ORIGINAL PAPER

Two-parameter link functions, with applications to negative binomial, Weibull and quantile regression

V. F. Miranda-Soberanis[1] 🔾 · **Thomas W. Yee**[2]

© The Author(s) 2022

Abstract

One-parameter link functions play a fundamental role in regression via generalized linear modelling. This paper develops the general theory for *two*-parameter links in the very large class of vector generalized linear models by using total derivatives applied to a composite log-likelihood within the Fisher scoring/iteratively reweighted least squares algorithm. We solve a four-decade old problem with an interesting history as our first example: the canonical link for negative binomial regression. The remaining examples are fitting Weibull regression using both the mean and quantile directly compared to GAMLSS, and performing quantile regression based on the Gaussian distribution. Numerical examples based on real and simulated data are given. The methods described here are implemented by the VGAM and VGAMextra R packages, available on CRAN. Supplementary materials for this article are available online.

1 Introduction

For 50 years now generalized linear models (GLMs; Nelder and Wedderburn 1972) have been the foundational building block of generalized regression. Its central formula is

---

✉ V. F. Miranda-Soberanis
victor.miranda@aut.ac.nz

1 Department of Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

2 Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

$$\mathcal{G}(\mu; \boldsymbol{x}) \;=\; \eta \;=\; \boldsymbol{\beta}^T \boldsymbol{x}, \tag{1.1}$$

with $\mu = \mathbb{E}(Y)$ as the mean response and link function $\mathcal{G}$ having strict monotonicity and being twice-differentiable in the range of μ. In its original formulation (1.1) was restricted to Y belonging to the *one*-parameter exponential family (see, e.g., Efron 1978), and Fisher scoring/iteratively reweighted least squares (IRLS; see, e.g., Green 1984) has become the standard GLM algorithm. The purpose of this paper is to develop its natural extension to *two* parameters θ_1 and θ_2, written as

$$\mathcal{G}_j(\theta_1, \theta_2; \boldsymbol{x}) \;=\; \eta_j \;=\; \boldsymbol{\beta}_j^T \boldsymbol{x}, \qquad j = 1, 2 \, (= M), \tag{1.2}$$

to also be solved by IRLS. Here, θ_j are general parameters and not necessarily a mean. The core problem is to derive the working weight matrices associated with this. It will be seen that total derivatives with chain rule expressions hold the key to handling multi-parameters. In (1.2) we say the parameters are intertwined in a *composite* linear predictor.

The most celebrated example of (1.2) is the canonical link of the negative binomial distribution (NBD),

$$\mathcal{G}_1(\mu, k; \boldsymbol{x}) \;=\; \log\left(\mu/(\mu + k)\right), \tag{1.3}$$

where $Y \sim \mathrm{NB}(\mu, k)$ with variance $\mu(1 + \mu/k)$, and $\mathcal{G}_2(\mu, k; \boldsymbol{x}) = \log k$. This model is often abbreviated NB-C by some authors, e.g., Hilbe (2011), and it can be treated as an ordinary GLM if k is known. Prior to this work, estimation of (1.3) was deemed too difficult by some or it was 'solved' defectively by others; we recount some of its history in Sect. 4.1 and solve this four-decade old problem as our first application.

Given a suitable computational framework, (1.1)–(1.3) can also be used for GLM-like extensions such as the class of generalized additive models (GAMs; Hastie and Tibshirani 1986, 1990; Wood 2017) and Bayesian additive models for location scale and shape (Umlauf et al. 2018). This paper is motivated by the work of Miranda-Soberanis and Yee (2019) on one-parameter quantile link functions and uses the vector generalized linear and additive model classes (VGLMS/VGAMs; Yee 2015) as the computational and statistical framework. VGLMs can be loosely thought of as multivariate GLMs beyond the exponential family and this allows diversification to different data types such as categorical data (Yee 2010), extremes (Yee and Stephenson 2007), quantile regression (Yee 2004) and time series analysis (Miranda-Soberanis 2018). VGLMs/VGAMs handle M linear predictors η_j so that the present work is focussed on $M = 2$. Our work should not be confused with the composite link functions of Thompson and Baker (1981) who extended GLMs by associating more than one η_j to each mean value μ_i ($i = 1, \ldots, n$). Our approach works in the opposite direction by associating $M = 2$ parameters to each η_j with both sides being *i*-specific. The linear/additive predictors thus involve two interlaced parameters so that the likelihood that now can be handled is what might be described as a composite likelihood.

This work confers several immediate benefits. First, as (1.2) allows $\boldsymbol{x}$ to affect both parameters directly, it is much more flexible than (1.1) operating twice separately. Consequently, a second benefit is that we can now fit models previously unimplemented

such as the NBD with its canonical link. Indeed, the methodology can be applied to any function of $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ with a tractable inverse, as defined in Sect. 3.1. In this paper we present three diverse applications, e.g., the mean (μ_W) and the quantile functions of the two-parameter Weibull distribution, and the normal distribution's quantile function. The *new* link functions have the form $\eta = \mathcal{G}(\mu_W(\theta_1, \theta_2))$ or $\eta = \mathcal{G}(\tau^{\text{th}}(\theta_1, \theta_2) \text{ quantile})$. However, our approach is broader still. In theory, one can directly model, say, the variance as an alternative to study a model's underlying homoscedasticity assumptions. A third benefit is that the methodology can be readily extended to *three*-parameter VGLMs, and theoretically to *M*-parameter link functions, and so introducing more flexibility by handling statistical measures of distributions having location, scale and shape parameters (or LSS–links). The basic motivation for using LSS–links rather than simple one–parameter links is that the stochastic relationship between predictors and response can be modelled much better and with greater accuracy. Thus this work points towards $\eta_j = \mathcal{G}_j(\boldsymbol{\theta}; \boldsymbol{x})$, $j = 1, \ldots, M$, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)^T$, as being the ultimate set of link functions.

An outline of this paper is as follows. Sect. 2 summarizes skeletal details of VGLMs needed here. Sect. 3 gives the general theory for two-parameter links. Sect. 4 solves the NBD canonical link problem and proposes two-parameter Weibull mean and quantile regression as well as Gaussian quantile regression. Numerical examples involving real and simulated data including a comparison to generalized additive models for location, scale and shape (GAMLSS) appear in Sect. 5. The paper ends with a discussion. An R script file and online appendices are available as supplementary material.

Notationally, we use '∂' for partial derivatives and 'd' for total derivatives (see, e.g., Loomis and Sternberg 1990). The digamma and trigamma functions are denoted by ψ and ψ'. The Hadamard (element-by-element) and Kronecker products of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are $\boldsymbol{A} \circ \boldsymbol{B} = [(a_{js} \cdot b_{js})]$ and $\boldsymbol{A} \otimes \boldsymbol{B} = [(a_{js} \cdot \boldsymbol{B})]$ respectively.

2 VGLM review

VGLMs operate on data $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, independently with $\boldsymbol{y}_i$ a Q-dimensional response and covariates $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})^T$ with $x_{i1} = 1$ if there is an intercept. A VGLM is defined in terms of M linear predictors as a model where the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ has the form (dropping the subscript i for simplicity)

$$\mathcal{F}(\boldsymbol{y}|\boldsymbol{x}; \boldsymbol{B}) = \mathcal{F}(\boldsymbol{y}; \eta_1, \ldots, \eta_M) \tag{2.1}$$

for some function $\mathcal{F}(\cdot)$, where $\boldsymbol{B}$ is $d \times M$ with $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{x}) =$

$$\begin{pmatrix} \eta_1(\boldsymbol{x}) \\ \vdots \\ \eta_M(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1^T \boldsymbol{x} \\ \vdots \\ \boldsymbol{\beta}_M^T \boldsymbol{x} \end{pmatrix} = \sum_{k=1}^{d} \boldsymbol{\beta}_{(k)} x_k = \boldsymbol{B}^T \boldsymbol{x}.$$

In the usual case the jth linear predictor is $\eta_j(\boldsymbol{x}) = \boldsymbol{\beta}_j^T \boldsymbol{x}$ and can be tied in to the parameters θ_j of any distribution as $\mathcal{G}_j(\theta_j) = \eta_j = \boldsymbol{\beta}_j^T \boldsymbol{x}$, $j = 1, \ldots, M$.

Fitting VGLMs produce full maximum likelihood estimates (MLEs). The 'general' VGLM log-likelihood is

$$\ell(\boldsymbol{\eta};\boldsymbol{x}) = \sum_{i=1}^{n} w_i \, \ell_i \big\{ \eta_1(\boldsymbol{x}_i), \dots, \eta_M(\boldsymbol{x}_i) \big\}, \tag{2.2}$$

for known fixed positive prior weights w_i. A Newton-like algorithm for maximizing (2.2) has the form $\boldsymbol{\beta}^{(a)} = \boldsymbol{\beta}^{(a-1)} + \mathcal{I}\big(\boldsymbol{\beta}^{(a-1)}\big)^{-1} \boldsymbol{U}\big(\boldsymbol{\beta}^{(a-1)}\big)$ at iteration a . For VGLMs, the vector of coefficients $\boldsymbol{\beta}^{(a)}$ is obtained by iteratively regressing the *working* responses $\boldsymbol{z}^{(a-1)}$ on the 'big' model matrix $\boldsymbol{X}_{\mathrm{VLM}} = \big(\boldsymbol{X}_1^T, \dots, \boldsymbol{X}_n^T\big)^T$, which has the form

$$\boldsymbol{z}^{(a-1)} = \boldsymbol{X}_{\mathrm{VLM}} \, \boldsymbol{\beta}^{(a)} + \boldsymbol{W}^{-1 \, (a-1)} \, \boldsymbol{U}^{(a-1)}, \tag{2.3}$$

where $\boldsymbol{X}_{\mathrm{VLM}} = \boldsymbol{X}_{\mathrm{LM}} \otimes \boldsymbol{I}_M$, known as the 'large' model matrix, has dimension $nM \times Md$, and $\boldsymbol{X}_i$ is the $M \times (Md)$ block-matrix $\boldsymbol{X}_i = \boldsymbol{x}_i \otimes \boldsymbol{I}_M$.

In (2.3), $\boldsymbol{U} = \big(\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\big)^T$ comprises the individual *score vector* $\boldsymbol{u}_i$ whose jth element is $\big(\boldsymbol{u}_i\big)_j = \dfrac{\partial \ell_i}{\partial \eta_j}$ and $\boldsymbol{W}$ is the matrix $\boldsymbol{W} = \mathrm{diag}\big(w_1 \boldsymbol{W}_1^{-1}, \dots, w_n \boldsymbol{W}_n^{-1}\big)$. The $\boldsymbol{W}_i$s are the 'general' $M \times M$ *working weight matrices*, with (j, k)th element $\big[\boldsymbol{W}_i\big]_{j,k} = -w_i \, \mathbb{E}\left(\dfrac{\partial^2 \ell_i}{\partial \eta_j \partial \eta_k}\right)$.

At convergence, the estimated variance-covariance matrix is $\widehat{\mathrm{Var}}\big(\widehat{\boldsymbol{\beta}^*}\big) = \widehat{\phi}\big(\boldsymbol{X}_{\mathrm{VLM}}^T \boldsymbol{W}^{(a)} \boldsymbol{X}_{\mathrm{VLM}}\big)^{-1}$. Using the observed information matrix (OIM) corresponds to the Newton-Raphson algorithm. Fisher scoring is primarily used within the VGLM framework over Newton-Raphson because the expected information matrices (EIMs) are positive-definite over a larger portion of the parameter space.

With multiple η_j one can enforce linear relationships between them via

$$\boldsymbol{\eta}(\boldsymbol{x}_i) = \sum_{k=1}^{d} \boldsymbol{\beta}_{(k)} \, x_{ik} = \sum_{k=1}^{d} \boldsymbol{H}_k \, \boldsymbol{\beta}_{(k)}^* \, x_{ik}, \tag{2.4}$$

for known *constraint* matrices $\boldsymbol{H}_k$ of full column-rank (i.e., rank $\boldsymbol{H}_k = \mathcal{R}_k = \mathtt{ncol}(\boldsymbol{H}_k)$), and $\boldsymbol{\beta}_{(k)}^*$ is a possibly reduced set of regression coefficients to be estimated. The matrices $\boldsymbol{H}_k$ can constrain the effect of a covariate over some η_j and to have no effect for others. Trivial constraints are attained with $\boldsymbol{H}_k = \boldsymbol{I}_M$ for all k, where $\boldsymbol{X}_{\mathrm{VLM}} = \boldsymbol{X}_{\mathrm{LM}} \otimes \boldsymbol{I}_M$ is preserved. Other common examples include parallelism ($\boldsymbol{H}_k = \boldsymbol{1}_M$), exchangeability, and *intercept-only* parameters $\eta_j = \beta_{(j)1}^*$. For a general review on VGAMs see Yee (2015, Sec. 1.3.2).

3 General theory for two-parameter links

3.1 Problem formulation

Given the preceding background, our attention focusses on $M = \dim(\boldsymbol{\Theta}) = 2$ transformations of $\boldsymbol{\Theta} = \{(\theta_1, \theta_2)^T ; \theta_j \in A_j \subset \mathbb{R}\}$ of the form

$$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1, \theta_2), \qquad \eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_1, \theta_2), \tag{3.1}$$

for known bivariate functions $\mathcal{G}_k : \boldsymbol{\Theta} \to B \subseteq \mathbb{R}$, $k = 1, \dots M$. Here, B, and $A_j, j = 1, 2$ are open intervals. Consequently, the VGLM log-likelihood (2.2) reduces to

$$\ell(\boldsymbol{\eta}; \boldsymbol{x}) = \ell\left(\eta_{\mathcal{G}_1}, \eta_{\mathcal{G}_2}; \boldsymbol{x}\right) = \sum_{i=1}^{n} w_i \, \ell_i\left(\eta_{\mathcal{G}_1}(\boldsymbol{x}_i), \eta_{\mathcal{G}_2}(\boldsymbol{x}_i)\right)$$

where $\boldsymbol{\eta} = \left(\eta_{\mathcal{G}_1}, \eta_{\mathcal{G}_2}\right)^T$.

Equation (3.1) is central to this article and gives rise to the three cases

$$\eta_{\mathcal{G}_1}(\boldsymbol{x}) = \mathcal{G}_1(\theta_1, \theta_2), \qquad \eta_{\mathcal{G}_2}(\boldsymbol{x}) = \mathcal{G}_2(\theta_2); \tag{3.2}$$

$$\eta_{\mathcal{G}_1}(\boldsymbol{x}) = \mathcal{G}_1(\theta_1), \qquad \eta_{\mathcal{G}_2}(\boldsymbol{x}) = \mathcal{G}_2(\theta_1, \theta_2); \tag{3.3}$$

$$\eta_{\mathcal{G}_1}(\boldsymbol{x}) = \mathcal{G}_1(\theta_1, \theta_2), \qquad \eta_{\mathcal{G}_2}(\boldsymbol{x}) = \mathcal{G}_2(\theta_1, \theta_2). \tag{3.4}$$

We assume that $\eta_{\mathcal{G}_j}$ can be solved at least iteratively for θ_j (cf. (3.9)), allowing (3.2)–(3.4) to be expressed as

$$\theta_1 = \theta_1\left(\eta_{\mathcal{G}_1}, \theta_2\right), \qquad \theta_2 = \theta_2\left(\eta_{\mathcal{G}_2}\right); \tag{3.5}$$

$$\theta_1 = \theta_1\left(\eta_{\mathcal{G}_1}\right), \qquad \theta_2 = \theta_2\left(\theta_1, \eta_{\mathcal{G}_2}\right); \quad \text{and} \tag{3.6}$$

$$\theta_1 = \theta_1\left(\eta_{\mathcal{G}_1}, \theta_2\right), \qquad \theta_2 = \theta_2\left(\theta_1, \eta_{\mathcal{G}_2}\right). \tag{3.7}$$

The interdependency between the θ_j enforces a reparametrized log-likelihood. For example, for the general case (3.7) above we have

$$\ell(\boldsymbol{\theta}; \boldsymbol{x}) = \ell\left(\theta_1\left(\eta_{\mathcal{G}_1}, \theta_2\right), \theta_2\left(\theta_1, \eta_{\mathcal{G}_2}\right); \boldsymbol{x}\right). \tag{3.8}$$

We call (3.8) the *modified* VGLM log-likelihood as it reflects the composite structure of two-parameter linear predictors. Fisher scoring must be consequently adjusted.

Before detailing its solution, we list the assumptions made over $\mathcal{G}_k$ in (3.1) required.

Assumptions

(i)   For each $\mathcal{G}_j$ in (3.1) the derivatives $\partial\mathcal{G}_j/\partial\theta_k$, $\partial^2\mathcal{G}_j/\partial\theta_k^2$ and $\partial^2\mathcal{G}_j/\partial\theta_l\partial\theta_k$ exist over either A_j or $\boldsymbol{\Theta}$ accordingly.

(ii) For each equation in (3.1), $\eta_{\mathcal{G}_j}$ can be solved for θ_j, in the form

$$\theta_1 = \theta_1\big(\eta_{\mathcal{G}_1},\theta_2\big) = \mathcal{G}_1^*(\eta_{\mathcal{G}_1},\theta_2) \text{ and } \theta_2 = \theta_2\big(\theta_1,\eta_{\mathcal{G}_2}\big) = \mathcal{G}_2^*\big(\theta_1,\eta_{\mathcal{G}_2}\big), \tag{3.9}$$

for known functions $\mathcal{G}_j^*$, $j = 1, 2$.

(iii) For every $\boldsymbol{\eta}^0 = \Big(\eta_{\mathcal{G}_1}^0, \eta_{\mathcal{G}_2}^0\Big)^T \in B \times B$, the system (3.1) has unique solution, denoted $\boldsymbol{\theta}_{\mathcal{G}} = \Big(\theta_{\mathcal{G}}^1, \theta_{\mathcal{G}}^2\Big)^T$, given by

$$\theta_{\mathcal{G}}^1 = \theta_1\Big(\eta_{\mathcal{G}_1}^0, \theta_2\Big) = \mathcal{G}_1^*(\eta_{\mathcal{G}_1}^0, \theta_2) \text{ and } \theta_{\mathcal{G}}^2 = \theta_2\Big(\theta_1, \eta_{\mathcal{G}_2}^0\Big) = \mathcal{G}_2^*\Big(\theta_1, \eta_{\mathcal{G}_2}^0\Big). \tag{3.10}$$

(iv) From (3.9),

(a) the derivatives

$$\frac{\partial\theta_j}{\partial\eta_{\mathcal{G}_j}} \text{ and } \frac{\partial^2\theta_j}{\partial\eta_{\mathcal{G}_j}^2}, \quad j = 1, 2, \tag{3.11}$$

exist on B, and

(b) the derivatives

$$\frac{\partial\theta_j}{\partial\theta_k}, \qquad j, k = 1, 2, \ j \neq k, \tag{3.12}$$

exist over A_k.

When $\mathcal{G}_j^*, j = 1, 2$, in (3.1) are not tractable then $\mathcal{G}_j^*$ and $\boldsymbol{\theta}_{\mathcal{G}}$ in (3.10) can be computed via iterative methods. Optionally, *implicit* differentiation applied to (3.1) can be used to obtain (3.11) and (3.12) if (3.1) cannot be explicitly solved for θ_j. In particular, $\partial^2\theta_j/(\partial\eta_{\mathcal{G}_1}\partial\eta_{\mathcal{G}_2}) = 0$.

3.2 Solution

Expressions (3.5)–(3.7). shed light on the interdependence between θ_j and θ_k and its pivotal role when computing

$$\frac{\partial\ell}{\partial\theta_j}, \frac{\partial^2\ell}{\partial\theta_j^2}, \text{ and } \frac{\partial^2\ell}{\partial\theta_j\,\partial\theta_k}, \qquad j, k = 1, 2, \text{ for } j \neq k, \tag{3.13}$$

from the modified log-likelihood (3.8). Note that

$$\frac{\partial \theta_j}{\partial \theta_k} \neq 0. \tag{3.14}$$

The solution to incorporating two-parameter linear predictors to the VGLM framework relies on computing (3.13) by appropriately embedding (3.8) and (3.14). Total derivatives are necessary since θ_1 and θ_2 now vary simultaneously, and the solution produces a new set of expressions for the score vector and working weight matrices.

For the score vector, each component is a total derivative $\mathrm{d}\ell / \mathrm{d}\theta_j$ given by

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta_1} = \left(1, \frac{\partial \theta_2}{\partial \theta_1}\right) \cdot \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \end{pmatrix} = \frac{\partial \ell}{\partial \theta_1} + \frac{\partial \ell}{\partial \theta_2} \cdot \frac{\partial \theta_2}{\partial \theta_1}, \quad \text{and}$$

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta_2} = \left(\frac{\partial \theta_1}{\partial \theta_2}, 1\right) \cdot \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \end{pmatrix} = \frac{\partial \ell}{\partial \theta_1} \cdot \frac{\partial \theta_1}{\partial \theta_2} + \frac{\partial \ell}{\partial \theta_2}.$$

which apply to VGLMs with the two linear predictors as in case (3.4). For the special cases (3.2) and (3.3), at least one linear predictor is univariate, that is, a function of either θ_1 or θ_2 (but not both), hence $\partial \theta_2 / \partial \theta_1 = 0$ or $\partial \theta_1 / \partial \theta_2 = 0$. Table 9 in Appendix A shows the expressions for the three cases (3.2)–(3.4).

Likewise, the EIMs for VGLMs with $M = 2$ linear predictors $\boldsymbol{\eta}$ is given by $\mathcal{I}_E(\boldsymbol{\eta}) = -\mathbb{E}\left[\partial^2 \ell / (\partial \boldsymbol{\eta} \, \partial \boldsymbol{\eta}^T)\right]$ where

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\eta} \, \partial \boldsymbol{\eta}^T} = \begin{pmatrix} \dfrac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_1}^2} & \dfrac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_1} \partial \eta_{\mathcal{G}_2}} \\ \dfrac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_2} \partial \eta_{\mathcal{G}_1}} & \dfrac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_2}^2} \end{pmatrix}.$$

Applying the chain rule with the inclusion of total derivatives,

$$\frac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_j}^2} = \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_1^2} \cdot \left(\frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_j}}\right)^2 + 2 \cdot \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_1 \mathrm{d}\theta_2} \cdot \frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_j}} \cdot \frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_j}}$$

$$+ \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_2^2} \cdot \left(\frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_j}}\right)^2 + \left\{ \frac{\mathrm{d}\ell}{\mathrm{d}\theta_1} \cdot \frac{\partial^2 \theta_1}{\partial \eta_{\mathcal{G}_j}^2} + \frac{\mathrm{d}\ell}{\mathrm{d}\theta_2} \cdot \frac{\partial^2 \theta_2}{\partial \eta_{\mathcal{G}_j}^2} \right\}, \quad \text{for } j = 1, 2 \text{ and}$$

$$\tag{3.15}$$

$$\frac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_j} \partial \eta_{\mathcal{G}_k}} = \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_1^2} \cdot \frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_1}} \cdot \frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_2}}$$

$$+ \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_1 \mathrm{d}\theta_2} \cdot \left(\frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_1}} \cdot \frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_2}} + \frac{\partial \theta_1}{\partial \eta_{\mathcal{G}_2}} \cdot \frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_1}} \right)$$

$$+ \frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_2^2} \cdot \frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_1}} \cdot \frac{\partial \theta_2}{\partial \eta_{\mathcal{G}_2}} + \left\{ \frac{\mathrm{d}\ell}{\mathrm{d}\theta_1} \cdot \frac{\partial^2 \theta_1}{\partial \eta_{\mathcal{G}_1} \partial \eta_{\mathcal{G}_2}} \cdot \frac{\mathrm{d}\ell}{\mathrm{d}\theta_2} \cdot \frac{\partial^2 \theta_2}{\partial \eta_{\mathcal{G}_1} \partial \eta_{\mathcal{G}_2}} \right\}.$$

$$(3.16)$$

Under regularity conditions, the terms in braces vanish after taking the expectation, giving place to the new expressions for the EIMs:

$$-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_j}^2} \right] = -\mathbb{E}\left[\frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_j^2} \right] \left(\frac{\partial \theta_j}{\partial \eta_{\mathcal{G}_j}} \right)^2,$$

$$-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \eta_{\mathcal{G}_j} \partial \eta_{\mathcal{G}_k}} \right] = -\mathbb{E}\left[\frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_j \mathrm{d}\theta_k} \right] \left(\frac{\partial \theta_j}{\partial \eta_{\mathcal{G}_j}} \right) \left(\frac{\partial \theta_k}{\partial \eta_{\mathcal{G}_k}} \right), \quad j, k = 1, \dots, M.$$

$$(3.17)$$

Again, new complementary expressions for $\mathrm{d}^2 \ell / \mathrm{d}\theta_j^2$ and $\mathrm{d}^2 \ell / \mathrm{d}\theta_j \mathrm{d}\theta_k$, $j, k = 1, 2$, are required. For the the general case (3.4), where each linear predictor depends on θ_1 and θ_2, these are

$$\frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_j^2} = \frac{\partial^2 \ell}{\partial \theta_j^2} + 2 \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_j} \cdot \frac{\partial \theta_k}{\partial \theta_j} + \frac{\partial \ell}{\partial \theta_k} \cdot \frac{\partial^2 \theta_k}{\partial \theta_j^2} + \frac{\partial^2 \ell}{\partial \theta_k^2} \cdot \left(\frac{\partial \theta_k}{\partial \theta_j} \right)^2 \quad \text{and}$$

$$\frac{\mathrm{d}^2 \ell}{\mathrm{d}\theta_j \mathrm{d}\theta_k} = \frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} + \frac{\partial^2 \ell}{\partial \theta_j^2} \cdot \frac{\partial \theta_j}{\partial \theta_k} + \frac{\partial^2 \ell}{\partial \theta_k^2} \cdot \frac{\partial \theta_k}{\partial \theta_j},$$

respectively. Tables 10 and 11 outline the expressions for the three cases (3.2)–(3.4).

4 Some applications

We present direct applications of the previous section, tying it in with software implementations both new and old. The first shows that the NB-C can be straightforwardly estimated as a VGLM by adjusting the score vector and working weight matrices of the composite log-likelihood. The second considers two variants of Weibull regression: the mean parameterization coincides with WEI3() in gamlss. dist whereas the quantile parameterization is novel. The third sketches the details for directly fitting the two-parameter quantile function of the normal distribution, as implemented by a new link function normalQlink() from VGAMextra 0.0-5.

4.1 Negative binomial canonical link

The estimation of the NB-C has a somewhat interesting history. If we take the problem as beginning with the admission of McCullagh and Nelder (1983, p.195) [see also McCullagh and Nelder (1989, p.374)] that the NB is little used in applications and has a "problematical" canonical link then to our knowledge only one other publicized attempt has been made since to solve the problem seriously. It eluded McCullagh and Nelder because the model makes η a function of a parameter of the variance. Hilbe (2011, pp. 210,309,315–6) sheds more light on the problem as well as proposing an adhoc method that tends to work in most cases. He writes "In discussing this statement with Nelder, I found that the foremost problem they had in mind was the difficulty they experienced in attempting to estimate the model. They do not state it in their text, but the problem *largely disappears* when k is entered into the GLM estimating algorithm as a constant" [italics added]. He notes that having k in the link and variance can result in estimation difficulties in a full MLE algorithm, with it being sensitive to initial values and having tedious convergence with Newton-Raphson-type algorithms. To our knowledge his R package COUNT function `ml.nbc()` is the only other NB-C MLE-implementation. It treats k as constant at each iteration but it is iteratively estimated in the process. It also treats k as an *additional* scalar parameter to be estimated, hence is constrained to be intercept-only. A general optimizer `stats::optim()` is invoked and it has by default some prechosen constants for initial values ($k^{(0)} = 2$ and $\mu^{(0)} = e^{-1} \approx 0.368$) so that it will be unreliable for large μ and k. In contrast, we believe our solution to be nondefective. Prior to this work, the NB-C had also been unsatisfactorily implemented in VGAM 1.0-3 or earlier (see, e.g., Yee 2015, Sec.11.3.4) but has since been corrected in functions `negbinomial()` and `nbcanlink()`—more details are given in Miranda-Soberanis (2018).

Ordinarily, NB regression operates with

$$\eta_{\mathcal{G}_1}(\boldsymbol{x}) = \log \mu \ \text{ and } \ \eta_{\mathcal{G}_2}(\boldsymbol{x}) = \log k, \tag{4.1}$$

and adopting the R parametrization

$$\Pr(Y = y; \mu, k) = \binom{y + k - 1}{y} \left(\frac{\mu}{\mu + k}\right)^y \left(\frac{k}{k + \mu}\right)^k, \ y = 0, 1, \dots. \tag{4.2}$$

with $0 < \mu$ and $0 < k$, if the ancillary parameter k is intercept-only then this is referred to as the NB-2 model. The NB-C is of the first case of Table 9, and is

$$\eta_{\mathcal{G}_1}(\mu, k; \boldsymbol{x}) = \log \frac{\mu}{\mu + k}, \quad \eta_{\mathcal{G}_2}(k; \boldsymbol{x}) = \log k. \tag{4.3}$$

The *i*th contribution to the log-likelihood is given by

$$\ell_i(\mu_i, k_i; y_i) = \log \Gamma(y_i + k_i) - \log \Gamma(y_i + 1) - \log \Gamma(k_i)$$
$$+ y_i \log \left(\frac{\mu_i}{\mu_i + k_i} \right) + k_i \log \left(\frac{k_i}{\mu_i + k_i} \right). \tag{4.4}$$

With the ordinary η_j of (4.1), one has

$$\frac{\partial \ell_i}{\partial k_i} = \psi(y_i + k_i) - \psi(k_i) + \log \left(\frac{k_i}{k_i + \mu_i} \right) - \frac{y_i - \mu_i}{\mu_i + k_i}. \tag{4.5}$$

However, with the η_j as in (4.3), the relationship $\mu = k/[e^{-\eta_1} - 1]$ implies (the subscript i is dropped for simplicity)

$$\frac{d\ell}{dk} = \frac{\partial \ell}{\partial \mu} \frac{\partial \mu}{\partial k} + \frac{\partial \ell}{\partial k}. \tag{4.6}$$

The partial derivatives here are directly computed from (4.4), resulting in

$$\frac{\partial \mu}{\partial k} = \frac{\mu}{k}, \qquad \frac{\partial \ell}{\partial \mu} = \frac{y}{\mu} - \frac{y + k}{\mu + k}, \tag{4.7}$$

and $\partial \ell / \partial k$ as in (4.5). Substituting these into (4.6) gives

$$\frac{d\ell}{dk} = \psi(y + k) - \psi(k) + \log \frac{k}{\mu + k}, \tag{4.8}$$

which is (4.5) without its last term. Also, $d\ell / d\mu = \partial \ell / \partial \mu$ is given in (4.7), completing the score vector.

The new EIMs are fully specified by (3.17) (see Table 11). Unlike other NB variants the NB-C EIM is nondiagonal. The adjusted nondiagonal component is given by

$$-\mathbb{E} \left[\frac{d^2 \ell}{d\mu \, dk} \right] = \mathbb{E} \left[\frac{d}{d\mu} \log(\mu + k) \right] = \frac{1}{\mu + k}.$$

Moreover, while $\mathbb{E}[d^2 \ell / d\mu^2]$ remains as usual, the other diagonal element needs adjustment:

$$-\mathbb{E} \left[\frac{d^2 \ell}{dk^2} \right] = \mathcal{W}_{kk} + \frac{\mu}{k(\mu + k)} = \psi'(k) - \mathbb{E}[\psi'(Y + k)],$$

where $\mathcal{W}_{kk}$ is the usual 2–2 element. Combining everything, the working weight matrix is

$$\left(\begin{matrix} \dfrac{1}{\mu} - \dfrac{1}{\mu + k} & \dfrac{1}{\mu + k} \\ \dfrac{1}{\mu + k} & \psi'(k) - E[\psi'(Y + k)] \end{matrix} \right) \circ \left(\begin{matrix} \left(\dfrac{\mathrm{d}\mu}{\mathrm{d}\eta_1} \right)^2 & \dfrac{\mathrm{d}\mu}{\mathrm{d}\eta_1} \dfrac{\mathrm{d}k}{\mathrm{d}\eta_2} \\ \dfrac{\mathrm{d}\mu}{\mathrm{d}\eta_1} \dfrac{\mathrm{d}k}{\mathrm{d}\eta_2} & \left(\dfrac{\mathrm{d}k}{\mathrm{d}\eta_2} \right)^2 \end{matrix} \right)$$

$$= \left(\begin{matrix} \mu \left(1 + \dfrac{\mu}{k} \right) & \mu \\ \mu & k^2 \{ \psi'(k) - E[\psi'(Y + k)] \} \end{matrix} \right)$$

when $\eta_{\mathcal{G}_2} = \log k$. Because the EIM of all NB variants except for the NB-C is diagonal the alternating algorithm adopted by Hilbe and `MASS::glm.nb()` is less prone to failure. But with NB-C, μ and k are asymptotically dependent therefore the alternating algorithm is more likely to fail.

4.2 Weibull regression

The **VGAM** family function `weibullR()` follows from base **R**'s `[dpqr]` `weibull()` parameterization $f(y) = (s/b)(y/b)^{s-1} \exp(-(y/b)^s)$ for `shape` parameter $s > 0$ and `scale` parameter $b > 0$ on $y \in (0, \infty)$. Its mean is $b \, \Gamma(1 + s^{-1})$. For $\theta = (b, s)^T$, `weibullR(lss = TRUE)` has EIM

$$\mathcal{I}(\theta) = \left(\begin{matrix} \mathcal{W}_{11} & \mathcal{W}_{12} \\ \mathcal{W}_{21} & \mathcal{W}_{22} \end{matrix} \right) = \left(\begin{matrix} (s/b)^2 & -(1-\gamma)/b \\ -(1-\gamma)/b & [\pi^2 + 6(\gamma-1)^2]/(6s^2) \end{matrix} \right) \quad (4.9)$$

where $\gamma = -\psi(1)$ is the Euler–Mascheroni constant (Yee 2015, Table 12.3). For this, $\eta_1 = \log b$ and $\eta_2 = \log s$.

Adapted for the Weibull, the methodology of this paper results in the addition of two new 2–parameter links for mean and quantile modelling, called `weibullMlink()` and `weibullQlink()`. To our knowledge the latter is first implementation to exist in such general form. They are used in conjunction with the newly written `weibullRff()` in **VGAMextra 0.0-5**. These links apply only to η_1. Table 1 gives the Weibull distribution's new links and their inverses $\eta_1^{-1} = b(s) = b_s$ as a function of s, hence the composite $\ell(b(s), s; y) = \log s + (s-1) \log y - (y/b_s)^s - s \log b_s$. The new EIMs have form

Table 1 New Two–parameter link functions for Weibull regression handled by `weibull-Rff()` in **VGAMextra 0.0-5**. Here $\tau = (\tau_1, \tau_2, \ldots)^T$ is a vector of quantiles

Link	weibullMlink()	weibullQlink()
$\eta_1 = \mathcal{G}(b \mid x, s)$	$\log \left[b \cdot \Gamma(1 + 1/s) \right]$	$\log \left\{ b \left[-\log(1-\tau) \right]^{1/s} \right\}$
$b = \eta_1^{-1}(s)$	$\exp \left[\eta_1 - \log \Gamma(1 + 1/s) \right]$	$\exp \left[\eta_1 + \log(-\log\{1-\tau\})^{-1/s} \right]$
$\partial b / \partial s$	$-b \cdot \psi(1 + 1/s)$	$\log(-\log\{1-\tau\}) \cdot b/s^2$

$$\mathcal{I}^{\star}(\boldsymbol{\theta} = (b_s, s)^T) = \begin{pmatrix} \mathcal{W}_{11}^{\star} & \mathcal{W}_{12}^{\star} \\ \mathcal{W}_{21}^{\star} & \mathcal{W}_{22}^{\star} \end{pmatrix}, \quad \text{where } \mathcal{W}_{11}^{\star} = \mathcal{W}_{11},$$

$$\mathcal{W}_{21}^{\star} = \mathcal{W}_{21} + \frac{\partial b}{\partial s} \left(\frac{s}{b} \right)^2 \quad \text{and}$$

$$\mathcal{W}_{22}^{\star} = \mathcal{W}_{22} - \left(\frac{2\,\psi(2)}{b} \right) \frac{\partial b}{\partial s} + \left(\frac{s}{b} \frac{\partial b}{\partial s} \right)^2.$$

It is noted that Noufaily and Jones (2013) concerns parametric quantile regression like ours but based on a generalized gamma distribution.

4.3 Distribution–specific quantile regression: the normal distribution

Another example is a two-parameter link for the quantiles of the normal distribution defined as

$$\mathcal{G}(\mu; \sigma, \boldsymbol{x}) = \mu + \sigma\, \Phi^{-1}(\boldsymbol{\tau}), \tag{4.10}$$

where Φ is the standard normal CDF and pre-specified quantiles of interest $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_s)^T$. Its implementation is `uninormalQlink()`, available in VGAMextra 0.0-5.

Ordinarily, $\boldsymbol{\theta} = (\mu, \sigma)^T$ can be estimated with `VGAM::uninormal()` whose default linear predictors are $\eta_1(\mu; \boldsymbol{x}) = \mu$ and $\eta_2(\sigma; \boldsymbol{x}) = \log \sigma$. For `uninormal(var.arg = FALSE)` the EIMs are

$$\mathcal{I} = \begin{pmatrix} \mathcal{W}_{11} & \mathcal{W}_{12} \\ \mathcal{W}_{21} & \mathcal{W}_{22} \end{pmatrix} = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}.$$

With `uninormalQlink()` we allow η_1 of `uninormal()` to handle (4.10):

$$\eta_1(\mu; \sigma, \boldsymbol{x}) = \eta_{\mathcal{G}}(\mu; \sigma, \boldsymbol{x}) = \mu + \sigma\, \Phi^{-1}(\boldsymbol{\tau}), \qquad \eta_2(\sigma; \boldsymbol{x}) = \log \sigma, \tag{4.11}$$

implemented in the brand–new `VGAMextra::uninormalff()`. Using $\frac{\partial \mu}{\partial \sigma} = -\Phi^{-1}(\boldsymbol{\tau})$, the new EIMs are

$$\mathcal{I}^{\star} = \begin{pmatrix} \mathcal{W}_{11}^{\star} & \mathcal{W}_{12}^{\star} \\ \mathcal{W}_{21}^{\star} & \mathcal{W}_{22}^{\star} \end{pmatrix}, \quad \text{with } \mathcal{W}_{11}^{\star} = \mathcal{W}_{11},$$

$$\mathcal{W}_{21}^{\star} = \mathcal{W}_{21} + \frac{\partial \mu}{\partial \sigma} \left(\frac{1}{\sigma^2} \right), \quad \text{and} \quad \mathcal{W}_{22}^{\star} = \mathcal{W}_{22} + \left(\frac{\partial \mu}{\partial \sigma} \right)^2 \left(\frac{1}{\sigma^2} \right).$$

Our work on quantile regression have the following advantages over Koenker and Bassett (1978):

(i) Parametric quantile regression provides more accurate inference when the data comes from the stipulated distribution. In contrast, theirs is a nonparametric L_1–norm method based on linear programming which is less familiar

Table 2 Fitted regression coefficients for η_1 of a NB-C fitted to simulated data. The LHS and RHS are from VGAM and COUNT respectively

	Estimate	Std. Error	z value	Estimate	Std. Error	z value
(Intercept):1	−2.680	0.218	−12.290	−1.511	0.009	−175.723
x2	1.204	0.041	29.321	0.871	0.008	107.366
x3	0.097	0.009	10.658	0.051	0.007	7.240

Table 3 Approximate 95% Wald confidence intervals for the regression coefficients for η_1 of a NB-C fitted to simulated data is the upper table. The lower table is for η_2. The LHS and RHS are from VGAM and COUNT respectively

	2.5%	97.5%	2.5%	97.5%
(Intercept):1	−3.107	−2.253	−1.528	−1.494
x2	1.123	1.284	0.855	0.887
x3	0.079	0.115	0.037	0.065
k	503.398	1230.022	214.945	220.375

to statisticians than L_2 methods like IRLS. Moreover, VGAM computes confidence intervals based on the well–known Wald and score tests. In theirs the confidence intervals are based on piecewise linear approximations and when using large number of predictors the algorithm may become unstable, see, e.g., Kneib (2013).

(ii) The VGLM/VGAM framework can circumvent the quantile crossing problem by choosing appropriate constraint matrices $\boldsymbol{H}_k$ (Eq. (2.4); see also Yee (2015, Sec. 3.3)). Indeed, for some VGLMs such as the exponential and Maxwell the η_j are naturally parallel and so constraints need not to be enforced specifically (Miranda-Soberanis and Yee 2019).

(iii) While their methodology is less sensitive to extreme values than least squares, ours is more amenable to skewed distributions such as the lognormal (theoretically, we can now implement quantile link functions for as often as needed), thus providing a more suitable framework to handle various asymmetric data such as income or wealth.

(iv) VGAM has infrastructure to accommodate spline modelling, so matching Koenker and et al. (2020) which includes linear as well as nonlinear effects.

5 Examples

5.1 NB canonical link

In this section we present two numerical examples of NB-C regression. The first aims to demonstrate the instability of Fisher scoring without total derivative adjustment. We also compare our results to quantile regression using quantreg (Koenker

| **Table 4** Sorted AIC of four NB variants fitted to the `trapO` data	NB-C	NB-H	NB-1	NB-2
	3617.36	3620.74	3645.49	3714.15

and et al. 2020). The second uses the data set medpar from COUNT to compare the Hilbe's and our software. We aim to demonstrate the advantages conferred by `neg-binomial()` and its ability to handle the size parameter as a covariate-specific linear predictor. All the R code is available as a supplement.

5.1.1 Simulated data

To illustrate the unreliability of `ml.nbc()` we adapted code from the COUNT online help and generated $n = 1000$ random variates for $\boldsymbol{\beta}_1 = (-3, 1.25, 0.1)^T$, $k = e^7 \approx 1100$, and x_2 and $x_3 \sim \text{Unif}(0, 1)$ independently. Upon fitting the model, several warnings were issued by `ml.nbc()` from attempts to evaluate (4.2) outside the parameter space. When $k = e^8 \approx 2980$ was attempted `ml.nbc()` issued an error message. A brief comparison of the fits is as follows. The log-likelihood of the `ml.nbc()` fit, -3917.8, is grossly suboptimal compared to VGAM (-3880.2). Also, its estimated regression coefficients differ much from the true values (Table 2). In comparison, the results of VGAM fare well in Table 2 as well as the Wald 95% confidence intervals (Table 3) which cover the true values. For the latter, `ml.nbc()` fails on all counts. From experience, about 5–8 iterations is typical for well-fitting VGLMs and the VGAM fit took 5 iterations to converge, and $\log \widehat{k} \approx 6.66082$ with SE 0.22791. In VGAM 1.0-6 or earlier, 'convergence' was not achieved within 30 iterations.

5.1.2 Lake Otamangakau trout data

Dedual et al. (2000) describe a quantitative study of the ecology of brown (*Salmo trutta*) and rainbow trout (*Oncorhynchus mykiss*) in the central North Island of New
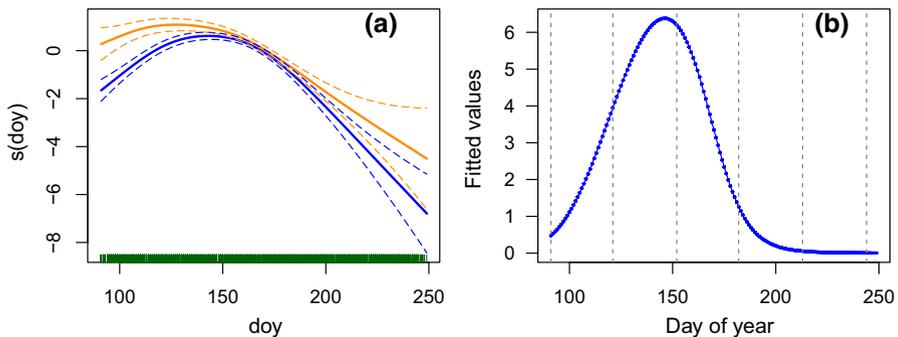
Fig. 1 **a** Centred component functions (overlaid) with pointwise ± 2SE bands of a fitted NB-H VGAM: $\widehat{\eta}_1$ is blue, $\widehat{\eta}_2$ is orange; **b** Fitted values $\widehat{\mu}_i$ of the NB-C using regression splines. The vertical lines denote the first day of the months April to September (color figure online)

Zealand. We fit four NB-variants to closely related data, which may be found in the `trapO` data frame in **VGAMdata**. Comprising 1226 rows of daily captures of the two species by gender at the Te Whaiau Trap at Lake Otamangakau, the data were collected during the main spawning period over 8 consecutive years by the Department of Conservation. The primary aim of this analysis is to model Y = the number of male brown trout captured as a function of the day of the year (e.g., 1 = Jan 1, 244 = Sep 1), i.e., `doy` is the sole (primary) explanatory variable. It is well known that spawning peaks around the second half of May.

The variance-to-mean ratio of 12.12 indicates overdispersion relative to the Poisson. Following the same nomenclature as Hilbe (2007), the four NB-variants fitted here are abbreviated NB-1, NB-2, NB-H and NB-C. The former has $\text{var}(Y) \propto \mu$, and the NB-H has $\eta_2 = \log k = \boldsymbol{\beta}_2^T \boldsymbol{x}$ while NB-2 has an intercept-only η_2. All four models have $\eta_1 = \boldsymbol{\beta}_1^T \boldsymbol{x}$. Because μ is clearly unimodal, we fit a NB-H VGAM to explore the data; the component functions are overlaid in Fig. 1a. Interestingly, both have an approximate quadratic shape with different peaks and curvature. It was considered safest to model the nonlinearity in both η_j nonparametrically.

To compare the variants more rigorously we replace the cubic (vector) smoothing splines (Yee and Wild 1996) by regression splines because inference is more standard (Figure 1). In R this was the term `bs(doy)` used by `vglm()` rather than `s(doy)` within `vgam()`. The term offers 3 degrees of freedom excluding an intercept.

Table 4 summarizes the results ranked by AIC. The NB-C was superior followed by the NB-H, and this suggests that the day of the year strongly affects both μ and k—the other models are too simple. That the NB-C performed best suggests that the relationship between the mean and variance is not as simple as what a basic loglinear relationship can allow. Indeed, μ appears to be coupled with k in the more complex nonlinear manner provided by the canonical link. Other NB analyses, including the NB-C, on real data are presented in Yee (2020).

5.2 Distribution-specific quantile regression: the Weibull distribution

We use the Munich rental guide data (Fahrmeir et al. 2013, Page 5) available in the **gamlss** R package (data set **rent99**) which has been analysed using a Box-Cox Cole and Green (BCCG) regression with GAMLSS by Rigby et al. (2013) (Table 5).

As an alternative, we re-analyse this data using Weibull quantile regression with `weibullQlink()` for two reasons: as a demonstration of the new link function, and to show that the Weibull distribution appears to work well for the target variable **rent**, even though it is traditionally used more in survival analysis. The proposed GAMLSS model is

Table 5 Variables involved in Model (5.1); dataset **rent99** from the **gamlss** package

Variable	Description
rent	The monthly net rent per month (in Euro)
area	Living area in Square meters

Table 6 AICs and BICs from BCCG and Weibull regressions on the rent data, as in (5.1)

Modelling framework	Modelling function	Distribution	Family function	AIC	BIC
GAMLSS	gamlss()	BCCG	BCCG()	39332.02	39356.15
VGAM	vgam()	Weibull	weibullR()	39315.03	39351.23
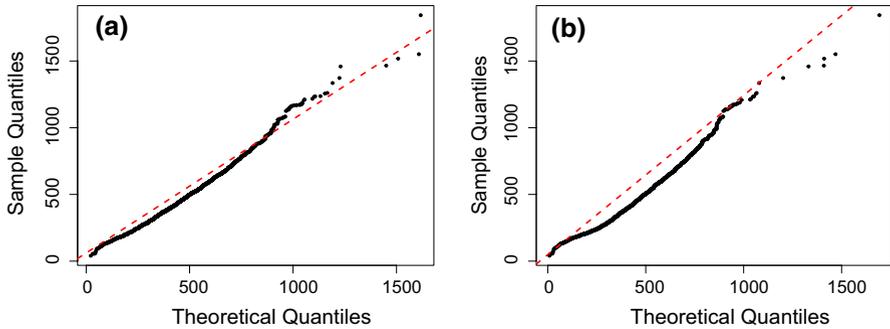
Fig. 2 Estimated **a** Weibull and **b** BCCG QQ–plots of the response RENT
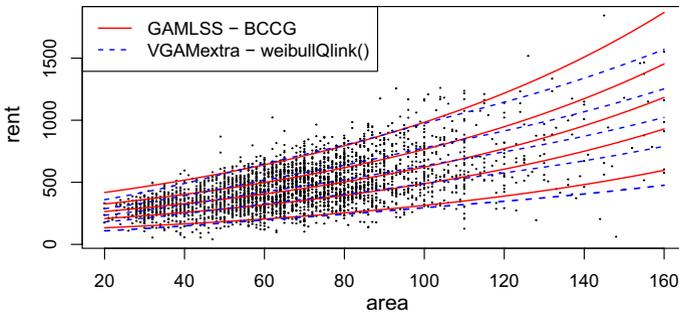
Fig. 3 Quantile curves from BCCG regression with GAMLSS and Weibull quantile regression with `weibullQlink()`

Table 7 Empirical quantiles from GAMLSS using `BCCG()` and Weibull quantile regression with `weibullQlink()` from VGAMextra

	5%	25%	50%	75%	95%
GAMLSS	5.19	25.67	50.49	73.46	95.2
VGAMextra	4.83	25.05	51.1	74.82	95.07

$$\text{rent} \sim \mathcal{G}(\mu, \sigma, \nu), \quad \log \mu = b_{10} + s_{11}(\text{area}), \quad \log \sigma = b_{20}, \quad \log \nu = b_{30}.$$
$$(5.1)$$

where $\mathcal{G}$ is either the BCCG or the Weibull distribution. Intercept–only models for σ and ν are set for comparison purposes.

To check on the distributional assumptions we fit BCCG and Weibull regressions to the data as in (5.1). The estimated BCCG and Weibull QQ-plots shown in Figure 2, and AICs and BICs from both models in Table 6 show the Weibull distribution performs as well as or better than the BCCG method.

We now compare Weibull quantile regression with $\boldsymbol{\tau} = c(0.5, 0.25, 0.50, 0.75, 0.95)^T$ and `weibullQlink()` from **VGAMextra** to the τ quantile of rent under GAMLSS given by $y_\tau = \mu \cdot \text{qBCCG}(\tau, 1, \sigma, \nu)$, where $\text{qBCCG}(\tau, 1, \sigma, \nu)$ is the τ quantile of the BCCG distribution. Figure 3 shows the estimated quantile curves and Table 7 gives the empirical quantiles. Both methods are effectively comparable and perform equally well.

## 5.3 Quantile modelling with the normal distribution

This example is based on Koenker and Hallock (2001) who performed conditional quantile regression of covariates associated to birth weight of live babies to demonstrate its better performance over ordinary least squares (OLS) when estimating the effects on the lower tail of the skewed birthweight distribution. This relationship was initially explored by Abreveya (2001). Koenker and Hallock (2001) used a sample ($n = 198, 377$) of the June 1997 *Detailed Natality Data* published by the National Centre for Health Statistics (NCHS) that contains information from live, singleton births, mothers aged 18–45 and residing in USA.

5.3.1 An example using `uninormalQlink()`

To test the quantile regression framework introduced in Sect. 4.3 we carry out VGLM quantile modelling with `uninormaQlink()` and `uninormalff()`, and compare our results to Koenker and Hallock (2001) via `rq()` from **quantreg**. Due to availability constraints, we restrict ourselves to a $n = 50, 000$ subsample of the the 1997 NCHS data (Koenker and Hallock 2001) stored in the file *BWeights.csv* and incidentally obtained from the SAS® file **Sashelp.BWeights**. Section D of the Online Supplements gives a short description of *BWeights.csv* and SAS® code used to generate it, as well as supplementary code pertaining this section. *BWeights.csv* is available in the Supplementary Materials.

Following Koenker and Hallock (2001) and Abreveya (2001), the response is *birthweight* (recorded in grames) and the remaining factors are added into the model as covariates including a quadratic term for the mother's age and the mother's reported weight gain during pregnancy:

$$
\begin{aligned}
Birth\_weight \sim & \ Black + Married + Boy + Visit + MomEdLevel \\
& + MomSmoke + CigsPerDay + MomAge + MomWtGain \\
& + MomAge^2 + MomWtGain^2.
\end{aligned}
$$

$$(5.2)$$
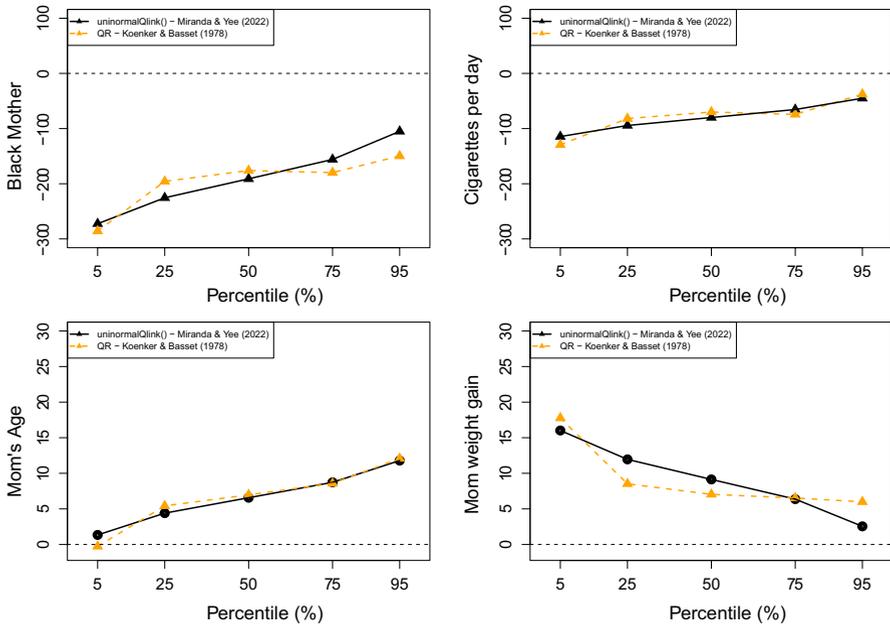
Fig. 4 Regression estimates by quantile level for the *birthweight* model using `uninormalQlink()` (orange) and `quantreg` (black), c.f. Koenker and Hallock (2001, p. 150) (color figure online)

Table 8 Summary (minimum, maximum, mean and quartiles) of `Mom weight gain` (kgs) centered around its median

Minimum	1st Qu.	Median	Mean	3rd Qu.	Maximum
−30	−8	0	0.7092	9	68

We set $\tau = (0.05, 0.25, 0.50, 0.75, 0.95)$.

The trace output from `vglm()` looks natural and correct. We have set `zero = NULL` to allow both linear predictors to be regressed on all the covariates. Note, $\eta_{\mathcal{G}_1}$ in (4.11) is managed by the new link function `uninormalQlink()`. The function `Q.reg()` is required to create a multiple responses matrix spanning `r = length(mytau)` columns.

Figure 4 shows the estimated parameters by quantile level from **VGAMextra** and **quantreg** for a few relevant covariates. Results conform with Koenker and Hallock (2001, p.150) to a great extent showing similar trends, except by *Mom weight gain* perhaps due the large range covered by this covariate (largest range among all covariates), as shown in Table 8.

Infants born to black mothers appear to weigh less, by between 100 and 300 grams, than newborns from white mothers across 5% to 95% quantiles. Likewise, although to a smaller extent, 'smoking' (this is `Cigarettes per day`

(CigsPerDay) coded as '1' if the mother smokes more than five cigarettes p/day and '0' otherwise) is associated with newborns' weight loss. With smaller effects, maternal age appears to top birth weight up incrementally, by 4–6 grams, from the 5% quantile to the 95% quantile, while every kilogram gained during pregnancy is associated with gradual reductions in the infant's weight (4–5 grams). The model summary is in Section D of the Online Supplements.

6 Discussion

This paper has developed the general methodology for two-parameter link functions and used the NB-C and quantile modelling as the primary examples. The approach taken here is the first of two general options: (i) offer one family function that handles several different link parameterizations; (ii) offer several family functions corresponding to several different parameterizations of the same distribution. Each option has its advantages and disadvantages. Writing a new link function arguably involves less work and also the EIM may be too difficult to derive directly. However, given a choice, it is likely that many practitioners would choose a particular family function and use that solely.

Over many years the NB-C has been widely referred to in mathematical statistics because of its connections with core concepts such as sufficient statistics,GLMs and variance functions. Despite this, Hilbe (p. 315, 2011) writes concerning its practical use: "Little work as been done with NB-C models. No research has been published using an NB-C model on data." This might be partially explained by the observation that almost all practitioners use existing implementations written by others and that the NB-C is a model with practical shortcomings such as $\eta_{G_1} < 0$ in (4.3) rather than being unbounded. It is disappointing that, after several decades, our implementation is the first to fit the NB-C by a 'proper' algorithm because some packages simply call a general optimizer such as optim(). Our solution is naturally flexible too, as seen by the NB-C-H VGAM fitted in Sect. 5.1.2.

Our results on quantile regression showing that it is loosely 'regression estimated on multiple quantiles' are obviously dependent on a strong distributional assumption, however there are real practical benefits and realistic applications, as discussed in Sects. 4.3 and 5. The Weibull distribution is used invariably to model observed failures in survival analysis and reliability, and the normal distribution is almost foundational for many natural phenomena. Covariates can also be included and their effects on the distributions examined. We believe that drawbacks from a distribution-specific framework are ameliorated by smoothing-based infrastructure capable of identifying nonlinearity automatically and graphically, as well as the handling of a very broad range of response types such as categorical and survival data.

It is not surprising that this work is seemingly related to that of others. For example, Efron (1986) and Smyth and Verbyla (1999) model the mean and dispersion simultaneously in what are called 'double exponential families' and 'double GLMs' respectively. However, both apply separate ordinary links to each of the first two moments rather than including a single link function of both parameters. Likewise, Cepeda-Cuervo et al. (2014) develop 'double GLMs' with random-effects utilizing the same type of ordinary link functions.

There is scope for future work. Slightly more interpretable than the NB-C link might be $\eta_G = \log\left[-\log\{\mu/(\mu + k)\}\right]$, and practically, other alternatives include $\text{logit}\{\mu/(\mu + k)\} = \log\{\mu/k\}$ and $\Phi^{-1}(\mu/(\mu + k))$; they could easily be estimated with the present methodology. More generally, NB regression has the limitation that it does not handle underdispersion relative to the Poisson, hence other alternatives such as the Conway-Maxwell-Poisson distribution have gained popularity (see, e.g., Sellers and Shmueli 2010). Of course, the methodology could be generalized to $M > 2$ parameters and in particular, the $M = 3$ case would match distributions having a location, scale and shape parameter.

Quantile and mean modelling is another area to be further exploited in the short-term. We are applying this methodology to several two–parameter distributions which will soon have, as the Weibull example, links of the form *distribution* Qlink() or *distribution* Mlink() alluding the 'quantile' and 'mean' link respectively. We have already commenced work in this direction, e.g., with the 'mean link' for the 2–parameter gamma distribution, viz. VGAMextra::gammaRMlink(). However, the VGLM framework is broader, with further options such as additive models (Yee and Wild 1996) and reduced–rank regression (Yee and Hastie 2003) over the same $\boldsymbol{\eta}$. We hope to investigate these options too, including nlrq() from quantreg for nonlinear quantile regression.

Appendix: Expressions for complementary derivatives

This appendix summarizes the new complementary expressions for $d\ell/d\theta_j$, $d^2\ell/d\theta_j^2$ and $d^2\ell/d\theta_j d\theta_k$, in Table 9, Table 10 and Table 11 respectively, for the three cases (3.2)-(3.4).

Table 9 New complementary expressions for the total derivatives $d\ell/d\theta_j$ for cases (3.2)-(3.4)

Linear predictors, $\boldsymbol{\eta}_G = \left(\eta_{G_1}, \eta_{G_2}\right)^T$	Composite–ℓ	$d\ell/d\theta_j$
$\eta_{G_1} = G_1(\theta_1, \theta_2),$	$\ell\left(\theta_1\left(\eta_{G_1}, \theta_2\right), \theta_2(\eta_{G_2}); \mathbf{y}\right)$	$\dfrac{d\ell}{d\theta_1} = \dfrac{\partial\ell}{\partial\theta_1,}$
$\eta_{G_2} = G_2(\theta_2)$		$\dfrac{d\ell}{d\theta_2} = \dfrac{\partial\ell}{\partial\theta_1} \cdot \dfrac{\partial\theta_1}{\partial\theta_2} + \dfrac{\partial\ell}{\partial\theta_2.}$
$\eta_{G_1} = G_1(\theta_1),$	$\ell\left(\theta_1(\eta_{G_1}), \theta_2\left(\theta_1, \eta_{G_2}\right); \mathbf{y}\right)$	$\dfrac{d\ell}{d\theta_1} = \dfrac{\partial\ell}{\partial\theta_1} + \dfrac{\partial\ell}{\partial\theta_2} \cdot \dfrac{\partial\theta_2}{\partial\theta_1,}$
$\eta_{G_2} = G_2(\theta_1, \theta_2)$		$\dfrac{d\ell}{d\theta_2} = \dfrac{\partial\ell}{\partial\theta_2.}$
$\eta_{G_1} = G_1(\theta_1, \theta_2),$	$\ell\left(\theta_1\left(\eta_{G_1}, \theta_2\right), \theta_2\left(\theta_1, \eta_{G_2}\right); \mathbf{y}\right)$	$\dfrac{d\ell}{d\theta_1} = \dfrac{\partial\ell}{\partial\theta_1} + \dfrac{\partial\ell}{\partial\theta_2} \cdot \dfrac{\partial\theta_2}{\partial\theta_1,}$
$\eta_{G_2} = G_2(\theta_1, \theta_2)$		$\dfrac{d\ell}{d\theta_2} = \dfrac{\partial\ell}{\partial\theta_1} \cdot \dfrac{\partial\theta_1}{\partial\theta_2} + \dfrac{\partial\ell}{\partial\theta_2.}$

Table 10 New complementary expressions for the total derivatives $\mathrm{d}^2\ell/\mathrm{d}\theta_j^2$ in (3.17) for cases (3.2)–(3.4)

Linear predictors, $\boldsymbol{\eta}_\mathcal{G} = \left(\eta_{\mathcal{G}_1}, \eta_{\mathcal{G}_1}\right)^T$	Composite–ℓ	$\mathrm{d}^2\ell/\mathrm{d}\theta_j^2$
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1, \theta_2),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}, \theta_2), \theta_2(\eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1^2} = \dfrac{\partial^2\ell}{\partial\theta_1^2},$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_2)$		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2^2} = \dfrac{\partial^2\ell}{\partial\theta_2^2} + 2 \cdot \dfrac{\partial^2\ell}{\partial\theta_1\partial\theta_2} \cdot \dfrac{\partial\theta_1}{\partial\theta_2} +$
		$\dfrac{\partial\ell}{\partial\theta_1} \cdot \dfrac{\partial^2\theta_1}{\partial\theta_2^2} + \dfrac{\partial^2\ell}{\partial\theta_1^2} \cdot \left(\dfrac{\partial\theta_1}{\partial\theta_2}\right)^2.$
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}), \theta_2(\theta_1, \eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1^2} = \dfrac{\partial^2\ell}{\partial\theta_1^2} + 2 \cdot \dfrac{\partial^2\ell}{\partial\theta_2\partial\theta_1} \cdot \dfrac{\partial\theta_2}{\partial\theta_1} +$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_1, \theta_2)$		$\dfrac{\partial\ell}{\partial\theta_2} \cdot \dfrac{\partial^2\theta_2}{\partial\theta_1^2} + \dfrac{\partial^2\ell}{\partial\theta_2^2} \cdot \left(\dfrac{\partial\theta_2}{\partial\theta_1}\right)^2,$
		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2^2} = \dfrac{\partial^2\ell}{\partial\theta_2^2}.$
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1, \theta_2),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}, \theta_2), \theta_2(\theta_1, \eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1^2} = \dfrac{\partial^2\ell}{\partial\theta_1^2} + 2 \cdot \dfrac{\partial^2\ell}{\partial\theta_2\partial\theta_1} \cdot \dfrac{\partial\theta_2}{\partial\theta_1} +$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_1, \theta_2)$		$\dfrac{\partial\ell}{\partial\theta_2} \cdot \dfrac{\partial^2\theta_2}{\partial\theta_1^2} + \dfrac{\partial^2\ell}{\partial\theta_2^2} \cdot \left(\dfrac{\partial\theta_2}{\partial\theta_1}\right)^2,$
		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2^2} = \dfrac{\partial^2\ell}{\partial\theta_2^2} + 2 \cdot \dfrac{\partial^2\ell}{\partial\theta_1\partial\theta_2} \cdot \dfrac{\partial\theta_1}{\partial\theta_2} +$
		$\dfrac{\partial\ell}{\partial\theta_1} \cdot \dfrac{\partial^2\theta_1}{\partial\theta_2^2} + \dfrac{\partial^2\ell}{\partial\theta_1^2} \cdot \left(\dfrac{\partial\theta_1}{\partial\theta_2}\right)^2.$

Table 11 New complementary expressions for the total derivatives $\mathrm{d}\ell^2/\mathrm{d}\theta_j\partial\theta_k$ in (3.17) for cases (3.2)–(3.4)

Linear predictors, $\boldsymbol{\eta}_{\mathcal{G}} = \left(\eta_{\mathcal{G}_1}, \eta_{\mathcal{G}_2}\right)^T$	Composite–ℓ	$\mathrm{d}^2\ell/\mathrm{d}\theta_j\mathrm{d}\theta_k$ in (3.15)–(3.16) †
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1, \theta_2),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}, \theta_2),\ \theta_2(\eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1\,\mathrm{d}\theta_2} = \dfrac{\partial^2\ell}{\partial\theta_1\,\partial\theta_2} + \dfrac{\partial^2\ell}{\partial\theta_1^2}\cdot\dfrac{\partial\theta_1}{\partial\theta_2},$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_2)$		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2\,\mathrm{d}\theta_1} = \dfrac{\partial^2\ell}{\partial\theta_2\,\partial\theta_1} + \dfrac{\partial^2\ell}{\partial\theta_1^2}\cdot\dfrac{\partial\theta_1}{\partial\theta_2}.$
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}),\ \theta_2(\theta_1, \eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1\,\mathrm{d}\theta_2} = \dfrac{\partial^2\ell}{\partial\theta_1\,\partial\theta_2} + \dfrac{\partial^2\ell}{\partial\theta_2^2}\cdot\dfrac{\partial\theta_2}{\partial\theta_1},$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_1, \theta_2)$		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2\,\mathrm{d}\theta_1} = \dfrac{\partial^2\ell}{\partial\theta_2\,\partial\theta_1} + \dfrac{\partial^2\ell}{\partial\theta_2^2}\cdot\dfrac{\partial\theta_2}{\partial\theta_1}.$
$\eta_{\mathcal{G}_1} = \mathcal{G}_1(\theta_1, \theta_2),$	$\ell\left(\theta_1(\eta_{\mathcal{G}_1}, \theta_2),\ \theta_2(\theta_1, \eta_{\mathcal{G}_2}); \boldsymbol{y}\right)$	$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_1\,\mathrm{d}\theta_2} = \dfrac{\partial^2\ell}{\partial\theta_1\,\partial\theta_2} + \dfrac{\partial^2\ell}{\partial\theta_1^2}\cdot\dfrac{\partial\theta_1}{\partial\theta_2} +$
$\eta_{\mathcal{G}_2} = \mathcal{G}_2(\theta_1, \theta_2)$		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2^2}\cdot\dfrac{\partial\theta_2}{\partial\theta_1},$
		$\dfrac{\mathrm{d}^2\ell}{\mathrm{d}\theta_2\,\mathrm{d}\theta_1} = \dfrac{\partial^2\ell}{\partial\theta_2\,\partial\theta_1} + \dfrac{\partial^2\ell}{\partial\theta_1^2}\cdot\dfrac{\partial\theta_1}{\partial\theta_2} +$
		$\dfrac{\partial^2\ell}{\partial\theta_2^2}\cdot\dfrac{\partial\theta_2}{\partial\theta_1}.$

† The EIMs are symmetric so that $\partial^2\ell/[\partial\theta_1\,\partial\theta_2] = \partial^2\ell/[\partial\theta_2\,\partial\theta_1]$ is expected to hold over the entire parameter space

References

Abreveya J (2001) The effects of demographics and maternal behaviour on the distribution of birth outcomes. J Econ 26(1):247–257

Cepeda-Cuervo E, Migon HS, Garrido L, Achcar JA (2014) Generalized linear models with random effects in the two-parameter exponential family. J Stat Comput Simul 84(3):513–525

Dedual M, Maxwell ID, Hayes JW, Strickland RR (2000) Distribution and movements of brown (Salmo trutta) and rainbow trout (Oncorhynchus mykiss) in Lake Otamangakau, central North Island, New Zealand. New Zealand J Marine Freshwater Res 34(4):615–627

Efron B (1978) Geometry of exponential families. Ann Statist 6(2):362–376

Efron B (1986) Double exponential families and their use in generalized linear regression. J Amer Statist Assoc 81(395):709–721

Fahrmeir L, Kneib T, Lang S, Marx B (2013) Regression: models. Springer, Methods and Applications

Green PJ (1984) Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. J Roy Statist Soc Ser B 46(2):149–192

Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1(3):297–318

Hastie T, Tibshirani R (1990) Generalized additive models. Chapman & Hall, London, UK

Hilbe JM (2007) Negative binomial regression. Cambridge University Press, Cambridge

Hilbe JM (2011) Negative binomial regression, 2nd edn. Cambridge University Press, Cambridge, UK

Kneib T (2013) Beyond mean regression. Stat Model 13(2):275–303

Koenker R, Bassett JG (1978) Regression Quantiles. Econom J Econom Soc 46(1):33–50

Koenker R, et al (2020) *quantreg: Quantile Regression*. URL https://CRAN.R-project.org/package=quantreg. R package version 5.67

Koenker R, Hallock K (2001) Quantile regression. J Econ Perspect 15(4):143–156

Loomis LH, Sternberg S (1990) Advanced calculus. Jones and Bartlett Publishers, Boston, USA, revised edition edition

McCullagh P, Nelder JA (1983) Generalized linear models, 1st edn. Chapman & Hall, London, UK

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall, London, UK

Miranda-Soberanis, V (2018) Vector generalized linear time series models with an implementation in R. PhD thesis, Department of Statistics, University of Auckland, New Zealand

Miranda-Soberanis VF, Yee TW (2019) New link functions for distribution-specific quantile regression based on vector generalized linear and additive models. J Probab Stat 5:1–11

Nelder JA, Wedderburn RWM (1972) Generalized linear models. J Royal Stat Soc Series B 135(3):370–384

Noufaily A, Jones MC (2013) Parametric quantile regression based on the generalized gamma distribution. J Royal Stat Soc Series C 62(5):723–740

Rigby RA, Stasinopoulos DM, Voudouris V (2013) Discussion: a comparison of GAMLSS with quantile regression. Stat Model 13(4):335–348

Sellers KF, Shmueli G (2010) A flexible regression model for count data. Annals Appl Stat 4(2):943–961

Smyth GK, Verbyla AP (1999) Double generalized linear models: approximate REML and diagnostics. In: H. Friedl, A. Berghold, and G. Kauermann (eds) Statistical modelling: proceedings of the 14th international workshop on statistical modelling, Graz, Austria, July 19–23, pages 66–80, Technical University, Graz, Austria

Thompson R, Baker RJ (1981) Composite link functions in generalized linear models. J Roy Statist Soc Ser C 30(2):125–131

Umlauf N, Klein N, Zeileis A (2018) BAMLSS: Bayesian additive models for location, scale, and shape (and beyond). J Comput Graph Stat 27(3):612–627

Wood SN (2017) Generalized additive models: an introduction with R, 2nd edn. Chapman & Hall, New York, USA

Yee TW (2004) Quantile regression via vector generalized additive models. Stat Med 23(14):2295–2315

Yee TW (2010) The VGAM package for categorical data analysis. J Stat Softw 32(10):1–34

Yee TW (2015) Vector generalized linear and additive models with an implementation in R. Springer, New York, USA

Yee TW (2020) The VGAM package for negative binomial regression. Aust N. Z. J Stat 62(1):116–131

Yee TW, Hastie TJ (2003) Reduced-rank vector generalized linear models. Stat Model 3(1):15–41

Yee TW, Stephenson AG (2007) Vector generalized linear and additive extreme value models. Extremes 10(1):1–19

Yee TW, Wild CJ (1996) Vector generalized additive models. J Royal Soc Series B 58(3):481–493