

Article

Generalized Linear Model (GLM) Applications for the Exponential Dispersion Model Generated by the Landau Distribution

Shaul K. Bar-Lev ¹, Xu Liu ^{2,3,*} , Ad Ridder ⁴ and Ziyu Xiang ²

¹ Faculty of Industrial Engineering and Technology Management, Holon Institute of Technology, Holon 5810201, Israel; shaulb@hit.ac.il

² School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China; xiangziyu@stu.sufe.edu.cn

³ Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China

⁴ School of Business and Economics, Vrije University of Amsterdam, 1081 HV Amsterdam, The Netherlands; ad.ridder@vu.nl

* Correspondence: liu.xu@mail.shufe.edu.cn

Abstract: The exponential dispersion model (EDM) generated by the Landau distribution, denoted by EDM-EVF (exponential variance function), belongs to the Tweedie scale with power infinity. Its density function does not have an explicit form and, as of yet, has not been used for statistical aspects. Out of all EDMs belonging to the Tweedie scale, only two EDMs are steep and supported on the whole real line: the normal EDM with constant variance function and the EDM-EVF. All other absolutely continuous steep EDMs in the Tweedie scale are supported on the positive real line. This paper aims to accomplish an overall picture of all generalized linear model (GLM) applications belonging to the Tweedie scale by including the EDM-EVF. This paper introduces all GLM ingredients needed for its analysis, including the respective link function and total and scaled deviance. We study its analysis of deviance, derive the asymptotic properties of the maximum likelihood estimation (MLE) of the covariate parameters, and obtain the asymptotic distribution of deviance, using saddlepoint approximation. We provide numerical studies, which include estimation algorithm, simulation studies, and applications to three real datasets, and demonstrate that GLM using the EDM-EVF performs better than the linear model based on the normal EDM. An R package accompanies all of these.

Keywords: exponential dispersion model; generalized linear model; exponential variance function; small-dispersion asymptotics; saddlepoint approximation; analysis of the deviance

MSC: 62J12; 62F03; 62F12



Citation: Bar-Lev, S.K.; Liu, X.; Ridder, A.; Xiang, Z. Generalized Linear Model (GLM) Applications for the Exponential Dispersion Model Generated by the Landau Distribution. *Mathematics* **2024**, *12*, 2021. <https://doi.org/10.3390/math12132021>

Academic Editor: Tzong-Ru Tsai

Received: 6 June 2024

Revised: 26 June 2024

Accepted: 27 June 2024

Published: 28 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The (reproductive) TBE_γ , $\gamma \in \{(-\infty, 0), 0, 1, [2, \infty]\}$ class (known as the Tweedie class, cf., [1]) is composed of all exponential dispersion models (EDMs) with variance functions (VFs) of the form

$$V(m) = \phi m^\gamma, \phi > 0, m \in M_\gamma,$$

where m is the mean, M_γ is the mean parameter space, ϕ is the dispersion parameter, and γ is the power parameter (cf., [2,3] and the references cited therein).

Let $F_\gamma \in TBE_\gamma$ be an EDM belonging to the TBE class. Also, let C_γ and M_γ denote, respectively, the convex support and mean parameter space of F_γ . Among the TBE_γ class, the subclasses containing all absolutely continuous (with respect to the Lebesgue measure) models comprise the following cases (cf., [2]):

1. When $\gamma < 0$, F_γ is generated by a stable distribution with a stable index in $(1, 2)$ supported on $C_\gamma = \mathbb{R}$ with $M_\gamma = \mathbb{R}^+$, i.e., $M_\gamma = \mathbb{R}^+ \subset \text{int } C_\gamma = \mathbb{R}$, a proper subset of $\text{int } C_\gamma$ (interior of C_γ), for all $\gamma < 0$.
2. When $\gamma = 0$, F_0 is the normal EDM with $M_0 = \text{int } C_0 = \mathbb{R}$.
3. When $\gamma = 2$, F_2 is the gamma EDM with $M_2 = \text{int } C_2 = \mathbb{R}^+$.
4. When $2 < \gamma < \infty$, F_γ is generated by a positive stable distribution with a stable index in $(0, 1)$ supported on $C_\gamma = \mathbb{R}^+$ with $M_\gamma = \mathbb{R}^+$ for all $\gamma > 2$.
5. When $\gamma = \infty$, F_∞ is the EDM generated by the Landau distribution, supported on $C_\infty = \mathbb{R}$ with $M_\infty = \mathbb{R}$. It is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and is the limit of EDMs having power VFs (see [2,3] for further details).

Two important aspects related to the above TBE models should be remarked at this point:

1. **Complexity of the density function.** Except for the normal ($\gamma = 0$), gamma ($\gamma = 2$) and inverse Gaussian ($\gamma = 3$) EDMs, no other TBE_γ possesses an explicit density, in terms of algebraic functions. All such densities can only be expressed in terms of integral form or power series; hence, their evaluation becomes rather complicated. To resolve the problem, several studies have directly employed the saddlepoint approximation for density estimation on the TBE_γ scale for $2 < \gamma < \infty$, as discussed in [4–8]. The saddlepoint approximation can do so by substituting the part of the density that lacks a closed-form representation with a simple analytic expression. Additionally, the saddlepoint approximation can be utilized instead of traditional likelihood methods to derive the maximum likelihood estimation (MLE) of φ (cf., [6,9]). Dunn created and maintained the Tweedie R package [10], while [11] contributed to and maintained the statmod R package. In this frame, the function `tweedie.profile` in the tweedie R package practically enables the fit of TBE models. These packages can be extended to include the TBE_∞ as well.
2. **Steepness.** The model $F_\gamma \in \text{TBE}_\gamma$ is called *steep* if $M_\gamma = \text{int } C_\gamma$. Steepness is an essential property in two aspects: (1) First, it is related to the existence of the MLE of m . Indeed, if F_γ is steep and $Y_1, \dots, Y_n$ are n i.i.d. random variables drawn from F_γ then the MLE of m , denoted by $\hat{m} = \bar{Y}_n$ (sample average), exists with probability one, and is given by the gradient of the log-likelihood (cf., [12], Theorem 9.29). (2) Second, steepness is a necessary condition for applying generalized linear models (GLMs) methodology to EDMs (cf., [2,6,13]). Consequently, out of the absolutely continuous TBE_γ models described above, only those with $\gamma \in \{0, [2, \infty), \infty\}$ are steep, as their mean parameter space M_γ equals the interior of their convex support (i.e., $M_\gamma = \mathbb{R}$ for $\gamma = 0, \infty$ and $M_\gamma = \mathbb{R}^+$ for $\gamma \in [2, \infty)$). For any $\gamma < 0$, TBE_γ is not steep, as its mean parameter space $M_\gamma = \mathbb{R}^+$ is a proper subset of its interior convex support $\mathbb{R}$.

GLM applications for TBE_γ , $\gamma = 2, 3$, are straightforward and have been analyzed in various references (cf., [13] and the references cited therein). GLM applications for TBE_γ ($\gamma > 2, \gamma \neq 3$), are discussed and presented by [6], who also maintained an R package (see [10]) for these EDMs. Consequently, we are left with absolutely continuous TBE_γ models supported on the whole real line ($\gamma < 0, \gamma = 0, \gamma = \infty$). As already noted, the TBE_γ models for $\gamma < 0$ are not steep—a fact which precludes them from being candidates for GLM analysis. This is quite unfortunate, as such a subclass comprises an infinite set of absolutely continuous EDMs (with respect to the Lebesgue measure) that are supported on the whole real line. Thus, the only remaining steep EDMs supported on the whole real line are the normal (TBE_0) and the EDM generated by the Landau distribution (TBE_∞), where both are suitable for GLM applications. The normal EDM constitutes the classical linear regression model, whereas the TBE_∞ requires further analysis by GLM methodology, an analysis that establishes the core of this paper. Such an analysis would complement the results of [6] and accomplish a complete analysis of all absolutely continuous TBE models.

The paper is organized as follows. Section 2 presents some preliminaries on natural exponential families (NEFs), additive, and reproductive EDMs. Section 3 introduces the TBE_∞ —the EDM generated by the Landau distribution and the GLM ingredients needed

for its analysis. Mainly, we present its link function and total and scaled deviance. In Section 4, we study its analysis of deviance, derive the asymptotic properties of the MLEs of the covariate parameters β , and obtain the asymptotic distribution of deviance, using the saddlepoint approximation. Section 5 includes the estimation algorithm, a brief description of our R package, and simulation studies. In Section 6, we provide the analysis of real data, including applications to three real datasets. It is demonstrated there that the GLM using the TBE_∞ performs better than the linear model based on the normal distribution. Some concluding remarks are presented in Section 7. Proofs of statements (propositions, corollaries, and theorems) in this paper are relegated to Appendix A.

2. Preliminaries: NEFs, Mean Value Representation, and Additive and Reproductive EDMs

NEFs. The preliminaries in the sequel hold for any positive Radon measure $\mu(dx)$ on $\mathbb{R}$. Without loss of generality, we confine our introduction to $\mu(dx) = h(x)dx$, where $h(x)dx$ is an absolutely continuous positive Radon measure with respect to the Lebesgue measure on the real line. The Laplace transform of $h(x)$ and its effective domain are defined, respectively, by

$$L(\theta) = \int_{\mathbb{R}} h(x)e^{\theta x} dx \text{ and } D_h = \{\theta \in \mathbb{R}: L(\theta) < \infty\}.$$

Let $\Theta = \text{int } D_h$, and assume Θ is non-empty. Then, the NEF generated by h is defined by the densities of the form

$$h(x; \theta) = h(x) \exp\{\theta x - k(\theta)\}, \theta \in \Theta \subset \mathbb{R}, \quad (1)$$

where $k(\theta) = \ln L(\theta)$ is the cumulant transform of L . The cumulant transform $k(\theta)$ is a real analytic on Θ , implying that the r -th cumulant of $h(x; \theta)$ is given by $\kappa_r(\theta) = d^r k(\theta) / d\theta^r$. In particular, the mean, mean parameter space, and variance corresponding to (1) are given, respectively, by $m = k'(\theta)$, $M = k'(\Theta)$, and $k''(\theta)$, $\theta \in \Theta$. As k' is strictly increasing, its inverse mapping $\psi : M \rightarrow \Theta$ is well-defined. So, we denote by $V(m) = k''(\psi(m))$ the variance function (VF) corresponding to (1). The pair (V, M) uniquely defines the NEF generated by h within the class of NEFs (cf., [14]). Also, V is called the unit VF.

Mean value parameterization. For GLM applications and various other statistical aspects, it is necessary to express the NEF with densities (1) in terms of its mean rather than in terms of the artificial parameter θ (for details, see [3]). Indeed, given a VF (V, M) then $\theta(m)$ and $k(\theta(m))$ are the primitives of $1/V(m)$ and $m/V(m)$ and, thus, are given by

$$\theta(m) = \int \frac{dm}{V(m)} \text{ and } k(\theta(m)) = \int \frac{mdm}{V(m)}, \quad (2)$$

implying that the mean value representation of (1) is given by

$$h(x; \theta(m)) = h(x) \exp\{\theta(m)x - k(\theta(m))\}, m \in M.$$

Additive EDMs. The Jorgensen set related to (1) is defined by (cf., [2])

$$\Lambda = \{p \in \mathbb{R}^+ : pk(\theta) \text{ is a cumulant transform of some density } h(; p) \text{ on } \mathbb{R}\}.$$

The set Λ is not empty, due to convolution. Moreover, $\Lambda = \mathbb{R}^+$ if h is infinitely divisible, a valid property for all TBE members. Accordingly, the additive EDM (cf., [2]) is defined by densities of the form

$$h(x; \theta, p) = h(x; p) \exp\{\theta x - pk(\theta)\}, \theta \in \Theta, p \in \Lambda = \mathbb{R}^+, \quad (3)$$

where the VF corresponding to the additive EDM is given by $(pV(m/p), pM)$.

Reproductive EDMs. In general, for various statistical aspects, particularly for GLM applications, it is more effective to represent (3) by resembling the normal structure. Such a

representation, called the reproductive EDM, is obtained by mapping $x \rightarrow y = \varphi x$, where $\varphi = 1/p$. Then, the densities of this mapping have the form (cf., [2,6,15])

$$f_{\varphi}(y; \theta) = \varphi^{-1} h(y\varphi^{-1}; \theta, \varphi^{-1}) = h(y\varphi^{-1}; \varphi^{-1}) \exp\left\{\varphi^{-1}[\theta y - k(\theta)]\right\}, \quad (4)$$

where $y \in \varphi C_h$, $\theta \in \Theta$, $\varphi \in \mathbb{R}^+$, and C_h is the support of h . It is crucial to note that the structure in (4) is not suitable for the discrete case (counting measures on $\mathbb{N}$). This is because, for different φ 's, it alters the support C_h of h . In contrast, for the absolutely continuous case, the structure in (4) is appropriate. The VF of the reproductive EDM (4) is given by

$$(\varphi V(m), \varphi M), \quad (5)$$

where if $\text{int } C_h = \mathbb{R}, \mathbb{R}^+$ or $\mathbb{R}^-$ then $\varphi M = \mathbb{R}, \mathbb{R}^+$ or $\mathbb{R}^-$, respectively.

3. GLM Applications for the EDM Generated by the Landau Distribution—Some Basics

In this section, we provide some required components for GLM applications. We first give an expression for the TBE_{∞} density. We then present the related link function and scaled deviance.

3.1. Density Function

The TBE_{∞} distribution is the EDM generated by the Landau distribution, known as a Tweedie model with power infinity, and it possesses a simple unit VF of the form $(e^m, \mathbb{R})$. It is steep ($C = M = \mathbb{R}$), infinitely divisible, skewed to the right, leptokurtic (i.e., has fatter tails), and absolutely continuous, supported on the whole real line. It was surveyed in detail and further developed by [3], and was named there as EDM-EVF (exponential VF). Its reproductive EDM density, of the form (4), is

$$f(y; \theta, \varphi) = h_{\varphi}(y) \exp\left\{\varphi^{-1}[\theta y - k(\theta)]\right\}, y \in \mathbb{R}, (\theta, \varphi) \in \mathbb{R}^- \times \mathbb{R}^+, \quad (6)$$

where

$$h_{\varphi}(y) = \frac{1}{\pi} \int_0^{\infty} e^{(1-y)t - t \log t} \varphi^{-t} \sin(\pi t) dt, \quad (7)$$

$$k(\theta) = \theta - \theta \ln(-\theta), \quad (8)$$

and VF

$$(V, M) = (\varphi e^m, \mathbb{R}).$$

The expressions for $\theta(m)$ and $k(\theta(m))$, needed for its mean value parameterization, are

$$\theta(m) = \int \frac{dm}{V(m)} = -e^{-m} \text{ and } k(\theta(m)) = \int \frac{m dm}{V(m)} = -e^{-m}(m+1).$$

Thus, the density (6) can be written as

$$f(y; m, \varphi) = h_{\varphi}(y) \exp\left\{\varphi^{-1}[-e^{-m}y + e^{-m}(m+1)]\right\}, y \in \mathbb{R}, (m, \varphi) \in \mathbb{R} \times \mathbb{R}^+. \quad (9)$$

If $Y \sim f(y; m, \varphi)$ then we write this as $Y \sim \text{TBE}_{\infty}$ or we use the standard EDM notation and write $Y \sim \text{EDM-EVF}(m, \varphi)$. The mean, variance, and cumulants of such a Y are

$$E(Y) = m, \text{Var}(Y) = \varphi e^m, \kappa_r(m) = (r-2)! \varphi^{r-1} e^{(r-1)m}, r \geq 3. \quad (10)$$

3.2. Scaled Deviance and Link Function

We shall now consider two essential ingredients needed for GLM applications of EDM-EVFs (9): namely, the scaled deviance and the link function. These were introduced by [2,15] (see also [6,16]). For GLM applications, we need the following ingredients. Consider

$$t(y, m) = y\theta(m) - k(\theta(m)) = -ye^{-m} + e^{-m}(m + 1).$$

It is evident that $\arg \max_m f(y; m, \varphi) = \arg \max_m t(y, m)$. By taking the partial derivative of $t(y, m)$ and setting it to zero, we obtain

$$\frac{\partial t(y, m)}{\partial m} = ye^{-m} - me^{-m} \stackrel{\text{set}}{=} 0,$$

implying that $m = y$ maximizes $t(y, m)$ (since TBE_∞ is steep). Hence, the unit deviance

$$d(y, m) = 2[t(y, y) - t(y, m)] = 2[e^{-y} + e^{-m}(y - m - 1)] \quad (11)$$

can be considered as a distance measure with two properties: $d(y, y) = 0$ and $d(y, m) > 0$ for $y \neq m$. GLMs assume a systematic component with the linear predictor

$$\eta = \beta_0 + \sum_{j=1}^p \beta_j x_j.$$

This is linked to the mean m through a link function g , such that $g(m) = \eta$. For TBE_∞ , we choose the canonical (and simple) link function

$$\eta = g(m) = \theta(m) = -e^{-m}.$$

Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be a set of independent observations, where $y_i \sim \text{EDM-EVF}(m_i, \varphi)$ (assuming a single dispersion parameter) is associated with the link function $\eta_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j, i = 1, \dots, n$. Write $\mathbf{X} \in \mathbb{R}^{n \times (p+1)}$ for the set of covariates, in which case $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ and the total and scaled deviances are given, respectively, by

$$D(\mathbf{y}, \mathbf{m}) = \sum_{i=1}^n d(y_i, m_i)$$

and

$$D^*(\mathbf{y}, \mathbf{m}) = \frac{D(\mathbf{y}, \mathbf{m})}{\varphi} = \frac{2}{\varphi} \sum_{i=1}^n e^{-y_i} + e^{-m_i}(y_i - m_i - 1). \quad (12)$$

Consequently, the log-likelihood is

$$l(\mathbf{m}, \varphi; \mathbf{y}) = \sum_{i=1}^n \ln f(y_i; y_i, \varphi) - \frac{1}{2} D^*(\mathbf{y}, \mathbf{m}). \quad (13)$$

Let $\hat{\boldsymbol{\beta}}$ be the MLE of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$. As in linear models, we aim to estimate $\boldsymbol{\beta}$ and obtain its asymptotic behavior.

4. Asymptotic Properties

This section deals with the saddlepoint approximation and asymptotic behavior of the MLEs of the parameters involved. The section establishes the central core of the asymptotic behavior of all the statistics required for the appropriate analysis of the deviance.

4.1. Asymptotic Properties of MLE

Let us start with the saddlepoint approximation (14) below, which is essential in the asymptotic theory of GLMs. The exact distribution (9) is challenging to handle, due to the

cumbersome form of $h_\varphi(y)$ (7). The saddlepoint approximation neatly gets rid of it. For more details on this point, see Sections 1.5.3 and 3.5.1 in [2] and Section 5.4.3 in [6]. The following proposition presents the saddlepoint approximation for TBE_∞ .

Proposition 1. Let $Y \sim \text{TBE}_\infty$. Then, for sufficiently small φ , the saddlepoint approximation for the density of Y is given by

$$f(y; m, \varphi) \approx \frac{1}{\sqrt{2\pi\varphi V(y)}} \exp\left\{-\frac{1}{2\varphi} d(y, m)\right\}, \quad (14)$$

where $V(y) = e^y$ and $d(y, m) = 2[e^{-y} + e^{-m}(y - m - 1)]$.

Proof. See Appendix A. $\square$

The following corollary, an immediate consequence of Proposition 1, implies convergence to normality:

Corollary 1. Let $Y \sim \text{TBE}_\infty$; then,

$$\frac{Y - m}{\sqrt{\varphi}} \xrightarrow{d} N(0, V(m)), \text{ as } \varphi \rightarrow 0, \quad (15)$$

where $V(m) = e^m$ and $\xrightarrow{d}$ denotes convergence in distribution.

Proof. See Appendix A. $\square$

Corollary 1 provides the asymptotic normality for a single observation y . For $\mathbf{y} = (y_1, \dots, y_n)^\top$ with $y_i \sim \text{TBE}_\infty$, we have

$$\frac{\mathbf{Y} - \mathbf{m}}{\sqrt{\varphi}} \xrightarrow{d} N(\mathbf{0}_n, C), \text{ as } \varphi \rightarrow 0, \quad (16)$$

where

$$C = \text{diag}\{V(m_1), V(m_2), \dots, V(m_n)\} = \text{diag}\{e^{m_1}, e^{m_2}, \dots, e^{m_n}\}.$$

Using (16), the following theorem shows that the MLE of β is asymptotically normally distributed.

Theorem 1. Let $\hat{\beta}$ be the MLE of β and let $\mathbf{X} \in \mathbb{R}^{n \times (p+1)}$ be the design matrix. If $\mathbf{X}\mathbf{X}^\top$ has bounded eigenvalues, then

$$\frac{\hat{\beta} - \beta_0}{\sqrt{\varphi}} \xrightarrow{d} N(\mathbf{0}_{p+1}, (\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1}), \text{ as } \varphi \rightarrow 0, \quad (17)$$

where β_0 is the true parameter.

Proof. See Appendix A. $\square$

4.2. Analysis of the Deviance

With $\mathbf{m}$ and φ known, we consider the distribution of deviance. We claim that when the saddlepoint approximation holds (and it does for TBE_∞) then the scaled deviance distribution follows an approximate chi-square distribution.

Theorem 2. For the scaled deviance (12), we have

$$D^*(\mathbf{y}, \mathbf{m}) \xrightarrow{d} \chi_n^2, \text{ as } \varphi \rightarrow 0$$

at the true values of $\mathbf{m} = (m_1, \dots, m_n)^\top$.

Proof. See Appendix A. $\square$

When $\mathbf{m}$ is unknown, it is replaced by its MLE $\hat{\mathbf{m}}$. Thus, we define the residual and scaled residual deviances as

$$D(\mathbf{y}, \hat{\mathbf{m}}) = \sum_{i=1}^n d(y_i, \hat{m}_i)$$

and

$$D^*(\mathbf{y}, \hat{\mathbf{m}}) = \frac{D(\mathbf{y}, \hat{\mathbf{m}})}{\varphi}.$$

As the GLM considered in Section 3 is involved with $p + 1$ regression parameters, it follows that

$$D^*(\mathbf{y}, \hat{\mathbf{m}}) \xrightarrow{d} \chi_{n-p-1}^2, \text{ as } \varphi \rightarrow 0. \quad (18)$$

Generally, the deviance is most useful not as an absolute measure of goodness-of-fit, but rather for comparing two nested models. For example, one may want to test whether incorporating an additional covariate significantly improves the model fit. In this case, the deviance can be employed to compare two nested GLMs that are based on the same EDM but have different fitted systematic components:

$$\text{Model A: } g(\hat{\mathbf{m}}_A) = -e^{-\hat{\mathbf{m}}_A} = \hat{\beta}_0^A \mathbf{1}_n + \hat{\beta}_1^A \mathbf{x}_1 + \dots + \hat{\beta}_{p_A}^A \mathbf{x}_{p_A},$$

and

$$\text{Model B: } g(\hat{\mathbf{m}}_B) = -e^{-\hat{\mathbf{m}}_B} = \hat{\beta}_0^B \mathbf{1}_n + \hat{\beta}_1^B \mathbf{x}_1 + \dots + \hat{\beta}_{p_A}^B \mathbf{x}_{p_A} + \dots + \hat{\beta}_{p_B}^B \mathbf{x}_{p_B},$$

where $\hat{\beta}_i^A$ denotes the MLE of β_i under Model A, $\hat{\beta}_j^B$ denotes the MLE of β_j under Model B, and $\mathbf{x}_j$ is a covariate, $i = 0, 1, \dots, p_A, j = 0, 1, \dots, p_B$. Note that Model A is a special case of Model B, with $p_B > p_A$. Accordingly, we consider the following hypotheses, to determine if the simpler Model A is adequate to model the data:

$$H_0 : \beta_{p_A+1} = \dots = \beta_{p_B} = 0 \text{ versus } H_1 : \beta_j \neq 0 \quad \exists j \in \{p_A + 1, \dots, p_B\}. \quad (19)$$

We have previously observed that the total deviance captures that part of the log-likelihood that depends on $\mathbf{m}$. Therefore, the following theorem holds, from which it can be seen that (18) is a special case of Theorem 3:

Theorem 3. If φ is known, the likelihood ratio test (LRT) statistic for comparing Models A and B is

$$L = 2\{\ell_B - \ell_A\} = \frac{D(\mathbf{y}, \hat{\mathbf{m}}_A) - D(\mathbf{y}, \hat{\mathbf{m}}_B)}{\varphi}. \quad (20)$$

Then, under the null hypothesis in (19), $L \xrightarrow{d} \chi_{p_B-p_A}^2$ as $\varphi \rightarrow 0$.

Proof. See Appendix A. $\square$

Consider the two models in Theorem 3 with both $\mathbf{m}$ and φ unknown. Then, an estimate of φ is required. This is done in Theorem 4, which is deduced from Theorem 3:

Theorem 4. If φ is unknown, the appropriate statistic for comparing Model A with Model B is

$$F = \frac{[D(\mathbf{y}, \hat{\mathbf{m}}_A) - D(\mathbf{y}, \hat{\mathbf{m}}_B)] / (p_B - p_A)}{D(\mathbf{y}, \hat{\mathbf{m}}_B) / (n - p_B - 1)}, \quad (21)$$

where $D(\mathbf{y}, \hat{\mathbf{m}}_B) / (n - p_B - 1)$ is an estimate of φ based on Model B. Then, under the null hypothesis in (19),

$$F \xrightarrow{d} F(p_B - p_A, n - p_B - 1), \text{ as } \varphi \rightarrow 0.$$

Proof. It suffices to prove the asymptotic independence of $D(\mathbf{y}, \hat{\mathbf{m}}_A) - D(\mathbf{y}, \hat{\mathbf{m}}_B)$ and $D(\mathbf{y}, \hat{\mathbf{m}}_B)$. The proof is similar to Theorem 4.3 in reference [17]. $\square$

Note that our above statements about asymptotic distributions are all based on the assumption that $\varphi \rightarrow 0$. These results are called small-dispersion asymptotics, regardless of the sample size n . Large-sample asymptotics are also well-known and, hence, no further explanations are provided.

5. Simulation Studies

5.1. Implementation

Herein, we discuss the estimation of the unknown parameters in the TBE_∞ GLM: the covariate coefficients β and the dispersion parameter φ . For the estimation of β , we use iteratively reweighted least squares (IRLS). The score vector $\mathbf{U}$ for β is

$$\mathbf{U}(\beta) = \frac{1}{\varphi} \mathbf{X}^\top \mathbf{W} \mathbf{M} (\mathbf{y} - \mathbf{m}),$$

where $\mathbf{W} = \text{diag}(W_1, \dots, W_n)$ and $W_i = [V(m_i)(d\eta_i/dm_i)^2]^{-1}$ are called the working weights, and $\mathbf{M}$ is the diagonal matrix of the link derivatives $d\eta_i/dm_i = e^{-m_i}$. The Fisher information matrix for β is

$$\mathcal{I}(\beta) = \frac{1}{\varphi} \mathbf{X}^\top \mathbf{W} \mathbf{X}.$$

Thus, an iterative technique using the Newton–Raphson method yields

$$\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} + \mathcal{I}(\hat{\beta}^{(r)})^{-1} \mathbf{U}(\hat{\beta}^{(r)}),$$

where the Fisher information matrix $\mathcal{I}(\cdot)$ is used in place of the observed information matrix, and the superscript (r) denotes the r -th iterate. The iteration can be re-organized as IRLS (cf., [6]):

$$\hat{\beta}^{(r+1)} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{z},$$

where $\mathbf{z}$, the working response vector, is given by

$$\mathbf{z} = \hat{\eta} + \mathbf{M}(\mathbf{y} - \hat{\mathbf{m}}),$$

and all other quantities on the right-hand side are evaluated at $\hat{\beta}^{(r)}$.

For the estimation of φ , we use the mean deviance estimator in [6]. It is easy to show that the MLE of φ is the simple mean deviance $D(\mathbf{y}, \hat{\mathbf{m}})/n$ with the saddlepoint approximation density. When taking into account the estimation of β and the residual degrees of freedom, we obtain the mean deviance estimator of φ as

$$\hat{\varphi} = D(\mathbf{y}, \hat{\mathbf{m}})/(n - p - 1).$$

We summarize all of the above as Algorithm 1.

We have developed an R package named `TBEinf` [18], which is used in our numerical experiments and is publicly available at <https://github.com/xliusufe/TBEinf> (accessed on 28 April 2024).

The package includes a program for computing the density of TBE_∞ by a direct calculation (cf., [3]), saddlepoint approximation (cf., [6]), Fourier inversion (cf., [2,16]), and modified W-Transformation (cf., [16,19]). Specifically, the function `dTBEinf` in the package calculates density by real density when `method = "real"`, saddlepoint approximation when `method = "saddle"`, Fourier inversion when `method = "finverse"`, and modified W-transformation when `method = "mWtrans"`.

Algorithm 1 Estimating β and φ Based on Iteratively Reweighted Least Squares Estimation (IRLSE) and Mean Deviance

- 1: **Input:** Data $\{(\mathbf{y}, \mathbf{X})\}$, initial value of β , threshold $\tau = 10^{-8}$.
- 2: **Repeat:**
- 3: Step 1. Obtain $\eta = \mathbf{X}\beta$ and $\mathbf{m} = -\log(-\eta)$.
- 4: Step 2. Calculate

$$\beta_{new} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{z},$$

where $\mathbf{W} = \text{diag}\{e^{m_1}, e^{m_2}, \dots, e^{m_n}\}$ and $\mathbf{z} = (z_1, \dots, z_n)^\top$ with $z_i = \eta_i + (y_i - m_i)/e^{m_i}$, $i = 1, \dots, n$.

- 5: Step 3. Set $\beta = \beta_{new}$.
- 6: **Until:** $\|\beta_{new} - \beta\|_\infty \leq \tau$.
- 7: Step 4. Calculate the total deviance

$$D(\mathbf{y}, \mathbf{m}) = 2 \sum_{i=1}^n [e^{-y_i} + e^{-m_i} (y_i - m_i - 1)],$$

and $\varphi = D(\mathbf{y}, \mathbf{m}) / (n - p - 1)$.

- 8: **Output:** β and φ .
-

Also, the package applies GLMs methodology to TBE_∞ for estimation and prediction. The estimates of the covariate coefficients β and the dispersion parameter φ are obtained through Algorithm 1.

5.2. Simulation Studies

Firstly, we generated simulated data using (16). We let the sample sizes be $n = 100, 200, 400, 800$, respectively, and the true values be $\beta_0 = [-0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9, -1.0]$, $\varphi_0 = 0.01$. The first column of $\mathbf{X}$ was a 1-vector, and all the other elements were random numbers sampled from $U(0, 1)$. We set 1000 repetitions and then generated 1000 $\mathbf{y}$ for each n . We estimated β and φ according to Algorithm 1.

By applying Algorithm 1, the average value of estimated φ was $\bar{\varphi} = 3.59 \times 10^{-5}$ for $n = 100$, $\bar{\varphi} = 3.77 \times 10^{-5}$ for $n = 200$, 3.66×10^{-5} for $n = 400$, and 3.73×10^{-5} for $n = 800$. It can be seen that they were relatively small compared to the true value $\varphi_0 = 0.01$. This is because the total deviance was very small.

Table 1 lists the simulation results calculated by applying Algorithm 1 with varying sample sizes $n = 100, 200, 400, 800$. Herein, sd denotes the standard deviation (SD), which is computed by $\text{sd}(\hat{\beta}_{0,k}) = \sqrt{\sum_{j=1}^{1000} (\hat{\beta}_{0,k}^{(j)} - \bar{\beta}_{0,k})^2 / 999}$ for $k = 1, \dots, 10$, where $\hat{\beta}_{0,k}^{(j)}$ is the estimate of $\beta_{0,k}$ at the j -th repetition and $\bar{\beta}_{0,k}$ is the average of $\{\hat{\beta}_{0,k}^{(1)}, \dots, \hat{\beta}_{0,k}^{(1000)}\}$. Also, se denotes the standard error (SE) of $\hat{\beta}$, which can be calculated using (17) by $\text{se}(\hat{\beta}_{0,k}) = \sqrt{\bar{\varphi} b_k}$, where $\bar{\varphi}$ is the average value of estimated φ , and b_k is the k -th diagonal element of $(\mathbf{X}^\top \mathbf{C} \mathbf{X})^{-1}$. From Table 1, we can see that the average bias was around 10^{-5} and SD was around 10^{-3} , which demonstrated the estimation procedure performs well and stably. It is observed that the SDs were all close to the SEs, and they all decreased as n increased, which implies that the simulation results verify that the asymptotic properties are reasonable.

Table 1. Average of bias, sd, and se of estimated β for $n = 100, 200, 400, 800$. β_0 is the true value of β . $\beta_{0,1}$ is the intercept term.

β_0	$n = 100$			$n = 200$			$n = 400$			$n = 800$		
	bias ($\times 10^{-5}$)	sd	se	bias ($\times 10^{-5}$)	sd	se	bias ($\times 10^{-5}$)	sd	se	bias ($\times 10^{-5}$)	sd	se
$\beta_{0,1} = -0.1$	2.0	0.0056	0.0048	−16	0.0036	0.0032	−2.8	0.0029	0.0026	7.7	0.0020	0.0018
$\beta_{0,2} = -0.2$	−11	0.0037	0.0036	2.7	0.0026	0.0026	1.2	0.0018	0.0018	−10	0.0013	0.0012
$\beta_{0,3} = -0.3$	−3.8	0.0040	0.0039	−12	0.0026	0.0026	2.0	0.0017	0.0017	−0.43	0.0013	0.0012
$\beta_{0,4} = -0.4$	4.5	0.0036	0.0036	11	0.0025	0.0024	−5.1	0.0018	0.0017	−0.94	0.0012	0.0012
$\beta_{0,5} = -0.5$	−6.6	0.0040	0.0038	14	0.0026	0.0024	6.6	0.0018	0.0018	−4.1	0.0012	0.0012
$\beta_{0,6} = -0.6$	−7.3	0.0036	0.0034	2.5	0.0028	0.0026	0.22	0.0018	0.0017	2.5	0.0012	0.0012
$\beta_{0,7} = -0.7$	−3.1	0.0036	0.0034	5.8	0.0025	0.0025	−2.1	0.0018	0.0017	−6.2	0.0012	0.0012
$\beta_{0,8} = -0.8$	−6.8	0.0038	0.0036	14	0.0027	0.0026	−0.61	0.0017	0.0018	1.1	0.0013	0.0012
$\beta_{0,9} = -0.9$	12	0.0033	0.0032	−7.6	0.0025	0.0025	2.8	0.0018	0.0017	−2.8	0.0013	0.0012
$\beta_{0,10} = -1.0$	9.3	0.0038	0.0036	−0.84	0.0025	0.0024	−3.2	0.0018	0.0017	3.6	0.0013	0.0012

6. Real Data Analysis

We present the proposed estimation procedure through applications to three real datasets. The first and second datasets, grazing and hcrabs, are both from the R package ‘GLMsData’ (see [6,20]). The last one is Boston housing data.

6.1. Dataset “Grazing”

This dataset reveals the density of understorey birds across a series of sites located on either side of a stockproof fence, in two distinct areas. It has the potential to provide insights into the impact of habitat fragmentation on bird populations (cf., [20]):

1. Sample size: $n = 62$;
2. The number of variables: $p = 3$;
3. Variables description; see Table 2.

Table 2. Variables description of grazing dataset.

Variables	Meanings
Birds	the number of understorey birds; a numeric vector
When	when the bird count was conducted; a factor with levels Before (before herbivores were removed) and After (after herbivores were removed)
Grazed	which side of the stockproof fence; a factor with levels Reference (grazed by native herbivores) and Feral (grazed by feral herbivores, mainly horses)

To verify the appropriateness of the TBE_∞ GLM for the data, we evaluated the prediction performance of TBE_∞ GLM and compared it with a linear model. We conducted 500 random splits of the 62 observations. In each split, we randomly selected 80% as the training set and the rest as the testing set $\{(y_{test,i}, \mathbf{x}_{test,i}), i = 1, \dots, 13\}$, where 13 was the result of multiplying 62 by 20% and rounding up. We applied both TBE_∞ GLM and the linear model to the training set and estimated the coefficients.

By applying Algorithm 1 for the TBE_∞ GLM, the estimates of β and φ were

$$\hat{\beta} = (-0.392, 0.013, -0.017)^\top$$

and $\hat{\varphi} = 0.040$. Then, we predicted $y_{test,i}$ by $\hat{y}_i = -\log(-\hat{\eta}_i)$, where $\hat{\eta}_i = \mathbf{x}_{test,i}^\top \hat{\beta}$. We let $\hat{\epsilon}_i = y_{test,i} - \hat{y}_i$ and calculated the mean squared error (MSE) as $MSE(\hat{\mathbf{y}}) = \frac{1}{13} \sum_{i=1}^{13} \hat{\epsilon}_i^2$, where $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_{13})$. In the linear model, we estimated β as $\tilde{\beta}$, using least squares (without φ). Then, we predicted $y_{test,i}$ by $\tilde{y}_i = \mathbf{x}_{test,i}^\top \tilde{\beta}$. We let $\tilde{\epsilon}_i = y_{test,i} - \tilde{y}_i$ and calculated $MSE(\tilde{\mathbf{y}}) = \frac{1}{13} \sum_{i=1}^{13} \tilde{\epsilon}_i^2$, where $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_{13})$.

Thus, we could compute the average and sd of the predictions' MSEs of both models under 500 random splits. For TBE_{∞} GLM, we obtained the average and sd of the MSEs to be 0.111 and 0.017, respectively. For the linear model, we obtained the average and sd of the MSEs to be 0.760 and 0.238, respectively. Thus, it can be seen that the results for the TBE_{∞} GLM performed much better than the linear model for both average and sd. Additionally, we calculated the Bayesian information criterion (BIC) for both models, resulting in $BIC(TBE_{\infty}) = -267.028$ and $BIC(LM) = 385.949$, where LM denotes the linear model. It is evident that the BIC for TBE_{∞} GLM was significantly lower than that for the linear model, indicating a better model fit.

6.2. Dataset "Hcrabs"

This dataset describes the number of male crabs attached to female horseshoe crabs (cf., [20]):

1. Sample size: $n = 173$;
2. The number of variables: $p = 5$;
3. Variables description; see Table 3 below.

Table 3. Variables description of hcrabs dataset.

Variables	Meanings
Col	the color of the female; a factor with levels LM (light medium), M (medium), DM (dark medium) or D (dark)
Spine	the spine condition; a factor with levels BothOK, OneOK or NoneOK
Width	the carapace width of the female crab in cm; a numeric vector
Wt	the weight of the female crab in grams; a numeric vector
Sat	the number of male crabs attached to the female ('satellites'); a numeric vector

As we did with processing the first dataset above, we meticulously conducted 500 random splits of the 173 observations. In each split, we carefully selected 80% as the training set and the rest as the testing set $\{(y_{test,i}, \mathbf{x}_{test,i}), i = 1, \dots, 35\}$, where 35 was the result of multiplying 173 by 20% and rounding up. We applied both TBE_{∞} GLM and the linear model, ensuring the validity of our approach.

The estimates of β and φ for the TBE_{∞} GLM were

$$\hat{\beta} = (-1.276, 0.001, 0.002, 0.885, 0.009)^{\top}$$

and $\hat{\varphi} = 0.004$. Then, we predicted $y_{test,i}$ by $\hat{y}_i = -\log(-\hat{\eta}_i)$, where $\hat{\eta}_i = \mathbf{x}_{test,i}^{\top} \hat{\beta}$. We let $\hat{\epsilon}_i = y_{test,i} - \hat{y}_i$ and calculated $MSE(\hat{y}) = \frac{1}{35} \sum_{i=1}^{35} \hat{\epsilon}_i^2$, where $\hat{y} = (\hat{y}_1, \dots, \hat{y}_{35})$. In the linear model, we estimated β as $\tilde{\beta}$ using least squares. Then, we predicted $y_{test,i}$ by $\tilde{y}_i = \mathbf{x}_{test,i}^{\top} \tilde{\beta}$. We let $\tilde{\epsilon}_i = y_{test,i} - \tilde{y}_i$ and calculated $MSE(\tilde{y}) = \frac{1}{35} \sum_{i=1}^{35} \tilde{\epsilon}_i^2$, where $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_{35})$.

For TBE_{∞} GLM, we ascertained the average and sd of the MSEs to be 0.011 and 0.004, respectively. For the linear model, we ascertained the average and sd of the MSEs to be 0.837 and 0.276, respectively. Here, again, it can be seen that the results for the TBE_{∞} GLM performed much better than the linear model for both average and sd. We calculated the BIC for both models, resulting in $BIC(TBE_{\infty}) = 48.704$ and $BIC(LM) = 150.366$. It is clear that the BIC for TBE_{∞} GLM was lower than that for the linear model, indicating a superior model fit.

6.3. Dataset "Boston Housing"

This dataset is taken from Harrison Jr and Rubinfeld 1978, including 14 variables that were measured across 506 census tracts in the Boston area. The response variable can be the

logarithm of the median value of the houses in those census tracts of the Boston Standard Metropolitan Statistical Area:

1. Sample size: $n = 506$;
2. The number of variables: $p = 14$;
3. Variables description; see Table 4.

Table 4. Variables description of Boston housing dataset.

Variables	Meanings
CRIM	Crime rate by town
ZN	Proportion of residential land zoned for lots over 25,000 sq. ft
INDUS	Proportion of nonretail business acres per town
CHAS	Charles River dummy variable (1 if tract bounds the river)
NOX	Concentration of nitrogen oxides in parts per 10 million
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	Weighted mean of distances to five Boston employment centres
RAD	Index of accessibility to radial highways
TAX	Full-value property-tax rate per \$10,000
PTRATIO	Pupil/teacher ratio by town
B	The proportion of black people by town
LSTAT	Percentage of people of lower status

Again, we conducted 500 random splits of the 506 observations. In each split, we carefully selected 80% as the training set and the rest as the testing set $\{(y_{test,i}, \mathbf{x}_{test,i}), i = 1, \dots, 102\}$, where 102 was the result of multiplying 506 by 20% and rounding up. We applied both TBE_{∞} GLM and the linear model, comparing the performance.

For the TBE_{∞} GLM, the estimates of β and φ were

$$\hat{\beta} = (-0.142, -0.027, 0.010, 0.009, 0.008, -0.148, 0.270, 0.007, -0.078, 0.068, \\ -0.084, -0.241, 0.062, -0.155)^{\top}$$

and $\hat{\varphi} = 0.010$. Then, we predicted $y_{test,i}$ by $\hat{y}_i = -\log(-\hat{\eta}_i)$, where $\hat{\eta}_i = \mathbf{x}_{test,i}^{\top} \hat{\beta}$. We let $\hat{\epsilon}_i = y_{test,i} - \hat{y}_i$ and calculated $MSE(\hat{\mathbf{y}}) = \frac{1}{102} \sum_{i=1}^{102} \hat{\epsilon}_i^2$, where $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_{102})$. In the linear model, we estimated β as $\tilde{\beta}$, using least squares. Then, we predicted $y_{test,i}$ by $\tilde{y}_i = \mathbf{x}_{test,i}^{\top} \tilde{\beta}$. We let $\tilde{\epsilon}_i = y_{test,i} - \tilde{y}_i$ and calculated $MSE(\tilde{\mathbf{y}}) = \frac{1}{102} \sum_{i=1}^{102} \tilde{\epsilon}_i^2$, where $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_{102})$.

For TBE_{∞} GLM, we ascertained the average and sd of the MSEs to be 0.031 and 0.009, respectively. For the linear model, we ascertained the average and sd of the MSEs to be 0.041 and 0.010, respectively. It can be seen that this dataset was appropriate for the linear model, and that our TBE_{∞} GLM could also fit well, which, to some extent, reflects the wide application of TBE_{∞} GLM. It can also be seen that the results for the TBE_{∞} GLM were slightly better than the linear model for both average and sd. We computed the BIC for both models, yielding $BIC(TBE_{\infty}) = -332.293$ and $BIC(LM) = -140.174$. The lower BIC of TBE_{∞} GLM compared to the linear model indicated a superior model fit.

7. Conclusions

In this paper, we were interested in GLM methodology applied to the TBE_{∞} —the EDM generated by the Landau distribution, an EDM supported on the real line. We introduced its density function, deviance, and link function. We considered the saddlepoint approximation approach for $Y \sim TBE_{\infty}$ and then deduced the convergence of Y to normality. Based on the small dispersion and saddlepoint approximation, we derived that the asymptotic distribution of MLE for $\hat{\beta}$ was normal. The analysis of deviance was also studied, considering different situations of φ and m . In numerical studies, we first estimated β and φ using Algorithm 1 and then evaluated its estimation performance. We reported averages of bias, standard deviations (SDs), and standard errors (SEs) in a simulation study. We demonstrated that the biases and SDs were relatively small and that the SDs were close to the SEs. As for applications to the three datasets of real data, the results for TBE_{∞} GLM showed much better performance than the linear models. To some extent, this indicated the widespread applications of TBE_{∞} . We also composed an R package for GLM applications of TBE_{∞} .

We trust that the proposed TBE_{∞} GLM will be well utilized in modeling more real data and various statistical purposes.

Author Contributions: Conceptualization, S.K.B.-L.; methodology, S.K.B.-L., X.L. and Z.X.; software, X.L. and Z.X.; validation, X.L. and Z.X.; formal analysis, X.L. and Z.X.; investigation, S.K.B.-L., X.L., A.R. and Z.X.; resources, S.K.B.-L. and X.L.; data curation, S.K.B.-L., X.L. and Z.X.; writing—original draft preparation, S.K.B.-L., X.L. and Z.X.; writing—review and editing, S.K.B.-L., X.L., A.R. and Z.X.; visualization, X.L. and Z.X.; supervision, S.K.B.-L. and X.L.; project administration, S.K.B.-L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research of Liu and Xiang was funded by the National Natural Science Foundation of China (12271329, 72331005), the Program for Innovative Research Team of SUFE, the Shanghai Research Center for Data Science and Decision Technology, the Open Research Fund of the Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, and the Open Research Fund of the Key Laboratory of Analytical Mathematics and Applications (Fujian Normal University), Ministry of Education, P. R. China. The research of Bar-Lev and Ridder was funded by STAR (Stochastics—Theoretical and Applied Research), one of the four mathematics clusters within the Dutch Research Council (NWO).

Data Availability Statement: All real datasets used in this manuscript are explicitly displayed in the paper.

Acknowledgments: We thank two reviewers for helpful comments and suggestions.

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

Abbreviations

The following abbreviations are used in this manuscript:

EDM	exponential dispersion model
EVF	exponential variance function
VF	variance function
TBE	Tweedie, Bar-Lev, and Enis
NEF	natural exponential family
LRT	likelihood ratio test
IRLS	iteratively reweighted least squares
SD	standard deviation
SE	standard error

Appendix A. Proofs

Proof of Proposition 1. For $Y \sim \text{TBE}_\infty$, by steepness and (6), the characteristic function of Y is

$$\begin{aligned}\Phi(t; \theta, \varphi) &= E[\exp(ity)] \\ &= \int_{-\infty}^{\infty} \exp(ity) h_\varphi(y) \exp\left\{\varphi^{-1}[\theta y - k(\theta)]\right\} dy \\ &= \int_{-\infty}^{\infty} h_\varphi(y) \exp\left\{\varphi^{-1}[(\theta + it\varphi)y - k(\theta)]\right\} dy \\ &= \exp\left\{\frac{k(\theta') - k(\theta)}{\varphi}\right\} \int_{-\infty}^{\infty} h_\varphi(y) \exp\left\{\frac{\theta'y - k(\theta')}{\varphi}\right\} dy \\ &= \exp\left\{\frac{k(\theta') - k(\theta)}{\varphi}\right\},\end{aligned}$$

where $\theta' = \theta + it\varphi$. The last equation holds, since the integrand is an EDM density function written in terms of θ' rather than θ . If $\Phi(t; \theta, \varphi)$ is absolutely integrable, then by the Fourier inversion theorem the probability density function of Y is

$$\begin{aligned}f(y; m, \varphi) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(t; \theta, \varphi) \exp(-ity) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{\frac{k(\theta + it\varphi) - k(\theta)}{\varphi} - ity\right\} dt \\ &= \frac{1}{2\pi\varphi} \int_{-\infty}^{\infty} \exp\left\{\frac{k(\theta + is) - k(\theta) - isy}{\varphi}\right\} ds,\end{aligned}$$

where i is the complex imaginary unit and $s = t\varphi$.

Since $m = k'(\theta)$ and ψ is the inverse mapping of k' , we have $\theta = \psi(m)$. Let $\tilde{\theta} = \psi(y)$; then, $y = k'(\tilde{\theta})$. Since the integrand is analytic, we may move the path of integration from $(-\infty, \infty)$ to $i(\theta - \tilde{\theta}) + (-\infty, \infty)$. The density then becomes

$$f(y; m, \varphi) = \frac{1}{2\pi\varphi} \int_{-\infty}^{\infty} \exp\left\{\frac{k(\tilde{\theta} + is) - (\tilde{\theta} + is)y + \theta y - k(\theta)}{\varphi}\right\} ds. \quad (\text{A1})$$

We introduce the unit deviance,

$$\begin{aligned}d(y, m) &= 2[t(y, y) - t(y, m)] \\ &= 2[y\theta(y) - k(\theta(y)) - y\theta(m) + k(\theta(m))] \\ &= 2[y\psi(y) - k(\psi(y)) - y\theta + k(\theta)] \\ &= 2[y\tilde{\theta} - k(\tilde{\theta}) - y\theta + k(\theta)].\end{aligned}$$

Let $\varphi \rightarrow 0$; then, for every fixed t , $s = t\varphi \rightarrow 0$. By expanding k around $\tilde{\theta}$, we obtain

$$\begin{aligned}k(\tilde{\theta} + is) &= k(\tilde{\theta}) + isk'(\tilde{\theta}) + \frac{1}{2}(is)^2 k''(\tilde{\theta}) + o((is)^2) \\ &= k(\tilde{\theta}) + isy - \frac{1}{2}s^2 k''(\tilde{\theta}) + o(s^2) \\ &= k(\tilde{\theta}) + isy - \frac{1}{2}s^2 V(y) + o(s^2),\end{aligned}$$

where $k''(\tilde{\theta}) = V(y)$, since $V(m) = k''(\psi(m))$ and $\psi(y) = \tilde{\theta}$.

We now consider the term in curly brackets in (A1). By introducing the unit deviance and expanding k around $\tilde{\theta}$, this term becomes

$$\begin{aligned} & \varphi^{-1} [k(\tilde{\theta} + is) - (\tilde{\theta} + is)y + \theta y - k(\theta)] \\ &= \varphi^{-1} [k(\tilde{\theta} + is) - (\tilde{\theta} + is)y + \tilde{\theta}y - k(\tilde{\theta}) - d(y, m)/2] \\ &\approx \varphi^{-1} \left[-\frac{1}{2}s^2 V(y) - \frac{1}{2}d(y, m) \right], \end{aligned}$$

where high-order terms of s^2 are discarded. From the result

$$\int_{-\infty}^{\infty} \exp \left\{ -\frac{V(y)s^2}{2\varphi} \right\} ds = \sqrt{\frac{2\pi\varphi}{V(y)}},$$

we obtain the approximation for $\varphi > 0$ small enough,

$$\begin{aligned} f(y; m, \varphi) &\approx \frac{1}{2\pi\varphi} \int_{-\infty}^{\infty} \exp \left\{ \varphi^{-1} \left[-\frac{1}{2}s^2 V(y) - \frac{1}{2}d(y, m) \right] \right\} ds \\ &= \frac{1}{2\pi\varphi} \sqrt{\frac{2\pi\varphi}{V(y)}} \exp \left\{ -\frac{1}{2\varphi}d(y, m) \right\} \\ &= \frac{1}{\sqrt{2\pi\varphi V(y)}} \exp \left\{ -\frac{1}{2\varphi}d(y, m) \right\}. \end{aligned}$$

This completes the proof of Proposition 1 (cf., [2]). $\square$

Proof of Corollary 1. Before proving this, we prove a lemma: the unit scaled deviance behaves approximately as the normal unit deviance near its minimum $m = y$ (cf., [2]). Let

$$d^*(y, m) = \frac{d(y, m)}{\varphi} = \frac{2[e^{-y} + e^{-m}(y - m - 1)]}{\varphi}.$$

For $Y \sim \text{TBE}_{\infty}$, we first show that

$$\frac{\partial^2 d^*}{\partial y^2}(m, m) = \frac{\partial^2 d^*}{\partial m^2}(m, m) = -\frac{\partial^2 d^*}{\partial m \partial y}(m, m). \quad (\text{A2})$$

By a simple calculation, we have

$$\begin{aligned} \frac{\partial d^*}{\partial y}(y, m) &= \frac{2}{\varphi}(e^{-m} - e^{-y}), \text{ i.e., } \frac{\partial d^*}{\partial y}(m, m) = 0. \\ \frac{\partial d^*}{\partial m}(y, m) &= \frac{2}{\varphi}(-e^{-m}(y - m)), \text{ i.e., } \frac{\partial d^*}{\partial m}(m, m) = 0. \\ \frac{\partial^2 d^*}{\partial y^2}(y, m) &= \frac{2}{\varphi}e^{-y}, \text{ i.e., } \frac{\partial^2 d^*}{\partial y^2}(m, m) = \frac{2}{\varphi}e^{-m}. \\ \frac{\partial^2 d^*}{\partial m^2}(y, m) &= \frac{2}{\varphi}e^{-m}(y + 1 - m), \text{ i.e., } \frac{\partial^2 d^*}{\partial m^2}(m, m) = \frac{2}{\varphi}e^{-m}. \\ \frac{\partial^2 d^*}{\partial m \partial y}(y, m) &= -\frac{2}{\varphi}e^{-m}, \text{ i.e., } -\frac{\partial^2 d^*}{\partial m \partial y}(m, m) = \frac{2}{\varphi}e^{-m}. \end{aligned}$$

Thus, (A2) holds. Then, the unit variance function satisfies the relationship

$$\varphi V(m) = \varphi e^m = \frac{2}{\frac{\partial^2 d^*}{\partial y^2}(m, m)} = \frac{2}{\frac{\partial^2 d^*}{\partial m^2}(m, m)} = -\frac{2}{\frac{\partial^2 d^*}{\partial m \partial y}(m, m)}. \quad (\text{A3})$$

Furthermore, (A3) implies the following second-order Taylor expansion of d^* near its minimum ($\delta^2 \rightarrow 0$):

$$d^*(m_0 + a\delta, m_0 + b\delta) = \frac{\delta^2}{\varphi V(m_0)}(a - b)^2 + o(\delta^2). \quad (\text{A4})$$

This expansion shows that the unit deviance behaves approximately as does the normal unit deviance near its minimum.

Let the mean of Y be m_0 and $Z = \frac{Y - m_0}{\sqrt{\varphi}}$. From the saddlepoint approximation (14) of Y , we ascertain, for $\varphi > 0$ small enough, that

$$f(y; m_0, \varphi) \approx \frac{1}{\sqrt{2\pi\varphi V(y)}} \exp\left\{-\frac{1}{2}d^*(y, m_0)\right\}.$$

Then, substituting (A4) and taking $\delta = \sqrt{\varphi}$, $b = 0$, we ascertain, for $\varphi > 0$ small enough, that

$$\begin{aligned} f(z; \varphi) &= \sqrt{\varphi} f(y; m_0, \varphi) \\ &\approx \sqrt{\varphi} \frac{1}{\sqrt{2\pi\varphi V(\sqrt{\varphi}z + m_0)}} \exp\left\{-\frac{1}{2}d^*(\sqrt{\varphi}z + m_0, m_0)\right\} \\ &= \frac{1}{\sqrt{2\pi V(\sqrt{\varphi}z + m_0)}} \exp\left\{-\frac{1}{2} \frac{\varphi}{\varphi V(m_0)} z^2 + o(\varphi)\right\} \\ &\approx \frac{1}{\sqrt{2\pi V(m_0)}} \exp\left\{-\frac{1}{2V(m_0)} z^2\right\}. \end{aligned}$$

The last “ $\approx$ ” holds because of the continuity of $V(x) = e^x$. $\square$

Proof of Theorem 1. First, we prove the consistency of $\hat{\beta}$, i.e., $\hat{\beta} \xrightarrow{P} \beta_0$ as $\varphi \rightarrow 0$, where β_0 is the true parameter. We shall consider the behavior of the log-likelihood $l(\beta)$ on the sphere Q_h with center at the true point β_0 and radius h . We will show that for any sufficiently small h , the probability of

$$l(\beta) < l(\beta_0)$$

tends to 1 at all points β on the surface of Q_h , i.e., $\|\beta - \beta_0\|_2 = h$ (cf., [17]). Note that this method also handles the proof of the MLE’s consistency in large-sample asymptotics.

By (13), we have

$$l(\beta) = l(m(\beta); \varphi, y) = \sum_{i=1}^n \ln f(y_i; y_i, \varphi) - \frac{1}{2}D^*(y, m).$$

Thus, we obtain

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^n \frac{\partial l}{\partial m_i} \frac{\partial m_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \varphi^{-1} (y_i - m_i) e^{-m_i} (-\eta_i^{-1}) x_{ij} \\ &= \sum_{i=1}^n \varphi^{-1} (y_i - m_i) e^{-m_i} e^{m_i} x_{ij} \\ &= \sum_{i=1}^n \varphi^{-1} (y_i - m_i) x_{ij}. \end{aligned}$$

This can be written in a matrix form as

$$\frac{\partial l}{\partial \beta} = \varphi^{-1} \mathbf{X}^\top (y - m). \quad (\text{A5})$$

Additionally, we have

$$\begin{aligned}\frac{\partial^2 l}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}^\top} \left(\frac{\partial l}{\partial \boldsymbol{\beta}} \right) = \frac{\partial}{\partial \mathbf{m}^\top} \left(\frac{\partial l}{\partial \boldsymbol{\beta}} \right) \frac{\partial \mathbf{m}}{\partial \boldsymbol{\beta}^\top} \\ &= \frac{\partial}{\partial \mathbf{m}^\top} \left(\frac{\partial \mathbf{m}^\top}{\partial \boldsymbol{\beta}} \frac{\partial l}{\partial \mathbf{m}} \right) \frac{\partial \mathbf{m}}{\partial \boldsymbol{\beta}^\top} \\ &= \frac{\partial \mathbf{m}^\top}{\partial \boldsymbol{\beta}} \frac{\partial^2 l}{\partial \mathbf{m}^\top \partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \boldsymbol{\beta}^\top}.\end{aligned}$$

We denote

$$\begin{aligned}\mathbf{m}_0 &= \mathbf{m}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}, \\ M(\boldsymbol{\beta}) &= \frac{\partial \mathbf{m}}{\partial \boldsymbol{\beta}^\top}, M(\boldsymbol{\beta}_0) = M(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}, \\ K(\mathbf{m}) &= \frac{\partial^2 l}{\partial \mathbf{m}^\top \partial \mathbf{m}}, K(\mathbf{m}_0) = K(\mathbf{m})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = K(\mathbf{m})|_{\mathbf{m}=\mathbf{m}_0}.\end{aligned}$$

Through differential operations, we obtain

$$K(\mathbf{m}) = \text{diag} \left\{ \varphi^{-1}(m_1 - y_1 - 1)e^{-m_1}, \dots, \varphi^{-1}(m_n - y_n - 1)e^{-m_n} \right\}.$$

Let

$$W(\mathbf{m}_0) = K(\mathbf{m}_0)|_{\mathbf{y}=\mathbf{m}_0} = \text{diag} \left\{ -\varphi^{-1} \exp(-m_1^0), \dots, -\varphi^{-1} \exp(-m_n^0) \right\},$$

where $\mathbf{m}_0 = (m_1^0, \dots, m_n^0)$. Obviously, $W(\mathbf{m}_0)$ is negative definite. By (16), we know $\mathbf{Y} \xrightarrow{P} \mathbf{m}_0$ as $\varphi \rightarrow 0$ (by use of the Chebyshev inequality); then

$$K(\mathbf{m}_0) \xrightarrow{P} W(\mathbf{m}_0). \quad (\text{A6})$$

For sufficiently small h , by expanding $l(\boldsymbol{\beta})$ around the true point $\boldsymbol{\beta}_0$ and multiplying by φ , we have

$$\begin{aligned}& \varphi[l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)] \\ &= \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \frac{\partial l}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \frac{1}{2} \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \frac{\partial^2 l}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(\varphi \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2) \\ &= \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \frac{\partial l}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \frac{1}{2} \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top M(\boldsymbol{\beta}_0)^\top K(\mathbf{m}_0) M(\boldsymbol{\beta}_0) (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(h^2) \\ &:= S_1 + S_2 + o(h^2).\end{aligned} \quad (\text{A7})$$

We now consider S_1 . Suppose that $\mathbf{X}\mathbf{X}^\top$ has bounded eigenvalues and its maximum eigenvalue is $\lambda_{\max}$. By (A5) and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}|S_1| &= |(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \mathbf{X}^\top (\mathbf{y} - \mathbf{m}_0)| \\ &\leq \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \|\mathbf{X}^\top (\mathbf{y} - \mathbf{m}_0)\|_2 \\ &= h \|\mathbf{X}^\top (\mathbf{y} - \mathbf{m}_0)\|_2 \\ &= h \left[(\mathbf{y} - \mathbf{m}_0)^\top \mathbf{X} \mathbf{X}^\top (\mathbf{y} - \mathbf{m}_0) \right]^{1/2} \\ &\leq h \lambda_{\max}^{1/2} \|\mathbf{y} - \mathbf{m}_0\|_2 \\ &\leq h \lambda_{\max}^{1/2} \|\mathbf{y} - \mathbf{m}_0\|_1.\end{aligned}$$

Consider $\|\mathbf{y} - \mathbf{m}_0\|_1$. For each $i \in \{1, 2, \dots, n\}$, by (16) and the Chebyshev inequality we obtain, by letting $\varphi \rightarrow 0$,

$$P\left(\varphi^{-1/2}|y_i - m_i^0| \geq \frac{h^2}{n}\right) \leq \frac{n^2 \exp(m_i^0)}{h^4};$$

that is,

$$P\left(|y_i - m_i^0| \geq \frac{h^2}{n}\right) \leq \frac{\varphi n^2 \exp(m_i^0)}{h^4},$$

implying that

$$\begin{aligned} P(\|\mathbf{y} - \mathbf{m}_0\|_1 \geq h^2) &= P\left(\sum_{i=1}^n |y_i - m_i^0| \geq h^2\right) \\ &\leq P\left(\bigcup_{i=1}^n \left\{|y_i - m_i^0| \geq \frac{h^2}{n}\right\}\right) \\ &\leq \sum_{i=1}^n P\left(|y_i - m_i^0| \geq \frac{h^2}{n}\right) \\ &\leq \frac{\varphi n^2}{h^4} \sum_{i=1}^n \exp(m_i^0). \end{aligned}$$

The last term of the above inequality tends to 0, since $\varphi \rightarrow 0$, and the other terms are constants. Thus, we have

$$P(\|\mathbf{y} - \mathbf{m}_0\|_1 < h^2) \geq 1 - \frac{\varphi n^2}{h^4} \sum_{i=1}^n \exp(m_i^0).$$

That is, $\|\mathbf{y} - \mathbf{m}_0\|_1 < h^2$ with probability tending to 1. Returning to $|S_1|$, we have, with probability tending to 1,

$$|S_1| \leq \lambda_{\max}^{1/2} h^3. \quad (\text{A8})$$

The above argument about $\|\mathbf{y} - \mathbf{m}_0\|_1$ is based on the result of convergence in the distribution of $\frac{\mathbf{Y} - \mathbf{m}_0}{\sqrt{\varphi}} \xrightarrow{d} N(\mathbf{0}_n, C)$. In fact, the same conclusion can be achieved by the convergence in probability of $\mathbf{Y} \xrightarrow{P} \mathbf{m}_0$. It is crucial to note that the argument above demonstrates how convergence in probability can be achieved through convergence in distribution when $\varphi \rightarrow 0$. This allows us to directly apply convergence in probability and omit any mention of convergence in distribution in terms of S_2 below.

We now consider S_2 . We have

$$\begin{aligned} 2S_2 &= \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top M(\boldsymbol{\beta}_0)^\top (K(\mathbf{m}_0) - W(\mathbf{m}_0))M(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &\quad + \varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top M(\boldsymbol{\beta}_0)^\top W(\mathbf{m}_0)M(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0). \end{aligned}$$

For the first term, we note that by (A6), $\varphi M(\boldsymbol{\beta}_0)^\top (K(\mathbf{m}_0) - W(\mathbf{m}_0))M(\boldsymbol{\beta}_0) \xrightarrow{P} \mathbf{0}$. We use an argument analogous to that used for S_1 but replace the Chebyshev inequality with the definition of convergence in probability. Thus, the absolute value of the first term is less than a constant multiple of h^4 with probability tending to 1. The second term is a negative quadratic form in $M(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Let $M(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = (a_1, a_2, \dots, a_n)^\top$. Then, by a straightforward calculation, we have

$$\varphi(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top M(\boldsymbol{\beta}_0)^\top W(\mathbf{m}_0)M(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = -\sum_{i=1}^n a_i^2 \exp(-m_i^0) = O(h^2).$$

Thus, $2S_2$ is negative and we obtain

$$|S_2| = O(h^2). \quad (\text{A9})$$

So, with (A7)–(A9), we have

$$l(\beta) - l(\beta_0) < 0 \quad (\text{A10})$$

for sufficiently small h .

Because $l(\beta)$ is continuous and differentiable on Q_h , there must be a local maximum point $\hat{\beta}$ that satisfies

$$\frac{\partial l}{\partial \beta} \Big|_{\beta=\hat{\beta}} = 0.$$

Combining this with (A10), we obtain

$$P(\|\hat{\beta} - \beta_0\| < h) \rightarrow 1.$$

So, when $\varphi \rightarrow 0$, we have

$$\hat{\beta} \xrightarrow{P} \beta_0.$$

Now, we will show that $\hat{\beta} - \beta_0 \xrightarrow{d} N(\mathbf{0}_{p+1}, S)$, where S is a covariance matrix. Denote

$$l'(\beta) = \frac{\partial l}{\partial \beta}.$$

By expanding $l'(\beta)$ around the true point β_0 , we obtain

$$l'(\beta) = l'(\beta_0) + \frac{\partial^2 l}{\partial \beta^\top \partial \beta} \Big|_{\beta=\beta_0} (\beta - \beta_0) + \cdots,$$

where higher-order terms are ignored. Replace β with $\hat{\beta}$ (we can do this since $\hat{\beta} \xrightarrow{P} \beta_0$), and then note that the left side of the equation is $\mathbf{0}_{p+1}$. Rearranging this equation, we have

$$l'(\beta_0) = -M(\beta_0)^\top K(\mathbf{m}_0)M(\beta_0)(\hat{\beta} - \beta_0). \quad (\text{A11})$$

Consider $l'(\beta_0)$ and $-M(\beta_0)^\top K(\mathbf{m}_0)M(\beta_0)$ separately. From (A5) and (16) we have

$$\sqrt{\varphi}l'(\beta_0) \xrightarrow{d} N(\mathbf{0}_{p+1}, \mathbf{X}^\top C\mathbf{X}). \quad (\text{A12})$$

So, by a simple calculation, we find that

$$M(\beta_0) = \begin{bmatrix} e^{m_1^0} & e^{m_1^0}x_{11} & \cdots & e^{m_1^0}x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ e^{m_n^0} & e^{m_n^0}x_{n1} & \cdots & e^{m_n^0}x_{np} \end{bmatrix} = C\mathbf{X}.$$

Thus, we obtain

$$-M(\beta_0)^\top W(\mathbf{m}_0)M(\beta_0) = \mathbf{X}^\top C^\top W(\mathbf{m}_0)C\mathbf{X} = \varphi^{-1}\mathbf{X}^\top C\mathbf{X}.$$

By (A6), we obtain

$$-\varphi M(\beta_0)^\top K(\mathbf{m}_0)M(\beta_0) \xrightarrow{P} \mathbf{X}^\top C\mathbf{X}. \quad (\text{A13})$$

By (A11)–(A13), we obtain

$$\frac{\hat{\beta} - \beta_0}{\sqrt{\varphi}} \xrightarrow{d} N(\mathbf{0}_{p+1}, (\mathbf{X}^\top C\mathbf{X})^{-1}).$$

This completes the proof of Theorem 1. $\square$

Proof of Theorem 2. First, we show the unit deviance follows an approximate χ_1^2 . By Proposition 1, the moment-generating function (MGF) of the unit deviance is approximately

$$\begin{aligned} M_{d(y,m)}(t) &= E[\exp\{d(y,m)t\}] \\ &\approx \int_{-\infty}^{\infty} \exp\{d(y,m)t\} \frac{1}{\sqrt{2\pi\varphi V(y)}} \exp\left\{-\frac{1}{2\varphi}d(y,m)\right\} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\varphi V(y)}} \exp\left\{-\frac{1-2\varphi t}{2\varphi}d(y,m)\right\} dy \\ &= (1-2\varphi t)^{-1/2} \int_{-\infty}^{\infty} \frac{(1-2\varphi t)^{1/2}}{[2\pi\varphi V(y)]^{1/2}} \exp\left\{-\frac{1-2\varphi t}{2\varphi}d(y,m)\right\} dy \\ &= (1-2\varphi t)^{-1/2} \int_{-\infty}^{\infty} \frac{1}{[2\pi\varphi' V(y)]^{1/2}} \exp\left\{-\frac{1}{2\varphi'}d(y,m)\right\} dy, \end{aligned}$$

where $\varphi' = \varphi/(1-2\varphi t)$. Since the integrand is the (saddlepoint) density of the distribution with $\varphi' = \varphi/(1-2\varphi t)$, we have

$$M_{d(y,m)}(t) = (1-2\varphi t)^{-1/2}, \text{ i.e., } M_{d(y,m)/\varphi}(t) = (1-2t)^{-1/2},$$

which is identical to the MGF of a χ_1^2 . So, as $\varphi \rightarrow 0$ we have

$$d(y,m)/\varphi \xrightarrow{d} \chi_1^2.$$

For the set of observations $\mathbf{y} = (y_1, \dots, y_n)^\top$, where the y_i 's are independent, with $y_i \sim \text{TBE}_\infty$, we have

$$d(y_i, m_i)/\varphi \xrightarrow{d} \chi_1^2, \text{ as } \varphi \rightarrow 0.$$

Then, by independence, we obtain

$$D^*(\mathbf{y}, \mathbf{m}) \xrightarrow{d} \chi_n^2, \text{ as } \varphi \rightarrow 0,$$

which completes the proof of Theorem 2 (cf., [6]). $\square$

Proof of Theorem 3. We consider the four nested hypotheses (cf., [17]):

$H_0 : \mathbf{m} \in M$ (the saturated hypothesis);

$H_1 : \mathbf{m} = \mathbf{m}(\beta_0, \beta_1, \dots, \beta_p)$;

$H_2 : \mathbf{m} = \mathbf{m}(\beta_0, \beta_1, \dots, \beta_{p_B})$;

$H_3 : \mathbf{m} = \mathbf{m}(\beta_0, \beta_1, \dots, \beta_{p_A})$

of dimensions $n, p+1, p_B+1, p_A+1$, respectively, where $n > p+1 \geq p_B+1 > p_A+1$.

Since we proved the asymptotic normality of $\hat{\beta}$ in Theorem 1, just as in Theorems 10.3.1 and 10.3.3 of [21], we can prove that the likelihood ratio test (LRT) statistic follows asymptotically a chi-square distribution by starting from the simple hypothesis and moving on to the composite hypothesis. That is, for LRT $\lambda(Y)$, we have $-2 \ln \lambda(Y) \sim \chi_q^2$, where q is the corresponding degrees of freedom.

For testing H_0 vs. H_1 (H_1 is the null hypothesis), we ascertain that (18) holds. For testing H_2 vs. H_3 (H_3 is the null hypothesis), we ascertain that (20) holds. Note that their degrees of freedom are the difference in their numbers of dimensions. $\square$

References

1. Bar-Lev, S.K. Independent tough identical results: The Tweedie class on power variance functions and the class of Bar-Lev and Enis on reproducible natural exponential families. *Int. J. Stat. Probab.* **2020**, *9*, 30–35. [CrossRef]
2. Jørgensen, B. *The Theory of Dispersion Models*; Chapman and Hall: London, UK, 1997.
3. Bar-Lev, S.K. The Exponential Dispersion Model Generated by the Landau Distribution—A Comprehensive Review and Further Developments. *Mathematics* **2023**, *11*, 4343. [CrossRef]
4. Dunn, P.K.; Smyth, G.K. Tweedie Family Densities: Methods of Evaluation. In Proceedings of the 16th International Workshop on Statistical Modelling, Odense, Denmark, 2–6 July 2001.

5. Dunn, P.K.; Smyth, G.K. Series Evaluation of Tweedie Exponential Dispersion Model Densities. *Stat. Comput.* **2005**, *15*, 267–280. [CrossRef]
6. Dunn, P.K.; Smyth, G.K. *Generalized Linear Models with Examples in R*; Springer: Berlin/Heidelberg, Germany, 2018.
7. Hougaard, P. Nonlinear Regression and Curved Exponential Families. Improvement of the Approximation to the Asymptotic Distribution. *Metrika* **1995**, *42*, 191–202. [CrossRef]
8. Chen, Z.; Pan, E.; Xia, T.; Li, Y. Optimal degradation-based burn-in policy using Tweedie exponential-dispersion process model with measurement errors. *Reliab. Syst. Saf.* **2020**, *195*, 106748. [CrossRef]
9. Ricci, L.; Martínez, R. Adjusted R^2 -type measures for Tweedie models. *Comput. Stat. Data Anal.* **2008**, *52*, 1650–1660. [CrossRef]
10. Dunn, P.K. *Tweedie: Evaluation of Tweedie Exponential Family Models*; R Package Version 2.3.5. 2022. Available online: <https://cran.r-project.org/web/packages/tweedie/tweedie.pdf> (accessed on 12 September 2023).
11. Smyth, G.K. *Statmod: Statistical Modeling*; R Package Version 1.4.30. 2017. Available online: <https://CRAN.R-project.org/package=statmod> (accessed on 3 April 2024).
12. Barndorff-Nielsen, O. *Information and Exponential Families in Statistical Theory*; Wiley: New York, NY, USA, 1978.
13. Merz, M.; Wüthrich, M.V. *Statistical Foundations of Actuarial Learning and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2023.
14. Morris, C.N. Natural exponential families with quadratic variance functions. *Ann. Statist.* **1982**, *10*, 65–80. [CrossRef]
15. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 1989.
16. Dunn, P.K.; Smyth, G.K. Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Stat. Comput.* **2008**, *18*, 73–86. [CrossRef]
17. Jørgensen, B. Small dispersion asymptotics. *Braz. J. Probab. Stat.* **1987**, *1*, 59–90.
18. Liu, X.; Xiang, Z.; Bar-Lev, S.K.; Ridder, A. *TBEinf*; R Package Version 0.0.1. 2024. Available online: <https://github.com/xliusufe/TBEinf> (accessed on 28 April 2024).
19. Sidi, A. A user-friendly extrapolation method for oscillatory infinite integrals. *Math. Comput.* **1988**, *51*, 249–266. [CrossRef]
20. Dunn, P.K.; Smyth, G.K. *GLMsData: Generalized Linear Model Data Sets*; R Package Version 1.0.0. 2017. Available online: <https://CRAN.R-project.org/package=GLMsData> (accessed on 12 April 2024).
21. Casella, G.; Berger, R.L. *Statistical Inference*; Thomson Learning Inc.: Duxbury, MA, USA, 2002.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.