



Article

On a skew t–discrete gamma alternative to the normal-Poisson model for clustered count outcomes

Chénangnon Frédéric Tovissodé ^{1,*}  and Romain Glèlè Kakai ¹ 

¹ Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Abomey-Calavi (Benin); romain.glelekakai@fsa.uac.bj (R.G.K)

* Correspondence: chenangnon@gmail.com (C.F.T.)

Abstract: The normal and Poisson distribution assumptions in the normal-Poisson mixed effects regression model are often too restrictive for many real count data. Several works have independently relaxed the Poisson conditional distribution assumption for counts or the normal distribution assumption for random effects. This work couples some recent advances in these two regards to develop a skew t–discrete gamma regression model in which the count outcomes have full dispersion flexibility and random effects can be skewed and heavy tailed. Inference in the model is achieved by maximum likelihood using pseudo-adaptive Gaussian quadrature. The use of the proposal is demonstrated on a popular owl sibling negotiation data. It appears that, for this example, the proposed approach outperforms models based on normal random effects and the Poisson or negative binomial count distribution.

Keywords: Discrete gamma distribution ; correlated counts ; sparse-grid quadrature ; empirical Bayes estimators

1. Introduction

The Normal–Poisson (NP) model incorporating normal random effects in Poisson regression [1] is commonly used to analyze clustered, longitudinal or hierarchical count outcomes which arise routinely from applications in e.g. agricultural, environmental, biomedical, economic, and social sciences. Statistical softwares such as the R freeware [2], Stata [3] and SAS [4] provide routines returning for a count dataset, a NP model fit which gives a theoretical generating mechanism and provides statistics to guide decisions, relying on the relative flexibility of the NP model for handling a wide range of count data [5]. There is however a rich literature on the lack of robustness against departures from distributional assumptions underlying the model, namely the Poisson equidispersion assumption (variance equal to mean) and the normality of random effects [6–9].

For the control of dispersion in observed counts, overdispersion (variance greater than mean) models [6] are the most popular alternative count models among which, Poisson mixtures (e.g. the negative binomial [10], discrete Poisson-Lindley [11]) play a central role. Conway–Maxwell–Poisson (CMP) [12–14], double Poisson [15,16] and Consul's generalized Poisson [17] are also well known models for overdispersed count data. On the other hand, underdispersion (variance less than mean) is often considered rarer [18]. Indeed, underdispersion has received much less attention [19] and pure underdispersion models such as the positive binomial [20], Morse's [21] and Haight's generalized Poisson models [22,23] are rarely used. Sellers and Shmueli [24] have shown that underdispersion may have been hidden alongside Poisson and overdispersion models, and the rarity of underdispersion may thus be linked to the scarcity of statistical routines for fitting underdispersion and group-level dispersion models. This suggests that models for the control of dispersion should have full dispersion flexibility, i.e. allow under-, equi- as well as overdispersion. Although some existing models can be used to handle underdispersed, equidispersed as well as overdispersed data, they yield quasi-models for underdispersed counts (e.g. Quasi-Poisson [25], and extended Poisson-Tweedie [26]) or lack a simple mean parameterization (e.g. CMP [12], double Poisson [15], gamma count [27], semi nonparametric Poisson [28], Faddy [29] and discrete Weibull [30]) rendering inference only approximate [31,32]. An old but reviving



Citation: Tovissodé, C.F.; Glèlè Kakai, R. A skew t–discrete gamma alternative to the normal-Poisson model for clustered count outcomes. *Preprints* **2021**, *1*, 0. <https://doi.org/>

Received:
Accepted:
Published:

Publisher's Note: NA.

approach to flexible count models identification is the discretization of flexible continuous distributions [33]. In this frame, Tovissodé et al. [34] have recently introduced the discrete gamma (DG) regression model for counts using the balanced discretization method [32]. The DG model interestingly yields a full dispersion regression model free of the above mentioned shortcomings.

In count data with a natural structure (clustered, longitudinal or hierarchical), within-subject dispersion may differ from between-subject dispersion [35]. For the sake of flexibility in modelling between-subject variability, various distributions have been considered for random effects in mixed count models, e.g. Semi-nonparametric [36], skew normal [37], log normal [38], generalized log-gamma [39] and skew t [40]. In particular, skew t distributions [41] have been satisfactorily used in various areas of mixed modelling, e.g. linear mixed models [42], generalized linear mixed models [40,43] and nonlinear mixed models [44,45], allowing both skewness and kurtosis while keeping the normal distribution as a limiting special case. However, except works which treated overdispersion in Poisson regression by introducing observation-level random effects [46–49], there has been, so far, very few attempts to explicitly couple advances in the control of dispersion and the use of flexible random effects distributions. For instance, Khan et al. [50] introduced a CMP model with serial correlation and log-gamma random effects for clustered longitudinal counts. As noted above, the CMP distribution [12] has no simple closed form expression for the mean count in terms of model parameters, and thus leads to hardly interpretable regression coefficients [14,34,51].

In order to bridge advances in control of dispersion and relaxation of normality, this work proposes an extension of the NP model by introducing skew t [41] distributed random effects in the DG count regression model [34]. The resulting mixed DG regression model offers a very flexible approach to the modelling of correlated counts, allowing an unlimited dispersion level for the data and skewed as well as thick tailed random effects.

For clarity, we present in Section 2 some properties of the balanced discrete gamma and the skew t distributions. Then, we define in Section 3 the skew t–discrete gamma regression model and present a maximum likelihood estimation procedure. In Section 4, the proposed model is illustrated on a real count data and compared to some competing approaches. Concluding remarks are given in Section 5.

2. Preliminaries

In this section, we introduce the balanced discrete gamma (BDG) and the multivariate skew t (ST) distributions as ingredients to define the skew t–discrete gamma model. Let us denote \mathbb{N} , the set of natural numbers, \mathbb{R} , the set of reals, \mathbb{R}_+ , the set of positive reals, \mathbb{R}^* , the set of non null reals, and $[x]$, the integral part of any real x . For $a \in \mathbb{R}_+$ and $x \in \mathbb{R}_+$, let $\gamma(x, a) = \int_0^x u^{a-1} e^{-u} du / \Gamma(a)$ the lower incomplete gamma ratio. We shall denote $\mathcal{G}amma(a, b)$ the gamma distribution with scale $a \in \mathbb{R}_+$ and shape $b \in \mathbb{R}_+$, which has mean value b/a .

2.1. The balanced discrete gamma distribution

A random count variable Y is said to follow a balanced discrete gamma distribution denoted $\mathcal{DG}(\mu, a)$ with parameters $\mu \in \mathbb{R}_+$ and $a \in \mathbb{R}_+$, if for $y \in \mathbb{N}$ it has the probability mass function (pmf) [32]

$$f_{bg}(y|\mu, a) = (y-1)\gamma(a(y-1), a\mu) - 2y\gamma(ay, a\mu) + (y+1)\gamma(a(y+1), a\mu) - \mu[\gamma(a(y-1), a\mu+1) - 2\gamma(ay, a\mu+1) + \gamma(a(y+1), a\mu+1)] \quad (1)$$

If $Y \sim \mathcal{DG}(\mu, a)$, then Y has expected value μ . The parameter a controls the level of dispersion of Y . Accordingly, the variance and the index of dispersion (variance to mean ratio) of Y satisfy respectively [32]

$$\sigma_{bg}^2(\mu, a) = \mu a^{-1} + \zeta_{g_0}(\mu, a) \quad (2)$$

$$\text{ID}_Y(\mu, a) = a^{-1} + \mu^{-1} \zeta_{g_0}(\mu, a) \quad (3)$$

where $\zeta_{g_0}(\mu, a) = \sum_{z=0}^{\infty} \zeta_{g_0}(z, \mu, a)$ with

$$\begin{aligned} \zeta_{g_0}(z, \mu, a) &= \mu(2z+1)[\gamma(a(z+1), a\mu+1) - \gamma(az, a\mu+1)] \\ &\quad - \mu \frac{a\mu+1}{a} [\gamma(a(z+1), a\mu+2) - \gamma(az, a\mu+2)] \\ &\quad - z(z+1)[\gamma(a(z+1), a\mu) - \gamma(az, a\mu)]. \end{aligned}$$

The quantity ζ_{g_0} is bounded as $0 < \zeta_{g_0} < 1/4$ and can be approximated as

$$\widehat{\zeta}_{g_0}(\mu, a) = \sum_{z=z_i}^{z_f+1} \zeta_{g_0}(z, \mu, a) \quad (4)$$

where z_i and z_f are respectively the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\mathcal{Gamma}(a, a\mu)$, and α is a tolerance value in $(0, 1)$ controlling the precision of $\widehat{\zeta}_g(\mu, a)$ as the absolute error $|\widehat{\zeta}_g(\mu, a) - \zeta_g(\mu, a)|$ is less than $1 - \gamma(a(z_f+2), a\mu) + \gamma(az_i, a\mu)$ [32].

Figure 1 illustrates the flexibility of the balanced discrete gamma distribution relative to the Poisson distribution with mean $\mu = 5$. Whereas the Poisson distribution is constrained to equidispersion ($\text{ID}_Y = 1$), $\mathcal{DG}(5, a)$ with $a = 2.1428, 1.0345, 0.5085$ allow underdispersion ($\text{ID}_Y = 0.5$), equidispersion ($\text{ID}_Y = 1$) and overdispersion ($\text{ID}_Y = 2$) respectively.

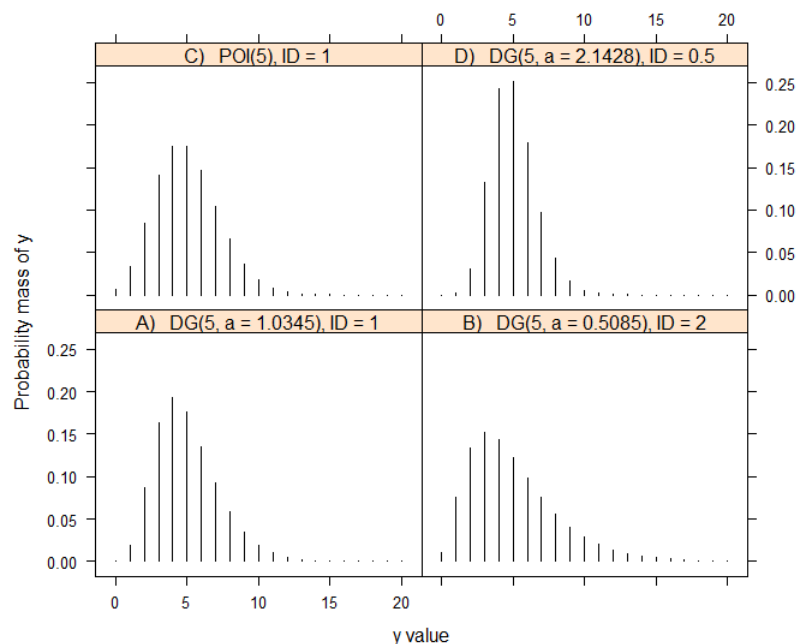


Figure 1. Probability mass of Poisson distribution (POI) with mean $\mu = 5$ with a variance to mean ratio $\text{ID}_Y = 1$ against balanced discrete gamma (BDG) distributions with mean $\mu = 5$ and scales $a = 2.1428, a = 1.0345$ and $a = 0.5085$ corresponding to $\text{ID}_Y = 0.5, \text{ID}_Y = 1$ and $\text{ID}_Y = 2$ respectively

2.2. The multivariate skew t distribution

A random variable \mathbf{Y} is said to follow a q -variate skew t (ST) distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^q$, scale $\boldsymbol{\Omega} \in \mathbb{R}^{q \times q}$ (positive definite), shape $\boldsymbol{\lambda} \in \mathbb{R}^q$ and degrees of freedom ν if for a value $\mathbf{y} \in \mathbb{R}^q$ of \mathbf{Y} , the pdf is [42]:

$$St_q(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu) = 2 t_q(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) T\left(\alpha \left(\frac{\nu+q}{\nu+\mathbf{y}_0^\top \mathbf{y}_0}\right)^{1/2} \mid \nu+q\right) \quad (5)$$

where $t_q(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Omega}, \nu) = \frac{\Gamma((\nu+q)/2)}{|\boldsymbol{\Omega}|^{1/2}(\pi\nu)^{q/2}\Gamma(\nu/2)} \left(1 + \frac{\mathbf{y}_0^\top \mathbf{y}_0}{\nu}\right)^{-(\nu+q)/2}$ is the pdf of the q -variate t distribution, $\mathbf{y}_0 = \boldsymbol{\Omega}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, $\alpha = \boldsymbol{\lambda}^\top \mathbf{y}_0$, and $T(\cdot|\nu)$ is the cdf of the standard univariate t distribution with ν degrees of freedom. A random vector \mathbf{Y} with the pdf (5) is denoted $\mathbf{Y} \sim \mathcal{ST}_q(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$ and has the stochastic representation [42]:

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}(\mathbf{d}Z_0 + \mathbf{Z}), \quad Z_0 \sim \mathcal{HN}(0, 1), \quad \mathbf{Z} \sim \mathcal{N}_q(0, \bar{\boldsymbol{\Omega}}) \quad (6)$$

where $U \sim \mathcal{Gamma}(\nu/2, \nu/2)$ with pdf $f_U(u_i|\nu) = (\nu/2)^{\nu/2} u_i^{\nu/2-1} \exp(-\nu u_i/2)/\Gamma(\nu/2)$ for $u_i > 0$, $\mathcal{HN}(0, 1)$ denotes a standard half normal variable, $\mathcal{N}_q(\boldsymbol{\mu}, \bar{\boldsymbol{\Omega}})$ denotes the q -variate normal distribution with pdf $\phi_q(\mathbf{x}|\boldsymbol{\mu}, \bar{\boldsymbol{\Omega}}) = |2\pi\bar{\boldsymbol{\Omega}}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \bar{\boldsymbol{\Omega}}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$ for $\mathbf{x} \in \mathbb{R}^q$, mean vector $\boldsymbol{\mu}$, and covariance matrix $\bar{\boldsymbol{\Omega}}$; $\bar{\boldsymbol{\Omega}} = \boldsymbol{\Omega} - \mathbf{d}\mathbf{d}^\top$, $\mathbf{d} = (1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda})^{-1/2} \boldsymbol{\Omega}^{1/2} \boldsymbol{\lambda}$; and Z_0 , \mathbf{Z} and U are independent. Setting $\tilde{U}_1 = (\frac{\nu}{2})^{1/2} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}$ for $\nu > 2$ and $c = \sqrt{2/\pi}$, the mean vector and the covariance matrix of \mathbf{Y} are respectively given by [42]:

$$\mathbb{E}\{\mathbf{Y}\} = \boldsymbol{\mu} + c\tilde{U}_1 \mathbf{d} \quad (7)$$

$$\text{Var}\{\mathbf{Y}\} = \frac{\nu}{\nu-2} \boldsymbol{\Omega} - c^2 \tilde{U}_1^2 \mathbf{d}\mathbf{d}^\top. \quad (8)$$

Figure 2 depicts univariate ST variables to illustrate the flexibility of the distribution. The multivariate skew t family indeed encompasses the multivariate t distributions ($\boldsymbol{\lambda} = \mathbf{0}$), the multivariate skew normal distributions ($\nu \rightarrow \infty$) and the multivariate normal distributions ($\boldsymbol{\lambda} = \mathbf{0}$ and $\nu \rightarrow \infty$).

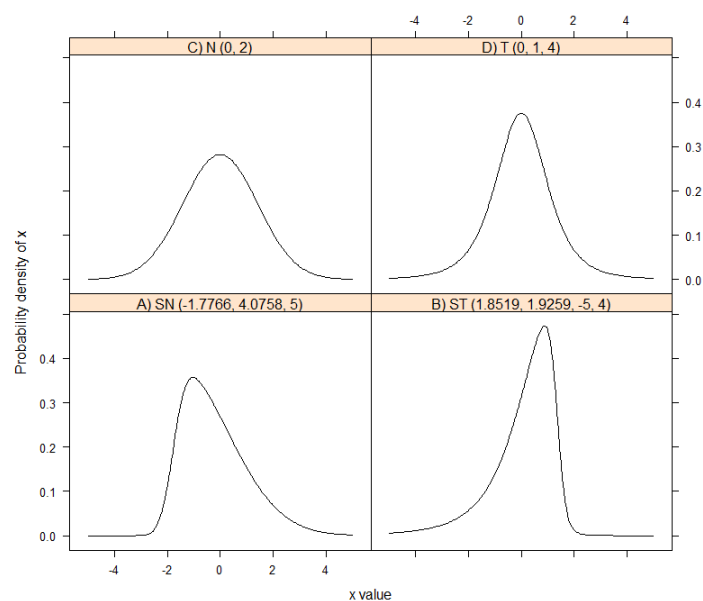


Figure 2. Probability density of (A) skew normal ($\lambda = 5$), (B) skew t ($\lambda = -5, \nu = 4$), (C) normal, and (D) Student t ($\nu = 4$) distributions all with null mean and variance 2

3. The skew t–discrete gamma model

We introduce in the section a flexible mixed count model which combines full dispersion flexibility in counts with a flexible distribution for random effects. We then describe a quadrature based procedure for maximum likelihood (ML) estimation of model parameters.

3.1. Model specification

Let us consider a population in which, the count response Y_{ij} from the j th measurement on an individual i has mean value μ_{ij} and a variance σ_{ij}^2 depending on μ_{ij} and a positive scalar a_i given a q -vector of random effects $\mathbf{b}_i = (b_{1i}, \dots, b_{qi})^\top$. The conditional mean count μ_{ij} is log-linearly related to a set of covariates (constants) $\mathbf{X}_{ij} = (1, x_{1ij}, \dots, x_{pij})^\top$ with fixed (population) effects and $\mathbf{W}_{ij} = (w_{1ij}, \dots, w_{qij})^\top$ with random (individual) effects as $\mu_{ij} = \exp(\eta_{ij})$ where $\eta_{ij} = \mathbf{X}_{ij}^\top \boldsymbol{\beta} + \mathbf{W}_{ij}^\top \mathbf{b}_i$. The $(p+1)$ -vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ contains population effects, $\beta_j \in \mathbb{R}$ representing the marginal effect of the covariate X_j on the log-average count, $\log \mu_i$, so that, conditional on \mathbf{b}_i , a unit change in X_{ij} induces *ceteris paribus* a $\beta_j\%$ variation in the mean count μ_i as in NP regression. In addition, the scalar a_i has the log-linear form $a_i = \exp(\mathbf{Z}_i^\top \boldsymbol{\delta})$ where $\mathbf{Z}_i = (1, z_{1i}, \dots, z_{ri})^\top$ and the $(r+1)$ -vector $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots, \delta_r)^\top$, with $\delta_j \in \mathbb{R}$, controls the dispersion of Y_{ij} conditional on \mathbf{b}_i .

For such a population, a skew t–discrete gamma (STDG) model is specified as:

$$Y_{ij} | \mathbf{b}_i \stackrel{\text{ind}}{\sim} \mathcal{DG}(\mu_{ij}, a_i) \quad \text{and} \quad \mathbf{b}_i \stackrel{\text{ind}}{\sim} \mathcal{ST}_q(-c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu) \quad (9)$$

where $c = \sqrt{2/\pi}$ and $\tilde{U}_1 = (\frac{\nu}{2})^{1/2} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}$, $\mathbf{D} = \bar{\mathbf{D}} + \mathbf{d}\mathbf{d}^\top$, $\boldsymbol{\lambda} = (1 + \mathbf{d}^\top \bar{\mathbf{D}}^{-1} \mathbf{d})^{1/2} \mathbf{D}^{-1/2} \mathbf{d}$, $\nu > 2$; $\mathbf{d} \in \mathbb{R}^q$, and $\bar{\mathbf{D}}$ is a positive definite scale matrix for random effects with lower triangular log-Cholesky factor \mathbf{R} , *i.e.* $\bar{\mathbf{D}} = \bar{\mathbf{R}}\bar{\mathbf{R}}^\top$ where $\bar{\mathbf{R}}$ has the same off diagonal elements as \mathbf{R} but has diagonal elements $\bar{\mathbf{R}}_{kk} = \exp(\mathbf{R}_{kk})$ (diagonal elements of $\bar{\mathbf{R}}$ constrained to be positive to ensure the uniqueness of \mathbf{R} for fixed $\bar{\mathbf{D}}$) [52].

The STDG model (9) has as special cases the Normal–discrete gamma (NDG) model obtained when $\mathbf{d} = \mathbf{0}$ and $\nu \rightarrow \infty$, the Student t–discrete gamma (TDG) model obtained with the restriction $\mathbf{d} = \mathbf{0}$, and the skew normal–discrete gamma (SNDG) model when $\mathbf{d} = \mathbf{0}$. Although the traditional Normal-Poisson mixed model [1] is not a special case of the model (9), it can be approximated with the special case obtained with the restrictions $\boldsymbol{\delta} = \mathbf{0}$, $\mathbf{d} = \mathbf{0}$ and $\nu \rightarrow \infty$ (Normal-discrete equidispersed gamma model). By (7), the random effects \mathbf{b}_i has a null mean vector, and on using (8), a covariance matrix

$$\boldsymbol{\Sigma}_b = \frac{\nu}{\nu-2} \bar{\mathbf{D}} + \left(\frac{\nu}{\nu-2} - c^2 \tilde{U}_1^2 \right) \mathbf{d}\mathbf{d}^\top. \quad (10)$$

3.2. Marginal likelihood

Given a sample of n independent individuals, let us denote $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$ the n_i -vector of the measurements from an individual i , $i = 1, 2, \dots, n$. The likelihood of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, \mathbf{d} , \mathbf{R} and ν is by Bayes' rule:

$$L(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b}_i | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu) d\mathbf{b}_i \quad (11)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top, \mathbf{d}^\top, \text{vech}(\mathbf{R})^\top, \nu)^\top$ with *vech* the usual vech operator which stacks the lower triangle elements of a matrix, $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top$, and since conditional on random effects \mathbf{b}_i , the individual measurements y_{ij} are independent, the joint conditional probability mass of \mathbf{y}_i is given by

$$f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) = \prod_{j=1}^{n_i} f_{bg}(y_{ij} | \mu_{ij}, a_i). \quad (12)$$

As for generalized linear mixed models in general, the q -dimensional integral in (11) is not analytically tractable and there is thus no closed form solution for the ML estimates for the fixed effects $\boldsymbol{\beta}$, the dispersion parameter $\boldsymbol{\delta}$, and the skewness parameter \mathbf{d} , the scale parameter \mathbf{R} and the degrees of freedom ν of random effects. We develop in the next section a pseudo-adaptive quadrature based estimation procedure.

3.3. Estimation based on pseudo-adaptive Gaussian quadrature

3.3.1. Likelihood approximation

Multi-dimensional Gaussian quadrature has proved useful for approximating multi-variable integrals of the form $\int_{\mathbb{R}^q} f(\mathbf{x}) \exp(\mathbf{x}^\top \mathbf{x}) d\mathbf{x}$ which emerges in the ML estimation of generalized linear mixed models [53,54]. To improve the accuracy of Gaussian quadrature while using a small number m of nodes per integration dimension, Pinheiro and Bates [55] have introduced adaptive Gaussian quadrature. However, this technique is impractical for models with complex structures such as the STDG, since it requires a repeated search of the mode of the integrand. We instead consider the pseudo-adaptive procedure introduced by Rizopoulos [56] where \mathbf{b}_i is centered and rescaled about its empirical Bayes estimates obtained from assuming normal random effects [56–59]. In this setting, we make the following change of variable:

$$\mathbf{b}_i = \tilde{\mathbf{b}}_i + \tilde{\mathbf{R}}_i \mathbf{r}_i \quad (13)$$

where, $\mathbf{r}_i \in \mathbb{R}^q$; $\tilde{\mathbf{b}}_i$ is the empirical Bayes estimate of \mathbf{b}_i and $\tilde{\mathbf{R}}_i$ is the lower triangular Cholesky factor of the covariance matrix estimate of \mathbf{b}_i given \mathbf{y}_i , obtained by fitting the NP model using e.g. the *glmer* routine of the package *lme4* [60] in the R freeware [2]. The likelihood function (11) can then be rewritten as

$$L(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n \int_{\mathbb{R}^q} g_y(\mathbf{y}_i | \mathbf{r}_i, \boldsymbol{\theta}) \phi_q(\mathbf{r}_i | \mathbf{0}, \mathbf{I}_q) d\mathbf{r}_i \quad (14)$$

where we have set

$$g_y(\mathbf{y}_i | \mathbf{r}_i, \boldsymbol{\theta}) = \left\{ \prod_{j=1}^{n_i} f_{bg}(y_{ij} | \mu_{ij}, a_i) \right\} |\tilde{\mathbf{R}}_i| \frac{St_q(\tilde{\mathbf{b}}_i + \tilde{\mathbf{R}}_i \mathbf{r}_i | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu)}{\phi_q(\mathbf{r}_i | \mathbf{0}, \mathbf{I}_q)}.$$

On using (14), the pseudo-adaptive Gaussian quadrature based approximation to the logarithm of the likelihood function (11) is

$$\hat{\ell}_0(\boldsymbol{\theta} | \mathbf{y}) = \sum_{i=1}^n \log \left[\sum_{k_1=1}^m \cdots \sum_{k_q=1}^m w_k g_y(\mathbf{y}_i | \mathbf{r}_k, \boldsymbol{\theta}) \right] \quad (15)$$

where m is a pre-specified number of nodes per integration dimension, and $w_k = \prod_{j=1}^q w_{kj}$ and $\mathbf{r}_k = (r_{k1}, r_{k2}, \dots, r_{kq})^\top$ are, respectively, the multidimensional integration nodes and the associated weights. The components r_{kj} and w_{kj} are available for instance in the R package *fastGHQuad* [61]. However, the total number of evaluations of the function $g_y(\mathbf{y}_i | \cdot, \boldsymbol{\theta})$ required in (15) increases exponentially in the number of random effects (m^q function evaluations). In order to further reduce the curse of dimensionality when $q \geq 3$, we consider sparse-grid integration based on the Smolyak rule [62]. The Smolyak rule is

a generic approach for extending any univariate quadrature rule to multivariate settings without the exponentially growing computational cost of the product rule in (15). It has been successfully used, for instance, on likelihood related estimation problems in econometrics [63–65], on Bayesian inverse problems [66] and in general optimization theory [67]. The advantage of this quadrature rule resides in the number of nodes which increases only polynomially in the number of dimensions [64]. Instead of the number m of nodes used per dimension in (15), the Smolyak rule requires an accuracy level $\varsigma \in \mathbb{N}$ which then determines the total number $Q(\varsigma)$ of nodes. The sparse-grid pseudo-adaptive quadrature based approximation to the logarithm of the likelihood function (11) is

$$\hat{\ell}_s(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log \left[\sum_{k=1}^{Q(\varsigma)} w_k g_y(\mathbf{y}_i|\mathbf{r}_k, \boldsymbol{\theta}) \right]. \quad (16)$$

Although the sparse-grid integration nodes \mathbf{r}_k and the associated weights w_k are not trivial to determine, they are available in the R package *SparseGrid* [64] and the total number $Q(\varsigma)$ of nodes increases only polynomially ($Q(m) < m^q$ for $q \geq 3$).

3.3.2. Maximum likelihood estimates, standard errors and model selection

The maximum likelihood estimate $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\delta}}^\top, \hat{\mathbf{d}}^\top, \text{vech}(\hat{\mathbf{R}}), \hat{\nu} \right)^\top$ of $\boldsymbol{\theta}$ is obtained by maximizing the (sparse-grid) pseudo-adaptive Gaussian approximation $\hat{\ell}$ to the log-likelihood function (11), using a numerical optimization software such as the *optim* function in R. Although such softwares run faster when analytical derivative of the log-likelihood function are supplied, we follow [53] and avoid the use of this option since numerical integration will be required for these derivatives too and they are unfortunately often more poorly approximated by quadrature based on the same nodes used for approximating the marginal likelihood itself [68]. It is worth noting that the approximate log-likelihood function may be unbounded in the degrees of freedom parameter ν . This can occur in Student t distribution related likelihood based inference [69], preventing the convergence of optimization routines and usually resulting in non-negative definite hessian of the log-likelihood function. In the event of such unboundedness issues, we advocate the profile likelihood approach described by [70]. It consists in setting a grid of feasible values of ν (e.g. $\nu_k = \nu_{k-1} + 0.5$; $k = 1, 2, \dots, 55$; $\nu_0 = 2.5$) and obtaining a sequence of estimates $\tilde{\boldsymbol{\theta}}_k$ while keeping $\nu = \nu_k$ during each maximization of (11). Then, the vector $\tilde{\boldsymbol{\theta}}_k$ which maximizes (11) is taken as the ML estimate $\hat{\boldsymbol{\theta}}$.

After the maximization of (11), the ML estimate of the scale matrix $\hat{\mathbf{D}}$ can be recovered as $\hat{\mathbf{D}} = \hat{\mathbf{R}}\hat{\mathbf{R}}^\top$ with $\hat{\mathbf{R}}$ the matrix $\hat{\mathbf{R}}$ with exponentiated diagonal elements (see (9)). The ML estimate of the random effects covariance matrix is then given by

$$\hat{\boldsymbol{\Sigma}}_b = \frac{\hat{\nu}}{\hat{\nu}-2} \hat{\mathbf{R}}\hat{\mathbf{R}}^\top + \left(\frac{\hat{\nu}}{\hat{\nu}-2} - c^2 \hat{U}_1 \right) \hat{\mathbf{d}}\hat{\mathbf{d}}^\top. \quad (17)$$

where $\hat{U}_1 = \left(\frac{\hat{\nu}}{2} \right)^{1/2} \frac{\Gamma(\frac{\hat{\nu}-1}{2})}{\Gamma(\hat{\nu}/2)}$. Interestingly, many optimization routines (e.g. *optim* in R) computes, on request, the hessian matrix $H_n(\boldsymbol{\theta})$ of the log-likelihood function evaluated at $\hat{\boldsymbol{\theta}}$. For inference on the model parameters, we assume that $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ asymptotically follows a normal distribution with null mean and covariance matrix $\hat{\mathbf{C}}_n(\hat{\boldsymbol{\theta}}) = -[H_n(\hat{\boldsymbol{\theta}})]^{-1}$.

The appropriate distribution (normal, Student t, skew normal or skew t) for random effects is *a priori* unknown for a specific dataset. In practice, the best fit can be determined after fitting the full version of the STDG model (9) as well as its three special cases (normal with the constraints $\mathbf{d} = \mathbf{0}$ and $\nu \rightarrow \infty$; Student t with $\mathbf{d} = \mathbf{0}$; and skew normal with $\nu \rightarrow \infty$). Indeed, when the boundary restriction $\nu \rightarrow \infty$ is not involved, an appropriate statistical test (e.g. Wald test [71] or likelihood ratio test [72]) can be used to identify the

best fit to a specific dataset. When on the contrary the restriction $\nu \rightarrow \infty$ is involved, we consider information criteria such as the Akaike's Information Criterion given for a fit with N_p estimated model parameters by $AIC = -2\ell(\hat{\boldsymbol{\theta}}) + 2N_p$ (the lower, the better) [73]. When the best fit to a specific dataset involves skew t ($\nu < \infty$) or Student t ($\boldsymbol{d} = \mathbf{0}$ and $\nu < \infty$) random effects, this fit can be used to identify possible outliers in the dataset (see Appendix A).

3.4. Estimation of random effects

In the Gaussian linear mixed models framework, Best Linear Unbiased Predictors (BLUPs) of random effects are routinely used in application fields like ecology, evolutionary biology and genetics [74,75] where they have become very important, for instance, for breeding values based selection. However, the validity of the BLUPs relies on, at least, an approximate adequation between the studied population and the assumed model [75–78]. BLUPs are equivalent to Empirical Bayes (EB) estimates in Gaussian models. As a result, in models with more complex structures designed to capture more latent data features than the traditional Gaussian model, EB estimates are developed as appropriate alternatives to BLUPs to give an overview on latent unobservable subject-specific quantities such as intercept and slopes [79–81]. Based on the STDG model (9), we provide in this section expressions for predicting random effects. From (11), the conditional density of the random effects \mathbf{b}_i given the observed data is

$$f_{b_i}(\mathbf{b}_i | \mathbf{y}_i, \boldsymbol{\theta}) = \frac{1}{f_{y_i}(\mathbf{y}_i)} f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b}_i | -c\tilde{U}_1 \boldsymbol{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu)$$

where $f_{y_i}(\mathbf{y}_i) = \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b}_i | -c\tilde{U}_1 \boldsymbol{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu) d\mathbf{b}_i$ is the contribution of the individual i to the sample likelihood (11). The minimum mean-squared error estimator of \mathbf{b}_i correspond to its conditional mean given \mathbf{y}_i [82] and can thus be approximated by a weighted average as

$$\bar{\mathbf{b}}_i(\boldsymbol{\theta}) = [\hat{f}_{y_i}(\mathbf{y}_i)]^{-1} \sum_{k=1}^{Q(s)} \mathbf{b}_{ik} w_k g_y(\mathbf{y}_i | \mathbf{r}_k, \boldsymbol{\theta}) \quad (18)$$

where $\mathbf{b}_{ik} = \tilde{\mathbf{b}}_i + \tilde{\mathbf{R}}_i \mathbf{r}_k$ and $\hat{f}_{y_i}(\mathbf{y}_i) = \sum_{k=1}^{Q(s)} w_k g_y(\mathbf{y}_i | \mathbf{r}_k, \boldsymbol{\theta})$. Similarly, the conditional covariance matrix of \mathbf{b}_i can be approximated as

$$\Sigma_{\mathbf{b}_i}(\boldsymbol{\theta}) = [\hat{f}_{y_i}(\mathbf{y}_i)]^{-1} \sum_{k=1}^{Q(s)} \mathbf{b}_{ik} \mathbf{b}_{ik}^\top w_k g_y(\mathbf{y}_i | \mathbf{r}_k, \boldsymbol{\theta}) - \bar{\mathbf{b}}_i(\boldsymbol{\theta}) \bar{\mathbf{b}}_i^\top(\boldsymbol{\theta}) \quad (19)$$

The EB estimator of \mathbf{b}_i is then obtained by substituting the ML estimate of $\boldsymbol{\theta}$ into (18), i.e. $\hat{\mathbf{b}}_i = \bar{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$. The covariance of these estimates are also obtained by substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ in (19). However, this error covariance matrix does not account for errors due to the estimation of $\hat{\boldsymbol{\theta}}$. This indicates that variances extracted from this estimated covariance matrix tend to underestimate the true error variances [80].

4. Application to the owl sibling negotiation data

4.1. Data and statistical analyses

To illustrate the use of the proposed model, we analyse in this section the owl sibling negotiation data from Roulin and Bersier [83]. This popular data is relative to the vocal begging behaviour before the owl parents bring prey to their nest. Young owls indeed use begging in the absence of parents to share information on the different levels of hunger between them, so that to save energy to compete for food when there is the highest probability of success, the least hungry young owl avoids competing for food against the

hungriest owls when a parent brings a prey [84]. Roulin and Bersier [83] investigated in nestling barn owls whether the intensity of vocal begging behaviour differs in the presence of the father as compared to the presence of the mother.

The interest response variable is sibling negotiation which is defined as the number of calls made by all offspring in the nearest 30 seconds interval before the arrival of a parent. The explanatory variables are the sex of the parent, the arrival time of the parent, and a food treatment defined as: half of nests were given extra prey ("food-satiated"), and prey were removed from the other half ("food-deprived"). The data collected from $n = 27$ nests is available in the R package *glmmADMB* [85] and some descriptive statistics are given in Appendix B. From analyses performed by Zuur et al. [84], the response variable (sibling negotiation) is overdispersed ($ID = 6.6$) and correlated within nest ($\rho = 0.51$ between two sequential parent arrivals). A dispersion model is indeed required to adequately model sibling negotiation. On the contrary, the need of a flexible random effect distribution is still undocumented. Following Zuur et al. [84], we consider the random intercept mixed effects model with conditional mean count μ_{ij} defined through

$$\begin{aligned} \log \mu_{ij} = & \log BroodSize_{ij} + \beta_0 + \beta_1 \times SexParent_{ij} + \beta_2 \times FoodTreatment_{ij} \\ & + \beta_3 \times ArrivalTime_{ij} + \beta_4 \times SexParent_{ij} \times FoodTreatment_{ij} \\ & + \beta_5 \times SexParent_{ij} \times ArrivalTime_{ij} + b_i \end{aligned} \quad (20)$$

where ij indexes observation j from a nest i , $BroodSize$ is the number of nestlings (used as a model offset) and b_i is the random intercept associated to all observations from nest i .

For comparison purposes, we fitted in addition to the STDG model (9) and its special cases including the NDG (Normal-discrete gamma), the TDG (Student t-discrete gamma) and the SNDG (Skew normal-discrete gamma) models; the Normal-Poisson (NP) and the Normal-negative binomial (NNB) models to the owl sibling negotiation data. The STDG model and its special cases were fitted following the procedures described in Section 3 and implemented in R software (see Appendix C for some linear algebra details on the implementation). The NP and NNB models were fitted using the R package *lme4* [60]. The significance level for statistical tests was set to 0.05.

4.2. Results

Tables 1-3 present the estimates of population parameters and AIC values for the six fitted models. Based on the AIC values in Tables 1-2, it appears that with normally distributed random effects, the discrete gamma model (AIC = 3420.8) is better than the Poisson (AIC = 5020.3) and the negative binomial (AIC = 3472.9) mixed models for the owl sibling negotiation data. There are important differences between estimates of fixed effects from these three models, but there is no switch in the 5% statistical significance of fixed effects. More importantly, the use of the Poisson distribution results in a random intercept variance estimate of $\hat{\sigma}_b^2 = 0.2$ whereas the use of flexible count distributions results in much lower random intercept variance estimates ($< 10^{-4}$). This indicates that much of the between-nests variability identified by the NP model is actually a within-nest variability.

Further considering discrete gamma models with flexible distributions for random effects, the AIC values in Tables 2-3, indicate that the skew normal model (AIC = 3394.1) provides the best fit as compared to the normal, the Student t (AIC = 3422.6) and the skew t models (AIC = 3396.1). It can be observed that the estimate ($\hat{\delta}_0 = -2.23$) of the skewness parameter of the nest related random intercept is significantly different from zero as indicated by the associated Wald test (p-value < 0.001). On the contrary, the random intercept distribution is not heavy tailed as indicated by the large degrees of freedom parameter estimates for both the Student t and the skew t models ($\hat{\nu} > 9987$). It can be noted that there are only slight and non significant differences between estimates of both fixed effects and random intercept variance from these three models. Based on the skew normal-discrete gamma model fit, it appears that sibling negotiation was significantly lower (82%) for food satiated young owls as compared to food deprived young owls, and each unit

Table 1: Estimated coefficients, standard errors (SE) and Akaike's Information Criterion (AIC) for Normal-Poisson (NPoisson) and Normal-negative binomial (NNB) mixed regression model fits to the owl sibling negotiation dataset

| Parameter | NP(AIC = 5020.31) | | | | NNB (AIC = 3472.92) | | | |
|---------------------|-------------------|-------|---------|---------|---------------------|------------------|---------|---------|
| | Estimate | SE | z-value | p-value | Estimate | SE | z-value | p-value |
| Intercept | 0.668 | 0.036 | 18.634 | <0.001 | 0.635 | 0.107 | 5.923 | <0.001 |
| SP=M | -0.014 | 0.042 | -0.320 | 0.749 | 0.057 | 0.136 | 0.417 | 0.677 |
| Treat.=S | -0.667 | 0.054 | -12.300 | <0.001 | -0.610 | 0.153 | -3.979 | <0.001 |
| AT | -0.119 | 0.014 | -8.395 | <0.001 | -0.134 | 0.038 | -3.510 | <0.001 |
| SP=M*Treat.=S | 0.132 | 0.069 | 1.925 | 0.054 | 0.155 | 0.199 | 0.780 | 0.436 |
| SP=M*AT | -0.016 | 0.018 | -0.901 | 0.367 | 0.026 | 0.050 | 0.515 | 0.606 |
| δ_0 | | | | | -0.211 | 0.076 | -2.795 | 0.005 |
| σ_b^2 (nest) | 0.199 | 0.058 | | | 2.310 ϵ | 0.575 ϵ | | |

Table notes: SP = Sex of parent; M = Male; Treat. = Treatment; S = Satiated; AT = Arrival time centered around its median value; δ_0 = natural logarithm of the dispersion parameter of the negative binomial family; $\epsilon = 10^{-5}$.

Table 2: Estimated coefficients, standard errors (SE), dispersion parameter (δ_0) and Akaike's Information Criterion (AIC) for Normal-discrete gamma (NDG) and Student t-discrete gamma (TDG) mixed regression model fits to the owl sibling negotiation dataset

| Parameter | NDG (AIC = 3420.83) | | | | TDG (AIC = 3422.63) | | | |
|---------------------|---------------------|------------------|---------|---------|---------------------|------------------|---------|---------|
| | Estimate | SE | z-value | p-value | Estimate | SE | z-value | p-value |
| Intercept | 0.637 | 0.086 | 7.397 | <0.001 | 0.637 | 0.086 | 7.392 | <0.001 |
| SP=M | 0.188 | 0.099 | 1.890 | 0.059 | 0.188 | 0.100 | 1.885 | 0.059 |
| Treat.=S | -0.839 | 0.136 | -6.170 | 0.000 | -0.838 | 0.136 | -6.163 | <0.001 |
| AT | -0.130 | 0.038 | -3.469 | 0.001 | -0.131 | 0.037 | -3.505 | <0.001 |
| SP=M*Treat.=S | 0.039 | 0.172 | 0.228 | 0.820 | 0.038 | 0.172 | 0.222 | 0.824 |
| SP=M*AT | -0.005 | 0.046 | -0.108 | 0.914 | -0.005 | 0.046 | -0.108 | 0.914 |
| δ_0 | -2.233 | 0.080 | -28.016 | <0.001 | -2.233 | 0.080 | -28.015 | <0.001 |
| ν | | | | | 9992.62 | 2351.12 | | |
| σ_b^2 (nest) | 2.300 U | 0.573 ϵ | | | 2.300 ϵ | 0.573 ϵ | | |

Table notes: SP = Sex of parent; M = Male; Treat. = Treatment; S = Satiated; AT = Arrival time centered around its median value; δ_0 = dispersion parameter; $\epsilon = 10^{-5}$.

increase in the arrival time reduced sibling negotiation by 13% on average. But the sex of the owl parent did not affect significantly sibling negotiation.

Table 3: Estimated coefficients, standard errors (SE), dispersion parameter (δ_0) and Akaike's Information Criterion (AIC) for skew normal-discrete gamma (SNDG) and skew t-discrete gamma (STDG) mixed regression model fits to the owl sibling negotiation dataset

| Parameter | SNDG (AIC = 3394.08) | | | | STDG (AIC = 3396.06) | | | |
|---------------------|----------------------|-----------|---------|---------|----------------------|------------------|---------|---------|
| | Estimate | SE | z-value | p-value | Estimate | SE | z-value | p-value |
| Intercept | 0.636 | 0.086 | 7.354 | <0.001 | 0.637 | 0.086 | 7.392 | <0.001 |
| SP=M | 0.189 | 0.099 | 1.898 | 0.058 | 0.187 | 0.109 | 1.716 | 0.086 |
| Treat.=S | -0.824 | 0.138 | -5.970 | <0.001 | -0.827 | 0.136 | -6.083 | <0.001 |
| AT | -0.130 | 0.037 | -3.478 | 0.001 | -0.131 | 0.038 | -3.492 | 0.000 |
| SP=M*Treat.=S | 0.039 | 0.173 | 0.227 | 0.821 | 0.039 | 0.172 | 0.228 | 0.820 |
| SP=M*AT | -0.005 | 0.047 | -0.106 | 0.915 | -0.005 | 0.046 | -0.107 | 0.915 |
| δ_0 | -2.231 | 0.079 | -28.256 | <0.001 | -2.230 | 0.079 | -28.071 | <0.001 |
| d | -0.012 | 0.006 | -2.162 | 0.031 | -0.010 | 0.005 | -1.869 | 0.062 |
| ν | | | | | 9987.84 | 854.83 | | |
| σ_b^2 (nest) | 1.210 ϵ | 0.017 U | | | 0.991 ϵ | 0.012 ϵ | | |

Table notes: SP = Sex of parent; M = Male; Treat. = Treatment; S = Satiated; AT = Arrival time centered around its median value; δ_0 = dispersion parameter; $\epsilon = 10^{-5}$.

5. Conclusion

Ignoring data dispersion or shape may lead to biased, inefficient and inconsistent estimates of population parameters in mixed count regression models. This work introduces a large class of mixed effects models for nested, clustered or longitudinal count data. The

proposal combines flexible count distribution with a flexible distribution for random effects, and estimates model parameters using pseudo-adaptive Gaussian quadrature. The application on owl sibling negotiation data demonstrates the usefulness of the approach in detecting and controlling dispersion in counts and skewness in random effects distribution. Further investigations will target the performance of the proposal when data comes from various count and random effects distributions in use in agricultural, environmental, biomedical, economic, and social sciences.

Author Contributions: Conceptualization, C.F.T.; methodology, C.F.T. and R.G.K.; software, C.F.T.; validation, C.F.T. and R.G.K.; formal analysis, C.F.T.; resources, R.G.K.; writing—original draft preparation, C.F.T.; writing—review and editing, C.F.T. and R.G.K.; visualization, C.F.T.; supervision, R.G.K.; project administration, R.G.K.; funding acquisition, R.G.K. All authors read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Appendix A. Individual-specific weights

Appendix A.1. Estimation of individual weights for outlier detection

When the best fit to a specific dataset involves skew t ($\nu < \infty$) or Student t ($\mathbf{d} = \mathbf{0}$ and $\nu < \infty$) random effects, this fit can be used to identify possible outliers in the dataset. Indeed, by (6), the STDG model (9) has the stochastic representation

$$\begin{aligned} Y_{ij} | \mathbf{b}_i, U_i = u_i, V_i = v_i &\stackrel{ind}{\sim} \mathcal{DG}(\mu_{ij}, \kappa_i) \\ \mathbf{b}_i | U_i = u_i, V_i = v_i &\stackrel{ind}{\sim} \mathcal{N}_q \left((v_i u_i^{-1/2} - c\tilde{U}_1) \mathbf{d}, u_i^{-1} \mathbf{D} \right) \\ U_i | V_i = v_i &\stackrel{ind}{\sim} \mathcal{Gamma}(\nu/2, \nu/2) \\ V_i &\stackrel{ind}{\sim} \mathcal{HN}(0, 1). \end{aligned} \quad (\text{A1})$$

The mixing variable U_i in (A1) is an individual-specific stochastic weight which serves for outlier detection using models built under the Student t and related distributions [43,81,86]. Using (A1), the conditional mean of U_i given \mathbf{y}_i turns out to be (see (A5))

$$\mathbb{E}\{U_i | \mathbf{y}_i, \boldsymbol{\theta}\} = \frac{1}{f_{\mathbf{y}_i}(\mathbf{y}_i)} \int_{\mathbb{R}^q} f_{\mathbf{y}}(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b}_i | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}^*, \boldsymbol{\lambda}, \nu + 2) d\mathbf{b}_i \quad (\text{A2})$$

where $\mathbf{D}^* = \frac{\nu}{\nu+2} \mathbf{D}$. Thus, on setting

$$g_{\mathbf{y}}^*(\mathbf{y}_i | \mathbf{r}_i, \boldsymbol{\theta}) = \left\{ \prod_{j=1}^{n_i} f_{bg}(y_{ij} | \mu_{ij}, a_i) \right\} |\tilde{\mathbf{R}}_i| \frac{St_q(\tilde{\mathbf{b}}_i + \tilde{\mathbf{R}}_i \mathbf{r}_i | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}^*, \boldsymbol{\lambda}, \nu + 2)}{\phi_q(\mathbf{r}_i | \mathbf{0}, \mathbf{I}_q)},$$

the conditional mean $\mathbb{E}\{U_i | \mathbf{y}_i, \boldsymbol{\theta}\}$ can be approximated as

$$\bar{U}_i(\boldsymbol{\theta}) = [\hat{f}_{\mathbf{y}_i}(\mathbf{y}_i)]^{-1} \sum_{k=1}^{Q(\varsigma)} \mathbf{b}_{ik} w_k g_{\mathbf{y}}^*(\mathbf{y}_i | \mathbf{r}_k, \boldsymbol{\theta}). \quad (\text{A3})$$

The EB estimator of the weight U_i of individual i is then given by $\hat{u}_i = \bar{U}_i(\hat{\boldsymbol{\theta}})$. Unusual or outlying individuals are indicated by relatively low weights ($\hat{u}_i < 1$) [43,81].

Appendix A.2. Conditional moments of stochastic weights

We derive the more general result $\mathbb{E}\{U_i^t | \mathbf{y}_i, \boldsymbol{\theta}\} = \int_0^\infty u_i^t f_{U_i | \mathbf{Y}_i}(u_i | \mathbf{y}_i) du_i$ so that the weight $\mathbb{E}\{U_i | \mathbf{y}_i, \boldsymbol{\theta}\}$ can be obtained as the special case when $t = 1$. We have by Bayes's rule:

$$\begin{aligned}
\mathbb{E}\{U_i^t | \mathbf{y}_i, \boldsymbol{\theta}\} &= \frac{1}{f_{\mathbf{y}_i}(\mathbf{y}_i)} \int_0^\infty u_i^t \int_{\mathbb{R}^q} f_{\mathbf{Y}_i, \mathbf{b}, U_i}(\mathbf{y}_i, \mathbf{b}, u_i) d\mathbf{b} du_i \\
&= \frac{1}{f_{\mathbf{y}_i}(\mathbf{y}_i)} \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\delta}) \int_0^\infty u_i^t f_{\mathbf{b}, U_i}(\mathbf{b}, u_i) du_i d\mathbf{b} \\
&= \frac{1}{f_{\mathbf{y}_i}(\mathbf{y}_i)} \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\delta}) f_b(\mathbf{b} | \mathbf{d}, \bar{\mathbf{D}}, \nu) \int_0^\infty u_i^t f_{U_i | \mathbf{b}}(u_i | \mathbf{b}) du_i d\mathbf{b} \\
\mathbb{E}\{U_i^t | \mathbf{y}_i, \boldsymbol{\theta}\} &= \frac{1}{f_{\mathbf{y}_i}(\mathbf{y}_i)} \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\delta}) f_b(\mathbf{b} | \mathbf{d}, \bar{\mathbf{D}}, \nu) u_t d\mathbf{b} \quad (\text{A4})
\end{aligned}$$

where $f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta})$ is given by (12), $f_b(\mathbf{b} | \mathbf{d}, \bar{\mathbf{D}}, \nu) = St_q(\mathbf{b} | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu)$ is given by (5), $f_{\mathbf{y}_i}(\mathbf{y}_i) = \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b}_i | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \boldsymbol{\lambda}, \nu) d\mathbf{b}_i$ and $u_t = \mathbb{E}\{U_i^t | \mathbf{b}_i\}$. Using the stochastic representation (A1), the conditional moment u_t is given by [45]

$$u_t = \frac{t_q(\mathbf{b} | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}, \nu)}{f_b(\mathbf{b} | \mathbf{d}, \bar{\mathbf{D}}, \nu)} \frac{2\Gamma\left(\frac{\nu+2t+q}{2}\right)}{\left(\frac{\nu}{2}\right)^t \Gamma\left(\frac{\nu+q}{2}\right)} \left(1 + \frac{\mathbf{b}_0^\top \mathbf{b}_0}{\nu}\right)^{-t} T\left(\alpha \sqrt{\frac{\nu+2t+q}{\nu + \mathbf{b}_0^\top \mathbf{b}_0}} \mid \nu + 2t + q\right)$$

where $\mathbf{b}_0 = \mathbf{D}^{-1/2}(\mathbf{b} + c\tilde{U}_1 \mathbf{d})$ and $\alpha = \boldsymbol{\lambda}^\top \mathbf{b}_0$. Replacing the last expression in (A4) and setting $\mathbf{D}^* = \frac{\nu}{\nu+2t} \mathbf{D}$ straightforwardly leads to

$$\mathbb{E}\{U_i^t | \mathbf{y}_i, \boldsymbol{\theta}\} = \frac{\left(\frac{\nu}{2}\right)^{-t} \Gamma\left(\frac{\nu+2t}{2}\right)}{f_{\mathbf{y}_i}(\mathbf{y}_i) \Gamma\left(\frac{\nu}{2}\right)} \int_{\mathbb{R}^q} f_y(\mathbf{y}_i | \mathbf{b}, \boldsymbol{\beta}, \boldsymbol{\delta}) St_q(\mathbf{b} | -c\tilde{U}_1 \mathbf{d}, \mathbf{D}^*, \boldsymbol{\lambda}, \nu + 2t) d\mathbf{b}. \quad (\text{A5})$$

Setting $t = 1$ in (A5) yields (A2).

Appendix B. Summary statistics for the owl sibling negotiation data

Table 4 presents some descriptive statistics for the owl sibling negotiation data.

Table 4: Summary statistics for the owl sibling negotiation data (N = 599)

| Categorical variables | | | Numerical variables | | | | |
|-----------------------|---------------|----------------|----------------------|------|------|------|-----|
| Variable | Category | Percentage (%) | Variable | Min | Max | Mean | SD |
| Sex of parent | Female | 40.9 | Brood size | 1 | 7 | 4.4 | 1.2 |
| | Male | 59.1 | Arrival time (hours) | 21.7 | 29.3 | 24.8 | 1.9 |
| Treatment | Food deprived | 53.4 | sibling negotiation | 0 | 32 | 6.7 | 6.7 |
| | Food satiated | 46.6 | | | | | |

Notes: Min = minimum; Max = maximum, SD = standard deviation.

Appendix C. Linear algebra details

This Appendix gives some linear algebra details useful for the implementation of models based on the skew t distribution defined by (5). Indeed, the skew t pdf for $\mathbf{y} \in \mathbb{R}^q$ involves the quantities $\mathbf{y}_0^\top \mathbf{y}_0$, $\alpha = \boldsymbol{\lambda}^\top \mathbf{y}_0$ and $|\mathbf{D}|^{1/2}$ where $\mathbf{y}_0 = \mathbf{D}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$. For the STDG model (9), we additionally have $\mathbf{D} = \bar{\mathbf{D}} + \mathbf{d}\mathbf{d}^\top$, $\boldsymbol{\lambda} = (1 + \mathbf{d}^\top \bar{\mathbf{D}}^{-1} \mathbf{d})^{1/2} \mathbf{D}^{-1/2} \mathbf{d}$, and $\bar{\mathbf{D}} = \bar{\mathbf{R}}\bar{\mathbf{R}}^\top$. We thus have by the Sherman-Morrison identity [87],

$$\mathbf{D}^{-1} = \bar{\mathbf{D}}^{-1} - \frac{1}{1 + \mathbf{d}^\top \bar{\mathbf{D}} \mathbf{d}} \bar{\mathbf{D}}^{-1} \mathbf{d} \mathbf{d}^\top \bar{\mathbf{D}}^{-1} \quad (\text{A6})$$

$$= \bar{\mathbf{R}}^{-\top} \left[\mathbf{I}_q - \frac{\bar{\mathbf{R}}^{-1} \mathbf{d} (\bar{\mathbf{R}}^{-1} \mathbf{d})^\top}{1 + (\bar{\mathbf{R}}^{-1} \mathbf{d})^\top \bar{\mathbf{R}}^{-1} \mathbf{d}} \right] \bar{\mathbf{R}}^{-1}. \quad (\text{A7})$$

This results on setting $\mathbf{z} = \bar{\mathbf{R}}^{-1} \mathbf{y}_0$ and $\mathbf{z}_\alpha = \bar{\mathbf{R}}^{-1} \mathbf{d}$ in

$$\mathbf{y}_0^\top \mathbf{y}_0 = \mathbf{z}^\top \mathbf{z} - \left(\mathbf{z}^\top \mathbf{z}_\alpha \right)^2 / \left(1 + \mathbf{z}_\alpha^\top \mathbf{z}_\alpha \right) \quad (\text{A8})$$

$$\alpha = \left[\left(1 + \mathbf{z}_\alpha^\top \mathbf{z}_\alpha \right)^{1/2} - \mathbf{z}_\alpha^\top \mathbf{z}_\alpha \right] \mathbf{z}^\top \mathbf{z}_\alpha. \quad (\text{A9})$$

Likely, from the identity $\mathbf{D} = \bar{\mathbf{D}} + \mathbf{d} \mathbf{d}^\top = \bar{\mathbf{R}} (\mathbf{I}_q + \bar{\mathbf{R}}^{-1} \mathbf{d} \mathbf{d}^\top \bar{\mathbf{R}}^{-\top}) \bar{\mathbf{R}}$, the determinant of \mathbf{D} is given by $|\mathbf{D}| = |\bar{\mathbf{R}}|^2 \times |\mathbf{I}_q + \bar{\mathbf{R}}^{-1} \mathbf{d} \mathbf{d}^\top \bar{\mathbf{R}}^{-\top}|$. Hence, by the Sherman-Morrison formula for the determinant [87],

$$|\mathbf{D}|^{1/2} = |\bar{\mathbf{R}}| \sqrt{1 + \mathbf{z}_\alpha^\top \mathbf{z}_\alpha}. \quad (\text{A10})$$

Finally, note that $|\bar{\mathbf{R}}|$ is the product of the diagonal elements of $\bar{\mathbf{R}}$ (since $\bar{\mathbf{R}}$ is lower triangular).

References

1. McCulloch, C.E.; Searle, S.R. *Generalized, Linear, and Mixed Models*, 1 ed.; Vol. 1, *Wiley Series in Probability and Statistics*, Wiley-Interscience, 2001.
2. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
3. StataCorp. *Stata Statistical Software: Release 16*. StataCorp LLC, Texas, USA, 2019.
4. Institute, S. *SAS user's guide: Statistics*, 1985.
5. McCulloch, C.E.; Neuhaus, J.M. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical science* **2011**, *28*, 388–402.
6. Jorgensen, B. *The theory of dispersion models*; CRC Press, 1997.
7. Jowaheer, V. Model misspecification effects in clustered count data analysis. *Statistics & Probability Letters* **2006**, *76*, 470–478.
8. Jeong, K.M. Effects on Regression Estimates under Misspecified Generalized Linear Mixed Models for Counts Data. *The Korean Journal of Applied Statistics* **2012**, *25*, 1037–1047. doi:http://dx.doi.org/10.5351/KJAS.2012.25.6.1037.
9. Gad, A.M.; El Kholy, R.B. Generalized linear mixed models for longitudinal data. *International Journal of Probability and Statistics* **2012**, *1*, 41–47.
10. Hilbe, J.M. *Negative binomial regression*; Cambridge University Press, 2011; p. 573.
11. Sankaran, M. 275. note: The discrete poisson-lindley distribution. *Biometrics* **1970**, *26*, 145–149.
12. Conway, R.W.; Maxwell, W.L. A queuing model with state dependent service rates. *Journal of Industrial Engineering* **1962**, *12*, 132–136.
13. Shmueli, G.; Minka, T.P.; Kadane, J.B.; Borle, S.; Boatwright, P. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **2005**, *54*, 127–142.
14. Huang, A. Mean-parametrized Conway–Maxwell–Poisson regression models for dispersed counts. *Statistical Modelling* **2017**, *17*, 359–380.
15. Efron, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **1986**, *81*, 709–721.
16. Zou, Y.; Geedipally, S.R.; Lord, D. Evaluating the double Poisson generalized linear model. *Accident Analysis & Prevention* **2013**, *59*, 497–505.
17. Consul, P.; Famoye, F. Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* **1992**, *21*, 89–109.
18. Kokonendji, C.C. Over-and underdispersion models. *Methods and Applications of Statistics in Clinical Trials* **2014**, *2*, 506–526.
19. Sellers, K.F.; Morris, D.S. Underdispersion models: Models that are “under the radar”. *Communications in Statistics-Theory and Methods* **2017**, *46*, 12075–12086.

20. King, G. Variance specification in event count models: From restrictive assumptions to a generalized estimator. *American Journal of Political Science* **1989**, pp. 762–784.
21. Morse, P.M. *Queues, Inventories and Maintenance*; John Wiley and Sons, New York, 1958.
22. Haight, F.A. The generalized Poisson distribution. *Annals of the Institute of Statistical Mathematics* **1959**, *11*, 101–105.
23. Mauro, R. *Traffic and random processes- an introduction*; Springer, 2015.
24. Sellers, K.F.; Shmueli, G. Data dispersion: now you see it... now you don't. *Communications in Statistics-Theory and Methods* **2013**, *42*, 3134–3147.
25. Wedderburn, R.W. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **1974**, *61*, 439–447.
26. Bonat, W.H.; Jørgensen, B.; Kokonendji, C.C.; Hinde, J.; Demétrio, C.G. Extended Poisson–Tweedie: properties and regression models for count data. *Statistical Modelling* **2018**, *18*, 24–49.
27. Winkelmann, R. A count data model for gamma waiting times. *Statistical Papers* **1996**, *37*, 177–187.
28. Cameron, A.C.; Johansson, P. Count data regression using series expansions: with applications. *Journal of Applied Econometrics* **1997**, *12*, 203–223.
29. Faddy, M.J. Extended Poisson process modelling and analysis of count data. *Biometrical Journal* **1997**, *39*, 431–440.
30. Klakattawi, H.; Vinciotti, V.; Yu, K. A simple and adaptive dispersion regression model for count data. *Entropy* **2018**, *20*, 142.
31. Weiß, C.H. Integer-valued autoregressive models for counts showing underdispersion. *Journal of Applied Statistics* **2013**, *40*, 1931–1948.
32. Tovissodé, C.F.; Honfo, S.H.; Doumatè, J.T.; Glèlè Kakaï, R. On the Discretization of Continuous Probability Distributions Using a Probabilistic Rounding Mechanism. *Mathematics* **2021**, *9*, 555.
33. Chakraborty, S. Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *Journal of Statistical Distributions and Applications* **2015**, *2*, 6.
34. Tovissodé, C.F.; Glèlè Kakaï, R. A Discrete Gamma Model Approach to Flexible Count Regression Analysis: Maximum Likelihood Inference. *Preprints* **2021**.
35. Grunwald, G.K.; Bruce, S.L.; Jiang, L.; Strand, M.; Rabinovitch, N. A statistical model for under-or overdispersed clustered and longitudinal count data. *Biometrical Journal* **2011**, *53*, 578–594.
36. Chen, J.; Zhang, D.; Davidian, M. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics* **2002**, *3*, 347–360.
37. Hosseini, F.; Eidsvik, J.; Mohammadzadeh, M. Approximate Bayesian inference in spatial GLMM with skew normal latent variables. *Computational Statistics & Data Analysis* **2011**, *55*, 1791–1806.
38. Broström, G.; Holmberg, H. Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics & Data Analysis* **2011**, *55*, 3123–3134.
39. Fabio, L.C.; Paula, G.A.; de Castro, M. A Poisson mixed model with nonnormal random effect distribution. *Computational Statistics & Data Analysis* **2012**, *56*, 1499–1510.
40. Fu, Y.z.; Chu, P.x.; Lu, L.y. A Bayesian approach of joint models for clustered zero-inflated count data with skewness and measurement errors. *Journal of Applied Statistics* **2015**, *42*, 745–761.
41. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **2001**, *79*, 99–113.
42. Lachos, V.H.; Labra, F.V.; Ghosh, P. Multivariate skew-normal/independent distributions: properties and inference. *Pro Mathematica* **2014**, *28*, 11–53.
43. Tovissodé, C.F.; Diop, A.; Glèlè Kakaï, R. Inference in skew generalized t-link models for clustered binary outcome via a parameter-expanded EM algorithm. *Plos one* **2021**, *16*, e0249604.
44. Dagne, G.; Huang, Y. Bayesian inference for a nonlinear mixed-effects Tobit model with multivariate skew-t distributions: application to AIDS studies. *The international journal of biostatistics* **2012**, *8*.
45. Pereira, M.A.A.; Russo, C.M. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. *Journal of Applied Statistics* **2019**, *46*, 1602–1620.
46. Molenberghs, G.; Kenward, M.; Verbeke, G.; Iddi, S.; Efendi, A. On the Connections Between Bridge Distributions, Marginalized Multilevel Models, and Generalized Linear Mixed Models. *International Journal of Statistics and Probability* **2013**, *2*, 1–21.
47. Iddi, S.; Molenberghs, G. A combined overdispersed and marginalized multilevel model. *Computational Statistics & Data Analysis* **2012**, *56*, 1944–1951.
48. Molenberghs, G.; Verbeke, G.; Demétrio, C.G. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime data analysis* **2007**, *13*, 513–531.
49. Alanko, T. *Statistical models for estimating the distribution function of alcohol consumption: A parametric approach*; Vol. 44, Finnish Foundation for Alcohol Studies, Helsinki, Finland, 1997.
50. Khan, N.M.; Soobhug, A.D.; Jannoo, Z. Modeling road traffic accidents in Mauritius using clustered longitudinal COM-Poisson with gamma random effects. *Communications in Statistics: Case Studies, Data Analysis and Applications* **2020**, pp. 1–15.

51. Sellers, K.F.; Premeaux, B. Conway–Maxwell–Poisson regression models for dispersed count data. *Wiley Interdisciplinary Reviews: Computational Statistics* **2020**, p. e1533.
52. Pinheiro, J.C.; Bates, D.M. Unconstrained parametrizations for variance-covariance matrices. *Statistics and computing* **1996**, *6*, 289–296.
53. Rabe-Hesketh, S.; Skrondal, A.; Pickles, A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* **2002**, *2*, 1–21.
54. Pan, J.; Thompson, R. Gauss-Hermite quadrature approximation for estimation in generalised linear mixed models. *Computational statistics* **2003**, *18*, 57–78.
55. Pinheiro, J.C.; Bates, D.M. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics* **1995**, *4*, 12–35.
56. Rizopoulos, D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis* **2012**, *56*, 491–501.
57. Vock, D.M.; Davidian, M.; Tsiatis, A.A.; Muir, A.J. Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics* **2012**, *13*, 61–73.
58. Vock, D.M.; Davidian, M.; Tsiatis, A.A. SNP_NLMM: A SAS macro to implement a flexible random effects density for generalized linear and nonlinear mixed models. *Journal of statistical software* **2014**, *56*, 2.
59. Wang, T.; Graves, B.; Rosseel, Y.; Merkle, E.C. Computation and application of generalized linear mixed model derivatives using lme4. *arXiv preprint arXiv:2011.10414* **2020**.
60. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **2015**, *67*, 1–48. doi:10.18637/jss.v067.i01.
61. Blocker, A.W. *fastGHQuad: Fast 'Rcpp' Implementation of Gauss-Hermite Quadrature*, 2018. R package version 1.0.
62. Smolyak, S.A. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics Doklady* **1963**, *4*, 240–243.
63. Winschel, V.; Krätzig, M. Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality. *Econometrica* **2010**, *78*, 803–821.
64. Heiss, F.; Winschel, V. Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics* **2008**, *144*, 62–80.
65. Malin, B.A.; Krueger, D.; Kubler, F. Solving the multi-country real business cycle model using a Smolyak-collocation method. *Journal of Economic Dynamics and Control* **2011**, *35*, 229–239.
66. Schillings, C.; Schwab, C. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems* **2013**, *29*, 065011.
67. Kieslich, C.A.; Boukouvala, F.; Floudas, C.A. Optimization of black-box problems using Smolyak grids and polynomial approximations. *Journal of Global Optimization* **2018**, *71*, 845–869.
68. Lesaffre, E.; Spiessens, B. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **2001**, *50*, 325–335.
69. Fernandez, C.; Steel, M.F. Multivariate Student-t regression models: Pitfalls and inference. *Biometrika* **1999**, *86*, 153–167.
70. Staudenmayer, J.; Lake, E.; Wand, M. Robustness for general design mixed models using the t-distribution. *Statistical Modelling* **2009**, *9*, 235–255.
71. Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **1943**, *54*, 426–482.
72. Rao, C.R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press, 1948, Vol. 44, pp. 50–57.
73. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* **1986**, *81*, 26853.
74. Robinson, G.K.; others. That BLUP is a good thing: the estimation of random effects. *Statistical science* **1991**, *6*, 15–32.
75. Hadfield, J.D.; Wilson, A.J.; Garant, D.; Sheldon, B.C.; Kruuk, L.E. The misuse of BLUP in ecology and evolution. *The American Naturalist* **2010**, *175*, 116–125.
76. McGilchrist, C. Estimation in generalized mixed models. *Journal of the Royal Statistical Society: Series B (Methodological)* **1994**, *56*, 61–69.
77. Medeiros, F.A.; Zangwill, L.M.; Weinreb, R.N. Improved prediction of rates of visual field loss in glaucoma using empirical Bayes estimates of slopes of change. *Journal of glaucoma* **2012**, *21*, 147.
78. Hwang, J.T. Empirical Bayes estimation for the means of the selected populations. *Sankhyā: The Indian Journal of Statistics, Series A* **1993**, pp. 285–304.
79. Prates, M.O.; Costa, D.R.; Lachos, V.H. Generalized linear mixed models for correlated binary data with t-link. *Statistics and Computing* **2014**, *24*, 1111–1123.
80. Lin, T.I.; Lee, J.C. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in medicine* **2008**, *27*, 1490–1507.

-
81. Meza, C.; Osorio, F.; De la Cruz, R. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. *Statistics and Computing* **2012**, *22*, 121–139.
 82. Lachos, V.H.; Ghosh, P.; Arellano-Valle, R.B. Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica* **2010**, *20*, 303–322.
 83. Roulin, A.; Bersier, L.F. Nestling barn owls beg more intensely in the presence of their mother than in the presence of their father. *Animal Behaviour* **2007**, *74*, 1099–1106.
 84. Zuur, A.F.; Ieno, E.N.; Walker, N.; Saveliev, A.A.; Smith, G.M. *Mixed Effects Models and Extensions in Ecology with R*, 1 ed.; Statistics for Biology and Health, Springer-Verlag New York, 2009.
 85. Skaug, H.; Fournier, D.; Bolker, B.; Magnusson, A.; Nielsen, A. *Generalized Linear Mixed Models using 'AD Model Builder'*, 2016-01-19. R package version 0.8.5.
 86. Pinheiro, J.C.; Liu, C.; Wu, Y.N. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* **2001**, *10*, 249–276.
 87. Kéri, G. The Sherman-Morrison formula for the determinant and its application for optimizing quadratic functions on condition sets given by extreme generators. In *Optimization Theory*; Giannessi, F.; Pardalos, P.; T, R., Eds.; Springer: Boston, 2001; pp. 119–138.