.3212	0.8004
10	0.0243	−0.2173	0.0238	1.0112	1.0400	0.9847	1.0850	−0.0388	0.9744	1.1698	−0.3251	0.8798
11	−0.0044	−0.2560	−0.0046	1.0181	1.0488	1.0032	1.2489	0.0140	1.0970	1.7721	−0.2085	1.3381
12	−0.0398	−0.2538	−0.0381	0.9745	1.0296	0.9511	0.8840	−0.0380	0.7806	0.5482	−0.4332	0.3003
13	−0.0670	−0.3078	−0.0641	0.9572	1.0285	0.9328	1.1419	−0.0571	0.9733	1.4541	−0.2718	0.9966
14	0.0434	−0.2018	0.0431	1.0196	1.0434	1.0044	1.1190	−0.0618	0.9772	1.3927	−0.2649	1.0077
15	0.0335	−0.2275	0.0310	1.0548	1.0662	1.0233	1.1033	0.0519	1.0104	1.1712	−0.5602	0.9200
16	−0.0280	−0.2797	−0.0269	0.9978	1.0400	0.9800	1.2202	0.0170	1.0566	1.8418	−0.3072	1.3151
17	−0.0786	−0.3355	−0.0763	0.9710	1.0383	0.9467	1.1916	0.0270	1.0297	1.5911	−0.3855	1.1109
18	0.0472	−0.1847	0.0453	1.0313	1.0588	1.0002	0.9745	−0.1008	0.8750	0.8118	−0.2650	0.5907
19	0.0579	−0.1931	0.0562	1.0403	1.0593	1.0149	1.0319	−0.0463	0.9250	0.8743	−0.3993	0.6241
20	−0.0049	−0.2136	−0.0073	1.0074	1.0394	0.9587	0.8862	0.0134	0.8140	0.5088	−0.5098	0.3754
21	−0.1951	−0.4552	−0.1689	0.9736	1.1166	0.8616	0.9830	−0.1202	0.8286	1.0077	−0.3096	0.6065
22	0.0570	−0.1967	0.0560	1.0409	1.0551	1.0235	1.0913	−0.0007	0.9649	1.1634	−0.4341	0.8314
23	0.0643	−0.1635	0.0620	1.0729	1.0719	1.0470	0.9652	−0.0001	0.8688	0.8063	−0.4161	0.5635
24	0.0578	−0.1561	0.0571	1.0357	1.0446	1.0123	0.9089	−0.0156	0.8136	0.6062	−0.4052	0.3901
25	0.0559	−0.1771	0.0540	1.0592	1.0703	1.0342	1.0613	−0.0404	0.9528	1.2137	−0.2377	0.9341
26	0.0624	−0.1613	0.0604	1.0590	1.0676	1.0318	0.9458	−0.0259	0.8515	0.7540	−0.3824	0.5307
27	0.0724	−0.1926	0.0694	1.0702	1.0731	1.0468	1.0881	0.0216	0.9785	1.0999	−0.5273	0.8026
28	−0.0610	−0.2829	−0.0571	0.9281	1.0002	0.8965	0.9891	−0.1037	0.8868	0.8720	−0.2565	0.6463
29	0.0468	−0.2085	0.0457	1.0385	1.0554	1.0217	1.1541	−0.0216	1.0146	1.3559	−0.2962	0.9869
30	−0.0928	−0.3137	−0.0850	0.9592	1.0384	0.9066	1.0092	−0.0420	0.8788	1.0390	−0.2298	0.7176

Table A4 demonstrates the behavior when the response variable of the model follows a negative binomial distribution. The results are consistent with those observed in cases where the distributions are continuous, displaying nonzero skewness for both ordinary and standardized combined residuals, and nonzero kurtosis for deviance residuals.

Lastly, we present another simulation where the response variable follows a double Poisson distribution, which serves as an alternative to the Poisson model in situations where the data exhibit overdispersion. In this case, the response variable is assumed to follow a double Poisson (DP) distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, denoted as $Y_i \sim \text{DP}(\mu_i, \phi_i)$. Here, $y_1, \ldots, y_{30}$ are independent random variables, each following a

double Poisson distribution with mean μ_i and dispersion parameter ϕ_i^{-1}, for $i \in \{1, \ldots, 30\}$. In this simulation, the mean of the response variable and the dispersion parameter are assumed to adhere to the functional relationships outlined in (A9), with the parameters set to $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = -0.4$, and $\gamma_2 = -0.2$.

Table A5. Empirical means, standard errors, skewness, and kurtosis of the standardized ordinary residuals r_p^{β}, deviance $d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$, and combined residuals $r_p^{\beta\gamma}$ for the model: $Y_i \sim \mathrm{DP}(\mu_i; \phi_i)$ with $\log(\mu_i) = \beta_1 + x_i^{\beta_2}$ and $\log(\phi_i) = \gamma_i + z_i^{\gamma_2}$, $\beta_1 = 1$, $\beta_2 = 0.8$, $\gamma_1 = -0.4$, $\gamma_2 = -0.2$, x_i are values from $X \sim \mathrm{U}(0.1; 1.1)$, z_i are values from $Z \sim \mathrm{U}(0.4; 1.4)$, and $i \in \{1, \ldots, 30\}$.

	Mean			**Standard Error**			**Skewness**			**Kurtosis**		
i	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$	r_p^{β}	$d_\mu(\widehat{\mu}_i; \widehat{\phi}_i)$	$r_p^{\beta\gamma}$
1	0.0233	-0.1658	1.4118	3.3167	3.5686	2.6543	0.5352	−0.1450	0.8083	0.8123	0.4426	1.0319
2	−0.1037	−0.3751	0.8548	3.8579	4.3109	2.3268	0.8648	−0.0387	1.2593	1.7674	0.6244	2.1948
3	0.0615	−0.1496	1.3403	3.5216	3.7657	2.7926	0.6165	−0.0839	0.8689	0.8094	0.3001	1.0851
4	0.0463	−0.1527	1.4415	3.9185	4.1153	3.0228	0.9254	−0.0206	1.2355	3.3578	1.8486	3.0421
5	−0.0016	−0.2054	1.6672	4.3049	4.6379	3.3868	0.5709	−0.1810	0.9803	1.5157	1.0457	1.8564
6	0.0635	−0.1361	1.4354	3.7256	3.9687	2.9319	0.6426	−0.1530	1.0031	1.4686	0.9935	1.6238
7	−0.0933	−0.3979	0.9768	4.9688	5.2090	3.0480	2.1098	0.3302	1.7778	21.2966	3.3476	8.8820
8	−0.0647	−0.3235	0.8992	3.4578	3.7909	2.3544	0.8915	0.0818	1.1611	1.6709	0.1812	2.1290
9	−0.3374	−0.5158	1.6586	4.1449	4.7605	2.7663	0.1114	−0.8286	1.0090	3.9539	4.7466	2.9466
10	−0.0323	−0.2191	1.5739	3.6045	3.8846	2.9438	0.5688	−0.1215	0.8459	1.0055	0.6263	1.2320
11	0.0069	−0.2295	1.1594	3.6421	3.9128	2.7782	0.8287	0.0206	1.0989	1.6480	0.4310	2.0542
12	−0.0201	−0.2809	1.0686	4.7487	5.1345	2.8904	1.3168	−0.1040	1.5475	7.6618	4.5142	4.3919
13	−0.0649	−0.3008	0.9345	3.1557	3.4906	2.2213	0.7113	−0.0139	0.9614	0.7537	0.0196	1.0298
14	0.0895	−0.1376	1.2945	3.7493	3.9578	2.9200	0.7406	−0.0080	0.9939	1.3571	0.3503	1.6857
15	0.0194	−0.2279	1.7977	6.6654	7.0479	4.5443	1.2710	−0.0740	1.5437	10.9736	5.2700	5.3047
16	−0.0592	−0.3287	1.0744	4.2024	4.5966	2.9267	0.9173	0.0138	1.2347	2.3228	0.6937	2.4515
17	0.0102	−0.2943	1.1006	5.2006	5.5565	3.3215	1.1066	0.0598	1.4255	3.5479	1.2652	3.3694
18	0.0069	−0.1714	1.5834	3.4867	3.7450	2.8046	0.4870	−0.1907	0.7964	0.8338	0.6903	0.9775
19	−0.0149	−0.2235	1.5777	4.2101	4.4967	3.3186	0.7519	−0.0853	1.1014	2.1757	1.1462	2.4559
20	−0.0655	−0.2604	1.6440	5.5428	5.7653	3.4208	2.0548	0.0715	1.7183	23.1532	12.7579	7.2832
21	−0.1503	−0.4306	0.7868	3.7096	4.2006	2.1232	0.8248	0.0206	1.1638	1.3389	0.1311	1.9946
22	0.0535	−0.2025	1.4024	5.0639	5.3102	3.6818	1.1248	0.0734	1.3769	4.7469	1.8224	3.6726
23	0.0967	−0.1154	1.4131	4.2417	4.3887	3.1568	1.0820	0.0249	1.3385	4.8203	2.4179	3.5112
24	0.1074	−0.1237	1.3676	5.3277	5.1619	3.4436	2.7325	0.6202	1.9467	27.4171	9.0764	9.4828
25	0.0627	−0.1290	1.4403	3.5540	3.7567	2.8391	0.7231	−0.0904	1.0249	2.0506	1.0417	2.0577
26	0.0864	−0.1182	1.4713	4.0637	4.2597	3.1574	0.8106	−0.0747	1.1975	2.3497	1.4990	2.5252
27	0.1782	−0.1030	1.7318	7.1841	7.4618	4.8472	1.1150	−0.0332	1.4564	5.2160	3.0318	3.6466
28	−0.1599	−0.3145	1.5901	3.0414	3.3301	2.4662	0.4244	−0.1989	0.7015	0.7759	0.6846	0.7803
29	0.0484	−0.2111	1.3164	4.6417	4.9355	3.4256	0.9671	0.0090	1.2822	2.9733	1.1306	3.0575
30	−0.0717	−0.3348	0.8960	3.8000	4.1193	2.4258	1.0736	0.0992	1.3362	3.3738	0.8143	3.2316

Table A5 illustrates the behavior of the model when the response variable follows a double Poisson distribution, revealing high deviations from the standard normal distribution across all types of residuals analyzed. A critique of the double Poisson model, where the variable of interest follows this distribution, is provided in [53]; for further discussion of this, see [54]. These sources discuss the approximation nature of the mean and variance in the double Poisson distribution. Moreover, challenges associated with the estimator for the normalization constant, as proposed in [55], are highlighted. These challenges can lead to nonconvergence in the optimization methods used for parameter estimation in this model. Due to these limitations, it is advisable to consider alternative models when analyzing count data that exhibit overdispersion.

References

1. Paula, G.A. *Regression Models with Computational Support*; IME-USP: São Paulo, Brazil, 2024. Available online: https://www.ime.usp.br/~giapaula/texto_2024.pdf (accessed on 27 August 2024). (In Portuguese)
2. Draper, N.R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, NY, USA , 1998.
3. Belsley, D.A.; Kuh, E.; Welsch, R.E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; Wiley: New York, NY, USA, 2005.

4. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; Wiley: New York, NY, USA, 2012.
5. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Chapman and Hall: London, UK, 1989.
6. Pinho, L.G.B.; Nobre, J.S.; de Freitas, S.M. On linear mixed models and their influence diagnostics applied to an actuarial problem. *Chil. J. Stat.* **2012**, *3*, 57–73.
7. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [CrossRef]
8. McCulloch, C.E.; Searle, S.R. *Generalized, Linear, and Mixed Models*; Wiley: New York, NY, USA, 2000.
9. Dobson, A.J.; Barnett, A.G. *An Introduction to Generalized Linear Models*; CRC Press: Boca Raton, MA, USA, 2018.
10. Akdur, H.T.K.; Kilic, D.; Bayrak, H. Residual diagnostic methods for bell-type count models. *Stat. Biosci.* 2024, *in press*. [CrossRef]
11. Araripe, P.P.; de Lara, I.A.R. Ordinal data and residual analysis: Review and application. *Braz. J. Biom.* **2023**, *1*, 287–310. [CrossRef]
12. Hussain, Z.; Akbar, A. Diagnostics through residual plots in binomial regression addressing chemical species data. *Math. Probl. Eng.* **2022**, *2022*, 4375945. [CrossRef]
13. Rivera, P.A.; Gallardo, D.I.; Venegas, O.; Gómez-Déniz, E.; Gómez, H.W. Reparameterized scale mixture of Rayleigh distribution regression models with varying precision. *Mathematics* **2024**, *12*, 1982. [CrossRef]
14. Imran, M.; Akbar, A. Diagnostics via partial residual plots in inverse Gaussian regression. *J. Chemom.* **2020**, *34*, e3235. [CrossRef]
15. Heit, D.R.; Ortiz-Calo, W.; Poisson, M.K.P. Generalized nonlinearity in animal ecology: Research, review, and recommendations. *Ecol. Evol.* **2024**, *14*, e11387. [CrossRef]
16. Barros, M.; Paula, G.A.; Leiva, V. An R implementation for generalized Birnbaum-Saunders distributions. *Comput. Stat. Data. Anal.* **2009**, *53*, 1511–1528. [CrossRef]
17. Cavalcante, T.; Ospina, R.; Leiva, V.; Martin-Barreiro, C.; Cabezas, X. Weibull regression and machine learning survival models: Methodology, comparison, and application to biomedical data related to cardiac surgery. *Biology* **2023**, *11*, 1394. [CrossRef]
18. Sanchez, L.; Leiva, V.; Galea, M.; Saulo, H. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Appl. Stoch. Model. Bus. Ind.* **2021**, *37*, 53–73. [CrossRef]
19. Saulo, H.; Dasilva, A.; Leiva, V.; Sanchez, L.; de la Fuente, H. Log-symmetric quantile regression models. *Stat. Neerl.* **2022**, *76*, 124–163. [CrossRef]
20. Leiva, V.; Ferreira, M.; Gomes, M.I.; Lillo, C. Extreme value Birnbaum-Saunders regression models applied to environmental data. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1045–1058. [CrossRef]
21. Marchant, C.; Leiva, V; Cavieres, M.F.; Sanhueza, A. Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Rev. Environ. Contam. Toxicol.* **2013**, *223*, 1–31.
22. Azevedo, C.; Leiva, V.; Athayde, E.; Balakrishnan, N. Shape and change point analyses of the Birnbaum-Saunders-t hazard rate and associated estimation. *Comput. Stat. Data. Anal.* **2012**, *56*, 3887–3897. [CrossRef]
23. Palacios, C.A.; Reyes-Suarez, J.A.; Bearzotti, L.A.; Leiva, V.; Marchant, C. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy* **2021**, *23*, 485. [CrossRef]
24. Chatterjee, A.; Gupta, S.; Lahiri, S.N. On the residual empirical process based on the ALASSO in high dimensions and its functional oracle property. *J. Econom.* **2015**, *186*, 317–324. [CrossRef]
25. Espinheira, P.L.; de Oliveira Silva, A. Residual and influence analysis to a general class of simplex regression. *Test* **2020**, *29*, 523–552. [CrossRef]
26. Liu, J. Feature screening and variable selection for partially linear models with ultrahigh-dimensional longitudinal data. *Neurocomputing* **2016**, *195*, 202–210. [CrossRef]
27. Mi, G.; Di, Y.; Schafer, D.W. Goodness-of-fit tests and model diagnostics for negative binomial regression of RNA sequencing data. *PLoS ONE* **2015**, *10*, e0119254. [CrossRef]
28. Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V.; Barros, M. Reparameterized Birnbaum-Saunders regression models with varying precision. *Electron. J. Stat.* **2016**, *10*, 2825–2855. [CrossRef]
29. Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A.; Barros, M. Birnbaum-Saunders statistical modelling: A new approach. *Stat. Model.* **2014**, *14*, 21–48. [CrossRef]
30. Espinheira, P.L.; da Silva, L.C.M.; Silva, A.O.; Ospina, R. Model selection criteria on beta regression for machine learning. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 26. [CrossRef]
31. Santana, T.V.F.; Ortega, E.M.M.; Cordeiro, G.M. Generalized beta Weibull linear model: Estimation, diagnostic tools and residual analysis. *J. Stat. Theory Pract.* **2019**, *13*, 531–546. [CrossRef]
32. Warton, D.I.; Lyons, M.; Stoklosa, J.; Ives, A.R. Three points to consider when choosing a LM or GLM test for count data. *Methods Ecol. Evol.* **2016**, *7*, 882–890. [CrossRef]
33. Bar-Lev, S.K. The exponential dispersion model generated by the Landau distribution—S comprehensive review and further developments. *Mathematics* **2023**, *11*, 4343. [CrossRef]
34. Martinez-Flórez, G.; Tovar-Falón, R.; Leiva, V.; Castro, C. Skew-normal inflated models: Mathematical characterization and applications to medical data with excess of zeros and ones. *Mathematics* **2024**, *12*, 2486. [CrossRef]
35. Carcamo, E.; Marchant, C.; Ibacache-Pulgar, G.; Leiva, V. Birnbaum-Saunders semi-parametric additive modeling: Estimation, smoothing, diagnostics, and application. *REVSTAT-Stat. J.* **2024**, *22*, 211–237.
36. Landgraf, A.J.; Lee, Y. Generalized principal component analysis: Projection of saturated model parameters. *Technometrics* **2020**, *62*, 507–518. [CrossRef]

37. Svetliza, C.F.; Paula, G.A. Diagnostics in nonlinear negative binomial models. *Commun. Stat. Theory Methods* **2003**, *32*, 1227–1250. [CrossRef]
38. Smyth, G.K. Generalized linear models with varying dispersion. *J. R. Stat. Soc. B* **1989**, *51*, 47–60. [CrossRef]
39. Kidzinski, L.; Hui, F.K.C.; Warton, D.I.; Hastie, T.J. Generalized matrix factorization: Efficient algorithms for fitting generalized linear latent variable models to large data arrays. *J. Mach. Learn. Res.* **2022**, *23*, 1–29.
40. Korkmaz, M.C.; Leiva, V.; Martin-Barreiro, C. The continuous Bernoulli distribution: mathematical characterization, fractile regression, computational simulations, and applications. *Fractal Fract.* **2023**, *7*, 386. [CrossRef]
41. de Oliveira, J.S.C.; Ospina, R.; Leiva, V.; Figueroa-Zúñiga, J.; Castro, C. Quasi-Cauchy regression modeling for fractiles based on data supported in the unit interval. *Fractal Fract.* **2023**, *7*, 667. [CrossRef]
42. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
43. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: New York, NY, USA, 1993.
44. Efron, B. *Exponential Families in Theory and Practice*; Cambridge University Press: Cambridge, UK, 2022.
45. Rice, J.A. *Mathematical Statistics and Data Analysis*; Duxbury Press: Belmont, CA, USA, 1995.
46. Casella, G.; Berger, R.L. *Statistical Inference*; Duxbury Press: Pacific Grove, CA, USA, 2002.
47. Hardin, J.W.; Hilbe, J.W. *Generalized Linear Models and Extensions*; Stata Press: College Station, TX USA, 2012.
48. Lawless, J.F. Negative binomial and mixed Poisson regression. *Can. J. Stat.* **1987**, *15*, 209–225. [CrossRef]
49. Atkinson, A.C. *Plots, Transformations, and Regression an Introduction to Graphical Methods of Diagnostic Regression Analysis*; Oxford University Press: Oxford, UK, 1985.
50. Ridout, M.; Hinde, J.; Demétrio, C.G. A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. *Biometrics* **2001**, *57*, 219–223. [CrossRef]
51. Bélisle, C.J. Convergence theorems for a class of simulated annealing algorithms on Rd. *J. Appl. Probab.* **1992**, *29*, 885–895. [CrossRef]
52. Lehmann, E.; Casella, G. *Theory of Point Estimation*; Springer: New York, NY, USA, 1998.
53. Hilbe, J.M. *Negative Binomial Regression*; Cambridge University Press: Cambridge, UK, 2011.
54. Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*; Cambridge University Press: Cambridge, UK, 2013.
55. Efron, B. Double exponential families and their use in generalized linear regression. *J. Am. Stat. Assoc.* **1986**, *81*, 709–721. [CrossRef]