KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Public Health and Primary Care

Principled hierarchical models for dealing with correlated, overdispersed, and/or underdispersed data

Martial LUYTS

Promoter:	Prof. Dr. Geert Molenberghs	
Co-promoter:	Prof. Dr. Geert Verbeke	
Chair:	Prof. Dr. Pascal Borry	
Secretary:	Prof. Dr. Tim Verdonck	
Jury members:	Prof. Dr. Ingeborg Stalmans	Dissertation presented
	Prof. Dr. Koenraad Matthijs	in partial fulfillment
	Prof. Dr. Christel Faes	of the requirements
	Prof. Dr. John Hinde	for the degree of Doctor
	Prof. Dr. Clarice G. B. Demétrio	in Biomedical Sciences

28 September 2021

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Public Health and Primary Care

Principled hierarchical models for dealing with correlated, overdispersed, and/or underdispersed data

Martial LUYTS

Promoter:	Prof. Dr. Geert Molenberghs	
Co-promoter:	Prof. Dr. Geert Verbeke	
Chair:	Prof. Dr. Pascal Borry	
Secretary:	Prof. Dr. Tim Verdonck	
Jury members:	Prof. Dr. Ingeborg Stalmans	Dissertation presented
	Prof. Dr. Koenraad Matthijs	in partial fulfillment
	Prof. Dr. Christel Faes	of the requirements
	Prof. Dr. John Hinde	for the degree of Doctor
	Prof. Dr. Clarice G. B. Demétrio	in Biomedical Sciences

28 September 2021

Acknowledgment

Several persons have contributed towards the fulfillment of this work and they deserve a word of thanks.

First, gratitude is given to my promotor, Prof. dr. Geert Molenberghs, and co-promotor, Prof. dr. Geert Verbeke, for their great insights and help during my PhD. Dear Geert and Geert, many thanks is given to both of you for the great and instructive guidance throughout my PhD, not only professionally, but also personally. During the 6 years of my PhD, I had the enjoyment of working with both of you in the research and educational field of biostatistics, contributing to statistical research and teaching in the field of medicine and beyond, and even be involved in the management field of I-Biostat, department of public health and primary care and the Master of Statistics and Data Science program. It has been a great pleasure to have such amazing promoters on my side, and eventually, becoming great friends of me.

From my work as ombuds in the Master of Statistics and Data Science program, I express thanks to Prof. dr. An Carbonez, Prof. dr. Ingrid Van Keilegom, Ms. Eveline Declercq and Ms. Cristina Cametti, for all the help they have provided me. Dear An, Ingrid, Eveline and Cristina, thank you for all the assistance on this task. Together with Geert M., it is a pleasure working with you.

Towards my dearest colleagues from the L-Biostat unit, which ensured that I came with a big smile to my office: Thanks for being such great colleagues! You all have contributed towards this work by making sure that I feel at home within the unit, and sharing happy moments together (BBQ at Ann's place, etc.) that were unforgettable!

Credits are given towards my dearest friends, Kris Peremans, Caro Lievens, Jeroen Sichien, Maarten Coemans, and many others. Over my professional and personal career, they have contributed a lot in the person that I have became today. My mathematical friends from UHasselt, from who we monthly do an activity together, and my football friends from Leuven, from who we weekly met.

The achievement of this manuscript would not have been realized without the help of my family. To my parents, and two brothers, many thanks is given for all

the help and assistance. In my entire life, they have supported me in everything, and have always been ready when help was needed.

At last, let me express my appreciation and love towards my girlfriend, Elien Beets. Elien, in the couple of years that I know you, you have become one of the most important persons in my life. The moments that we spend so far have been quite amazing. The bets that I have won against you, living together with you, and even sharing my deepest feelings towards you have been astonishing. Thank you Elien, for always being there for me and supporting me, even when times are rough and not pleasant!!

List of Papers

The contents of this manuscript are based on the following original papers:

Chapter 2: Luyts, M., Tibaldi, F., Van de Putte, B., Geys, H., Matthijs, K., Molenberghs, G., and Verbeke, G. (2021). Two clustered time-to-event models to study inheritance in historic survival data. In preparation.
Contribution: Supplementary to the examination of familial transmission between father, mother and first born child within familial clusters in Moerzeke, where focus is placed on the inheritance of survival, the presence of underdispersion is discovered by using a specific member of the combined modeling framework (Molenberghs et al., 2010). By using descriptive statistics to examine the within-household variability, we have seen that many households encompass low within-household variability on life expectancy. These findings could be linked with the geographical isolated nature of Moerzeke. Precisely, a specific feature of an isolate is that the population is more homogeneous than is generally the case. This, in turn, may induce underdispersion in event times than would be the case in the general population. While these discoveries have not been examined in detail yet within the Moerzeke study, more broadly, within demographic clustered data, and since aspects like negative variance components have come forward in the discussion, new research possibilities open in the exploration of underdispersion within hierarchical data structures.

Chapter 3: Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Ribeiro Jr., E. E., Demétrio, C. G. B., and Hinde, J. (2020). A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures. *Statistical Modelling*, **19(5)**, 569–589.
Contribution: Starting from an existing (univariate) discrete framework, i.e., the so-called Weibull-count approach (Nakagawa and Osaki, 1975), we have proposed an extended version that can handle both under- and

overdispersed, and hierarchical discrete data structures. Our approach assumes that extra dispersion is captured in the pre-specified distribution, and differs from that in Molenberghs et al. (2010) where the extra dispersion is captured by an additional random effect. In addition, various settings (heavy tails, zero-inflation/-deflation) in combination with dispersion and correlation are examined, and compared with other well-known count models. Results showed that the model fits the data well, for both under- and overdispersed case studies, i.e., the Moerzeke and epilepsy dataset, respectively. The approach used is able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data, similar to the combined model approach of Molenberghs et al. (2010). In addition, the approach is capable of modeling some low overdispersed regions with zero-deflation and even underdispersed data, regions that cannot be captured within the (hierarchical) combined modeling framework. Due to the presence of a closed-form median expression, interpretations of the parameters can directly be related to the median profile, which is of particular interest when modeling skewed data.

Chapter 4: Luyts, M., Molenberghs, G., Verbeke, G., and Matthijs, K. (2021). A flexible finite mixture model family for analyzing over- and underdispersed, and/or zero-inflated and -deflated discrete data, with possibly negative weights. In preparation.
Contribution: We extended the traditional finite mixture framework (Newcomb, 1886; Prentice, 1988) in count data to account for both over- and underdispersion by considering a flexible class of elementary components, and lifting the non-negativity constraint for mixture probabilities in order to add a certain amount of flexibility. Here, focus is placed on the practical use of the approach within underdispersed data. Applying our framework to data from an underdispersed demographic setting (i.e., the Moerzeke dataset) showed advantageous regarding goodness-of-fit when comparing it to existing univariate frameworks. Negative and positive weights are encountered in our data analysis, underscoring the practical relevance of this extension.

Chapter 5: Luyts, M., Molenberghs, G., Verbeke, G., and Matthijs, K. (2021). Fitting negative variance components in a mixed approach for counts: A flexible marginalized estimation approach, In preparation.
Contribution: We have constructed second-order estimating equations (Zhao and Prentice, 1990; Liang et al., 1992) based on the specific first- and second-order marginal moments of generalized linear mixed models for counts. Specifically, conditionally on the random effects, here assumed

normally distributed, a log-linear Poisson model is used for the outcomes. The second-order nature of the moment-based method allows for proper marginal inferences on the variance components, allowing the fit of negative variance components and underdispersion. Proof of concept is given on the Moerzeke dataset; clear significant presence of a negative variance component is observed here.

The author has also been involved in the following original publications:

Sichien, J., Lesaffre, E., **Luyts, M.**, Swinkels, S., and Verbeke, G. (2021). Analysis of growth trial data: Which statistical model to choose? In preparation.

Breda, J. B., Pinto, L. A., Van Keer, K., Ivanova, A., **Luyts, M.**, Molenberghs, G., Willekens, K., Vandewalle, E., Sousa, A. R., and Stalmans, I. (2021) Glaucoma progression – The Leuven eye study. In preparation.

Lemmens, S., **Luyts, M.**, Gerrits, N., Ivanova, A., Landtmeeters, C., Peeters, R., Simons, A.-S., Vercauteren, J., Sunaric-Mégevand; G., Van Keer, K., Molenberghs, G., De Boever, P., and Stalmans, I. (2021) Age-related changes in the fractal dimension of the retinal microvasculature, effects of cardiovascular risk factors and smoking behavior. In preparation.

Verbeeck, J., **Luyts, M.**, Smolnik, R., Reimitz, P.-E., Tijssen, J., Goette, A., Vranckx, P., and Molenberghs G. (2021) Repeated bleeding event analysis with a Weibull-Poisson-binomial model. In preparation.

Delporte, M., De Coninck, D., D'Haenens, L., **Luyts, M.**, Verbeke, G., Molenberghs, M., and Matthijs, K. (2021). A longitudinal perceived vulnerability to disease during the COVID-19 pandemic in Belgium. In preparation.

Table of Contents

Abbreviations

Here, we give a list of the most often used abbreviations in the thesis.

AED	Anti-epileptic drug.
AGHQ	Adaptive Gauss-Hermite quadrature.
AIC	Akaike's information criterion.
BIC	Bayesian information criterion.
CM	Combined model.
COM	Conway-Maxwell-Poisson.
COMMM	Combined marginalized multilevel model.
COMN	Conway-Maxwell-Poisson-normal.
DE	Discrete exponential.
DEN	Discrete exponential-normal.
DI	Dispersion index.
DR	Discrete Rayleigh.
DP	Double Poisson.
DPN	Double Poisson-normal.
DW	Discrete Weibull.
DWN	Discrete Weibull-normal.
EDM	Exponential dispersion model.
EF	Exponential family.
EM	Expectation-maximization.
FMM	Finite mixture model.
GEE	Generalized estimating equations.
GEE2	Second-order generalized estimating equations.
GHQ	Gauss-Hermite quadrature.
GLM	Generalized linear model.
GLMM	Generalized linear mixed model.
GMM	Gaussian mixture model.
GP	Generalized Poisson.

GWPS	:	Generalized Weibull power series.
HT	:	Heavy-tail index.
IND	:	Uncorrelated random effects.
LMM	:	Linear mixed model.
ML	:	Maximum likelihood.
MMM	:	Marginalized multilevel model.
MQL	:	Marginal quasi-likelihood.
NB	:	Negative binomial model.
NEF	:	Natural exponential family.
P	:	Poisson.
PD	:	Plackett-Dale.
PGN	:	Poisson-gamma-normal.
PL	:	Pseudo-likelihood.
PMF	:	Probability mass function.
PN	:	Poisson-normal.
PQL	:	Penalized quasi-likelihood.
UN	:	Correlated random effects.
W	:	Weibull.
WG	:	Weibull-gamma.
WN	:	Weibull-normal.
WGN	:	Weibull-gamma-normal.
ZI	:	Zero-inflation index.
ZICOM	:	Zero-inflated Conway-Maxwell-Poisson.
ZINB	:	Zero-inflated negative binomial.
ZIP	:	Zero-inflated Poisson.

List of Tables

List of Figures

Chapter 1

General Introduction

1.1 Introduction

Life science studies regularly deal with complex data structures. Examples are longitudinally and/or hierarchically collected data, where subjects/clusters are repeatedly measured. While these structures are often observed in practice, statistical inferences on these studies are frequently challenging. In particular, statistical modeling frameworks are needed that (perfectly) reflect the design and characterizations of the study to avoid bias in the results. These include the correct treatment of continuous (e.g., cholesterol, diastolic blood pressure), binary (e.g., disease yes/no), count (e.g., number of epileptic seizures), and/or time-to-event (e.g., time untill having the COVID-19 disease) data, together by taking into account the full heterogeneity/homogeneity of the data.

In case of Gaussian outcomes, linear mixed models (LMMs) have become an attractive instrument in the literature (Laird and Ware, 1982; Verbeke and Lesaffre, 1996; Verbeke and Molenberghs, 2000), due to the flexibility of so-called random effects and a versatile range of possible covariance structures. For non-Gaussian outcomes, usually modeled with exponential family (EF; Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Jørgensen, 1987) members, generalized linear mixed models (GLMMs; Thall and Vail, 1990; Dean, 1991; Engel and Keen, 1994) are commonly preferred; these combine ideas of generalized linear models (GLMs; Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Jørgensen, 1987) with random effects. Unfortunately, due to the predefined mean-variance link in non-Gaussian GLMs, the variance in the data is sometimes insufficiently specified by the models. Two phenomena can occur: Over- and underdispersion; i.e., when the variability from the data is greater and lower, respectively, than the theoretical variance (limited by the mean-variance association) under the assumed model. Apart from these, other aspects like zero-inflation and -deflation can occur as well, alongside data that embrace heavy-tails.

We speak about zero-inflation if the data possess a surplus of zero responses than was foreseen from the model. Zero-deflation is defined when the data has less zero outcomes than was expected from the model.

Handling these phenomena in modeling strategies are crucial due to the possible flaws in point and precision estimation on key parameters (Paul and Plackett, 1978; Cox, 1983; Breslow, 1990), hereby leading to potentially erroneous inferences. In medicine, for example, a significant treatment might be rated as non-significant, and vice versa. The standard errors obtained can be under- or overestimated. We explore existing methodology in the literature, highlighting some of the limitations that these models and their inferences have. New approaches are constructed to tackle these constraints.

1.2 Motivating Case Studies

The motivation for conducting this research originates from sociological and life science studies where phenomena like dispersion and zero-inflation, together with longitudinal and/or clustered structures are present. Two motivating case studies are examined, i.e., the Moerzeke and epilepsy study. In what follows, a brief introduction is given around these studies.

1.2.1 The Moerzeke Study

In the Moerzeke study, i.e., a historical population study of Moerzeke, a small town based in Flanders, Belgium, civil and parochial registers were available including all persons who were born, died, or married in Moerzeke between 1700 and 1976. After reorganizing all available registers into a family database, a total of 5107 families were gathered, containing the primary and secondary generation, i.e., every person and his/her partner, and the children of the primary generation, respectively. Socio-demographic features such as age, sex, migration background of the children, and marital status were collected as well.

Main focus of this study is to examine the familial transmittance of survival among household members. Since individuals are included in households, i.e., a clustered data structure, the examination of homogeneity/heterogeneity in survival remains unexplored in the literature.

1.2.2 The Epilepsy Study

A randomized, double-blinded, parallel group multi-center study is carried out aiming to compare placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. At the start of the study, 45 patients were assigned to the placebo group, and 44 to the active (new) treatment group. Patients were repeatedly followed for several weeks - during which the number of epileptic seizures experienced in the last week - were counted, i.e., since the last time the outcome was measured. The main research question is whether or not the new treatment reduces the number of epileptic seizures. A full description of the epilepsy dataset is provided in Faught et al. (1996).

1.3 The Generalized Linear Model

The (univariate) GLMs (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Jørgensen, 1987) are often used as golden standard for analyzing (non-)Gaussian outcomes in many research domains. Warne (2020) and Myers et al. (2012), for example, pointed out the practical relevance of using GLMs in the domains of social sciences and engineering & physical sciences, respectively. In this framework, the response variable $Y_i, i = 1, \ldots, N$, follows an EF distribution that takes the form

$$f(y_i) \equiv f(y_i \mid \nu, \phi) = \exp\{\phi^{-1} \cdot [y_i \cdot \nu - \psi(\nu)] + c(y_i, \phi)\}. \tag{1.1}$$

Here, ν ("natural parameter") and ϕ ("dispersion parameter") defines a distinct collection of unknown parameters, and $\psi(\cdot)$ and $c(\cdot, \cdot)$ are known functions.

The mean and variance are expressed by

$$\mathsf{E}(\mathbf{Y}) = \mu = \psi'(\nu), \tag{1.2}$$

$$\mathsf{Var}(\mathbf{Y}) = \sigma^2 = \phi \cdot \psi''(\nu), \tag{1.3}$$

respectively, entailing a mean-variance relationship as follows: $\sigma^2 = \phi \cdot \psi''(\nu) = \phi \cdot \psi''[\psi'^{-1}(\mu)] = \phi \cdot v(\mu)$, with $v(\cdot) = \psi''(\psi'^{-1}(\cdot))$ the so-called variance function. The EF includes many well-known distributions such as the normal (for continuous data), Poisson (for count data), binomial/Bernoulli (for binary data), and exponential and Weibull (for time-to-event data) distribution. Remark that the Weibull distribution does not traditionally belong to the EF, though in a contrived manner by replacing Y with Y^ρ, where $\rho = 1$ expresses the more simple exponential distribution. In case $\phi = 1$, the EF is also referred as the natural exponential family (NEF).

In GLMs, the mean function μ is often associated with covariates $\boldsymbol{x}_i$ through a so-called link function h, i.e., $h(\mu) = \eta = \boldsymbol{x}_i \cdot \boldsymbol{\xi}$, with η referred to as the linear predictor, indicating the possibility to make inferences on a p-dimensional vector of unknown regression parameters $\boldsymbol{\xi}$. These inferences are commonly achieved by maximum likelihood or quasi-likelihood principles. In the latter case, only the identification of the first two moments (1.2) and (1.3) is needed instead of a full likelihood specification, implying that no full distributional assumptions need to be made. In those instances, a particular set of estimating equations is needed, where the solution is entitled as the quasi-likelihood estimates.

1.4 Dispersion Model Frameworks

In practice, the restricted mean-variance relationship in most GLMs are often insufficient to adequately capture the variability of the data, causing serious flaws in point and precision estimations on covariates/factors relevant for clinicians, sociologists, etc. In clinical trial studies, for example, a non-significant treatment might be assessed as significant, and vice versa.

For these and other reasons (e.g., zero-inflation/-deflation and heavy tails within the data), alternative/extended frameworks have been suggested and discussed in the literature. A first, and straight-forward step to embrace dispersion is to permit the dispersion parameter ϕ in Eq. (1.3) vary across values, in agreement with the so-called moment-based approach. Implementations with fully parametric assumptions can be carried out as well. Another way is through the so-called two-stage approach, i.e., by putting distributional assumptions on the model- or latent-parameter θ_i (also called a random effect), denoted by $f(\theta_i)$. Conditional on the random effect, a model is considered for the response, i.e., $f(y_i \mid \theta_i)$. The so-called marginal model $f(y_i)$ is then obtained through the equation

$$f(y_i) = \int_\theta f(y_i \mid \theta_i) f(\theta_i) d\theta_i. \tag{1.4}$$

In statistical terminology, $f(y_i \mid \theta_i)$, together with $f(\theta_i)$, is often referred as the conditional/hierarchical notation, while Eq. (1.4) presents the marginal notation. Throughout this thesis, the same terminology will be used.

Examples in the literature can be found on both methods. In count modeling, for example, the quasi-Poisson and negative binomial model exist and can be allocated in the former and latter procedure, respectively. The latter model even enjoys the property of conjugacy (Cox and Hinkley, 1979; Lee et al., 2018), i.e., where akin algebraic formulations are used for the hierarchical and random effects densities so that a general closed-form expression can be determined from Eq. (1.4), making it an interesting framework to explore analytically. Details can be found in Supplementary Material S.4.1 and S.4.2. For time-to-event outcomes, examples like the exponential-gamma and Weibull-gamma exist, where the property of conjugacy is also retained (Table 1.1).

In our motivating case studies (Section 1.2), counts and time-to-event outcomes are of interest. Therefore, existing and new approaches are examined for

Table 1.1*: Hierarchical & its marginal formulation for the exponential-gamma and Weibull-gamma models, with corresponding mean and variance expression.* $\Gamma(.)$ *expresses the gamma function.*

Element	Notation	Exponential-gamma	Weibull-gamma
Hier. model	$f(y_i \mid \theta_i)$	$\varphi \cdot \theta_i \cdot e^{-\varphi \cdot \theta_i \cdot y_i}$	$\varphi \cdot \theta_i \cdot \rho \cdot y_i^{\rho-1} \cdot e^{-\varphi \cdot \theta_i \cdot y_i^{\rho}}$
RE model	$f(\theta_i)$	$\frac{\theta_i^{\alpha-1} \cdot e^{-\theta_i/\beta}}{\beta^{\alpha} \cdot \Gamma(\alpha)}$	$\frac{\theta_i^{\alpha-1} \cdot e^{-\theta_i/\beta}}{\beta^{\alpha} \cdot \Gamma(\alpha)}$
Marg. model	$f(y_i)$	$\frac{\varphi \cdot \alpha \cdot \beta}{(1+\varphi \cdot \beta \cdot y_i)^{\alpha+1}}$	$\frac{\varphi \cdot \rho \cdot y_i^{\rho-1} \cdot \alpha \cdot \beta}{(1+\varphi \cdot \beta \cdot y_i^{\rho})^{\alpha+1}}$
Mean	$\mathsf{E}(\mathbf{Y})$	$[\varphi \cdot (\alpha - 1) \cdot \beta]^{-1}$	$\frac{\Gamma(\alpha-\rho^{-1}) \cdot \Gamma(\rho^{-1}+1)}{(\varphi \cdot \beta)^{1/\rho} \cdot \Gamma(\alpha)}$
Variance	$\mathsf{Var}(\mathbf{Y})$	$\alpha \cdot [\varphi^2 \cdot (\alpha-1)^2 \cdot (\alpha-2) \cdot \beta^2]^{-1}$	$\frac{1}{\rho \cdot (\varphi \cdot \beta)^{1/\rho} \cdot \Gamma(\alpha)} \cdot \Big[2 \cdot \Gamma(\alpha - 2 \cdot \rho^{-1})$ $\cdot \Gamma(2 \cdot \rho^{-1}) - \frac{\Gamma(\alpha-\rho^{-1})^2 \cdot \Gamma(\rho^{-1})^2}{\rho \cdot \Gamma(\alpha)}\Big]$

these type of responses. Two techniques will be contrasted, i.e., the two-stage approach, where extra dispersion is captured within the random effect structure, and a class of flexible distributions that offers enough flexibility to apprehend dispersion. Principles like zero-inflation/-deflation and heavy-tailed profiles will be investigated as well for most models considered in this thesis.

1.5 Models for Correlated Data

Alternatively, random effects can be entered directly in the linear predictor η of GLMs, giving rise to the so-called family of GLMMs (Thall and Vail, 1990; Dean, 1991; Engel and Keen, 1994; Wolfinger and O'Connell, 1993) where LMMs are a special case of. Including random effects in the linear predictor have the advantage of (1) taking into account the longitudinal and hierarchical data structures, and (2) examining between- and within-subject correlations, through a single modeling framework. In what follows, we will give a brief discussion on these basic frameworks, i.e., the LMM and more general GLMM framework, and some of their extensions.

1.5.1 Linear Mixed Models

LMMs have become an attractive approach for analyzing longitudinal and/or hierarchical Gaussian data. Random effects in these models are often assumed normally distributed, but can be considered different as well. In the hierarchical and marginal notation, the classical LMM framework can be formulated as

$$\mathbf{Y}_i \mid \mathbf{b}_i \sim N(\overbrace{\mathbf{X}_i \cdot \boldsymbol{\xi} + \mathbf{Z}_i \cdot \mathbf{b}_i}^{\boldsymbol{\mu}_i^c}, \Sigma_i), \tag{1.5}$$

$$\mathbf{b}_i \sim N\left(\mathbf{0}, D\right); \tag{1.6}$$

$$\mathbf{Y}_i \sim N(\underbrace{\mathbf{X}_i \cdot \boldsymbol{\xi}}_{\boldsymbol{\mu}_i^m}, \mathbf{Z}_i \cdot D \cdot \mathbf{Z}_i^{'} + \Sigma_i), \tag{1.7}$$

respectively. $\mathbf{Y}_i$ is the n_i-dimensional vector of repeated outcomes for cluster/-subject i, $i = 1, \ldots, N$. $\boldsymbol{\xi}$ and $\mathbf{b}_i$ are the p- and q-dimensional vector of unknown regression parameters and random effects, respectively, where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the known corresponding design matrices with dimensions of respectively n_i x p and n_i x q. Σ_i illustrates a n_i x n_i covariance matrix for the residuals and D presents a q x q covariance matrix for the random effects. At last, $\boldsymbol{\mu}_i^c (= \mathsf{E}(\mathbf{Y}_i \mid \mathbf{b}_i))$ and $\boldsymbol{\mu}_i^m (= \mathsf{E}(\mathbf{Y}_i))$ are used to express the conditional and marginal vector of mean structures, respectively.

1.5.2 Generalized Linear Mixed Models

GLMMs can be seen as a natural extension of LMMs towards GLM structures, allowing researchers to conduct inferences on non-Gaussian responses within hierarchical/longitudinal designs. Conditional on the q-dimensional random effects vector $\mathbf{b}_i$ (1.6), the outcomes are assumed to follow EF densities, expressed by

$$f_i(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \phi) = \exp\{\phi^{-1} \cdot [y_{ij} \cdot \nu_{ij} - \psi(\nu_{ij})] + c(y_{ij}, \phi)\}, \tag{1.8}$$

with

$$h[\psi'(\nu_{ij})] = h(\mu_{ij}^c) = \eta_{ij} = \boldsymbol{x}_{ij}' \cdot \boldsymbol{\xi} + \boldsymbol{z}_{ij}' \cdot \boldsymbol{b_i}, \tag{1.9}$$

where $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the p- and q-dimensional vector of known covariate values corresponding to the p- and q-dimensional vector of unknown fixed regression parameters $\boldsymbol{\xi}$ and random effects $\boldsymbol{b_i}$, respectively. For counts and time-to-event outcomes, examples like the Poisson[(1.8)]–normal[(1.6)] and Weibull[(1.8)]–normal[(1.6)] models are standard used, with the exponential[(1.8)]–normal[(1.6)] model as special case for the latter approach.

1.5.3 Combined Models

While GLMMs can be seen as a general class of models to deal with longitudinal/hierarchical data, accounting for some of the dispersion as well, their given flexibility towards dispersion is sometimes limited in practical data settings. Molenberghs et al. (2007) remarked that GLMMs often inadequately fit data when dispersion and correlation between repeated measurements occur simultaneously. Therefore, Molenberghs et al. (2007) and Molenberghs et al. (2010) developed the so-called combined model (CM) framework, by appending an additional random effect vector $\boldsymbol{\theta}_i = (\theta_{i1}, ..., \theta_{in_i})$ within the GLMM structure that directly enters the mean via a multiplicative factor:

$$f_i(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \boldsymbol{\theta}_i, \phi) = \exp\{\phi^{-1} \cdot [y_{ij} \cdot \nu_{ij} - \psi(\nu_{ij})] + c(y_{ij}, \phi)\}, \qquad (1.10)$$

with conditional mean

$$\psi'(\nu_{ij}) = \mu_{ij}^c = \mathsf{E}(Y_{ij} \mid \mathbf{b}_i, \boldsymbol{\xi}, \theta_{ij}) = \theta_{ij} \cdot \kappa_{ij}, \qquad (1.11)$$

Here, θ_{ij} follows a predefined distribution Ξ_{ij} with mean ζ_{ij} and variance ϖ_{ij}, and $\kappa_{ij} = h^{-1}(\eta_{ij}) = h^{-1}(\mathbf{x}_{ij}' \cdot \boldsymbol{\xi} + \mathbf{z}_{ij}' \cdot \mathbf{b}_i)$ for some link function $h(.)$. Assumptions for the random effects vector $\mathbf{b}_i$ are held the same as before, i.e., Eq. (1.6). All elements in the distribution of θ_{ij} are subscripted by i and j, for generality of notation. Though, one might determine common distributions for a particular measurement j, or even common through i and j. Several choices for θ_{ij} can be made. First, concerning the elements θ_{ij} of $\boldsymbol{\theta}_i$, different assumptions can be drawn: (1) Independency among them; (2) dependency among them, indicating that the class of univariate distributions Ξ_{ij} must be modified by a multivariate distribution; and (3) they correspond to each other, beneficial in applications with exchangeable responses Y_{ij}. Second, it is natural to assume independency among random effects θ_{ij} and $\mathbf{b}_i$, even though this is not strictly necessary. Third, by selecting conjugate distributions for θ_{ij}, Molenberghs et al. (2010) showed that closed-forms could be derived for the whole marginal distribution. For counts and time-to-events, where the classical Poisson–normal and Weibull–normal approach exist in the GLMM case, (conjugate) gamma distributions can be used for θ_{ij}, giving rise to the so-called Poisson–gamma–normal (PGN) and Weibull–gamma–normal (WGN) models.

1.6 Inferences in Mixed Models

Inferences in mixed models are essential to reach conclusions on research questions of interest. In clinical trial research, for example, researchers are sometimes interested in the evaluation of the difference in response between a treated and an untreated subject with the same level for their random effects, implying subject-specific interpretations, while other researchers are more interested in the average (global) treatment effect, indicating a population-averaged interpretation. Therefore, two viewpoints can be taken in mixed models, depending on the research question. In what follows, we will review these angles, together with several estimation strategies that are currently available in the literature.

1.6.1 Subject-specific and Population-averaged Inferences

In the hierarchical LMM approach (1.5)–(1.7), the fixed parameters $\boldsymbol{\xi}$ have, next to a subject-specific interpretation, also a population-average interpretation, i.e., $\mathsf{E}(Y_{ij}) = \mathsf{E}[\mathsf{E}(Y_{ij} \mid \mathbf{b}_i)] = \mathbf{X}_i \cdot \boldsymbol{\xi}$. This property does not generally hold in the hierarchical GLMM framework (1.6) & (1.8)–(1.9) for non-Gaussian responses, where a subject-specific interpretation is standard retrieved for these effects. In the binary approach, for instance, where predictors and random effects parameters are connected with the mean of a binary outcome Y_{ij} through the logit link,

$$\mathsf{E}(Y_{ij}) = \mathsf{E}[\mathsf{E}(Y_{ij} \mid \mathbf{b}_i)] = \mathsf{E}\left[\frac{\mathsf{e}^{\mathbf{x}_{ij}^{'}\cdot\boldsymbol{\xi}+\mathbf{z}_{ij}^{'}\cdot\mathbf{b}_i}}{1+\mathsf{e}^{\mathbf{x}_{ij}^{'}\cdot\boldsymbol{\xi}+\mathbf{z}_{ij}^{'}\cdot\mathbf{b}_i}}\right] \neq \frac{\mathsf{e}^{\mathbf{x}_{ij}^{'}\cdot\boldsymbol{\xi}}}{1+\mathsf{e}^{\mathbf{x}_{ij}^{'}\cdot\boldsymbol{\xi}}}. \tag{1.12}$$

This hampers the ability to derive population-averaged interpretations within GLMMs for non-Gaussian outcomes, notwithstanding these are often of interest.

Preferably, a direct marginal procedure may be used to ensure population-averaged inferences. Liang and Zeger (1986) suggested so-called generalized estimation equations (GEE), that expand GLMs by permitting for correlation within cluster/subject via a so-called working correlation. Parameter estimates remain consistent and asymptotically normal when misspecifying this working correlation. This approach, however, lacks a full likelihood foundation, precluding various inferential paths.

Combining ideas on GLMMs and GEE, Heagerty (1999) and Heagerty and Zeger (2000) developed the so-called marginalized multilevel model (MMM) by bringing the power of both together, so that the marginal mean $\boldsymbol{\mu}_i^m$ instead of the conditional mean $\boldsymbol{\mu}_i^c$, i.e., the mean conditional on random effects, is directly regressed with predictors. Following the same terminology as before, where μ_{ij}^c denotes the conditional mean of EF densities (1.8), their MMMs can be formulated as

$$g(\mu_{ij}^m) = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\xi}^m, \tag{1.13}$$

$$h(\mu_{ij}^c) = \Delta_{ij} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i. \tag{1.14}$$

The covariate vector $\boldsymbol{\xi}^m$ is superscripted to identify its directly marginal interpretation, and Δ_{ij} presents the so-called connector function between Eq. (1.13) and (1.14), with a similar link function defined for the marginal and conditional component, i.e., $h(.) = g(.)$. Griswold and Zeger (2004) extended this work by allowing different link functions, and providing connections between marginal and conditional models transparent. Specifically, the connector function Δ_{ij} is expressed as the solution of the equation

$$g^{-1}(\mathbf{x}_{ij}' \cdot \boldsymbol{\xi}^m) = \int_b h^{-1}(\Delta_{ij} + \mathbf{z}_{ij}' \cdot \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i. \tag{1.15}$$

Extensions toward CMs are achieved as well, referred as the combined marginalized multilevel model (COMMM; Iddi and Molenberghs, 2012; Efendi et al., 2014) framework. Equivalent towards MMMs, this approach can be expressed as

$$\mu_{ij}^m = g^{-1}(\mathbf{x}_{ij}' \cdot \boldsymbol{\xi}^m), \tag{1.16}$$

$$\mu_{ij}^c = \theta_{ij} \cdot \kappa_{ij} = \theta_{ij} \cdot h^{-1}(\Delta_{ij} + \mathbf{z}_{ij}' \cdot \mathbf{b}_i), \tag{1.17}$$

where the connecter function Δ_{ij} is the solution of the function

$$g^{-1}(\mathbf{x}_{ij}' \cdot \boldsymbol{\xi}^m) = \mathsf{E}(\theta_{ij}) \cdot \int_b h^{-1}(\Delta_{ij} + \mathbf{z}_{ij}' \cdot \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \cdot d\mathbf{b}_i. \tag{1.18}$$

Examples in both frameworks can be found in the literature. In the CM paradigm, for example, Iddi and Molenberghs (2012) and Efendi et al. (2014) expressed closed-form solutions for Δ_{ij}, for the hierarchical PGN and WGN models, respectively. In the WGN version (Efendi et al., 2014), for example, the following expression was obtained for Δ_{ij}:

$$\Delta_{ij} = -\log(\alpha_j \cdot \beta_j) + \mathbf{x}'_{ij} \cdot \boldsymbol{\xi}^m - \frac{\mathbf{z}_{ij}' \cdot D \cdot \mathbf{z}_{ij}}{2}. \tag{1.19}$$

Here, α_j and β_j are the conventional gamma parameters, depending on measurement j. Ultimately, these models will be used in data analysis later on (Section 2.5.2).

1.6.2 Negative Variance Components

For the random effects parameters, often referred as variance components, a similar reasoning can be followed regarding their interpretation in GLMMs and CMs. In a hierarchical viewpoint, these variance components are interpreted as variances, obvious. Hence, these ought to be non-negative, implying the need for a one-sided testing paradigm on their inferences (Verbeke and Molenberghs, 2003). In the marginal case, however, the interpretation of variance disappears and the components only contribute in the marginal model. This implies that negative values can be obtained for these components with necessity that the marginal variance-covariance matrix remains positive definite. This matrix is composed of elements from the random effects, alongside with the underlying mean-variance association (if any) from the GLM. Evidently, a two-sided testing framework is desired here.

This phenomenon is reasonable well understood in the literature for LMMs, whereas such negative variance components relate to negative within-unit correlations in many cases (Molenberghs and Verbeke, 2011), a situation that often occurs in practice. In agricultural sciences, for example, studies exist where only fixed resources such as water and/or nutrients are present in a plot and the plants within this plot need to compete for it, implying the intraclass correlations to become negative. In medicine, for instance, triplets inside the mother's womb need to battle for the same food sources.

Pryseley et al. (2011) discussed this phenomenon in GLMMs for counts and binary outcomes, while Oliveira et al. (2017) explored it for the PGN model, where the negative variance components emerge from either the conjugate random effects $\boldsymbol{\theta}_i$, the normal random effects $\mathbf{b}_i$, or both, making the exploration of it more broader. Specifically, there exists various reasons in non-Gaussian frameworks like GLMMs and CMs where the phenomenon of negative variance components can be observed, e.g., correlation, underdispersion, simultaneous occurrence of high correlation & low overdispersion. A full discussion around negative variance components in the PGN model can be found in Oliveira et al. (2017), but can easily be expanded towards the WGN model (due to conjugacy principles).

When inferences for variance components are obligated, it is crucial to determine a prior their range of values, i.e., either positive or also negative. In addition, adequate estimation tools are needed to allow the fit of these components.

1.6.3 Estimation Strategies

Estimators of the mixed model parameters are derived by maximizing the marginal likelihood

$$L(\boldsymbol{\Upsilon}, \boldsymbol{\Omega}) = \prod_{i=1}^{N} L_i(\boldsymbol{\Upsilon}, \boldsymbol{\Omega}), \tag{1.20}$$

where $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Omega}$ present the list of fixed and random parameters, respectively, and $L_i(\boldsymbol{\Upsilon}, \boldsymbol{\Omega})$ is the marginal likelihood contribution of cluster/subject i, expressed by

$$L_i(\boldsymbol{\Upsilon}, \boldsymbol{\Omega}) = \int_b \left(\prod_{j=1}^{n_i} f_i(y_{ij} \mid \boldsymbol{\Upsilon}, \mathbf{b}_i) \cdot f(\mathbf{b}_i \mid D) \right) \cdot d\mathbf{b}_i, \tag{1.21}$$

$$L_i(\boldsymbol{\Upsilon}, \boldsymbol{\Omega}) = \int_b \left[\prod_{j=1}^{n_i} \left(\underbrace{\int_\theta f_i(y_{ij} \mid \boldsymbol{\Upsilon}, \mathbf{b}_i, \boldsymbol{\theta}_i) \cdot f(\boldsymbol{\theta}_i \mid \zeta_{ij}, \varpi_{ij}) \cdot d\boldsymbol{\theta}_i}_{f_i(y_{ij} \mid \boldsymbol{\Upsilon}, \zeta_{ij}, \varpi_{ij}, \mathbf{b}_i)} \right) \right.$$
$$\left. \cdot f(\mathbf{b}_i \mid D) \right] \cdot d\mathbf{b}_i, \tag{1.22}$$

for the GLMM and CM framework, respectively. Maximizing Eq. (1.20) is often cumbersome due to the intractable integrals in Eq. (1.21)–(1.22). Indeed, unlike LMMs, where an analytical closed-form expression exists for this integral, there is often a lack of these in Eq. (1.21)–(1.22) for non-Gaussian outcomes. Even when these exist, problems like infinite series can, for example, still be present, making the approach of maximization still intractable. A good example for the latter case is the WGN case, where the attendance of an analytically closed-form expression exist for the marginal distribution but composed of infinite series (Molenberghs et al., 2015). Therefore, different methods have been suggested in the literature for approximating the integral in Eq. (1.21). For the CM approach, Molenberghs et al. (2010) used partial marginalization ideas, by choosing conjugate distributions for $\boldsymbol{\theta}_i$ such that closed-forms exist for $f_i(y_{ij} \mid \boldsymbol{\Upsilon}, \zeta_{ij}, \varpi_{ij}, \mathbf{b}_i)$, making the problem of maximization identical to GLMMs. Therefore, we will focus the discussion here towards GLMMs.

In general, two types of solutions exist. In the first approach, the integral within Eq. (1.21) is approximated numerically such that the marginal likelihood can be calculated and optimized. For approximating the integral numerically, examples such as Gauss-Hermite quadrature (GHQ; Naylor and Smith, 1982), adaptive Gauss-Hermite quadrature (AGHQ; Rabe-Hesketh et al., 2002) and Monte Carlo integration exist, techniques that are standard integrated in many statistical software packages, e.g., the NLMIXED procedure (SAS Institute, 1999). To obtain the maximum of the marginal likelihood approximation, methods like the simplex algorithm (Nelder and Mead, 1965), the steepest descent method (Curry, 1944), the Newton-Raphson or (related) Fisher scoring algorithm (Longford, 1987), and quasi-Newton methods (Broyden, 1967) are often used. In the second approach, the integrand is approximated, such that a closed-form can be found for the integral of the approximation. These include methods like Laplace approximation (Tierney and Kadane, 1986) and quasi-likelihood ideas, e.g., the penalized quasi-likelihood (PQL; Breslow and Clayton, 1993; Schall, 1991), marginal quasi-likelihood (MQL; Goldstein, 1991) and several of their extensions (e.g., PQL2 and MQL2). A clear overview of these techniques, alongside their advantages and disadvantages can be found in Tuerlinckx et al. (2006).

Unfortunately, limitations can occur when dealing with, for example, negative variance components. In particular, when considering the marginal viewpoint and negative variance components are expected, the possibility exists that negative estimates cannot be guaranteed in the inferences due to the fact that the used algorithm does simply not allow for it (i.e., when the algorithm adopts a hierarchical specification in the estimation strategy). This aspect is avoided in the Laplace approximation technique, allowing for negative variance components as well. However, the reader should then be aware of the possible trade off between (1) the method's precision and (2) the ability of fitting negative variance components. In addition, the testing paradigm needs to be adapted, according to the work of Verbeke and Molenberghs (2003). Therefore, extensions and/or alternative routes are interesting to explore that enables the user to obtain unbiased inferences and easily allow for estimation of negative variance components as well.

1.7 Thesis Contribution

Many frameworks have been suggested in the literature that handle the longitudinal and hierarchical structures within the data, with prominent members like the random effects models as golden standard. These models have gained in popularity since the extension of LMMs for Gaussian data (Laird and Ware, 1982) towards GLMMs for non-Gaussian data (Breslow and Clayton, 1993, Wolfinger and O'Connell, 1993, and Engel and Keen, 1994) and their given flexibility in modeling different types of outcomes and covariances structures. In some extend, they also take into account some amount of variability in the data.

Understanding & extending these frameworks are essential in statistical research to make the right decisions and predictions needed for life science studies and beyond. For more than two decades, L-Biostat has extensively contributed to this research, resulted in several peer-reviewed articles in renowned international journals such as the Journal of the American Statistical Association and Computational Statistics & Data Analysis. The research in this field is still ongoing. Molenberghs et al. (2007) and Molenberghs et al. (2010) showed that, by using these GLMMs, the theoretical variability of the model often falls short to properly model the data variability. Therefore, extensions have been proposed. An example is the development of the CM (Molenberghs et al., 2010), to account for extra dispersion. In case of variability reduction, e.g., settings where more homogeneity is observed than expected, shortcomings sometimes arise when using these classical & extended models.

The contribution of this thesis is composed of 4 parts:

–1– *Dispersion detection in classical hierarchical frameworks & their shortcomings/boundaries in inferences*

In this part of the thesis, we apply a GLMM approach and corresponding CM extension for time-to-events, i.e., the WN and WGN model, respectively, to highlight the presence of underdispersion in a clustered sociological study. We demonstrate some of the boundaries that standard estimation techniques like Gaussian quadratures have, and how it affects the statistical inferences. The need to understand these complexities in practice, detecting them in early phase of the analysis and adapting the model design and/or estimation strategy for conducting valid inferences are essential and needed to be handled with great care.

–2– *Dispersion models in univariate structures*

To examine alternative routes, focus is given on the basic building blocks of GLMMs, i.e., GLMs, and the construction of several extensions/alternatives. In particular, we look at alternative building blocks that flexible handle basic problems like dispersion in univariate count data. These include the exploration of existing methodology in the literature, and development of new frameworks. In a further stage, these alternative methods are discussed and extended in the context of longitudinal/clustered frameworks.

To this end, the use of discrete versions of well-known continuous distributions and finite mixture models are examined in discrete data settings, and extended if needed. Both techniques will embrace both over- and underdispersed settings, together with a range of zero-inflation/-deflation and heavy-tail profiles.

–3– *Dispersion models in longitudinal/hierarchical structures*

In a further stage, the exploration of these (univariate) alternative models is extended towards hierarchical/longitudinal data settings by incorporating random effects in the model structure. These models are compared to the classical GLMM approach and CM extension for both overdispersed and underdispersed settings, putting particular focus on the so-called discrete Weibull (DW; Nakagawa and Osaki, 1975) model without and with normal random effects. Properties, limitations and advantages are examined in detail, motivating the use of it in our case studies, i.e., modeling the number of epilepsy attacks and (discrete) longevity within the epilepsy and Moerzeke data, respectively.

–4– Second-order generalized estimation equations in GLMMs

Because the GLMMs' genesis is subject-specific, their fixed-effects parameters have a subject-specific interpretation. To obtain population-average inferences, several authors have contributed to this. These include work done by Heagerty (1999), Heagerty and Zeger (2000) and Griswold and Zeger (2004). A similar reasoning applies for their variance components, which characterizes the homogeneity/heterogeneity of the model. In the subject-specific viewpoint, the random effects distribution is of interest, implying positive definiteness of the variance-covariance matrix. In the population-averaged case, the milder restriction of a positive definite marginal variance-covariance matrix is all that is needed, implying that some variance components in the mixed model may be negative, while still producing a valid marginal interpretation. Pryseley et al. (2011) investigated this phenomenon in the GLMM setting, and theoretically linked it with negative intraclass correlations, underdispersion, and others. Unfortunately, flexible estimation approaches in GLMMs are still lacking, indicating limitations when modeling these designs. These limitations are encountered in our analysis. Therefore, we construct flexible (closed-form) second-order estimating equations for the first and second central-moments of the marginal Poisson-normal mixed framework, allowing for the estimation of negative variance components in the marginal approach.

Summary

Classical existing models are explored that accounts for dispersion and/or correlated data. Limitations/boundaries are encountered in practical data settings, suggesting the need for extensions in modeling and estimation strategies. Alternative and new approaches are constructed, (partly) resolving these issues.

1.8 Outline of Thesis

This dissertation is composed of 6 chapters. Chapter 1 presents a general introduction of the objectives of this research, alongside the motivating case studies and a brief overview of existing methodology in this domain of research.

The next four chapters present the thesis' contribution. Chapter 2 explores details in line with objective 1. Existing and new modeling frameworks are considered and discussed in Chapter 3 and 4, corresponding to objective 3 and 2, respectively. Chapter 5 investigates marginal first- and second-order moments in a GLMM framework for counts, and uses these to construct second-order generalized estimating equations, consistent with objective 4.

In the last part of the thesis, i.e., Chapter 6, an overall conclusion is outlined to the reader, apace with a formal discussion of some of their limitations, and reflections on future research.

Chapter 2

Two Clustered Time-to-event Models to Study Inheritance in Historic Survival Data

This chapter is based upon:

Luyts, M., Tibaldi, F., Van de Putte, B., Geys, H., Matthijs, K., Molenberghs, G., and Verbeke, G. (2021). Two clustered time-to-event models to study inheritance in historic survival data. In preparation.

Abstract

The focus is on the study of inheritance of survival in historic cohorts, where data take the form of familial clusters. Apart from the time-to-event outcomes, also the association between them needs to be accommodated. Attention is given to the situation where data come from an isolate, which may induce underdispersion. Two modeling strategies are contrasted: (1) A multivariate Plackett-Dale model (Tibaldi et al., 2004), which is of a marginal nature and (2) a Weibull model with gamma frailties and normal random effects (Molenberghs et al., 2010), termed Weibull-gamma-normal, of a hierarchical nature.

The methodologies are used to model a historic cohort from a small Flemish village, a geographical isolate called Moerzeke (18^{th}–20^{th} century) (Matthijs et al., 2002). Life history data are obtained from administrative registers, available from the post-Napoleonic period. The main conclusion is that familial transmission mainly occurs from mother to daughter. Underdispersion is detected.

2.1 Introduction

Interest lies in examining historic inheritance in survival among family members in a small Flemish village called Moerzeke, geographically isolated at the time (18^{th}-20^{th} century). Data are extracted from a historic demographic database, by forming clusters of three members (father, mother, first child). The utilization of hierarchical modeling techniques, accommodating the association among family members, is necessary. Censoring will be taken into account, as well as extra-model variation in the event times. Because of the isolated nature of Moerzeke, one should not rule out underdispersion, meaning that the used model is not able to adequately accommodate the lower variability present within the data, so the models should ideally allow for this.

Two modeling strategies will be considered: A Plackett-Dale (PD) model (Dale, 1986; Plackett, 1965; Mardia, 1970; Molenberghs and Lesaffre, 1994; Tibaldi et al., 2004) and the Weibull-gamma-normal (WGN) model, an instance of the so-called combined modeling (CM) framework of Molenberghs et al. (2010).

The PD model allows for explicit modeling of the association structure and, because of it is marginal model nature, produces directly interpretable so-called global cross ratios (sometimes also called global odds ratios). Maximum likelihood (ML) estimation is often prohibitive, for which reason pseudo-likelihood (PL) principles can be applied if needed (Geys et al., 1997; Tibaldi et al., 2004; Molenberghs and Verbeke, 2005).

The WGN model flexibly allows for dependence between clustered outcomes as well as for over-/underdispersion. While of a hierarchical nature, it can be parameterized such that population–averaged parameters ensue.

The remainder of the chapter is organized as follows. In Section 2.2, a summary of the data, alongside the town from which it originates, is given. Section 2.3 sketches the underlying methodology of the PD model with the PL estimation and the WGN model. Adaptions of traditional hypothesis testing procedures (Wald, Score & likelihood ratio statistics) are needed in the PL paradigm. Details about this testing framework can be found in Supplementary Materials S.2. A brief discussion on familial transmittance of longevity is highlighted in Section 2.4. In Section 2.5, data from the Moerzeke study are analyzed with both methodologies, and conclusions, not only about the familial transmittance of longevity, but also about underdispersion, are covered in Section 2.6.

2.2 The Moerzeke Data

In the international research community, the demographic history of Belgium has been of interest for a long time. Both historic and demographic developments, as well as the contributions made by Belgian researchers, have added to this (Lescrauwaet et al., 2010; Vandresse, 2013).

Focus is placed on the analysis of historical data from Moerzeke, i.e., a small town in the center of Flanders, the Dutch speaking part of Belgium, within the province of East Flanders. It is a geographical isolate as it is almost completely surrounded by a meander in the river Scheldt and by the river Durme. De Beule (1962) and De Ridder (1986) detected that the population of Moerzeke, mainly populated by farmers until well into the 20^{th} century, rose from approximately 2000 in 1761 to 4706 in 1950, where fertility was traditionally high and dropped at the beginning of the 20^{th} century. Major mortality crises (mainly dysentery) occurred in the 18^{th} century. 24.8% of the children born in Moerzeke died within the first year, but this happened less frequently as the 18^{th} century progressed. Infant and childhood mortality was strikingly high. Infant mortality did not drop until the first decades of the 20^{th} century. The life expectancy at age 50 steadily rose for those born in the 19^{th} century, reaching a peak at the end of the observation period, i.e., those born after 1850. These times refer to the entire lifespan.

A subset of 457 families will be extracted, by choosing all fathers born in Moerzeke from 1750 to 1830, and then composing a household by including the mother and first born child. Children already selected are not included again as father or mother of new families, to avoid between-family correlation. All included were born and died in Moerzeke (the so-called "stable population"). To maintain data quality of the sample, inferences will be based on households in which the child reaches at least the age of 10. For the group under study, the mean age at death for those who were born and deceased in Moerzeke was 71.9 years for men and 71.7 for women. The upper deciles for the lifespan are 83.3 for mothers, 84.2 for fathers, 84.8 for sons, and 84.4 years for daughters, indicating that boys, on average, live longer than girls. A possible reason for this occurrence is that parents took extra care of boys. Knowing this research is about first-born children, and a boy is an ancestor, this was particularly important for farmers, the elite and the middle class. Figure 2.1 presents the bar plot of the lifespan, expressed in decades, for the different family members.

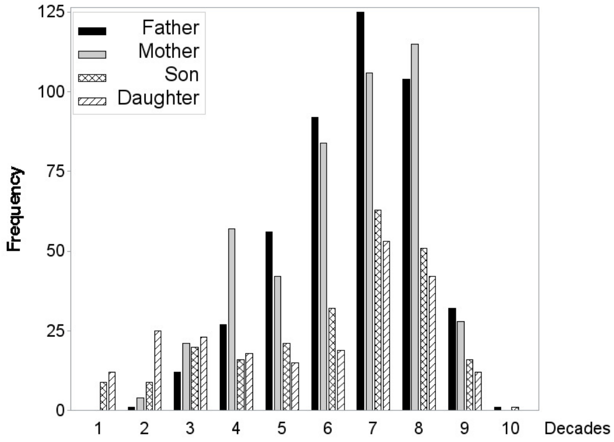


Figure 2.1: Moerzeke data. Bar plot of the lifespan, expressed in decades, for the different family members.

Civil and church registers were collected to construct the database, which can be categorized as good quality and suitable for population studies. The database holds information of individuals who were born, died or married in Moerzeke.

2.3 Modeling Framework

2.3.1 The Plackett-Dale Model

The multivariate Dale model for survival times (Tibaldi et al., 2004) will be reviewed. The model augments standard univariate survival distributions for each of the family members (mother, father, and child) separately, with global cross ratios, also termed global odds ratios, to describe the association between pairs of survival outcomes. The main advantage is that the univariate distributions derived from such a joint distribution are exactly equal to those that would be obtained, were univariate analyses done on each outcome separately. This property is not automatically obtained with frailty models (Clayton, 1978; Hougaard, 1986) or conventional generalized linear mixed models (GLMM; Molenberghs and Verbeke, 2005).

Another feature is the elegant way in which the association between the survival outcomes is modeled. This is important when one is interested in a separation between social and genetic aspects of longevity. Regarding the former, it is a strong asset that a number of covariates describing social and demographic aspects can be incorporated into the models for both survival values as well as for the association. Sociological explanations often have empirical implications in terms of sex and parity differences. For example, if inheritance of material products, such as a farm, is sex and birth-order related, this must be reflected in the sex and parity pattern of longevity inheritance.

The methodology will be applied to the Moerzeke data, introduced in Section 2.2. Consider the (unadjusted) survival times T_{ij}^* of mother, father, and first child $(j = 1, 2, 3)$ of $i = 1, \ldots, 457 (= N)$ families, and assume marginal Weibull distributions for each adjusted survival time T_{ij}, i.e., $T_{ij}^* - \{\text{given threshold}\}_j$. To avoid left-truncation, the threshold is held fixed at 10 years. The information of household i can be formulated in vector notation as $(T_{i1}, T_{i2}, T_{i3}, \Delta_{i1}, \Delta_{i2}, \Delta_{i3}, x_{i1}, \ldots, x_{in_i})$, with n_i the length of the covariate vector. Δ_{ij} indicates whether lifetime j in cluster i is observed (1) or censored (0). The data for family member j is denoted by $\boldsymbol{W}_{ij} = (T_{ij}, \Delta_{ij}, \mathbf{X}_i)$. $\mathbf{X}_i$ expresses the n_i-dimensional vector of covariates.

The multivariate Plackett distribution can be specified for any number of outcomes, using two-way and higher-order cross ratios to specify the associations (Molenberghs and Lesaffre, 1994). Such a specification is unavoidable should one choose for full ML inference. However, calculations quickly become very cumbersome. Therefore, PL estimation will be utilized, following ideas of Tibaldi et al. (2004), obviating the need to specify associations beyond the second order. The PL function is of the pairwise form (le Cessie and Van Houwelingen, 1994; Renard et al., 2004):

$$\ln p\ell(\boldsymbol{\Phi}) = \sum_{i=1}^{N} \sum_{(r,\ell)\in S} \ln f_{T_r,T_\ell}(\boldsymbol{W}_{ir}, \boldsymbol{W}_{i\ell}, \boldsymbol{\Phi}), \tag{2.1}$$

with $S = \{(r, \ell) \mid r < \ell, r = 1, 2, 3, \ell = 1, 2, 3\}$, $f_{T_r,T_\ell}(\boldsymbol{W}_{ir}, \boldsymbol{W}_{i\ell}, \boldsymbol{\Phi}) = \partial F_{T_r,T_\ell}(\boldsymbol{W}_{ir}, \boldsymbol{W}_{i\ell}, \boldsymbol{\Phi})/\partial \boldsymbol{W}_{ir} \partial \boldsymbol{W}_{i\ell}$ the density function of the bivariate PD distribution (Dale, 1986; Mardia, 1970; Tibaldi et al., 2004), where

$$F_{T_r,T_\ell}(\boldsymbol{W}_{ir},\boldsymbol{W}_{i\ell},\boldsymbol{\Phi}) = \\ \begin{cases} \frac{1+[F_{T_r}(\boldsymbol{W}_{ir})+F_{T_\ell}(\boldsymbol{W}_{i\ell})]\cdot(\theta_{r\ell}-1)-H[F_{T_r}(\boldsymbol{W}_{ir}),F_{T_\ell}(\boldsymbol{W}_{i\ell}),\theta_{r\ell}]}{2\cdot(\theta_{r\ell}-1)}, \text{if } \theta_{r\ell}\neq 1, \\ F_{T_r}(\boldsymbol{W}_{ir})\cdot F_{T_\ell}(\boldsymbol{W}_{i\ell}), \text{if } \theta_{r\ell}=1, \end{cases} \tag{2.2}$$

and

$$H[F_{T_r}(\boldsymbol{W}_{ir}),F_{T_\ell}(\boldsymbol{W}_{i\ell}),\theta_{r\ell}] = \\ \sqrt{\begin{array}{c}\{1+(\theta_{r\ell}-1)\cdot[F_{T_r}(\boldsymbol{W}_{ir})\cdot F_{T_\ell}(\boldsymbol{W}_{i\ell})]\}^2 \\ +4\cdot\theta_{r\ell}\cdot(1-\theta_{r\ell})\cdot F_{T_r}(\boldsymbol{W}_{ir})\cdot F_{T_\ell}(\boldsymbol{W}_{i\ell}),\end{array}} \tag{2.3}$$

with F_{T_r} and F_{T_ℓ} the univariate marginal distribution functions, here of a Weibull form. $\boldsymbol{\Phi} = (\boldsymbol{\theta}', \boldsymbol{\xi}', \boldsymbol{\lambda}', \boldsymbol{\rho}')'$ is the vector of parameters with $\boldsymbol{\theta}$, $\boldsymbol{\xi}$, and $(\boldsymbol{\lambda}', \boldsymbol{\rho}')$ the subvectors of association parameters $\theta_{r\ell}$, coefficients corresponding to the covariates $\mathbf{X}$, and parameters from the Weibull distribution, respectively.

The PL estimator $\widehat{\boldsymbol{\Phi}}$ is determined as the maximizer of Eq. (2.1). Appendix S.1 shows details on consistency and asymptotic normality. The PD model captures association between a pair of survival times T_r and T_ℓ, through global cross ratios $\theta_{r\ell}(\mathbf{t}_r, \mathbf{t}_\ell)$:

$$\theta_{r\ell}(t_r, t_\ell) = \frac{F_{T_r,T_\ell}\cdot[1-F_{T_r}-F_{T_\ell}+F_{T_r,T_\ell}]}{[F_{T_r}-F_{T_r,T_\ell}]\cdot[F_{T_\ell}-F_{T_r,T_\ell}]}. \tag{2.4}$$

$\theta = 1$ corresponds to independence and the range is from 0 to $+\infty$. The log odds ratio is often used to avoid range restrictions and for which, due to symmetry, standard errors are more readily interpretable. All plots will be based on the log odds ratio. Supplementary Materials S.1 gives a connection with Kendall's τ and Spearman's ρ.

The association between longevity of family members can be estimated with the PL method. In the case of ML estimation, several tools can be applied to test the parameters of the model such as Wald, score, or likelihood ratio tests. However, these tests need to be extended in our case as it was done by Geys et al. (1999) in the presence of clustered multivariate binary data. An overview for what is needed for our purpose is given in Supplementary Materials S.2.

2.3.2 The Weibull-Gamma-Normal Model

The CM framework (Booth et al., 2003; Molenberghs et al., 2007; Molenberghs et al., 2010) is considered as an alternative random effects and frailties (Duchateau and Janssen, 2007) based framework to model within-household association, using normal random effects like in a GLMM (Molenberghs and Verbeke, 2005) as well as dispersion (Hinde and Demétrio, 1998), through gamma random effects. Capturing extra-model dispersion is cumbersome in the PD framework.

For the Moerzeke data (Section 2.2), the WGN model takes the form:

$$\begin{aligned}
f_{T_i}(\boldsymbol{t}_i \mid \boldsymbol{\theta}_i, \boldsymbol{b}_i) &= \prod_{j=1}^{3} \lambda_j \cdot \rho_j \cdot t_{ij}^{\rho_j - 1} \cdot \theta_{ij} \cdot e^{\Omega_{ij}} \cdot e^{-\lambda_j \cdot t_{ij}^{\rho_j} \cdot \theta_{ij} \cdot e^{\Omega_{ij}}}, && (2.5)\\
\Omega_{ij} &= \boldsymbol{x}_{ij}^{'} \cdot \boldsymbol{\xi} + \boldsymbol{z}_{ij}^{'} \cdot \boldsymbol{b}_i, && (2.6)\\
f(\boldsymbol{\theta}_i) &= \prod_{j=1}^{3} \frac{1}{\beta_j^{\alpha_j} \cdot \Gamma(\alpha_j)} \cdot \theta_{ij}^{\alpha_j - 1} \cdot e^{-\theta_{ij}/\beta_j}, && (2.7)\\
f(\boldsymbol{b}_i) &= \frac{1}{(2\pi)^{q/2} \cdot \mid D \mid^{1/2}} \cdot e^{-\frac{1}{2} \cdot \boldsymbol{b}_i^{'} \cdot D^{-1} \cdot \boldsymbol{b}_i}. && (2.8)
\end{aligned}$$

Eq. (2.5) represents the conditional outcome distribution; Eq. (2.7) represents the conjugate distribution; Eq. (2.8) represents the normal random effects distribution. The Weibull parameters λ_j and ρ_j and the gamma parameters α_j and β_j are allowed to vary with family member, but evidently they do not have to. $\boldsymbol{\xi}$ are fixed parameters with corresponding covariates $\boldsymbol{x}_{ij}^{'}$; $\boldsymbol{b}_i$ are family-specific parameters with corresponding covariates $\boldsymbol{z}_{ij}^{'}$. D in Eq. (2.8) defines the variance-covariance matrix of the random-effects vector $\boldsymbol{b}_i$. For identification purposes, we enforce the constraint $\alpha_j \cdot \beta_j \equiv 1$ (Duchateau and Janssen, 2007) upon the gamma parameters.

Special cases of the WGN model are the Weibull-gamma (WG–), the Weibull-normal (W–N) and the standard Weibull (W—) model. Closed-forms expressions exists for the marginal mean and variance (Molenberghs et al., 2015), making it an attractive framework for mathematically examining relevant features such as dispersion and correlation.

2.4 The Longevity of Life Span

Gavrilova et al. (1998a), Korpelainen (1999), Gudmundsson et al. (2000), and Matthijs et al. (2002), have frequently been claimed that longevity or lifespan (the full life length) has a familial component, passed on from parent to offspring. We will focus on a specific aspect of longevity, in the context of three-member clusters made up of father, mother, and first born child. Various interesting questions surround this association: Is it sociological or biological, or both? Are the patterns sex-specific? We emphasize the effect of parental longevity on offspring mortality. While Matthijs et al. (2002) found that parental longevity had a relatively strong effect in women born in the early 19^{th} century, and that for the man-child the effect emerges only decades later, the precise mechanisms underlying the association remain unknown. We aim to shed more light on the sex and age-related patterns of the transmittance of longevity, while leaving an in-depth discussion of the many possible components that may contribute in the transmission of longevity of the first born child to others. Gavrilova and Gavrilov (2001), Gavrilov and Gavrilova (2001) and Salaris et al. (2013) provided a relevant discussion on familial transmission of human longevity.

We will examine whether differences arise in the influence of fathers and mothers on the transmission of lifespan or longevity and whether the transmission of mortality is related to the parents' age-at-death. Oftentimes, the analysis of longevity is limited to people over the age of 50, to eliminate phenotype variation stemming from contagious diseases, accidents, war, and environmental maternal effect during early childhood (Korpelainen, 1999). Causes of death such as accidents and pregnancy-related diseases dominate the mortality pattern under that age and may blur the picture. Genetic variability for survival is expected to increase with age, following the evolutionary aging theory in general and the mutation accumulation hypothesis in particular (Gavrilova et al., 1999). Of course, there are situations where a different choice of age cutoff is warranted. Gavrilova et al. (1998b), for example, motivate age 30 as a relevant cutoff.

2.5 Analysis of the Moerzeke Data

2.5.1 The Plackett-Dale Model

This is the first application of the PD model to data of a familial type. We address whether the familial transmittance of survival to daughters is mainly maternally or paternally transmitted (Matthijs et al., 2002) by way of the adjusted survival time T_{ij} (Section 2.3.1).

The model contains time-to-events for father, mother, and child, measured by the total number of days that he/she lived. The (continuous) year-of-birth of each family member and sex (0 for women, 1 for men) of the child are included as covariates. Note that, of course, each family members outcome is affected by a different year of birth, and a different gender-of-child effect is assumed for the family members. While odd at first sight, such an effect is useful to check whether gender of the child, e.g., as a result of testosterone production, has an impact on, for example, the mothers lifespan. This leads to the next parameters: ξ_{YB1} year of birth of the mother, ξ_{YB2} year of birth of the father, and ξ_{YB3} year of birth of the child. The sex of the child was included in all marginal Weibull parts, where corresponding coefficients vary between members: ξ_{G1} for mother,

ξ_{G2} for father, and ξ_{G3} for child. When interest would be primarily focused on the marginal distributions, one could consider more complex forms of marginal regression like including non-linear effects of time. Because our main interest lies with the association, and thanks to (near) orthogonality of the association and marginal parts in the PD (Palmgren, 1989), we do not make the mean structure more complex.

Table 2.1 displays the PL parameter and precision estimates. The logarithm of the estimated association parameter between mother and child is 0.301 (95% confidence interval $[0.168; 0.434]$), indicating a positive association between them. For father-child the value comes out lower (-0.013; not statistically significant). Values of Spearman's and Kendall's coefficients, together with their confidence intervals, are presented in Table 2.2.

Table 2.1*: Moerzeke data. Plackett–Dale model for father, mother, and child (son or daughter). Pseudo-likelihood estimates (empirically corrected standard errors) of the survival times. The indices 1 refer to mother; 2 to father; 3 to child.*

Effect	Parameter	Est. (s.e.)
Association (1,2)	θ_{12}	1.138 (0.160)
Association (1,3)	θ_{13}	1.351 (0.180)
Association (2,3)	θ_{23}	0.987 (0.137)
Sex effect on mother	ξ_{G1}	-0.110 (0.069)
Sex effect on father	ξ_{G2}	-0.077 (0.074)
Sex effect on child	ξ_{G3}	-0.149 (0.056)
Year of birth of mother	ξ_{YB1}	-1.076 (1.318)
Year of birth of father	ξ_{YB2}	-0.880 (1.523)
Year of birth of child	ξ_{YB3}	-3.137 (1.179)
Shape parameter mother	ρ_1	4.794 (0.168)
Shape parameter father	ρ_2	5.768 (0.194)
Shape parameter child	ρ_3	2.081 (0.126)
Scale parameter mother	λ_1	0.217 (0.493)
Scale parameter father	λ_2	0.184 (0.473)
Scale parameter child	λ_3	3.068 (1.028)

Table 2.2: *Moerzeke data. Plackett-Dale model for father, mother, and child (son or daughter). Pseudo-likelihood estimates and inference for the association parameters θ, Kendall and Spearman coefficients (95% confidence intervals). Pseudo-likelihood tests and theirs p-values. The indices 1 refer to mother; 2 to father; 3 to child.*

(i,j)	$\log(\theta_{ij})$	Kendall's τ_{ij}	Spearman's ρ_{ij}
$(1,2)$	0.129 (-0.012;0.270)	0.029 (0.013;0.044)	0.043 (-0.049;0.135)
$(1,3)$	0.301 (0.168;0.434)	0.067 (0.052;0.081)	0.100 (0.014;0.186)
$(2,3)$	-0.013 (-0.152;0.125)	-0.003 (-0.018;0.012)	-0.004 (-0.095;0.086)

H_0	Wald	p-value	G^{*2}	p-value	$S^*(m.b.)$	p-value	$S^*(e.c.)$	p-value
$\theta_{12}=1$	0.861	0.389	0.943	0.332	1.200	0.273	0.840	0.359
$\theta_{13}=1$	1.950	0.051	5.402	0.020	4.237	0.040	5.275	0.022
$\theta_{23}=1$	-0.096	0.923	0.010	0.922	0.015	0.904	0.009	0.925

Inferences are drawn using the tests described in Supplementary Materials S.2. The null hypothesis of no association was tested in each case through the Wald, score, and PL ratio tests and the results are displayed in Table 2.2. Note that if the results are used for formal inferences, correction for multiple comparisons may be considered. Similar conclusions are obtained. The Wald statistics give the least significant p-values in most cases. The null hypotheses of no association between father's and mother's survival times and father's and child's survival times (i.e., $\theta_{12} = 1$ and $\theta_{23} = 1$, respectively) cannot be rejected by any of the tests, but the situation is different for mother and child. We reject $\theta_{13} = 1$. The latter was already reflected in the fact that the 95% confidence intervals for Kendall's and Spearman's coefficients contain the zero value for the first and third hypotheses but not for the second one.

To enhance insight, the model is applied to different subsets. First, distinction is made between sons and daughters. It is known that differences in mortality between men and women are not necessarily the same for all age groups. Figure 2.2 displays the estimated unadjusted survival curves for sons and daughters in three different groups (same selection as outlined above). The model is fitted for sons and daughters separately, a simple but insightful way to find out whether the association depends on the offspring's sex.

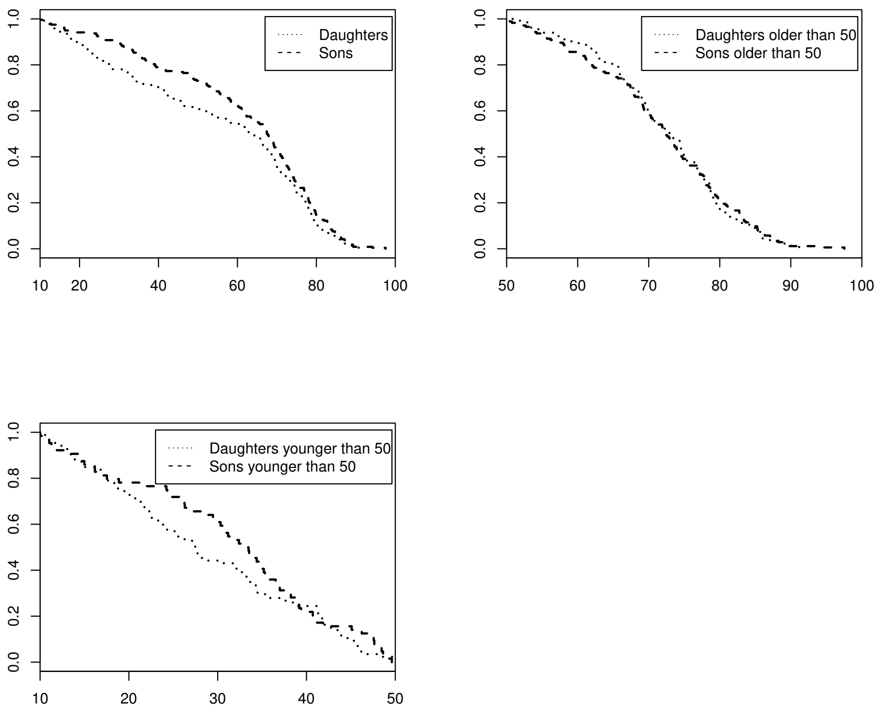


Figure 2.2: *Moerzeke data. Survival curves for sons and daughters with a cut-off point of 50 years. The horizontal axis refers to the (unadjusted) age in years and the vertical axis is the probability of survival.*

Table 2.3 displays sex-specific parameter estimates. No significant differences are present for sons (Table 2.4), while for daughters there seems to be a stronger association in case of mothers and daughters than for the rest of the association parameters (θ_{13}). It is of interest to see whether each of the pairwise associations is significantly different between sons and daughters. This is not the case, using Wald tests, which produce p values of 0.6915 for the mother-father, 0.2637 for the mother-child, and 0.6558 for the father-child association.

Table 2.3*: Moerzeke data. Plackett-Dale model for father, mother, and child (son and daughter separately). Pseudo-likelihood estimates (standard errors) of the survival times. The indices 1 refer to mother; 2 to father; 3 to child.*

Effect	Parameter	Sons	Daughters
Association (1,2)	θ_{12}	0.979 (0.200)	1.351 (0.266)
Association (1,3)	θ_{13}	1.280 (0.256)	1.403 (0.247)
Association (2,3)	θ_{23}	0.936 (0.198)	1.046 (0.194)
Year of birth of mother	ξ_{YB1}	-2.736 (1.719)	0.994 (2.068)
Year of birth of father	ξ_{YB2}	0.494 (2.180)	-2.360 (2.120)
Year of birth of child	ξ_{YB3}	-4.267 (1.691)	-2.271 (1.698)
Shape parameter mother	ρ_1	4.773 (0.242)	4.824 (0.234)
Shape parameter father	ρ_2	5.855 (0.268)	5.688 (0.282)
Shape parameter child	ρ_3	2.422 (0.241)	1.814 (0.132)
Scale parameter mother	λ_1	0.396 (0.645)	0.100 (0.769)
Scale parameter father	λ_2	0.119 (0.667)	0.295 (0.667)
Scale parameter child	λ_3	4.473 (1.236)	1.972 (1.717)

Table 2.4*: Moerzeke data. Plackett-Dale model for father, mother, and child (son and daughter separately). Pseudo-likelihood tests and theirs p-values for the association parameters for models from Table 2.2. The indices 1 refer to mother; 2 to father; 3 to child.*

Offspring	H_0	Wald	p-value	G^{*2}	p-value	$S^*(m.b.)$	p-value	$S^*(e.c.)$	p-value
Sons	$\theta_{12} = 1$	-0.105	0.917	0.013	0.910	0.040	0.841	0.036	0.850
	$\theta_{13} = 1$	1.094	0.274	1.759	0.185	1.444	0.229	1.536	0.215
	$\theta_{23} = 1$	-0.324	0.746	0.113	0.737	0.486	0.486	0.478	0.489
Daughters	$\theta_{12} = 1$	1.317	0.188	2.477	0.116	2.557	0.110	2.286	0.131
	$\theta_{13} = 1$	1.631	0.103	3.590	0.058	2.740	0.098	3.734	0.053
	$\theta_{23} = 1$	0.234	0.815	0.058	0.810	0.086	0.769	0.057	0.811

Next, it is explored whether associations within families can also depend on the age-at-death of the offspring. The PD model is fitted in six different groups: Overlapping sets of offspring, reaching at least the age of 10, 20, 30, 40, 50, and 60 years. This produces a decreasing number of observations, implying care is needed with interpretation. A graphical summary of the log of the estimated association values (and their 95 % confidence intervals) is given in Figure 2.3, where all three $\log(\theta)$ parameters are plotted for each group using as cut-off point the age of mortality of the offspring.

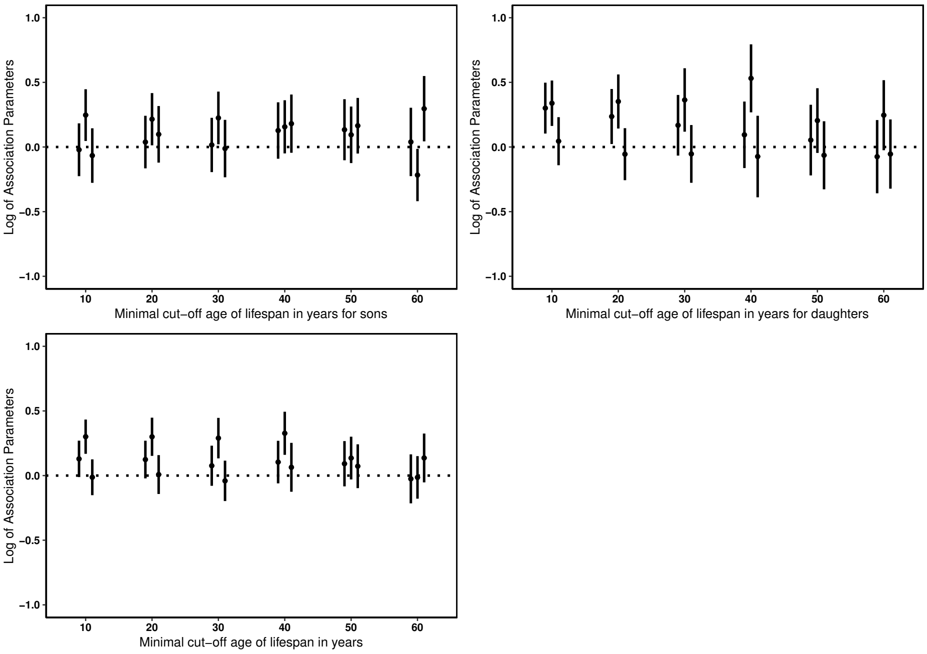


Figure 2.3: *Moerzeke data. Log of association parameters* θ_{12}, θ_{13}, *and* θ_{23} *(from left to right) for offspring mortality group, separated for sons (top-left), daughters (top-right) and all children (bottom-left). The horizontal axis refers to the households where the offspring survived at least the given threshold. The indices 1 refer to mother; 2 to father; 3 to child.*

Each set of parameters represents, from left to right, $\log(\theta_{12})$, $\log(\theta_{13})$, and $\log(\theta_{23})$. The second $\log(\theta)$ estimate in each group of three corresponds to the mother-child relationship; this particular relationship typically exhibits a stronger association than for the other pairs. The association between mother and daughter is not gradually becoming stronger when stepwise excluding those daughters who have died at an early age. On the other hand, parameters for those who have reached the age of 50 are generally lower. This finding is somewhat surprising in the light of, for example, the findings of Korpelainen (1999). The same findings are recovered for a set of French agricultural villages (Cournil et al., 2000). It must be added that the visibility of the association with daughters at an early age, is not against a social explanation of the transmittance of mortality that focuses on the position of women in intra-household resource competition. It should not be excluded that the weak position of women within some households has effects at almost all ages.

In the second panel, associations between parents and daughters were plotted and there clearly is a different structure as opposed to the other two. Larger values of $\log(\theta_{13})$ are observed in almost every group. The latter result implies, once more, higher associations between longevity of mothers and daughters. Some care has to be taken, given that the groups overlap. The differences observed in the length of the confidence intervals are due to the progressively decreasing sample sizes.

Finally, the influence of the age-at-death of the parents will be explored. Figure 2.4 offers a graphical display of their logarithms to ease interpretation of these values. These confirm the previous findings that the association between mothers and daughters is strongest and visible at all ages of the daughter. This suggests that adult mortality of woman family members is connected in a very general way, leading to associations in longevity between mothers and daughters, irrespective of the age thresholds for the offspring. Interpretation should be done cautiously due to relatively small sample sizes.

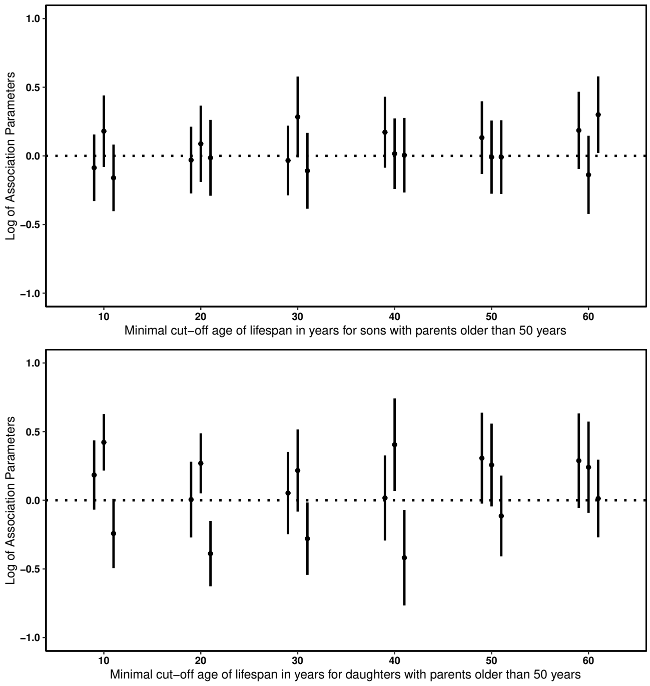


Figure 2.4: *Moerzeke data. Log of association parameters* θ_{12}, θ_{13}, *and* θ_{23} *(from left to right) for offspring mortality group, separated for sons (top) and daughters (bottom) with parents dying older than 50 years. The horizontal axis refers to households where the offspring survived at least the given threshold with parents dying older than 50 years. The indices 1 refer to mother; 2 to father; 3 to child.*

2.5.2 The Weibull-Gamma-Normal Model

The data will be examined using the WGN model (Section 2.3.2). The same data selection and covariates of Section 2.5.1 are considered to model the adjusted time-to-event outcome T_{ij}. Rather than following up on the discussion of familial transmittance of longevity (Section 2.5.1), the CM framework is merely used here to explore new aspects within the data.

The additional analyses result in three specific versions: (1) the W–N model, (2) the WGN model, and (3) the marginalized WGN model. While subject-specific interpretations ensue in the first and second case, population-averaged interpretations are available for the third. The gamma parameters $\alpha_j \equiv 1/\beta_j \equiv \alpha$ are held constant, and a random-intercept is included to take between household-variability into account ($D \equiv \sigma^2$). ML estimation is performed (Molenberghs et al., 2015). Estimates are gathered from the SAS procedure NLMIXED and listed in Tables 2.5 and 2.6. The source code for fitting these models in NLMIXED is available in Supplementary Material S.3.

Table 2.5: *Moerzeke data. Weibull-gamma-normal model for father, mother, and child (son or daughter). Maximum-likelihood estimates (standard errors) of the survival times in (1) the Weibull (W—) model and (2) Weibull-normal (W–N) model. The indices 1 refer to mother; 2 to father; 3 to child.*

		W—	**W–N**
Effect	Parameter	Est. (s.e.)	Est. (s.e.)
Sex effect on mother	ξ_{G1}	-0.115 (0.094)	-0.116 (0.094)
Sex effect on father	ξ_{G2}	-0.078 (0.094)	-0.078 (0.094)
Sex effect on child	ξ_{G3}	-0.150 (0.094)	-0.150 (0.094)
Year of birth of mother	ξ_{YB1}	-1.124 (1.886)	-0.895 (1.888)
Year of birth of father	ξ_{YB2}	-0.895 (2.010)	-1.112 (2.009)
Year of birth of child	ξ_{YB3}	-3.174 (1.932)	-3.173 (1.932)
Shape parameter mother	ρ_1	4.804 (0.187)	4.805 (0.187)
Shape parameter father	ρ_2	5.771 (0.218)	5.769 (0.218)
Shape parameter child	ρ_3	2.080 (0.087)	2.080 (0.087)
Scale parameter mother	λ_1	$6.93E-4$ $(2.36E-3)$	$4.58E-4$ $(1.56E-3)$
Scale parameter father	λ_2	$5.90E-5$ $(2.14E-4)$	$8.80E-5$ $(3.17E-4)$
Scale parameter child	λ_3	11.033 (38.875)	11.013 (38.803)
Std. dev. random effect	σ	$--$	$-1.8E-4$ (0.054)
-2 log-likelihood		5391.7	5391.7

Table 2.6: *Moerzeke data. Weibull-gamma-normal model for father, mother, and child (son or daughter). Maximum-likelihood estimates (standard errors) of the survival times in (3) Weibull-gamma-normal (WGN) model, also referred as the combined model (CM), and (4) the marginalized WGN model. The indices 1 refer to mother; 2 to father; 3 to child.*

		WGN	**Marg. WGN**
Effect	Parameter	Est. (s.e.)	Est. (s.e.)
Sex effect on mother	ξ_{G1}	-0.116 (0.094)	-0.117 (0.094)
Sex effect on father	ξ_{G2}	-0.078 (0.094)	-0.078 (0.094)
Sex effect on child	ξ_{G3}	-0.150 (0.094)	-0.150 (0.094)
Year of birth of mother	ξ_{YB1}	-0.898 (1.891)	-0.898 (1.891)
Year of birth of father	ξ_{YB2}	-1.113 (2.013)	-1.111 (2.013)
Year of birth of child	ξ_{YB3}	-3.176 (1.935)	-3.176 (1.935)
Shape parameter mother	ρ_1	4.809 (0.187)	4.809 (0.187)
Shape parameter father	ρ_2	5.775 (0.218)	5.775 (0.218)
Shape parameter child	ρ_3	2.081 (0.087)	2.081 (0.087)
Scale parameter mother	λ_1	$4.58E-4$ $(1.56E-3)$	$4.58E-4$ $(1.56E-3)$
Scale parameter father	λ_2	$8.70E-5$ $(3.15E-4)$	$8.70E-5$ $(3.14E-4)$
Scale parameter child	λ_3	11.073 (39.080)	11.080 (39.105)
Std. dev. random effect	σ	$-1E-4$ (0.054)	$-1E-5$ (0.054)
Dispersion parameter	α	459.02 (465.990)	459.02 (466.050)
-2 log-likelihood		5392.7	5392.7

From the fitted WGN model (Table 2.6), extra-dispersion parameter α is observed to be high (459.02), while the estimated variance component σ goes to 0. A similar result for σ is obtained in the W–N case in Table 2.5. To understand this phenomenon, note that in PROC NLMIXED a partially marginalized likelihood is specified: Analytically integrated over the gamma random effect but numerically over the normal random effect (Molenberghs et al., 2015), using adaptive Gaussian quadrature. It forces the random effects variance to be non-negative. This corresponds to a hierarchical viewpoint, whereas in a fully marginal one, the variance component could be negative. This might be desirable in several situations: Negative intraclass correlation, underdispersion, and others.

It is hypothesized that underdispersion is plausible. To motivate this, a descriptive exploration is done on the observed vs. fitted variances of WGN for longevity per household. Figure 2.5 (right) gives a graphical display of the observed (scatterplot) and fitted variances (line plot) of longevity per household consisting of the father, mother and first born child, resulting in approximately 64% (292 out of 457 – Figure 2.6) of the households with underdispersion. Overall, an observed variance of 3.19 and a fitted variance of 3.27 is seen, corroborating the likely presence of underdispersion.

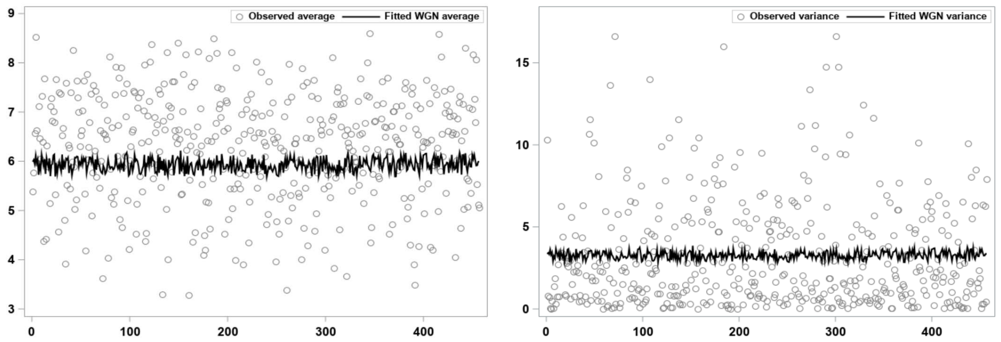


Figure 2.5: Moerzeke data. Mean (left) and variance (right) profiles for the observed data and fitted Weibull-gamma-normal (WGN) model. The horizontal and vertical axis refers to the household ID and lifespan, expressed in decades, respectively.

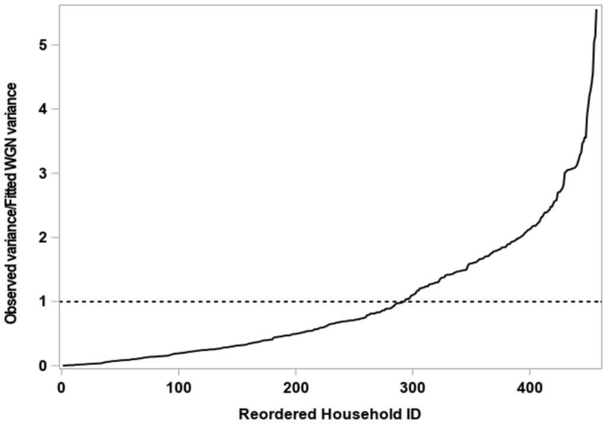


Figure 2.6: Moerzeke data. Variance ratio between the observed and fitted Weibull-gamma-normal (WGN) variance per household. The horizontal axis refers to the ordered household ID's, based on the increasing variance ratio.

While subject-specific interpretations are obtained from Eq. (2.5)–(2.8), population-averaged interpretations are often preferred. Efendi et al. (2014) constructed a so-called combined marginalized multilevel model (COMMM) version to this effect. Similar results were retrieved due to the estimated variance component σ^2 of approximately 0 (Table 2.6).

2.5.3 Accounting for Censoring

Matthijs et al. (2002) avoided censoring problems by limiting the analysis to a subset of the population with complete data on lifetimes. Family members with incomplete records were excluded from the analysis. However, information can be extracted from incomplete records.

Two sources of information were considered to recover additional observations. First, the date of marriage, if present, can be used as a censored observation for lifetime of both parents. Second, the date of birth of the last child of the family can be used to define censoring for mother's survival time. By combining these strategies, only 17 new families were incorporated and the analyses repeated. This at the same time is a valuable form of sensitivity analysis.

Table 2.7 contains the results of all fits for the multivariate PD model. The first column shows the results for sons and daughters together; the last two columns show the results for each of the sexes separately. Molenberghs et al. (2015) showed that fitting partially censored data for CMs can easily be done as well.

Table 2.7: *Moerzeke data. Plackett-Dale model for father, mother, and child (all children; son and daughter separately). Pseudo-likelihood estimates (standard errors) of the survival times. Censored observations are included. The indices 1 refer to mother; 2 to father; 3 to child.*

Effect	Parameter	All children	Sons	Daughters
Association (1,2)	θ_{12}	1.141 (0.161)	0.981 (0.200)	1.352 (0.266)
Association (1,3)	θ_{13}	1.332 (0.174)	1.305 (0.256)	1.346 (0.234)
Association (2,3)	θ_{23}	1.026 (0.141)	0.967 (0.205)	1.090 (0.200)
Sex effect on mother	ξ_{G1}	-0.116 (0.069)	-	-
Sex effect on father	ξ_{G2}	-0.083 (0.074)	-	-
Sex effect on child	ξ_{G3}	-0.148 (0.055)	-	-
Year of birth of mother	ξ_{YB1}	-0.719 (1.304)	-2.157 (1.684)	1.114 (2.062)
Year of birth of father	ξ_{YB2}	-0.794 (1.511)	0.315 (2.165)	-1.990 (2.102)
Year of birth of child	ξ_{YB3}	-3.114 (1.163)	-3.888 (1.685)	-2.596 (1.655)
Shape parameter mother	ρ_1	4.728 (0.162)	4.693 (0.234)	4.771 (0.225)
Shape parameter father	ρ_2	5.730 (0.191)	5.801 (0.266)	5.667 (0.274)
Shape parameter child	ρ_3	2.070 (0.121)	2.404 (0.230)	1.806 (0.128)
Scale parameter mother	λ_1	0.191 (0.495)	0.324 (0.643)	0.095 (0.775)
Scale parameter father	λ_2	0.180 (0.473)	0.125 (0.669)	0.264 (0.665)
Scale parameter child	λ_3	3.065 (1.020)	3.445 (1.241)	2.773 (1.683)

The estimated association between child and mother remains the highest of the three, even though a slight decrease is observed compared to the complete case analysis. All associations are somewhat higher in the case of daughters rather than sons, reflecting the same conclusions as before.

These techniques easily allow one to work with censored times. Whether or not to include censored observations depends, of course, on the context.

2.6 Conclusion

A marginal PD model for the joint analysis of a number of, possibly censored, survival outcomes is proposed. Parameter and precision estimation is performed using PL. Appropriate PL-based test statistics were proposed. While the statistical power of the Wald test is generally lower compared to PL and pseudo-score tests, it is the most easy one to implement. The WGN model was considered, a hierarchical model allowing for intra-cluster correlation as well as extra-model dispersion.

The methods were applied to a study on longevity inheritance in a small Flemish village (18^{th}–20^{th} century). The sample is relatively small (457 households). Importantly though, the main research interest lies with the estimation of association. The PD model includes a (log) odds ratio to describe association, which can be converted into Kendall's τ and Spearman's ρ. In addition, the association is allowed to be different for each of the three pairs of association (father↔mother; child↔mother; child↔father). Each of these associations can be made dependent on covariates. The association between child and mother is the highest of the three and also that all associations are somewhat higher in the case of daughters rather than sons, consistent with earlier findings.

There is no obstacle in applying it to larger sets of triplets given that the data retain their trivariate nature and the PL actually utilizes only bivariate joint models. The computation time will increase quadratically with the number of independent triplets.

The presence of underdispersion within the data was recorded, which is directly relevant to the variability structure of the Moerzeke data. It is a phenomenon that is rarely observed in practice when analyzing clustered demographical data. This has been noticed from the WGN. The gamma parameter was estimated to be high, with infinity corresponding to independence. A possible reason for this result is the occurrence of negative variance components, which we further linked to the presence of underdispersion. By using descriptive statistics to examine the within-household variability, we have seen that many households encompass low within-household variability. These findings could be linked to the geographical isolated nature of Moerzeke. While these discoveries have not been examined in detail yet within the Moerzeke study, more broadly, within demographic clustered data, we conclude this chapter by stating that further research possibilities open in the exploration of underdispersion within hierarchical data structures. Because of limiting this research to familial associations, the precise mechanisms behind these associations remains open for future investigation.

Chapter 3

A Weibull-count Approach for Handling Under- and Overdispersed Longitudinal/Clustered Data Structures

This chapter is based upon:

Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Ribeiro Jr., E. E., Demétrio, C. G. B., and Hinde, J. (2020). A Weibull-count approach for handling under- and overdispersed longitudinal/-clustered data structures. *Statistical Modelling*, **19(5)**, 569–589.

Abstract

A Weibull-model-based approach is examined to handle under- and overdispersed discrete data in a hierarhical framework. This methodology was first introduced by Nakagawa and Osaki (1975), and later examined for under- and overdispersion by Klakattawi et al. (2018) in the univariate case. Extensions to hierarchical approaches with under- and overdispersion were left unnoted, even though they can be obtained in a simple manner. This is of particular interest when analyzing clustered/longitudinal data structures, where the underlying correlation structure is often more complex compared to cross-sectional studies.

A random effects extension of the Weibull-count model is proposed and applied to two motivating case studies, originating from the clinical and sociological research fields. A goodness-of-fit evaluation of the model is provided through a comparison of some well-known count models, i.e., the negative binomial, Conway-Maxwell-Poisson, and double Poisson models.

Empirical results show that the proposed extension flexibly fits the data, more specifically, for heavy-tailed, zero-inflated, overdispersed and correlated count data. Discrete left-skewed time-to-event data structures are also flexibly modeled using the approach, with the ability to derive direct interpretations on the median scale, provided the complementary log-log link is used. Finally, a large simulated set of data is created to examine other characteristics such as computational ease and orthogonality properties of the model, with the conclusion that the approach behaves best for highly overdispersed cases.

3.1 Introduction

The analysis of count data has received considerable attention in the literature, with practical applications in public health, and social and behavioral sciences. Since the introduction of generalized linear models (GLMs) by Nelder and Wedderburn (1972), a GLM based on the Poisson distribution, a well-known member of the exponential family (EF), is a commonly applied statistical model for count data analysis. In spite of its many advantages, e.g., the ability of fitting skewed non-negative data, the model possesses a too restricted mean-variance relationship (equidispersion), a characteristic that is often violated in the data. In particular, two situations can occur: (1) The variability in the data is larger than the theoretical variance implied by the model (overdispersion), and (2) the variability in the data is smaller than the theoretical variance (underdispersion). For these and other reasons, e.g., zero-inflation (Iddi and Molenberghs, 2013) and heavy-tailed profiles (Zhu and Joe, 2009), many alternative and extended models have been proposed in the literature.

These models can often be classified as exponential dispersion models (EDMs), introduced by Jørgensen (1987), which include the GLM families as a special case. More specifically, EDMs increase the range of univariate/multivariate variance functions for which generalized linear type models exist. Kokonendji et al. (2004), for example, investigated two classes of EDMs for count data that is overdispersed compared to the Poisson distribution, i.e., the Poisson-Tweedie and Hinde-Demétrio classes. Efron (1986), on the other hand, proposed a different class of regression families, by introducing a second parameter in the EF that controls the dispersion independently of the mean while still carrying out the usual regression analysis in a GLM context. These are the so-called double-EFs because they enjoy EF properties simultaneously for the mean and dispersion parameters. A popular member is the double Poisson (DP) model (Appendix S.4.4). A general overview of some popular models is given in Supplementary Material S.4 for subsequent comparison (Section 3.4).

While most of these models find their origin back in the Poisson GLM framework, alternative approaches for modeling count data based on time-to-event distributions have recently been developed. These approaches are mainly built upon the direct relationship between the Poisson and exponential distributions (Conway and Maxwell, 1962). Zeviani et al. (2014), for example, focused on a discrete version of the gamma distribution to model count data by following the two-step approach of Winkelmann (1995): (1) Define the Poisson process as a sequence of i.i.d. exponentially distributed waiting times (Cox, 1962); and (2) replace the exponential distribution with a less restrictive (extended) non-negative distribution such as the gamma distribution. For the Weibull distribution, Morais and Barreto-Souza (2011) constructed count versions, i.e., the generalized Weibull power series (GWPS) class of distributions. Another, simple discrete approach based on the Weibull distribution, is that of Nakagawa and Osaki (1975). In particular, Klakattawi et al. (2018) recently pointed out that the corresponding regression model can model over- and underdispersed count data. Moreover, they showed that the model is able to adequately fit highly skewed count data with excessive zeros, without the need for introducing zero-inflated or hurdle components, in contrast to other existing methods, e.g., the zero-inflated Conway-Maxwell-Poisson (ZICOM) model (Sellers and Raim, 2016). A further generalization of this approach was introduced by Nekoukhou and Bidram (2015), where the exponentiated discrete Weibull (EDW) distribution is defined.

Apart from the presence of extra-dispersion, extended structures such as longitudinally collected data, where subjects/patients are repeatedly measured over time, and hierarchical structures, originating from hierarchical designs such as multi-center trials, can also be present. For the GLM framework, the generalized linear mixed model (GLMM), discussed by Engel and Keen (1994), Breslow and Clayton (1993), and Wolfinger and O'Connell (1993), has been suggested, and became a popular framework for taking into account hierarchical data structures. In these models, random effects are introduced to capture the association structure and to some extent dispersion. Molenberghs et al. (2007) extended this approach by introducing the so-called combined modeling (CM) framework, that was mainly developed to encompass both aspects, (1) overdispersion and (2) hierarchical/longitudinal structures, simultaneously, by adding an extra random effect into the GLMM framework.

In this chapter, we examine the existing (univariate) discrete Weibull(DW)-based approach of Nakagawa and Osaki (1975), and extend it with random effects to accommodate more complex data structures. This approach assumes that extra dispersion is captured in the pre-specified distribution, and differs from that in Molenberghs et al. (2007) where the extra dispersion is captured by an additional random effect. In addition, various settings (heavy-tails, zero-inflation) in combination with dispersion and correlation are examined, and compared with other well-known count models (Supplementary Materials S.4). Conclusions are supported with some characteristics of the model.

The remainder of this chapter is organized as follows. In Section 3.2, two motivating case studies are presented, stemming from patients with epileptic seizures, and historic data on household members from a Belgian town. An overview of the DW version of Nakagawa and Osaki (1975) is sketched in Section 3.3, alongside its extended version and characteristics. Section 3.4 is devoted to the analysis of the case studies, where a comparison is made between this approach and other count models (Appendix S.4). A simulation study is reported in Section 3.5 to investigate other characteristics of the framework, and concluding remarks are given in Section 3.6.

3.2 Case Studies

3.2.1 The Epilepsy Data

The epilepsy dataset comes from a randomized, double-blinded, parallel group multi-center study aimed at comparing placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. In total, 45 patients were assigned to the placebo group, and 44 to the active (new) treatment group. Patients were then followed for several weeks – during which the number of epileptic seizures experienced in the last week – were counted, i.e., since the last time the outcome was measured. The main research question is whether or not the new treatment reduces the number of epileptic seizures. A full description of the epilepsy dataset is provided in Faught et al. (1996). Figure 3.1 (top) displays the individual profiles with corresponding mean and median profiles of the seizure counts for every study week, and Figure 3.1 (bottom) shows the observed mean and variance of the seizure counts per patient ID, categorized for both treatment groups. The figure shows highly variable longitudinal count data with the presence of extreme values, zero-inflation, and very few observations available at some of the time-points, especially past week 20.

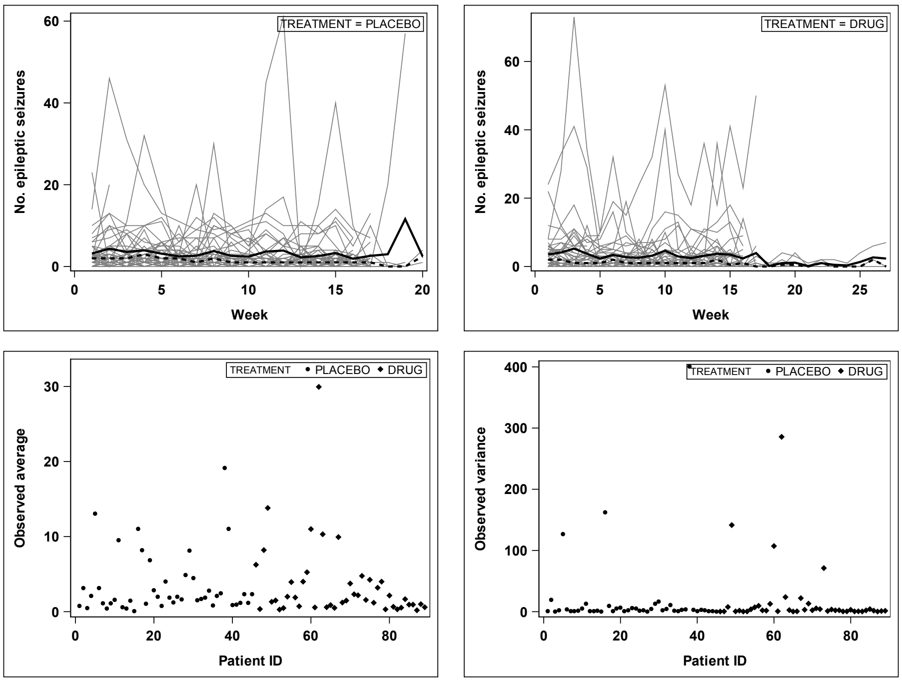


Figure 3.1: *Epilepsy data (Faught et al., 1996). Subject specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the number of epileptic attacks for every visit (top), and observed mean and variance of the seizure counts per patient ID (bottom), categorized for both treatments.*

3.2.2 The Moerzeke Data

The second dataset comes from a demographic, historical database of Moerzeke, a small village in the center of Flanders (the Dutch speaking part of Belgium) within the province of East Flanders. Information in the database is drawn from church and civil registers, which can be taken as high quality and appropriate for population studies, and includes all individuals who were born, married, or died in Moerzeke.

In this study, a sample of 457 families was taken - by selecting all fathers born between 1750 and 1830, and then forming a family by including also their first born children and the children's mothers. To avoid overlap, children already selected are not included again, as either father or mother of new families. For the group under study, the mean age at death for those who were born and deceased in Moerzeke was 71.9 years for men and 71.7 for women, respectively. The main interest lies in the exploration of different social and/or household characteristics (e.g., gender of first born child) on the (discrete) life expectancy of family members. Figure 3.2 (top) shows the household profiles with corresponding average and median profiles of the (discrete) life expectancy, and Figure 3.2 (bottom) shows the observed mean and variance of the (discrete) life expectancy per household ID, categorized for the gender of the first born child. On the average and median scales, a higher life expectancy of first born male children is observed compared to first born female children.

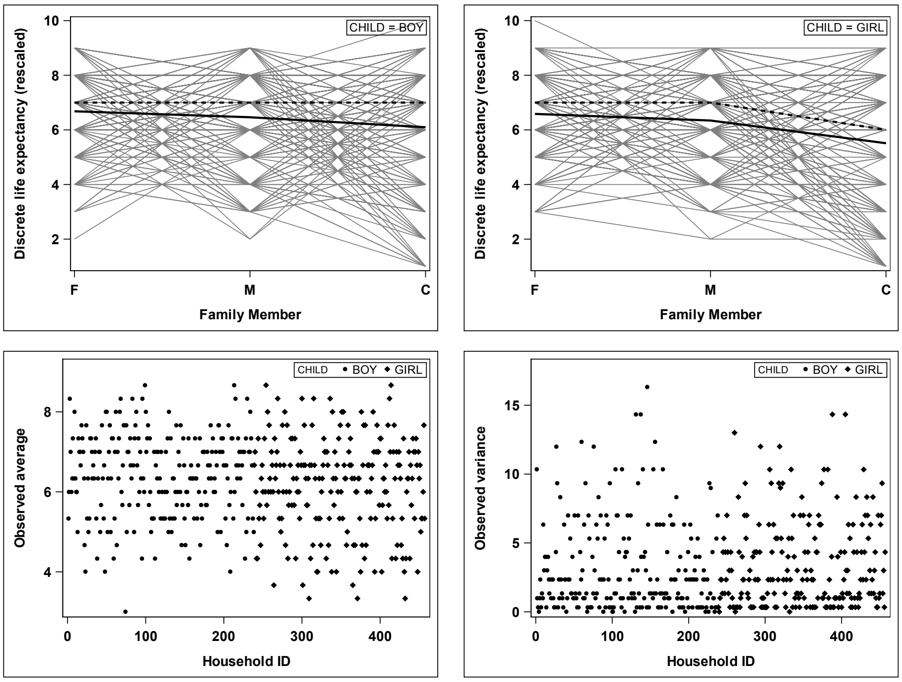


Figure 3.2: *Moerzeke data. Household specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the (discrete) life expectancy (rescaled) for every household member (top), and observed mean and variance of the (discrete) life expectancy (rescaled) per household ID (bottom), categorized for the gender of the first born child. The indices F, M, and C refer to father, mother, and first born child, respectively.*

3.3 The Weibull-count Approach

Due to the reproductive property of the gamma distribution, i.e., the sum of two gamma distributed random variables again follows a gamma distribution, Winkelmann (1995) pointed out that the gamma distribution is a useful choice for his two-step approach. Unfortunately, this reproductive property does not hold for the Weibull distribution. As an alternative, the discrete approach of Nakagawa and Osaki (1975), which is here referred to as the DW model, can be used instead and gives a simple and adaptable alternative for the Weibull case. In what follows, we will give a general overview of the DW approach of Nakagawa and Osaki (1975).

Let Y_i, $i = 1, \dots, n$, be (type 1) DW distributed (Nakagawa and Osaki, 1975) with parameters $0 < q < 1$ and $\rho > 0$. The probability mass function, cumulative distribution function, and hazard function are given by

$$\begin{gathered} P(Y_i = y_i) = q^{y_i^\rho} - q^{(y_i+1)^\rho}, \qquad F(y_i) = 1 - q^{(y_i+1)^\rho}, \\ h(y_i) = q^{y_i^\rho - (y_i+1)^\rho} - 1, \end{gathered} \tag{3.1}$$

respectively. Special cases result from this. When $\rho = 1$ and $q = 1 - p$, the geometric distribution follows. Particularly, when $\rho = 1$ and $q = e^{-\lambda}$, the discrete exponential (DE) distribution results (Sato et al., 1999), which is overdispersed relative to the standard Poisson distribution (Supplementary Material S.5). In addition, when $\rho = 2$ and $q = \theta$, the discrete Rayleigh (DR) distribution of Roy (2004) obtains. If $\rho \to +\infty$, the DW approaches a Bernoulli distribution with probability q. When q is small, an excessive zero case occurs (Klakattawi et al., 2018).

The mean and variance of the DW approach are given by

$$\mathsf{E}(Y_i) = \mu = \sum_{n=1}^{+\infty} q^{n^\rho}, \tag{3.2}$$

$$\mathsf{Var}(Y_i) = 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^{n^\rho} - \mu - \mu^2. \tag{3.3}$$

It can easily be shown that both of the infinite series in Eq. (3.2)–(3.3) converge (Supplementary Material S.6). Based on the integral test, general approximations can be found consisting of incomplete gamma functions, e.g., Englehardt and Li (2011). Closed-form expressions for lower and upper boundaries for Eq. (3.2)–(3.3), based on the integral test, can be found in Supplementary Material S.6.

To explore the characteristics of the DW model, we compute indexes for dispersion (DI), zero-inflation (ZI) and heavy-tail (HT), introduced by Puig and Valero (2006), which are respectively given by

$$
\begin{aligned}
\mathsf{DI} = \frac{\mathsf{Var}(Y_i)}{\mathsf{E}(Y_i)}, \qquad \mathsf{ZI} = 1 + \frac{\mathsf{log}P(Y_i = 0)}{\mathsf{E}(Y_i)}, \\
\mathsf{HT} = \frac{P(Y_i = y_i + 1)}{P(Y_i = y_i)}, \text{for } y_i \to \infty.
\end{aligned}
\tag{3.4}
$$

These indices are defined in relation to the Poisson distribution. Thus, the dispersion index indicates over-, under- and equidispersion for, respectively, DI >1, DI <1 and DI $=1$. The zero-inflation index indicates zero-inflation for ZI >0, zero-deflation for ZI <0 and no excess of zeros for ZI $=0$. Finally, the heavy-tail index indicates a heavy-tail distribution for HT $\to 1$ when $y_i \to \infty$. Figure 3.3 shows that the DW framework is able to model highly overdispersed, zero-inflated and heavy-tailed data. The approach also allows the fit of low overdispersed, zero-deflated data, and even some amount of underdispersion.

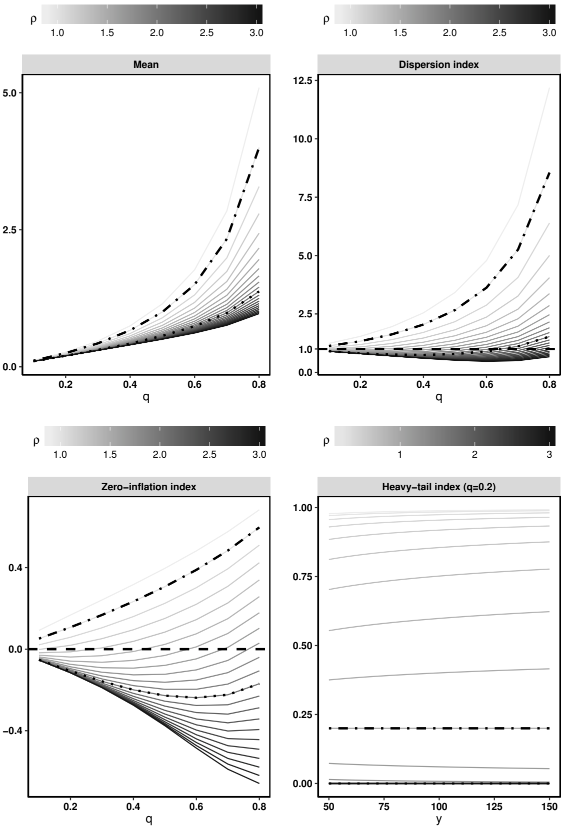


Figure 3.3: *Characteristic indexes of the discrete Weibull distribution related to the Poisson distribution. Dashed, dot dashed and dotted lines represent the Poisson, discrete exponential and discrete Rayleigh distribution, respectively.*

In a regression framework, Klakattawi et al. (2018) assumed that the response Y_i has a DW distribution, where a subject-specific parameter q_i is related to a p-dimensional vector of covariates $\mathbf{x}_i$ for ith observation through the complementary log-log link function

$$\ln[-\ln(q_i)] = \mathbf{x}_i^{'} \cdot \boldsymbol{\beta} \quad \Leftrightarrow \quad q_i = \mathrm{e}^{-\mathrm{e}^{\mathbf{x}_i^{'} \cdot \boldsymbol{\beta}}} (= \mathrm{e}^{-\lambda_i}) \tag{3.5}$$

Note that the complementary log-log link for q_i in Eq. (3.5) corresponds to a log link for λ_i. In addition, $\boldsymbol{\beta}$ in Eq. (3.5) represents the associated regression parameter vector, which can directly be interpreted in terms of the logarithm of the (closed-form) median. This is of particular interest when modeling, for example, highly skewed data, which often occurs in count data. Particular, by splitting the regression parameters $\boldsymbol{\beta}$ into $\{\beta_0\} \cup \{\beta_l \mid l = 1, \ldots, p\}$, it can easily be shown, thanks to the use of the complementary log-log link function (Klakattawi et al., 2018), that $\{\ln[\ln(2)] - \beta_0\}/\rho$ is related to the conditional median when all covariates are set to zero, whereas $-\beta_l/\rho$ $(l = 1, \ldots, p)$ can be related to the change in the median of the response corresponding to a one unit change of $\mathbf{x}_{li}$, keeping all other covariates constant.

In terms of estimation procedures, Klakattawi et al. (2018) and Kulasekera (1994) used maximum likelihood (ML) for parameter estimation, while Haselimashhadi et al. (2018) proposed a Bayesian approach for estimating the parameters.

3.3.1 The Extended Hierarchical Weibull-count Approach

If the discrete data are hierarchically structured, with Y_{ij} denoting the jth discrete outcome measured for cluster/subject i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, univariate models are often not appropriate to take into account the underlying correlation structure of the data. Therefore, mixed-effects models are often proposed where, in addition to fixed effects, random effects are added to the model to allow for the correlation structure of the data. These approaches have been studied extensively in the GLM framework, e.g., linear mixed model (LMM) and GLMM (Molenberghs and Verbeke, 2005), while little research has focussed on dispersion models outside of this framework. In our context, where the focus is on the DW approach, a dispersion model extension with random effects can simply be achieved as follows:

$$\ln[-\ln(q_{ij})] = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\beta} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i, \tag{3.6}$$

where $\mathbf{z}_{ij}$ represents a q-dimensional vector of known covariate values corresponding to the q-dimensional random effects vector $\mathbf{b}_i$ following a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix D.

In the following, we will analyze the epilepsy and Moerzeke datasets, introduced in Section 3.2. ML principles are used to obtain parameter estimates. The SAS procedure NLMIXED is used for the computations (Appendix S.7).

3.4 Analysis of the Case Studies

3.4.1 The Epilepsy Data

The epilepsy data of Section 3.2.1 will be analyzed with the DW and its nested DE model (Section 3.3), and compared with some conventional models from Supplementary Materials S.4, i.e., the classical Poisson log-linear (P), negative binomial (NB), Conway-Maxwell-Poisson (COM) and DP models. Previous work on this dataset was reported by Molenberghs and Verbeke (2005) and Molenberghs et al. (2007) in the context of generalized estimating equations (GEE; Liang and Zeger, 1986) and the CM framework, respectively.

Let Y_{ij} be the number of epileptic seizures that patient i experiences during week j of the follow-up period, and let t_{ij} be the time-point at which outcome Y_{ij} has been measured, i.e., $t_{ij} = 1, 2, \ldots,$ until at most 27. The following specific choice is made for the linear predictor:

$$\eta_{ij} = \beta_0 + b_i + \beta_0^{'} \cdot T_i + (\beta_1 + \beta_1^{'} \cdot T_i) \cdot t_{ij}, \tag{3.7}$$

where $T_i = 1$ if patient i receives the treatment, and 0 for placebo. Here, $\beta_0^{'}$ and $\beta_1^{'}$ represent differences between treatment and placebo in terms of intercept and slope, respectively. The link functions are $\eta_{ij} = \exp(\lambda_{ij})$ for the P, DE, NB, COM and DP models, and $\eta_{ij} = \ln[-\ln(q_{ij})]$ for the DW model. The random intercept b_i is assumed to be normally distributed with mean 0 and variance σ^2, reflecting the between-patient variability within the data. ML estimates and corresponding standard errors of the parameters are reported in Table 3.1 (for the univariate case, i.e., without the subject random effect) and Table 3.2 (for the clustered case, i.e., with the subject random effect).

In the univariate case, i.e., where clustering is ignored (Table 3.1), very large improvements is observed in the DE, NB, DP and DW models, in terms of the likelihood, relative to the classical P model. This, of course, is to be expected since the P model assumes equidispersion while the parameters α, ϕ and ρ (see Appendix S.4 for details) provide significant evidence of overdispersion. Furthermore, when a comparison is made between the DW and the conventional models, e.g., NB and DP, w.r.t. the Poisson model, we could consider the DW model as the better one in terms of log-likelihood. Indeed, similar to the NB approach (Figure S.8.1), the DW model is able to capture highly overdispersed, zero-inflated and heavy-tailed data (Figure 3.3), characteristics that are definitely present within the epilepsy dataset.

Table 3.1: *Epilepsy dataset. Parameter estimates (standard errors) for the (1) Poisson (P), (2) discrete exponential (DE), (3) negative binomial (NB), (4) Conway-Maxwell-Poisson (COM), (5) double Poisson (DP), and (6) the discrete Weibull (DW) model.*

		P	**DE**	**NB**	**COM**	**DP**	**DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	1.2662 (0.0424)	1.2601 (0.0864)	1.2594 (0.1119)	-0.5054 (0.0189)	1.2662 (0.1054)	0.7341 (0.1002)
Difference in intercepts	β_0'	0.1869 (0.0571)	0.2115 (0.1202)	0.2156 (0.1564)	0.0131 (0.0144)	0.1869 (0.1421)	0.0936 (0.1307)
Slope placebo	β_1	-0.0134 (0.0043)	-0.0126 (0.0086)	-0.0126 (0.0111)	-0.0011 (0.0012)	-0.0134 (0.0108)	-0.0174 (0.0095)
Difference in slopes	β_1'	-0.0195 (0.0058)	-0.0222 (0.0116)	-0.0227 (0.0150)	-0.0017 (0.0017)	-0.0195 (0.0144)	-0.0143 (0.0127)
Ratio of slopes	$1+\frac{\beta_1'}{\beta_1}$	2.4576 (0.8480)	2.7586 (1.9721)	2.8081 (2.6066)	2.5663 (3.1297)	2.4576 (2.1093)	1.8189 (1.1027)
	α	$--$	$--$	1.8961 (0.0918)	$--$	$--$	$--$
	τ	$--$	$--$	$--$	-0.1188 (0.0051)	$--$	$--$
	ϕ	$--$	$--$	$--$	$--$	0.1616 (0.0061)	$--$
	ρ	$--$	$--$	$--$	$--$	$--$	0.7383 (0.0172)
-2 loglik		11590.0	6502.5	6326.1	6256.2	6815.6	6291.3
AIC		11598.0	6510.5	6336.1	6266.2	6825.6	6301.3

Furthermore, we should mention that 'illegal' estimates were obtained for the COM model, implying that no valid conclusions can be made from it. Indeed, when looking at the fitted dispersion parameter τ, a negative estimate (-0.1188) is observed which is outside the parameter space (Appendix S.4.3). This, of course, can easily be explained by the fact that the COM distribution limits itself in flexibility towards underdispersed data with narrow flexibility to zero-inflation (Figure S.8.2). Bar charts of the fitted univariate models are given in Figure 3.4.

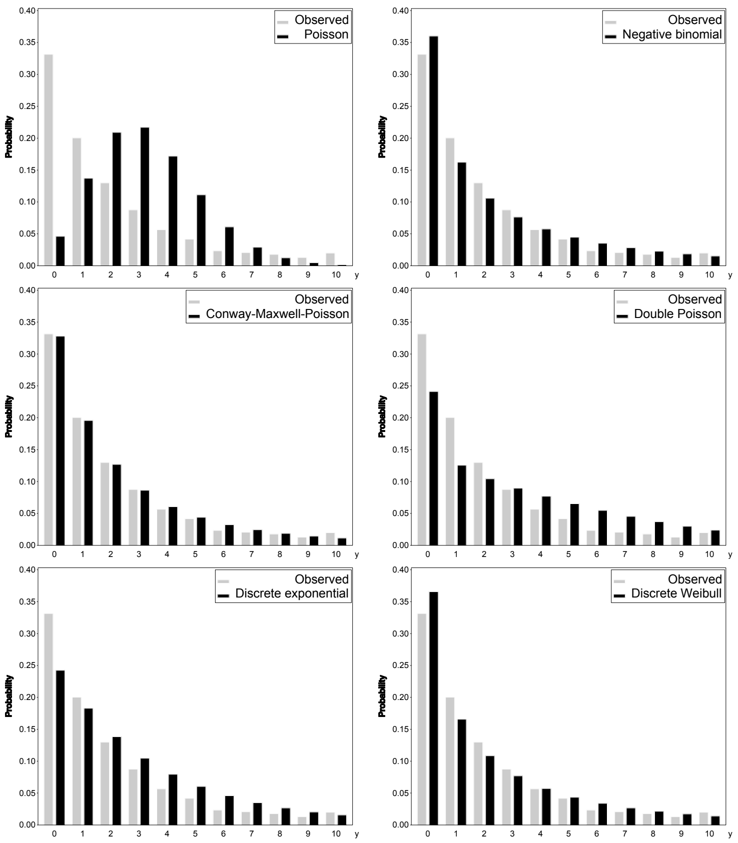


Figure 3.4: *Epilepsy data. Bar charts of fitted univariate models.*

For the clustered case, i.e., where a subject-specific random-intercept is added to account for correlation (Table 3.2), we find that the DWN is considerably better in terms of likelihood. Moreover, point and precision estimates of such key parameters as the slope difference and the slope ratio are strongly affected when a random effect is added to the models. This remark was also made by Molenberghs et al. (2010), who noted an impact on hypothesis testing. Surprisingly, a valid interpretation on the extended COM approach can now be given, while this was not possible in the univariate case. To explain this phenomenon, attention should be directed towards the limited flexibility of COM in terms of overdispersion and the multiplicity effect of the random effects. In particular, a limited number of highly overdispersed regions can be modeled with the COM approach (Figure S.8.2). By adding a random effect to the model, extra flexibility has been given towards capturing overdispersed regions. Indeed, since random effects are mainly used to capture the underlying correlation structure of the data, they are also able to seize a certain amount of dispersion. Therefore, more flexibility has been gained with the inclusion of random effects towards the modeling of overdispersed data. In addition, a much lower parameter estimate for σ was obtained for the Conway-Maxwell-Poisson-normal (COMN) case, compared to all other models. This directly results from the main disadvantage of the COM regression model, i.e., its location parameter does not correspond to the expectation, which complicates the interpretation of regression models towards the mean specified using this distribution (Sellers and Shmueli, 2010).

Table 3.2: *Epilepsy dataset. Parameter estimates (standard errors) for the (1) Poisson-normal (PN), (2) discrete exponential-normal (DEN), (3) combined model (CM), (4) Conway-Maxwell-Poisson-normal (COMN), (5) double Poisson-normal (DPN), and (6) the discrete Weibull-normal (DWN) model.*

		PN	**DEN**	**CM**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	0.8177 (0.1677)	0.9443 (0.1843)	0.9112 (0.1755)
Difference in intercepts	β_0'	−0.1705 (0.2387)	−0.2670 (0.2620)	−0.2556 (0.2500)
Slope placebo	β_1	−0.0143 (0.0044)	−0.0271 (0.0101)	−0.0248 (0.0077)
Difference in slopes	β_1'	0.0023 (0.0062)	0.0145 (0.0140)	0.0130 (0.0107)
Ratio of slopes	$1+\frac{\beta_1'}{\beta_1}$	0.8398 (0.3979)	0.4663 (0.3953)	0.4751 (0.3345)
Std. dev. random effect	σ	1.0755 (0.0857)	1.0436 (0.0888)	1.0626 (0.0871)
	α	−−	−−	0.4059 (0.0348)
	τ	−−	−−	−−
	ϕ	−−	−−	−−
	ρ	−−	−−	−−
−2 loglik		6271.9	5543.9	5417.0
AIC		6281.9	5553.9	5429.0

		COMN	**DPN**	**DWN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	−0.2384 (0.0779)	0.8314 (0.1721)	1.4319 (0.2183)
Difference in intercepts	β_0'	−0.0947 (0.1042)	−0.1582 (0.2451)	−0.2970 (0.3005)
Slope placebo	β_1	−0.0040 (0.0023)	−0.0146 (0.0067)	−0.0297 (0.0098)
Difference in slopes	β_1'	0.0005 (0.0032)	0.0018 (0.0093)	0.0180 (0.0135)
Ratio of slopes	$1+\frac{\beta_1'}{\beta_1}$	0.8646 (0.7451)	0.8778 (0.5980)	0.3947 (0.3382)
Std. dev. random effect	σ	0.4475 (0.0433)	1.0458 (0.0875)	1.2658 (0.1063)
	α	−−	−−	−−
	τ	0.1563 (0.0196)	−−	−−
	ϕ	−−	0.4355 (0.0169)	−−
	ρ	−−	−−	1.3074 (0.0340)
−2 loglik		5473.8	5652.2	5451.1
AIC		5485.8	5664.2	5463.1

Even though the CM is a more viable candidate in terms of likelihood (related to the P model), one should be aware of the restricted mean scale interpretation in this framework, especially when dealing with skewed data. In this setting, right skewed data (Figure 3.1) is observed, making the inferences less attractive from an interpretational point of view (similar to the DPN approach). The DW model avoids this problem by allowing inferences directly on the median scale (Section 3.3), making the approach more interesting here.

Finally, we expand our analysis with random slopes in the DWN model, i.e., considering two random effects instead of a single one to reflect the between- and within-patient variability of the data. The linear predictor becomes:

$$\eta_{ij} = \beta_0 + b_{1i} + \beta_0^{'} \cdot T_i + (\beta_1 + \beta_1^{'} \cdot T_i + b_{2i}) \cdot t_{ij}, \tag{3.8}$$

where the random effects vector $\mathbf{b}_i = (b_{1i}, b_{2i})^{'}$ is assumed to be multivariate normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix

$$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}. \tag{3.9}$$

A comparison with the random-intercept model will be made in two ways, i.e., (1) a random-slopes model with uncorrelated random effects ($\sigma_{12} = 0$; IND) and (2) a random-slopes model with correlated random effects ($\sigma_{12} \neq 0$; UN). ML estimates and corresponding standard errors of the parameters are reported in Table 3.3.

Table 3.3*: Epilepsy dataset. Parameter estimates (standard errors) for the discrete Weibull-normal (DWN) model with (1) random-intercept, (2) random-slope with uncorrelated random effects (IND), and (3) random-slope with correlated random effects (UN).*

		Random-intercept	**Random-slope (IND)**	**Random-slope (UN)**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	1.4319 (0.2183)	1.4973 (0.2183)	1.4947 (0.2287)
Difference in intercepts	$\beta_0^{'}$	−0.2970 (0.3005)	−0.2909 (0.2996)	−0.2984 (0.3150)
Slope placebo	β_1	−0.0297 (0.0098)	−0.0339 (0.0120)	−0.0327 (0.0126)
Difference in slopes	$\beta_1^{'}$	0.0180 (0.0135)	0.0169 (0.0168)	0.0167 (0.0176)
Ratio of slopes	$1+\frac{\beta_1^{'}}{\beta_1}$	0.3947 (0.3382)	0.5016 (0.3920)	0.4884 (0.4219)
Std. dev. random-intercept	σ_1	1.2658 (0.1063)	1.2553 (0.1114)	1.3333 (0.1302)
Std. dev. random-slope	σ_2	−− (−−)	0.0417 (0.0092)	0.0474 (0.0099)
Cov. between random-effects	σ_{12}	−− (−−)	−− (−−)	−0.0177 (0.0142)
	ρ	1.3074 (0.0340)	1.3393 (0.0362)	1.3463 (0.0366)
−2 loglik		5451.1	5439.6	5437.7
AIC		5463.1	5453.6	5453.7

A significant improvement in likelihood is observed when adding a random-slope to the model (likelihood ratio test $p = 0.0007$). However, there are no qualitative changes in the results of hypothesis testing for the main effects of interest. Furthermore, by comparing the independent random effects (IND) with correlated random effects (UN), no significant improvements were obtained (likelihood ratio test $p = 0.1692$). This extension at the same time illustrates the ease with which more than one random effect can be included.

3.4.2 The Moerzeke Data

While previous work on the Moerzeke data was provided by Matthijs et al. (2002) for the examination of historical mortality in terms of sociological and biological components, there has been no consideration of dispersion aspects. To this end, the DW and its nested DE models are considered in the analysis of the Moerzeke dataset (Section 3.2.2) and compared with the count models from Appendix S.4.

Let Y_{ij} represent the (discrete) life expectancy of the mother, father, and first born child ($j = 1, 2, 3$) in household $i = 1, \ldots, 457$. We assume the following predictor:

$$\eta_{ij} = \beta_0 \cdot I_{Cij} + \beta_0^{'} \cdot I_{Mij} + \beta_0^{''} \cdot I_{Fij} + b_i + (\beta_1 \cdot I_{Cij} + \beta_1^{'} \cdot I_{Mij} + \beta_1^{''} \cdot I_{Fij}) \cdot G_i, \quad (3.10)$$

where I_{Cij}, I_{Mij} and I_{Fij} are dummy variables for first born child, mother and father, respectively, and G_i is the binary indicator for the gender of the first born child, i.e., 1 for male and 0 for female. Similar to the epilepsy analysis, the link functions are $\eta_{ij} = \exp(\lambda_{ij})$ for the P, DE, NB, COM and DP models, and $\eta_{ij} = \ln[-\ln(q_{ij})]$ for the DW model. The random intercept b_i is used to capture between-household variability, which here is assumed normally distributed with mean 0 and variance σ^2. ML estimates and corresponding standard errors of the parameters are reported in Table 3.4 (for the univariate case without the random effect) and Table 3.5 (for the clustered case, including the random effect).

In the univariate case (Table 3.4), the COM, DP and DW models significantly improved the model fit, compared to the classical Poisson model, while, in terms of likelihood, a worse fit is observed for the DE case. Indeed, when considering the dispersion parameters (τ, ϕ and ρ), we observe the clear presence of underdispersion within the data. While the DW, COM and DP models are able to capture this phenomenon (Figures 3.3, Supplementary Material S.8.2, and Supplementary Material S.8.3, respectively), this is not the case for the DE (Supplementary Material S.5) and Poisson models. Therefore, it is fair to say that the DE model is completely wrong, not just in terms of underdispersion, but also in the fact that it fails to capture the unimodal shape, as expected from a geometric distribution. The underdispersion result can be explained by the

Table 3.4: *Moerzeke dataset. Parameter estimates (standard errors) for the (1) Poisson (P), (2) discrete exponential (DE), (3) Conway-Maxwell-Poisson (COM), (4) double Poisson (DP), and (5) discrete Weibull (DW) model.*

		P	**DE**	**COM**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0288)	1.7068 (0.0735)	3.0527 (0.1294)
Intercept father	β_0'	1.8473 (0.0268)	1.8473 (0.0727)	3.2891 (0.1373)
Intercept mother	β_0''	1.8847 (0.0263)	1.8847 (0.0725)	3.3522 (0.1395)
Gender effect on first born child	β_1	0.1009 (0.0390)	0.1009 (0.1014)	0.1697 (0.0509)
Gender effect on father	β_1'	0.0187 (0.0370)	0.0187 (0.1007)	0.0316 (0.0481)
Gender effect on mother	β_1''	0.0145 (0.0364)	0.0145 (0.1005)	0.0247 (0.0473)
	τ	$--$	$--$	1.7484 (0.0690)
	ϕ	$--$	$--$	$--$
	ρ	$--$	$--$	$--$
-2 loglik		5834.3	7985.1	5669.3
AIC		5846.3	7997.1	5683.3

		DP	**DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0225)	8.9228 (0.2301)
Intercept father	β_0'	1.8473 (0.0210)	9.0796 (0.2293)
Intercept mother	β_0''	1.8847 (0.0206)	9.1660 (0.2301)
Gender effect on first born child	β_1	0.1009 (0.0305)	0.1699 (0.0957)
Gender effect on father	β_1'	0.0187 (0.0290)	0.0831 (0.0955)
Gender effect on mother	β_1''	0.0145 (0.0285)	0.0350 (0.0954)
	τ	$--$	$--$
	ϕ	1.6333 (0.0624)	$--$
	ρ	$--$	4.5377 (0.1055)
-2 loglik		5693.3	5512.3
AIC		5707.3	5526.3

fact that Moerzeke has characteristics of a geographically isolated area as it is almost completely surrounded by a meander in the river Scheldt and by the river Durme. This was an important geographical limitation within the time bracket at which data were collected, and led to more genetic homogeneity than in the typical town. We observe that the DW model indicates the best fit, relative to the COM and DP models, in terms of likelihood compared to the Poisson model. A possible reason for this result is the presence of left-skewed discrete time-to-event data, which can flexibly be modeled with the DW approach due to the underlying Weibull connection. Bar charts of the fitted univariate models are given in Figure 3.5.

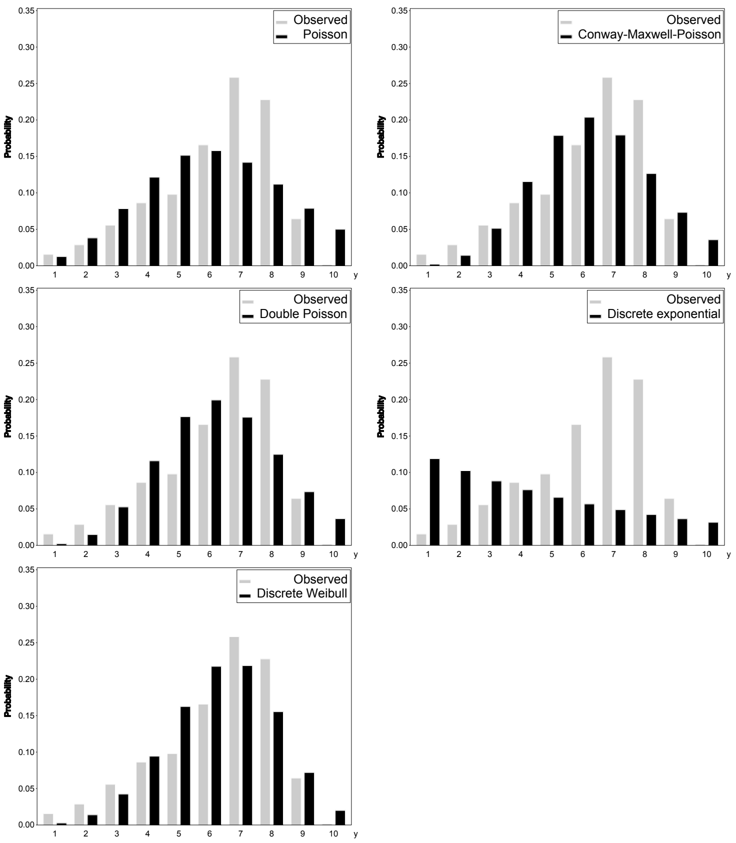


Figure 3.5: *Moerzeke data. Bar charts of fitted univariate models.*

In the clustered case (Table 3.5), noteworthy results were obtained for the estimated variance component σ^2. In all clustered models, the estimated component is very close to 0, leaving the standard errors estimates unchanged relative to the univariate cases. This phenomenon, while strange at first sight, is reasonably well understood in the literature. More specifically, partial marginalization is used here, in agreement with Molenberghs et al. (2010), where adaptive Gaussian quadrature principles are used to approximate the marginal likelihood obtained from integrating over the normal random effects. This automatically adopts a hierarchical perspective, implying the restriction that no negative estimates of σ^2 can be achieved, even though this could be present for several reasons (e.g., negative intraclass correlation, underdispersion, etc.). Molenberghs and Verbeke (2011) and Verbeke and Molenberghs (2003), for example, discussed this phenomenon in the context of LMMs. Pryseley et al. (2011) extended this discussion to non-Gaussian outcomes, while Oliveira et al. (2017) illustrated how such negative variance components play a natural role in modeling both the correlation between repeated measures on the same experimental unit and over- or underdispersion from a CM perspective. While a zero variance component could in principle also point to the absence of correlation, this is not something one would expect in view of these data.

Table 3.5: *Moerzeke dataset. Parameter estimates (standard errors) for the (1) Poisson-normal (PN), (2) discrete exponential-normal (DEN), (3) Conway-Maxwell-Poisson-normal (COMN), (4) double Poisson-normal (DPN), and (5) discrete Weibull-normal (DWN) model.*

		PN	**DEN**	**COMN**	**DPN**	**DWN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0288)	1.7068 (0.0735)	3.0529 (0.1294)	1.7068 (0.0225)	8.9228 (0.2301)
Intercept father	β_0'	1.8473 (0.0268)	1.8472 (0.0727)	3.2895 (0.1373)	1.8473 (0.0210)	9.0795 (0.2293)
Intercept mother	β_0''	1.8847 (0.0263)	1.8847 (0.0727)	3.3527 (0.1395)	1.8846 (0.0206)	9.1660 (0.2301)
Gender effect on first born child	β_1	0.1009 (0.0390)	0.1009 (0.1014)	0.1698 (0.0509)	0.1010 (0.0305)	0.1699 (0.0957)
Gender effect on father	β_1'	0.0187 (0.0370)	0.0187 (0.1007)	0.0317 (0.0481)	0.0187 (0.0290)	0.0831 (0.0955)
Gender effect on mother	β_1''	0.0145 (0.0364)	0.0145 (0.1005)	0.0245 (0.0473)	0.0145 (0.0285)	0.0350 (0.0954)
Std. dev. random effect	σ	$1.16E-4$ (0.0119)	$1.68E-4$ (0.0215)	$7.72E-4$ (0.1039)	$1.85E-4$ (0.0293)	$2.33E-4$ (0.0420)
	τ	−−	−−	1.7486 (0.0690)	−−	−−
	ϕ	−−	−−	−−	1.6333 (0.0624)	−−
	ρ	−−	−−	−−	−−	4.5376 (0.1055)
-2 loglik		5834.3	7985.1	5669.3	5693.3	5512.3
AIC		5848.3	7999.1	5685.3	5709.3	5528.3

To conclude, we should mention that, even though the DW model fits the data quite well in the context of underdispersed data for the univariate case, there is still scope for further research in the context of underdispersed clustered data. Even though it is not our scope to fully encounter this problem here, boundary issues are suggested for the variance component. Also note that the random effects variability is very different between the epilepsy and Moerzeke studies, underscoring that a large range of situations can be handled. Of course, this does not preclude further research towards underdispersion.

3.5 A Large Simulated Set of Data

To further explore the DW approach with dispersed count data, a large simulated set of data is obtained to examine the deviance surface under different dispersion situations. This highlights some other characteristics of the model such as the orthogonality and computational ease of estimating the parameters (q, ρ).

Figure 3.6 presents contour plots of the deviance surfaces for five different simulated DW data of size 1000, with expectation fixed at 1 and dispersion indices at 0.25 (very strong underdispersion), 0.5 (strong underdispersion), 1 (equidispersion), 5 (strong overdispersion) and 10 (very strong overdispersion). As a result, the figure indicates that the parameters are highly intra-related in the likelihood function, consequently the ML estimators for ρ and q are correlated. More specifically, a decreasing trend in the correlation seems to correspond with an increasing dispersion index. Based on the deviance surface, computational ease is combined with the ability to perform asymptotic (normally based) inferences in the regions with high dispersion, i.e., $DI \to \infty$. Note that this is not a genuine simulation study.

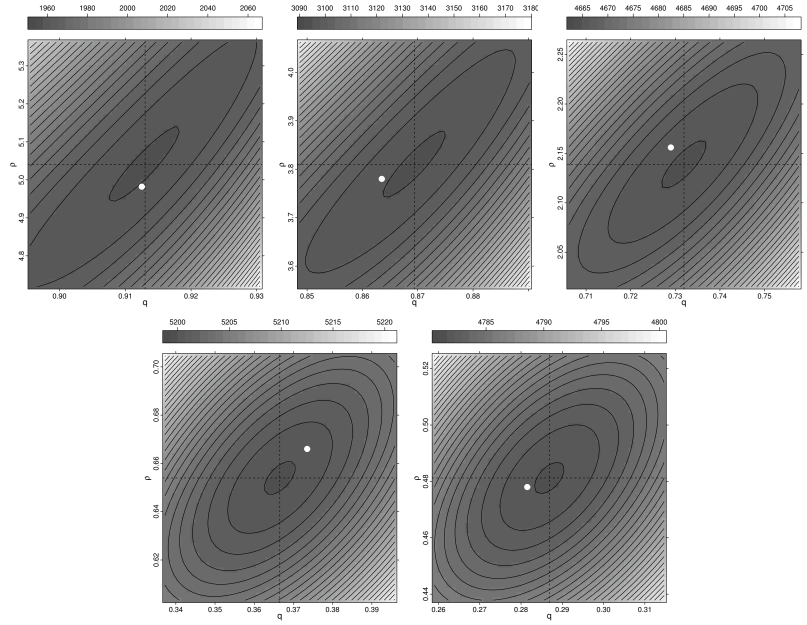


Figure 3.6*: Simulation study. Deviance surfaces for discrete Weibull model fitted to five simulated data with expectation 1 and dispersion 0.25 (top-left), 0.5 (top-middle), 1 (top-right), 5 (below-left) and 10 (below-right). Dotted lines are the maximum-likelihood estimates, and white points are the parameters used in the simulation.*

3.6 Conclusion

Starting from an existing univariate framework, we have proposed an extended version that can handle both under- and overdispersed, and hierarchical data structures. In both case studies, we showed that the model fits the data well, for both under- and overdispersed situations. More specifically, the approach used is able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data, similar to the CM approach. In addition, the approach is capable of modeling some low overdispersed regions with zero-deflation (e.g., the DR approach for small values of q) and even underdispersed data, regions that cannot be captured within the CM framework. Due to the presence of a closed-form median expression, interpretations of the parameters can directly be related to the median profile, which is of particular interest when modeling skewed data. Finally, orthogonality properties are examined through a large simulated set of data. The resulting outcome indicates the presence of correlation between ML estimators, related to the dispersion index.

Chapter 4

A Flexible Finite Mixture Model Family for Analyzing Over- and Underdispersed, and/or Zero-inflated and -deflated Discrete Data, with Possibly Negative Weights

This chapter is based upon:

Luyts, M., Molenberghs, G., Verbeke, G., and Matthijs, K. (2021). A flexible finite mixture model family for analyzing over- and underdispersed, and/or zero-inflated and -deflated discrete data, with possibly negative weights. In preparation.

Abstract

The possible occurrence of over- and/or underdispersed count data has been observed for decades, and the formulation of models to accommodate this phenomenon has an equally long history; prominent models are the negative binomial and Conway-Maxwell-Poisson models. While discrete finite mixture models offer a natural route to model overdispersion, their use for underdispersion is uncommon. Here, a general framework for discrete finite mixture models is proposed that flexibly handles both over- and underdispersed count data, with appealing interpretation in both of these cases. The flexibility of the proposal is twofold: (1) One can choose from a wide class of component distributions; and (2) mixture weights do not have to be all positive but certain negative values are allowed as well. As a result, a wide variety of dispersion structures can be modeled. Particular focus is placed on underdispersed settings. Applying the approach to an underdispersed demographic data setting shows important improvement in goodness-of-fit compared to existing models. Maximum likelihood principles are used to obtain the results. We allow for negative weights in mixtures to accommodate underdispersion. It will be shown that this approach, additionally, allows for zero-inflation as well as -deflation.

4.1 Introduction

In methodological and applied statistical research, the use of finite mixture models (FMMs; Newcomb, 1886; Prentice, 1988) has received considerable attention due to its flexibility when modeling data across a wide variety of distributions with various shapes, from symmetric to left- or right-skewed, from unimodal to multimodal structures, with varying kurtosis, etc. A well-known class of FMMs for continuous data is the Gaussian mixture model (GMM). In the case of discrete data, with prominent members such as count data, Poisson mixtures, binomial mixtures, etc. have been explored in detail by McLachlan and Peel (2004) and Everitt and Hand (1981).

Thanks to the additional parameters involved, combined with the flexibility of shape of the resulting marginal density function, FMMs have the potential to capture many combinations of mean, variance, and higher-moment functions. Therefore, Poisson and negative binomial mixtures, for example, are often suggested to capture high variability (i.e., overdispersion, going beyond what a single Poisson or negative binomial component is able to). A mixture of Conway-Maxwell-Poisson models (Conway and Maxwell, 1962), for example, has recently been explored to handle both situations separately in one framework (Sur et al., 2015). While substantial research is already available in the literature to the examination of FMMs in the context of overdispersed data, a sufficiently general framework of FMMs will be proposed here that can handle overdispersion as well as underdispersion (relative to the Poisson model, with mean and variance equal). Relaxing commonly used restrictions, we also allow for negative mixture weights. Perhaps surprisingly, the resulting mixture still allows for an intuitively appealing interpretation, in addition to having a meaningful mean and variance function. While the introduction of negative weights within FMMs have already been proposed by Bartholomew (1969), examining sufficient conditions for being a probability density function in a FMM of exponentials, Zhang and Zhang (2005) explored the presence of negative weights in Gaussian mixtures for cluster analysis. More general and recent work of FMMs with negative weights is available in Felgueiras et al. (2012) and Santos et al. (2016). These authors defined them as pseudo-convex mixtures. Their use within underdispersed discrete data, however, has not been examined before.

A related but different phenomenon is that of zero-inflation, where the frequency of zero counts is higher than predicted by the distribution under investigation. Commonly used models to this effect are the zero-inflated Poisson (ZIP) or zero-inflated negative-binomial (ZINB) model. Alternatively, so-called hurdle models can be used. Just like the occurrence of underdispersion, also zero-deflation can occur in practice. An overview, and extension to hierarchical data, is presented in Kassahun et al. (2015). We will show that the proposed model framework allows for both of these phenomena as well.

The remainder of this chapter is organized as follows. In Section 4.2, a motivating case study is presented, coming from historic data on household members from a Belgian town. Well-known univariate discrete distributions that allow the fit of under- and overdispersed data, our basic building blocks, are reviewed in Section 4.3, alongside our general FMM formulation. In addition, the added flexibility of using/obtaining negative weights towards underdispersed discrete data structures is sketched. Proof of concept is provided in Section 4.4, where the case study is examined within the proposed mixture model family.

4.2 Case Study: Historic Life Expectancy in Moerzeke

The Moerzeke dataset comes from a demographic, historical database of Moerzeke, a small village in the center of Flanders (the Dutch speaking part of Belgium) within the province of East Flanders. Information in the database is drawn from church and civil registers, which can be taken as high quality and appropriate for population studies, and includes all individuals who were born, married, or died in Moerzeke.

In this study, a sample of 457 families was taken – by selecting all fathers born between 1750 and 1830, and then forming a family by including also their first born child and the child's mother. To avoid overlap, children already selected are not included again, as either father or mother of new families. The main interest lies in the exploration of the discrete life expectancy of family members within the village, where the outcome has been discretized in decades. Figure 4.1 shows the bar chart of the discretized life expectancy, resulting in a clear presence of left skewed data. Furthermore, there is indication that the outcome can be explored by a mixture of two discrete approaches. The sample mean and variance equal 6.29 and 3.47, respectively, pointing to underdispersion.

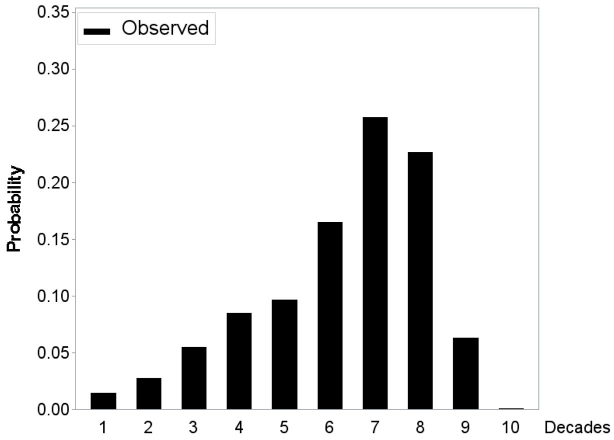


Figure 4.1: Moerzeke data. Bar chart of the rescaled discrete life expectancy

4.3 Finite Mixture Models

4.3.1 Elementary Components

Two key aspects of FMMs are the constituent components on the one hand, and the total number of components on the other. While several authors have contributed to the flexible use of the latter aspect (Huang et al., 2017, Zhang and Cheng, 2004), focus in this part is mainly on the former one. In particular, next to the Poisson case, a brief overview of existing count distributions will be given that can handle both over- and underdispersion.

Log-linear Poisson models, which possess the property of equidispersion (mean and variance equal), are often the standard or at least the basis for count data modeling. To allow for more general mean-variance relationships, extended and alternative approaches have been developed that can easily allow for over- and underdispersion. Examples are the negative binomial (NB), double Poisson (DP; Efron, 1986), and generalized Poisson (GP; Famoye, 1993) models.

There are many count-data distributions in the literature (Johnson et al., 2005; Chakraborty, 2015). Also the distributions described in Luyts et al. (2019) can be considered. All of these are candidate building blocks for finite mixtures, to be formalized in the next section. Evidently, this gives rise to a very large class, with arguably a lot of potential to flexibly model overdispersion as well as underdispersion, given that the aforementioned basic distributions contain members with overdispersion, underdispersion, or a combination of both (depending on the parameter values).

We consider a few of these here, i.e., the modified discrete normal (DN) distribution of Roy (2003), the discrete Weibull (DW) distribution (Nakagawa and Osaki, 1975) and the double Poisson (DP) approach of Efron (1986). Characteristics of these and other models that will be used in the data analysis (Section 4.4) are given in Table 4.1. All three models allow the fit of under- and overdispersed data (Klakattawi et al., 2018), in contrast to, for example, the NB approach, which is confined to overdispersion and with equidispersion as a limiting case. While a mean-variance association is given in the latter two distributions, this is not assumed in the DN approach, which is symmetric in shape. Finally, no analytical closed-form expressions for the mean and variance exist in the DW case, implying the need for numerical approximations when making inferences about these.

Table 4.1: *Characteristics of some discrete distributions.*

Element	**Notation**	**Distribution**	
Model		**Poisson**	**Discrete normal**
PMF	$p(y \mid \boldsymbol{\theta})$	$\frac{\mathsf{e}^{-\lambda}\lambda^{y}}{y!}$	$\Phi\left(\frac{x-\lambda+0.5}{\sigma}\right) - \Phi\left(\frac{x-\lambda-0.5}{\sigma}\right)$
Param.(s)	$\boldsymbol{\theta}$	$\lambda \geq 0$	$(\lambda; \sigma) \in \mathbb{R}$
Mean	$\mathsf{E}(\mathbf{Y})$	λ	λ
Var.	$\mathsf{Var}(\mathbf{Y})$	λ	$\sigma^2 + 0.083333$
Disp.		Only equi	Over/equi/under
Model		**Double Poisson**	**Discrete Weibull**
PMF	$p(y \mid \boldsymbol{\theta})$	$K(\lambda, \phi)\phi^{1/2}\mathsf{e}^{-\phi\lambda}\frac{\mathsf{e}^{-y}y^{y}}{y!}\left(\frac{\mathsf{e}\lambda}{y}\right)^{\phi y}$	$\lambda^{y^{\rho}} - \lambda^{(y+1)^{\rho}}$
Constant		$\frac{1}{K(\lambda,\phi)} \approx 1 + \frac{1-\phi}{12\phi\lambda}\left(1 + \frac{1}{\phi\lambda}\right)$	
Param.(s)	$\boldsymbol{\theta}$	$\lambda > 0; \phi \in \mathbb{R}$	$0 < \lambda < 1; \rho > 0$
Mean	$\mathsf{E}(\mathbf{Y})$	λ	$\sum_{n=1}^{+\infty} \lambda^{n^{\rho}}$
Var.	$\mathsf{Var}(\mathbf{Y})$	λ/ϕ	$2\sum_{n=1}^{+\infty} n\lambda^{n^{\rho}} - \mathsf{E}(\mathbf{Y}) - [\mathsf{E}(\mathbf{Y})]^2$
Disp.		Over/equi/under	Over/equi/under

4.3.2 General Finite Mixture Model Formulation

Suppose that a random variable $\mathbf{Y}$ follows a k-component FMM with parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_k)'$, in which $\boldsymbol{\theta}_j$ is the parameter vector for component j, $j = 1, \ldots, k$. Its probability mass function (PMF) can be written as

$$p(y \mid \boldsymbol{\theta}) = \sum_{j=1}^{k} \pi_j \cdot p_j(y \mid \boldsymbol{\theta}_j), \tag{4.1}$$

where π_j and $p_j(\cdot)$ represent the weight and PMF for the jth elementary component, respectively, subject to the usual constraints:

$$\pi_j \geq 0 \qquad \text{and} \qquad \sum_{j=1}^{k} \pi_j = 1. \tag{4.2}$$

Thus, from an interpretational point of view, the FMM assumes that the random variable $\mathbf{Y}$ is generated from k distinct random processes. Each of these processes is modeled by the PMF $p_j(\cdot)$, and π_j here defines the proportion of observations from this particular process. This implies that $\pi_j \in [0, 1]$, even though it is technically possible for values of π_j to lay outside of the unit interval, whilst maintaining a valid marginal distribution. As a result, the non-negative constraint in Eq. (4.2) can be lifted. To still ensure that $p(y \mid \boldsymbol{\theta})$ satisfies the constraint of a PMF, a new constraint is added:

$$p(y \mid \boldsymbol{\theta}) \geq 0, \forall y. \tag{4.3}$$

This extends the FMM framework into the use of negative weight components. A similar FMM framework was proposed by Zhang and Zhang (2005), with the aim of developing a new iterative estimation method for GMMs. Our purpose with the negative weights is to increase flexibility (Section 4.3.3) when using the elementary components from Section 4.3.1, in particular towards modeling underdispersion and/or zero-deflation (Section 4.4).

Further, we will allow the set of mixture component distributions to be either equal (e.g., all Poisson), or different (ranging over a class of base distributions).

4.3.3 Negative Weights

In the proposed FMM framework, the mean and variance expressions are functions of the model parameters originating from the elementary components. This suggests that negative weights can be allowed, while preserving the validity of the resulting PMF. In particular, we show the added flexibility negative weights can bring, within a mixture of two Poisson distributions, even though the Poisson itself is equidispersed and right-skewed.

(i) Case: Mixture of 2 Poisson distributions

Consider a mixture of 2 Poisson distributions, so that the resulting PMF, mean and variance functions take the form:

$$\begin{aligned} p(y \mid \lambda_1, \lambda_2) &= \pi_1 \frac{\mathrm{e}^{-\lambda_1} \lambda_1^y}{y!} + (1-\pi_1) \frac{\mathrm{e}^{-\lambda_2} \lambda_2^y}{y!}, & (4.4) \\ \mathrm{E}(\mathbf{Y}) &= \pi_1 \lambda_1 + (1-\pi_1)\lambda_2, & (4.5) \\ \mathrm{Var}(\mathbf{Y}) &= \pi_1 \lambda_1^2 + (1-\pi_1)\lambda_2^2 - [\pi_1 \lambda_1 + (1-\pi_1)\lambda_2]^2 & (4.6) \\ &\quad + \pi_1 \lambda_1 + (1-\pi_1)\lambda_2, \end{aligned}$$

respectively. By imposing the relaxed constraints from Section 4.3.2, the following boundaries can be found for weight π_1 to still ensure a PMF for $p(y \mid \lambda_1, \lambda_2)$:

$$\pi_1 \;\in\; [R1, R2] = \begin{cases} \left[0, \frac{\mathrm{e}^{-\lambda_2}}{\mathrm{e}^{-\lambda_2} - \mathrm{e}^{-\lambda_1}}\right] & \text{if } \lambda_1 > \lambda_2, \\ \left[\frac{\mathrm{e}^{-\lambda_2}}{\mathrm{e}^{-\lambda_2} - \mathrm{e}^{-\lambda_1}}, 1\right] & \text{if } \lambda_1 < \lambda_2. \end{cases} \qquad (4.7)$$

Calculations can be found in Supplementary Material S.9. Evidently, the bounds depend on λ_1 and λ_2.

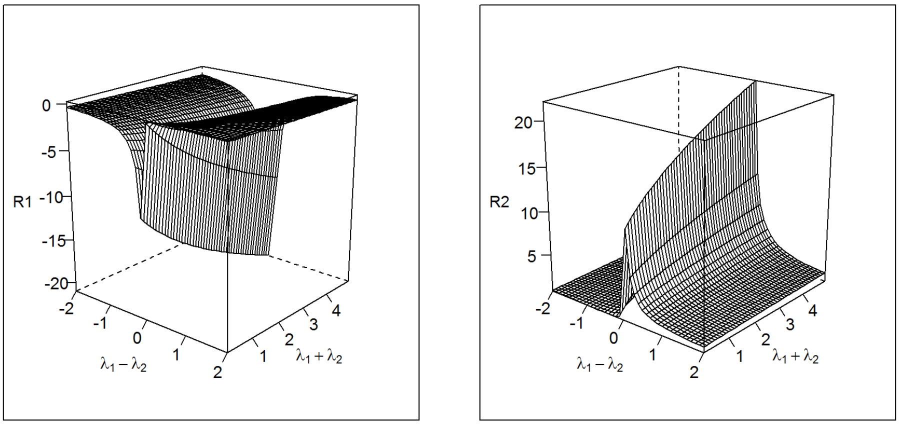
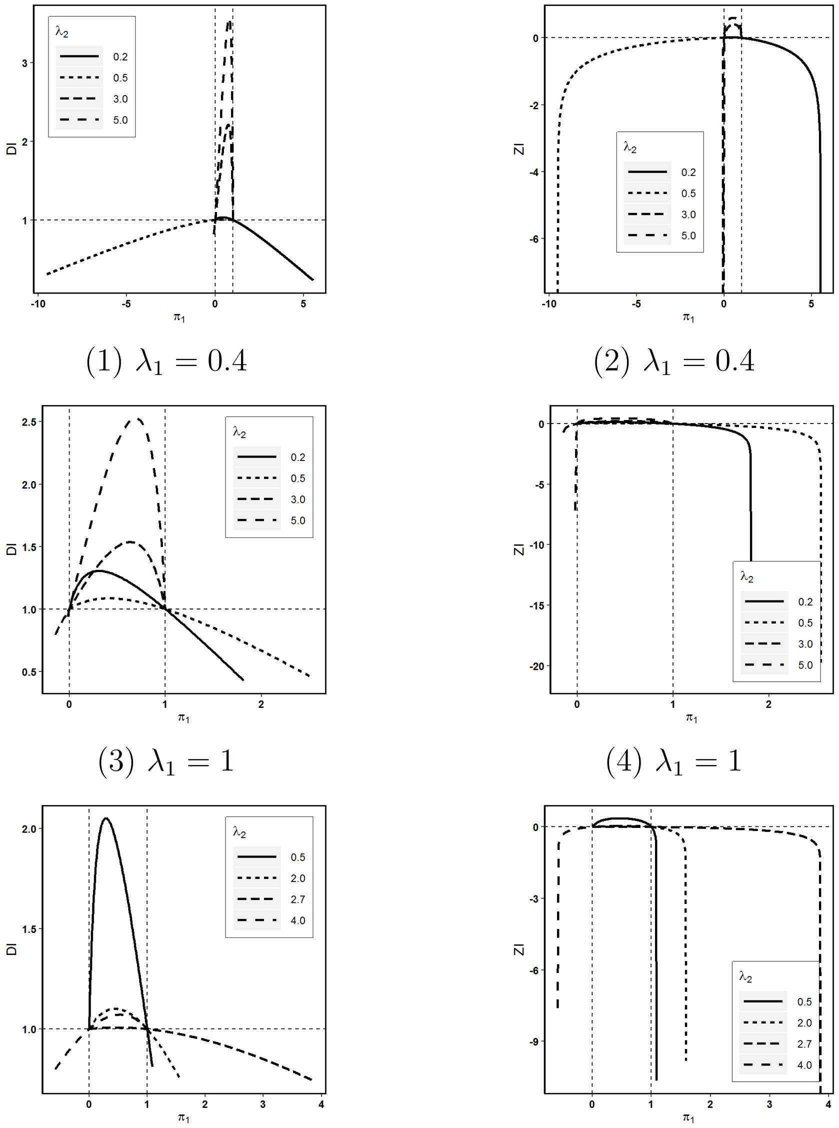


Figure 4.2: *Boundaries of* π_1 *in the 2-component Poisson mixture framework.*

Surprisingly, more homogeneous mean parameters imply a wider range of weight possibilities (Figure 4.2). Furthermore, it can easily be shown (Supplementary Material S.10) that $[0, 1] \subset [R1, R2]$ $(\forall \lambda_1, \lambda_2; \lambda_1 \neq \lambda_2)$. To explore some of the advantages of this extension, indexes for dispersion (DI) and zero-inflation (ZI), introduced by Puig and Valero (2006), are computed and given by:

$$\mathrm{DI} = \frac{\mathrm{Var}(\mathbf{Y})}{\mathrm{E}(\mathbf{Y})}, \qquad \mathrm{ZI} = 1 + \frac{\log[p(0)]}{\mathrm{E}(\mathbf{Y})}. \qquad (4.8)$$

Note that these indices are defined in relation to the Poisson distribution. Thus, the dispersion index indicates over-, under-, and equidispersion for, respectively, DI > 1, DI < 1 and DI $= 1$. The zero-inflation index indicates zero-inflation for ZI > 0, zero-deflation for ZI < 0 and no excess of zeros for ZI $= 0$. Figure 4.3 shows that, by examining the regions of DI and ZI for negative weights, underdispersion and/or zero-deflation can only be captured with negative weights.



(1) $\lambda_1 = 0.4$ (2) $\lambda_1 = 0.4$

(3) $\lambda_1 = 1$ (4) $\lambda_1 = 1$

(5) $\lambda_1 = 3$ (6) $\lambda_1 = 3$

Figure 4.3: *Dispersion (DI; left) and zero-inflated (ZI; right) index w.r.t.* π_1 *of the 2-component Poisson mixture framework, for varying* λ_1 *and* λ_2 *values.*

(ii) Interpretation of negative weights

Evidently, one view on a mixture with negative weights is a purely marginal one, i.e., the modeler is interested only in describing the marginal distribution but without specific focus on the mixture components and the interpretation that they could bring.

When weights are non-negative, a hierarchical interpretation is possible, i.e., that the total population is composed of two sub-populations, each representing a proportion π_j, $j = 1, 2$, of the total population.

Negative weights can be seen as occurring for compensatory purposes. For example, the probability of a large observation in one component is *compensated* by a similarly large observation from a second component, given that the two realizations are now not added but subtracted from one another. This reduces the probability for larger values, and hence the variance.

4.4 Analysis of the Case Study

The Moerzeke dataset of Section 4.2 will now be analyzed with mixtures of 2 similar and different elementary components from Section 4.3.1. Because the case study is used for illustrative purposes only, no distinction will be made between males, females, and children. No covariates were included in the models to explore the fitted (theoretical) frequency distribution with the empirical frequency distribution. Maximum likelihood estimation with quasi-Newton optimization is used to get the parameter estimates and standard errors. This technique allows the fit of negative weights, in contrary to, for example, the standard-used expectation-maximization (EM) algorithm in FMMs (Dempster et al., 1977; O'Hagan et al., 2012). Results are given in Tables 4.2–4.3, and Supplementary Materials S.11.

In the non-mixture models (Table 4.2 & Appendix S.11.1), large improvements in log-likelihood values can be observed for the DN, DP and DW model, when comparing them to the Poisson case. A significant dispersion parameter ϕ clearly indicates the presence of underdispersion within the data, as had also been discovered at the descriptive level (Section 4.2). Overall, the DW model can be considered as the 'best' in terms of log-likelihood among the non-mixture models, partially explained by the fact that an underlying time-to-event relationship is given here, i.e., a discretized time-to-event outcome, in contrary to the other models.

Table 4.2: *Moerzeke dataset. Maximum-likelihood estimates (standard errors) for the coefficients in the (1) Poisson (P), (2) discrete normal (DN), (3) double Poisson (DP), and (4) discrete Weibull (DW) distribution.*

		P	**DN**	**DP**	**DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intensity	λ	6.2852 (0.0677)	6.2850 (0.0503)	6.2852 (0.0540)	0.9999 ($2.5E-5$)
Std. dev.	σ	−− (−−)	1.8395 (0.0360)	−− (−−)	−− (−−)
Dispersion	ϕ	−− (−−)	−− (−−)	1.5710 (0.0600)	−− (−−)
	ρ	−− (−−)	−− (−−)	−− (−−)	4.5369 (0.1055)
−2 log-lik.		5867.6	5595.3	5746.6	5522.8
AIC		5869.6	5599.3	5750.6	5526.8
BIC		5874.8	5609.8	5761.1	5537.2

By considering the FMM approach with similar (Table 4.3 & Appendix S.11.2) and different components (Table 4.4 & Appendix S.11.3), further lowering of -2 log-likelihood, AIC, and BIC values can be observed in all models compared to their univariate components from Table 4.2, with lowest value given at the discrete Weibull FMM approach with similar components. Based on the significant parameter estimates and goodness-of-fit quantities, the Moerzeke data is best modeled with a mixture of 2 underdispersed components. Furthermore, a negative fit in weight $(1-\pi_1)$ is obtained in the Poisson FMM approach with similar components.

Table 4.3: *Moerzeke dataset. Maximum-likelihood estimates (standard errors) for the coefficients in their corresponding finite mixture model frameworks with similar components (MP, MDN, MDP, MDW). Components are the Poisson (P), discrete normal (DN), double Poisson (DP), and discrete Weibull (DW) distribution.*

		MP	**MDN**	**MDP**	**MDW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intensity 1	λ_1	5.4661 (0.1113)	4.5534 (0.2484)	4.6775 (0.1675)	0.9999 ($3.2E-8$)
Std. dev. 1	σ_1	−− (−−)	1.6430 (0.1174)	−− (−−)	−− (−−)
Dispersion 1	ϕ_1	−− (−−)	−− (−−)	1.3854 (0.1064)	−− (−−)
	ρ_1	−− (−−)	−− (−−)	−− (−−)	8.3960 (0.6059)
Intensity 2	λ_2	5.0918 (0.1307)	7.3614 (0.0626)	7.3376 (0.0484)	0.9956 (0.0014)
Std. dev. 2	σ_2	−− (−−)	0.8862 (0.0438)	−− (−−)	−− (−−)
Dispersion 2	ϕ_2	−− (−−)	−− (−−)	8.4796 (0.6980)	−− (−−)
	ρ_2	−− (−−)	−− (−−)	−− (−−)	3.3182 (0.2914)
Mixing prob.	π_1	3.1892 (1.2438)	0.3833 (0.0449)	0.3956 (0.0305)	0.7195 (0.0702)
−2 log-lik.		5814.1	5324.2	5358.9	5310.9
AIC		5820.1	5334.2	5368.9	5320.9
BIC		5835.7	5360.3	5395.0	5347.0

Table 4.4: *Moerzeke dataset. Maximum-likelihood estimates (standard errors) for the coefficients in their corresponding finite mixture model frameworks with different components (Mixt. P+DN, Mixt. P+DP, Mixt. P+DW, Mixt. DN+DP, Mixt. DN+DW, Mixt. DP+DW). Components are the Poisson (P), discrete normal (DN), double Poisson (DP), and discrete Weibull (DW) distribution.*

		Mixt. P+DN	**Mixt. P+DP**	**Mixt. P+DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intensity 1	λ_1	4.7133 (0.1601)	4.8153 (0.1504)	4.4450 (0.1958)
Std. dev. 1	σ_1	−− (−−)	−− (−−)	−− (−−)
Dispersion 1	ϕ_1	−− (−−)	−− (−−)	−− (−−)
	ρ_1	−− (−−)	−− (−−)	−− (−−)
Intensity 2	λ_2	7.2502 (0.0513)	7.2925 (0.0477)	0.9999 (3.8$E-8$)
Std. dev. 2	σ_2	0.9326 (0.0438)	−− (−−)	−− (−−)
Dispersion 2	ϕ_2	−− (−−)	8.2388 (0.7019)	−− (−−)
	ρ_2	−− (−−)	−− (−−)	8.1576 (0.4208)
Mixing prob.	π_1	0.3805 (0.0288)	0.4066 (0.0276)	0.3009 (0.0319)
−2 log-lik.		5353.4	5378.1	5329.6
AIC		5361.4	5386.1	5337.6
BIC		5382.3	5407.0	5358.5

		Mixt. DN+DP	**Mixt. DN+DW**	**Mixt. DP+DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intensity 1	λ_1	4.7703 (0.2174)	3.8696 (0.2838)	4.2241 (0.2533)
Std. dev. 1	σ_1	1.7161 (0.0978)	1.4198 (0.1466)	−− (−−)
Dispersion 1	ϕ_1	−− (−−)	−− (−−)	1.3731 (0.1504)
	ρ_1	−− (−−)	−− (−−)	−− (−−)
Intensity 2	λ_2	7.4187 (0.0559)	0.9999 (4.1$E-8$)	0.9999 (3.2$E-8$)
Std. dev. 2	σ_2	−− (−−)	−− (−−)	−− (−−)
Dispersion 2	ϕ_2	9.2706 (0.8534)	−− (−−)	−− (−−)
	ρ_2	−− (−−)	8.1644 (0.4689	8.2549 (0.4439)
Mixing prob.	π_1	0.4280 (0.0433)	0.2513 (0.0396)	0.2854 (0.0373)
−2 log-lik.		5345.2	5314.6	5319.1
AIC		5355.2	5324.6	5329.1
BIC		5381.3	5350.8	5355.2

To explore the negative weight estimation in more detail, focus is placed on the estimated mean and variance expressions (Table 4.5). Indeed, while the univariate Poisson case is limited to equidispersed data structures, extra flexibility is now gained in the variance expression towards underdispersed cases (Section 4.3.3).

Table 4.5: *Moerzeke dataset. Mean and variance estimates of all fitted univariate and finite mixture models. Components are the Poisson (P), discrete normal (DN), double Poisson (DP), and discrete Weibull (DW) distribution.*

	P	**DN**	**DP**	**DW***
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Mean	6.2852 (0.0677)	6.2850 (0.0503)	6.2852 (0.0540)	6.2985 (0.0462)
Variance	6.2852 (0.0677)	3.4672 (0.1324)	4.0007 (0.1566)	2.3914 (0.1100)
	MP	**MDN**	**MDP**	**MDW***
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Mean	6.2855 (0.0622)	6.2851 (0.0503)	6.2852 (0.0509)	6.2874 (0.0501)
Variance	4.4881 (0.1898)	3.4662 (0.1363)	3.5507 (0.1348)	2.8656 (0.1313)
	Mixt. P+DN	**Mixt. P+DP**	**Mixt. P+DW***	**Mixt. DN+DP**
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Mean	6.2850 (0.0533)	6.2852 (0.0538)	6.2874 (0.0520)	6.2852 (0.0503)
Variance	3.9324 (0.1322)	3.9639 (0.1263)	3.3015 (0.1494)	3.4353 (0.1364)
	Mixt. DN+DW*	**Mixt. DP+DW***		
	Est. (s.e.)	Est. (s.e.)		
Mean	6.2857 (0.0503)	6.2863 (0.0505)		
Variance	3.0073 (0.1406)	3.0782 (0.1437)		

*Approximations are used to obtain the estimates

4.5 Conclusion

In this chapter, the traditional FMM framework in count data is extended to account for both over- and underdispersion by considering a flexible class of elementary components, thus lifting the non-negativity constraint for mixture probabilities. Here, focus is placed on the practical use of the approach within underdispersed data. It is also useful to capture zero-deflation, in addition to zero-inflation. Applying the framework to underdispersed demographic data showed advantageous regarding fit when comparing it to existing univariate frameworks. Negative and positive weights are encountered in the data analysis, underscoring the practical relevance of this extension.

Extensions are possible and desirable in various ways. For example, when data are measured longitudinally or when otherwise hierarchical data are encountered, correlated-data versions can be constructed.

Chapter 5

Fitting Negative Variance Components in a Mixed Approach for Counts: A Flexible Marginalized Estimation Approach

This chapter is based upon:

Luyts, M., Molenberghs, G., Verbeke, G., and Matthijs, K. (2021). Fitting negative variance components in a mixed approach for counts: A flexible marginalized estimation approach. In preparation.

Abstract

When fitting Poisson-normal mixed-effects models to correlated non-negative discrete data, two viewpoints can be taken regarding the evaluation and interpretation of their parameters. In a hierarchical approach, subject-specific inferences are envisaged alongside population-averaged inferences; in a marginal approach, only the latter are of interest. In the hierarchical approach, the random-effects distribution is of interest, hence its variance-covariance matrix should be positive definite. In the marginal approach, this constraint does not apply and only the weaker condition of the marginal variance-covariance structure to be valid applies: The variance-covariance structure governing the random-effects does not have to be positive definite, as long as the overall variance-covariance structure is. For a single random effect, positive definiteness obviously corresponds to a positive random effect variance.

While this subject has already been examined in depth by Pryseley et al. (2011) in the generalized linear mixed model setting, where the occurrence of negative variance components is theoretically linked with negative intraclass correlations, underdispersion, and others, a flexible estimation approach in the hierarchical count data case is still lacking. Here, flexible (closed-form) second-order generalized estimating equations are proposed based on marginal moments up to order four of the Poisson-normal mixed framework, allowing for the estimation of negative variance components in the marginalized case. The methodology is not restricted to count data but can be applied to hierarchical time-to-event data as well. The approach allows for flexible modeling of the mean and variance-covariance structures. Model-based along with empirically corrected precision estimators are derived. Data from a historical demographic study are analyzed with a clustered discrete time-to-event outcome, that can be analyzed as a count.

5.1 Introduction

Poisson mixed models, well-known members of the generalized linear mixed model (GLMM; Thall and Vail, 1990; Dean, 1991; Engel and Keen, 1994) family, can accommodate both hierarchies in the data and are, too some extend, flexible in terms of dispersion modeling by adding random effects to the linear predictor. These random effects are considered to be drawn from a probabilistic, often normal, law. Applications range over medical, sociological, and psychological sciences, to name but a few. A medical example is the number of weekly epileptic attacks, recorded over several weeks. In psychological studies, the contact behavior among adults may be assessed repeatedly over time during the SARS-CoV-1 period.

Inferences in Poisson mixed models are frequently conducted via maximizing the marginal likelihood, i.e., the obtained likelihood by integrating over the random effects. It can be cumbersome because of non-linearity caused by the presence of a non-linear link function and, as a consequence, lack of a (convenient) closed-form expression or presence of infinite series. Therefore, approximation methods have been proposed for the integration that is required to derive the said marginal likelihood. Several such approximation methods exist, i.e., (1) those that approximate the integral numerically, and (2) those that approximate the integrand by Taylor-series expansion, such that a closed-form expression exists for the approximate integral. Examples like Gaussian quadrature (Pinheiro and Bates, 1995) and Monte Carlo integration (Zeger and Karim, 1991) have been proposed within the first approach, where the integral is substituted by a finite sum and maximized. For the latter approach, techniques like the Laplace approximation (Tierney and Kadane, 1986) and quasi-likelihood methods such as penalized quasi-likelihood (PQL; Schall, 1991; Breslow and Clayton, 1993), marginal quasi-likelihood (MQL; Goldstein, 1991) and several of their extensions (e.g., PQL2 and MQL2) have been constructed and applied. A full review of these methods can be found in Tuerlinckx et al. (2006). Numerical integration can be accurate but time-consuming, whereas PQL and especially MQL can be very inaccurate, especially when dealing with highly skewed data.

Because the models' genesis is subject-specific, also the fixed-effects parameters embrace a subject-specific interpretation. For example, a fixed treatment-effect parameter would describe the difference in response between a treated and an untreated subject with the same level for their random effects. This is not always what one wants. The average (marginal) treatment effect may be of higher scientific interest. A related observation is that the parameter space restrictions enforce a valid hierarchical interpretation; e.g., the distribution of the said random effects is valid, with a positive definite variance-covariance matrix. For a marginal interpretation, though, the milder restriction of a positive definite marginal variance-covariance matrix is all that is needed. In particular, some variance components in the hierarchical model may be negative, while still producing a valid marginal interpretation, but conventional estimation methods and their implementations preclude this. Several authors have investigated this phenomenon. These include Molenberghs and Verbeke (2011), Pryseley et al. (2011) and Oliveira et al. (2017), for the linear mixed models (LMMs; Laird and Ware, 1982), GLMMs and combined modeling (CM) framework of Molenberghs et al. (2010), respectively.

Alternatively, one can turn to methods that account for the hierarchical design of the data, without adding cluster-specific effects and therefore directly turn to investigating marginal inferences, with parameters that carry out a marginal, population-averaged, interpretation. Arguably, the most popular marginal method is the generalized estimating equations (GEE) approach by Liang and Zeger (1986), which extends the generalized linear model (GLM; Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Jørgensen, 1987) by conceding for within-cluster correlation, via a defined working correlation structure. Parameter estimators – obtained by solving the so-called GEE – remain consistent and asymptotically normal when misspecifying this working correlation. Unfortunately, when scientific interest also lies in the variance-covariance structure, the standard GEE approach falls short of this. To obtain these in the Poisson mixed framework, and maintain a marginal interpretation of the fixed-effects parameters, focus will here be placed on the second-order GEE (GEE2; Liang et al., 1992, Zhao and Prentice, 1990, Prentice and Zhao, 1991) approach. This method involves the derivation of the marginal moments up to order four, or approximations thereof. The so-obtained estimators allow for negative variance components and/or underdispersion.

The remainder of this chapter is organized as follows. We devote Section 5.2 to the introduction of a motivating case study where interest is with a clustered discretized time-to-event outcome. In these data, underdispersion is present. Negative variance components in the Poisson mixed model and the range thereof is examined in Section 5.3; particular focus is placed on the random-intercepts case. The basic principles of GEE2 and their application to the Poisson mixed case is the subject of Section 5.4. Again, the main focus is on the random-intercept model, although extension to, for example, random slopes, is definitely possible. The data are analyzed in Section 5.5. Concluding remarks are given in Section 5.6.

5.2 The Moerzeke Data

The database of Moerzeke has been analyzed before by Matthijs et al. (2002). In sociological and demographic historical studies, research around the transmission of longevity from parents to offspring has led to various insights (e.g., Van den Berg et al., 2019; Matthijs et al., 2002). Here, historical data of Moerzeke, a small Flemish village located in the province of East Flanders (Belgium), is explored over a period of three hundred years (18th–20th centuries). The outcome of interest is here the life expectancy among family members, expressed in decades, i.e., a discretized time-to-event outcome.

We focus on an initial sample of 1062 women with their relatives. Information was assembled from civil and church registration certificates. Several filters were applied to obtain our sample. The sample consists of not-remarrying women, to avoid overlap of children in different households. Only household members who were born and died in Moerzeke were used. Households with at least one "non-Moerzeke" member are therefore excluded. Households with no children were excluded. When more than one child was available in a family, only the oldest child is included. Only fathers born between 1750 and 1830 were selected. Applying all these criteria results in a sample of 457 families, with a mean age at death of 71.9 and 71.7 years for men and women, respectively.

5.3 The Poisson Mixed Approach

5.3.1 Methodology

Given the focus on longitudinal/hierarchical count and discretised time-to-event outcomes, it is common to choose, conditional on the random effects, the Poisson distribution as the base outcome distribution and normal random effects in the mean parameter to account for association, resulting in the Poisson mixed model. Consider Y_{ij} to be the jth non-negative discrete response in cluster/subject i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$. The Poisson mixed model takes the form:

$$P(Y_{ij} = y_{ij} \mid \mathbf{b}_i) = \frac{\mathrm{e}^{-\lambda_{ij}} \cdot \lambda_{ij}^{y_{ij}}}{y_{ij}!}, \tag{5.1}$$

$$\lambda_{ij} = \mathrm{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta} + \mathbf{z}_{ij}^{'}\mathbf{b}_i}, \tag{5.2}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \tag{5.3}$$

where $\boldsymbol{\beta}$ and $\mathbf{x}_{ij}^{'}$ are p-dimensional vectors of unknown fixed parameters and known covariate values, respectively, and $\mathbf{b}_i$ and $\mathbf{z}_{ij}^{'}$ are q-dimensional vectors of random effects and known covariate values, respectively. Furthermore, N describes the total amount of clusters/subjects, and n_i expresses the number of repeated measurements for cluster/subject i. D is the q x q covariance matrix for the random effects $\mathbf{b}_i$.

In this chapter, to gain useful insight from sufficiently simple expressions, particular focus will be placed on the random-intercept approach with no covariates, i.e., $\mathbf{x}_{ij}^{'}\boldsymbol{\beta} = \beta$, $\mathbf{z}_{ij}^{'}\mathbf{b}_i = b_i$ and $D = d$. Generalizations to more complex structures like random slopes in the Poisson mixed framework are straightforward.

5.3.2 Negative Variance Components

For the Poisson-normal model, closed-forms exist for the central and non-central marginal moments. Formulations of these moments up to the fourth order can be found in Appendix S.12. In case of the marginal mean and variance structure, the following expressions are obtained:

$$\mathsf{E}(Y_{ij}) = \mathrm{e}^{\beta + \frac{1}{2}d}, \tag{5.4}$$

$$\mathsf{Var}(Y_{ij}) = \mathrm{e}^{\beta + \frac{1}{2}d} + \left(\mathrm{e}^{\beta + \frac{1}{2}d}\right)^2 \left(\mathrm{e}^{d} - 1\right). \tag{5.5}$$

Evidently, negative estimates for d can easily be obtained while still maintaining a valid marginalized variance structure. In particular:

Theorem 1. *If β (or linear predictor) is sufficiently small, i.e., when $\beta \leq ln(\frac{3\sqrt{3}}{2})$, then all real values for d are valid to secure a proper marginalized variance structure. Otherwise, there is a finite interval where this is not attained.*

Proof. Let $\tilde{f}(d) = \mathrm{e}^{\beta+\frac{1}{2}d} + \mathrm{e}^{2\beta}\mathrm{e}^{d}(\mathrm{e}^{d}-1)$ and $\tilde{\tilde{f}}(\delta) = 1 + k\delta(\delta^2-1)$, where $k = \mathrm{e}^{\beta}$ and $\delta = \mathrm{e}^{\frac{1}{2}d}$. It can easily be shown that:

$$\tilde{f}(d) \geq 0 \Leftrightarrow \tilde{\tilde{f}}(\delta) \geq 0. \tag{5.6}$$

Furthermore, a global minimum exists for $\tilde{\tilde{f}}(\delta)$, since $\frac{\partial}{\partial\delta}\left[\tilde{\tilde{f}}(\delta_0)\right] = 0 \Leftrightarrow \delta_0 = \frac{\sqrt{3}}{3}$ and $\frac{\partial^2}{\partial\delta^2}\left[\tilde{\tilde{f}}(\delta)\right] = 2\sqrt{3}k \geq 0$ (concave up). Evaluating δ_0 in (5.6) results in the constraint $\beta \leq \ln(\frac{3\sqrt{3}}{2})$ for d to be valid for all values. Due to the convexity property, and when this constraint no longer applies, there is a finite interval where no valid values for d are observed on the real line (Figure 5.1). ☐

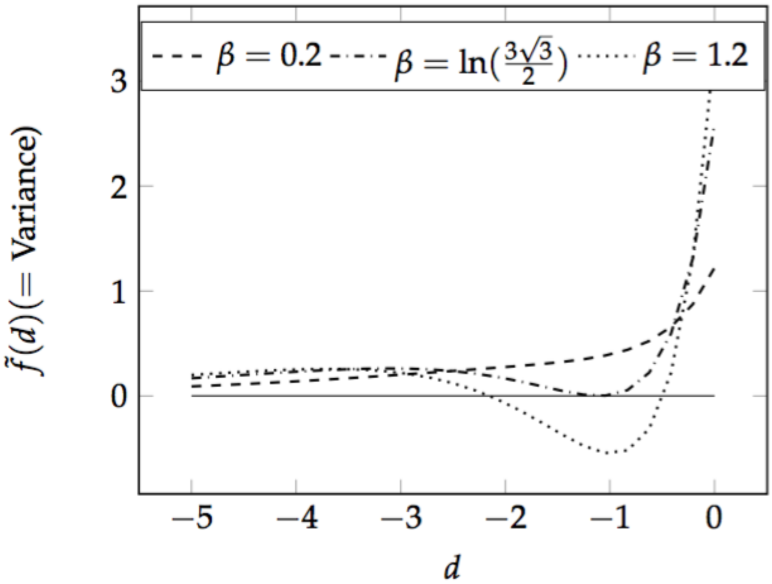


Figure 5.1: Marginal variance structure of the Poisson-normal random-intercept approach with no covariates w.r.t. variance component d, for varying β values.

5.4 Second-order Generalized Estimating Equations

Zhao and Prentice (1990) and Liang et al. (1992) proposed second-order GEE (GEE2) by using correlations and odds ratios, respectively. These are a natural extension of the Prentice (1988) approach, which modified the primary GEE of Liang and Zeger (1986). In what follows, we will propose and describe the GEE2 approach with GLMMs marginal moments, and study it explicitly for the Poisson-normal model of Section 5.3. Details of the Liang and Zeger (1986) and Prentice (1988) methods can be found in Molenberghs and Verbeke (2005).

5.4.1 Methodology

Consider $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ to be the vector of non-Gaussian outcomes for cluster/subject i, $i = 1, \ldots, N$, following exponential family densities, expressed by

$$f_i(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp\{\phi^{-1} \cdot [y_{ij} \cdot \nu_{ij} - \psi(\nu_{ij})] + c(y_{ij}, \phi)\}, \tag{5.7}$$

with

$$h[\psi'(\nu_{ij})] = h[\mathsf{E}(Y_{ij} \mid \mathbf{b}_i, \boldsymbol{\beta})] = \eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}, \tag{5.8}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, D). \tag{5.9}$$

The same terminology is used as in Section 5.3.1. Consider $\mathbf{S}_i$ to be the n_i-dimensional vector of empirical marginal variances defined by

$$\mathbf{S}_i = (S_{ij})'_{j \in [1, n_i]}, \tag{5.10}$$

where $S_{ij} = [Y_{ij} - \mathsf{E}(Y_{ij})]^2$. In what follows, estimating equations will be proposed for $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Omega}$, i.e., the vector of fixed and random parameters, respectively, originating from the marginal moments of Eq. (5.7)-(5.9).

Let Θ be the vector consisting of all parameters in $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Omega}$. The joint estimating equations are

$$\mathbf{Q}(\Theta) = \sum_i \mathbf{P}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{f}_i = \mathbf{0}, \tag{5.11}$$

where

$$\mathbf{P}_i = \begin{bmatrix} \mathbf{P}_{i11} & \mathbf{P}_{i12} \\ \mathbf{P}_{i21} & \mathbf{P}_{i22} \end{bmatrix}, \qquad \boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{i11} & \boldsymbol{\Sigma}_{i12} \\ \boldsymbol{\Sigma}_{i21} & \boldsymbol{\Sigma}_{i22} \end{bmatrix}, \qquad \mathbf{f}_i = \begin{bmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{S}_i - \boldsymbol{\eta}_i \end{bmatrix},$$

with $\boldsymbol{\mu}_i = \text{E}(\mathbf{Y}_i)$, $\boldsymbol{\eta}_i = \text{E}(\mathbf{S}_i)$, $\boldsymbol{\Sigma}_{i11} = \text{Var}(\mathbf{Y}_i)$, $\boldsymbol{\Sigma}_{i12} = \boldsymbol{\Sigma}_{i21} = \text{Cov}(\mathbf{Y}_i, \mathbf{S}_i)$, $\boldsymbol{\Sigma}_{i22} = \text{Var}(\mathbf{S}_i)$, $\mathbf{P}_{i11} = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\Upsilon}}$, $\mathbf{P}_{i12} = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\Omega}}$, $\mathbf{P}_{i21} = \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\Upsilon}}$, and $\mathbf{P}_{i22} = \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\Omega}}$.

The parameter estimates obtained from Eq. (5.11), denoted $\widehat{\boldsymbol{\Upsilon}}$ and $\widehat{\boldsymbol{\Omega}}$, can be derived either analytically (if a closed-form exists) or numerically by the Fisher scoring algorithm (Longford, 1987) where the iterative procedure at step $(k+1)$ takes the form:

$$\begin{bmatrix} \widehat{\boldsymbol{\Upsilon}}^{(k+1)} \\ \widehat{\boldsymbol{\Omega}}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\Upsilon}}^{(k)} \\ \widehat{\boldsymbol{\Omega}}^{(k)} \end{bmatrix} + \frac{1}{N}\left[\sum_i \widehat{\mathbf{P}}_i' \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\mathbf{P}}_i\right]^{-1} \left[\sum_i \widehat{\mathbf{P}}_i' \widehat{\boldsymbol{\Sigma}}_i^{-1} \widehat{\mathbf{f}}_i\right], \qquad (5.12)$$

where $\widehat{\mathbf{P}}_i$, $\widehat{\boldsymbol{\Sigma}}_i$ and $\widehat{\mathbf{f}}_i$ are the values of $\mathbf{P}_i$, $\boldsymbol{\Sigma}_i$ and $\mathbf{f}_i$, respectively, evaluated at $(\widehat{\boldsymbol{\Upsilon}}^{(k)}, \widehat{\boldsymbol{\Omega}}^{(k)})$. Since only marginal moments are used, there is no requirement that random effects variance-covariance matrix D needs to be positive definite. Furthermore, model-based and empirically corrected (sandwich) standard errors are derived as the square root of the diagonal entries of

$$U^* = \left[\sum_i \mathbf{P}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{P}_i\right]^{-1}, \qquad (5.13)$$

$$U^{**} = U^* \left[\sum_i \mathbf{P}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{f}_i \mathbf{f}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{P}_i\right] U^*, \qquad (5.14)$$

respectively.

5.4.2 Application: The Poisson-normal Model

In the Poisson-normal model of Section 5.3.1, the score equations become:

$$\mathbf{Q}([\beta, d]') = \frac{NB^2\Delta^2 K}{\det(\boldsymbol{\Sigma}_i)} \begin{bmatrix} \underbrace{\frac{1}{N}\sum_i \sum_j Y_{ij}}_{k_1} - B\Delta \\ \underbrace{\frac{1}{N}\sum_i \sum_j Y_{ij}^2}_{k_2} - (B^2\Delta^4 + B\Delta) \end{bmatrix} = \mathbf{0}, \qquad (5.15)$$

where $B = \mathrm{e}^{\beta}$, $\Delta = \mathrm{e}^{\frac{1}{2}d}$, and $K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$ with

$$K_{11} = B^3\Delta^7\left(\Delta^4-1\right)^2 + B^2\Delta^4\left(5\Delta^2-1\right)\left(\Delta^2-1\right) + 2B\Delta^3, \tag{5.16}$$

$$K_{12} = -B^2\Delta^4\left(\Delta^2-1\right)^2, \tag{5.17}$$

$$K_{21} = \frac{1}{2}\left[B^3\Delta^7\left(\Delta^4-1\right)\left(\Delta^4-3\right) + B^2\Delta^4\left(5\Delta^4-12\Delta^2+3\right)\right], \tag{5.18}$$

$$K_{22} = \frac{1}{2}\left[-B^2\Delta^4\left(\Delta^2-3\right)\left(\Delta^2-1\right) + 2B\Delta^3\right]. \tag{5.19}$$

Derivations can be found in Appendix S.13. Solving these equations analytically result in the following closed-form expressions:

$$\begin{cases} \widehat{\beta} &= 2\ln(k_1) - \frac{1}{2}\ln(k_2-k_1) \\ \widehat{d} &= \ln(k_2-k_1) - 2\ln(k_1) \end{cases} \tag{5.20}$$

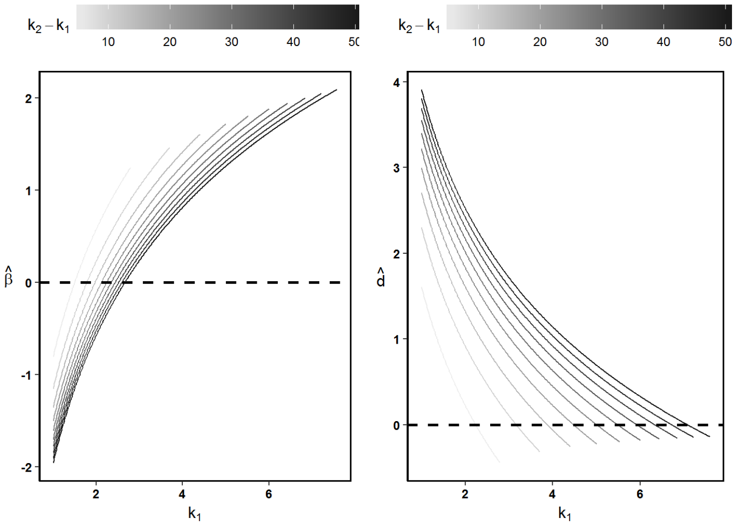


Figure 5.2: *Parameter estimates solution of the second-order generalized estimating equations for the Poisson-normal random-intercept approach with no covariates w.r.t. the non-central first sample moment k_1, for varying differences between the non-central second and first sample moment, i.e., $k_2 - k_1$.*

For calculating the model-based and sandwich standard errors, a PROC IML macro has been written (Appendix S.13.3). Due to the presence of expression $\mathbf{f}_i\mathbf{f}_i^T$, non-central sample moments up to the fourth order, i.e., k_1, k_2, $k_3(=\frac{1}{N}\sum_i\sum_j Y_{ij}^3)$, and $k_4(=\frac{1}{N}\sum_i\sum_j Y_{ij}^4)$, are needed for these calculations.

5.5 Analysis of the Moerzeke Data

The proposed GEE2 of Section 5.4.2 will be applied to the Moerzeke study. While Luyts et al. (2019) detected boundary issues in their analysis of Moerzeke, we try overcome this issue by allowing the fit of negative variance components. Particular focus is laid on the exploration of discretized longevity within household members, expressed in decades. It is a priori clear that the suggested method should flexibly allow for underdispersion, given that sample mean is 6.29 and sample variance 3.47.

Let Y_{ij} denote the discretized longevity for person j $(j = 1, 2, 3)$, i.e., father, mother, or first child, within household i $(i = 1, 2, \ldots, 457)$. The solution of the GEE2 approach for the Moerzeke data is given in Table 5.1. A graphical representation of the parameter estimates is presented in Figure 5.3.

Table 5.1: *Moerzeke data. Parameter estimates (model-based standard errors; empirically corrected standard errors) of the second-order generalized estimating equations solution for the Poisson-normal random-intercept approach with no covariates, based on the non-central first, second, third and fourth sample moments, i.e., k_1, k_2, k_3, and k_4, respectively.*

k_1	k_2	k_3	k_4
6.29	42.97	308.56	2290.46

Effect	Par.	Est. (m.b. s.e.; e.c. s.e.)	Z (m.b.)	p-value	Z (e.c.)	p-value
Fixed effect	β	1.8752 (0.0138; 0.0159)	135.88	$<$ 0.0001	117.94	$<$ 0.0001
Var. comp.	d	-0.0740 (0.0063; 0.0064)	-11.75	$<$ 0.0001	-11.56	$<$ 0.0001

As result, clear significant presence of a negative variance component is observed, underscoring our perception of the possible presence of negative variance components due to the feasible attendence of underdispersion in the Moerzeke data. In addition, the discrepancy between both model-based and empirically corrected standard errors is slight for both effects, implying that the working assumptions may well be sensible.

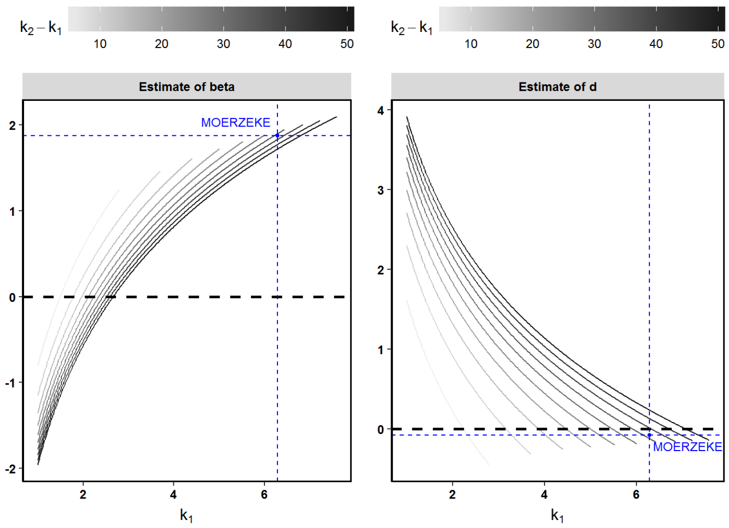


Figure 5.3: *Moerzeke data. Parameter estimates solution of the second-order generalized estimating equations for the Poisson-normal random-intercept approach with no covariates, for the observed non-central first and second sample moments.*

5.6 Conclusion

We have presented second-order estimating equations, based on the specific first- and second-order marginal moments of a generalized linear mixed model for count data, consisting of a Poisson model for the outcomes given random effects, and normally distributed random effects. The second-order nature of the method allows for proper inferences, not only on fixed effects, but also on variance components. At the same time, negative variance components, allowing for underdispersion, negative intraclass correlations, and others, are easily accommodated. Evidently, a negative random effects variance implies that a hierarchical interpretation is no longer possible. This is fine in applications were interest is on population-averaged quantities.

A PROC IML macro have been developed to derive the proposed method and make it applicable to other datasets as well. The macro reports both model-based and empirically corrected standard errors.

We should conclude by mentioning that our proposal also has restrictions. In this chapter, derivations and computations for analyzing the Moerzeke dataset were limited to a simple random-intercepts Poisson mixed model with no covariate structure. Extensions to a random-slope construction, with or without the inclusion of covariates, and consideration of marginal moments originating from other GLMM members can be considered as well. These generalizations are currently in development.

Chapter 6

Concluding Remarks

6.1 General Conclusion

In this manuscript, existing statistical developments have been explored for the analysis of two datasets originating from an observational clustered sociological study and longitudinal medical multi-center study. Due to the outcomes of interest, i.e., (discretized) longevity and number of epileptic attacks, respectively, particular focus was laid on counts and time-to-event outcomes. Under- and overdispersion were detected in the analysis. New statistical developments have been proposed, and applied to these studies. Significant improvement based on goodness-of-fit measurements are detected in our analysis compared to existing techniques, underscoring the relevance of this research. Characterizations of these new and existing methods have mathematically been investigated, highlighting the similarities and dissimilarities with some well-known methods in the literature.

The manuscript is divided into three parts. In the first part, i.e., Chapter 2, existing statistical methodologies were utilized for the analysis of a dispersed, clustered historic demographic data setting. Two models were considered, each with their own properties. A PD model with PL ideas was used to investigate the association structure of the inheritance in survival among family members within Moerzeke. A CM framework for time-to-event outcomes was fitted to reveal the presence of dispersion. A positive association between child and mother was observed within the PD model, with higher associations in case of daughters rather than sons. The plausible presence of negative variance components in the CM was highlighted, and theoretically linked with underdispersion. As a result, boundary issues were detected in the model fit when considering standard estimation tools. In the second part, i.e., Chapter 3 and Chapter 4, existing and new statistical frameworks were extended and developed, respectively, to handle dispersed, and/or correlated discrete data structures. In Chapter 3, a discrete Weibull-count approach was broadened with random effects to analyze the number of epileptic seizures and discretized longevity within Moerzeke, in a longitudinal and clustered setting, respectively. A comparison with some well-known count models was made, leading to the conclusion that the suggested approach properly fits the data compared to other conventional models. Characteristics of the approach demonstrated its practical relevance when considering zero-inflated, heavy-tailed, overdispersed and correlated skewed discrete data, and even underdispersed, zero-deflated skewed discrete data. In Chapter 4, a discrete FMM family with possible negative weights was proposed to account for over- and underdispersed, as well as zero-inflated/-deflated discrete data. In the data analysis of an underdispersed demographic setting, negative weights were fitted in a Poisson mixture of 2 components. In the final and third part of this thesis, i.e., Chapter 5, a marginal moment-based estimation procedure for counts was constructed to allow the fit of negative variance components originating from marginal moments of a hierarchical Poisson mixed framework. The estimation approach can easily be applied to applications in time-to-event outcomes as well.

Implementation of all models considered in this thesis were done in the statistical package SAS 9.4, using PROC NLMIXED. For matrix calculations, e.g., for calculating the model-based and empirically corrected standard errors in our GEE2 approach (Chapter 5), the PROC IML procedure was used. To avoid convergency problems on bounded parameters, appropriate transformations in combination with the delta method were used in the coding.

6.2 Limitations

Consistent with other (statistical) research in the literature, there is still capacity for growth and alternative routes in our suggestions and implementations.

Throughout this manuscript, a historic sample of the Moerzeke population consisting of complete data in lifespans was examined to explore our research topics of interest. Selection criteria such as the exclusion of childless households were applied. These restrictions were kept unchanged throughout this thesis, maintaining the underdispersed structure of the data but limiting our conclusions to this particular subset. Different subsets could have been explored as well, emphasizing different research questions, e.g., the comparison of native and foreign second generation child mortality within the 18^{th} century of Moerzeke. These questions remain open for further investigation.

In Chapter 2, two clustered time-to-event models were considered. Both models rely on the underlying Weibull distribution, even though other distributions can be considered as well. Gavrilov and Gavrilova (1991), for example, suggested the Gompertz-Makeham distribution as an appropriate alternative to explore mortality patterns in human populations between the ages of 20 and 80 years. Alternately, semi-parametric and/or non-parametric approaches could also have been considered, ruling out the problem of certain parametric choices. In this thesis, focus was placed on full-parametric models, with the aim of investigating certain model characteristics (e.g., DI, ZI and HT). Because of simplicity in interpretation, a PD model was used to analyze the pairs of familial survival outcomes. Model terminology with Plackett copulas could easily be substituted with other copula functions as well. The possible presence of underdispersion was detected with a WN and WGN model, opening new research possibilities that we explored further in this manuscript. Since these models are used for exploratory purposes only, we maintained ease in exploration in our analysis by opting for constant gamma parameters and a random-intercept structure.

In Chapter 3, a discrete version of the Weibull distribution is extended with random effects, and contrasted with some well-known count models that allow the fit of under- and/or overdispersed, correlated data, e.g., the CM. Inferences are interpreted on the median scale instead of the classical mean scale, making the comparison between model parameters with other well-known models cumbersome when highly skewed data is present. Additionally, Burger et al. (2020) pointed out that the DW approach is not necessary completely robust to excess zeros, and constructed a zero-inflated extension of our proposal. Maximum likelihood principles with numerical optimization techniques were used to obtain inferences. Only one level of hierarchy was present within the studied Moerzeke and epilepsy data. In most research, however, multiple hierarchical levels occur, and may pose computational challenges. Alternative estimation methods like Bayesian, semi-parametric or pseudo-likelihood and non-parametric approaches, can be used instead, as was done in Chapter 2 for the PD model.

In Chapter 4, an extended discrete FMM framework is proposed to model the discrete life expectancy of family members within the historic cohort Moerzeke. We limited the (mathematical) exploration of negative weights only to a mixture of two Poisson distributions, the most standard count distribution, leaving a whole area of discrete FMM models with possibly negative weights still open for investigation. No distinction was made between males, females, and children, an assumption that is often violated in practice. Therefore, extensions such as correlated-data versions are desirable in various ways. Properties like identifiability and convergence properties of various numerical optimization strategies were not covered in this chapter. The investigation of such properties remains open for further research.

In Chapter 5, a marginalized model-based estimation technique was proposed based on marginal moments of the Poisson mixed-effects model. Computations were considered for the random-intercept case only with no covariates, but can easily be extended with more complex structures as well, e.g., inclusion of covariates, random slopes with/without correlation, etc. While the method easily allows for the estimation of negative variance components and/or underdispersion, it involves the derivation of the marginal moments up to order four, or approximations thereof. In case of the extended DW model (Chapter 3), for example, no closed-form formulations exist for the first- and second-order moments, making the method not applicable here. Additionally, hands-on implementations in standard software packages are still lacking.

In general, the use of existing and new mixed models was considered. Normality was implicitly assumed for the underlying random effects distributions; an assumption that researchers often made when using mixed models. Diagnosis of the validity of this assumption, and, in general, towards model diagnostics, was not investigated in this manuscript.

6.3 Prospects for Future Research

The development and exploration of statistical models is done for centuries, with the aim of reducing bias in the data analysis. While the area of big data arises in every field of research, complexities increase in data structures (e.g., increased number of underdispersed data settings, level of hierarchies, etc.), indicating the need for new advanced techniques. Possible limitations in existing statistical tools are often not notified during the data analysis, making inferences possibly biased. In this thesis, we tried to address this problem by mentioning some drawbacks in classical techniques that are commonly used in dispersed, correlated data settings. Some new and alternative hierarchical frameworks are proposed in this thesis, putting particular focus on counts and time-to-event outcomes.

Several challenges are still open. While statistical practice is routinely performed in many research fields, a close collaboration between statisticians and clinicians, sociologists, etc. remains essential. Awareness needs to be provided when analyzing complex data with traditional techniques. This can be achieved by providing personal coaching activities, offering fundamental and advanced courses in statistics and data science, developing user-friendly software packages and setting-up close collaborations with statisticians.

Summary

The presence of longitudinal and/or hierarchical data structures in many fields (e.g., biomedicine, sociology, psychology, etc.) often imply the need for advanced statistical modeling strategies. These frameworks need to account for the underlying correlation structure, while still preserving simplicity in their interpretation. In this manuscript, focus is laid on models for counts and time-to-event outcomes.

In practice, these responses are traditionally modeled with members of the so-called exponential family, making them restricted in use due to their predefined mean-variance relation. There are two main reasons for expanding this family or propose alternative frameworks, i.e., (1) the occurence of dispersion, meaning that the variability of the data is not sufficiently captured under the assumed model, and (2) the longitudinal/hierarchical structure of the data. In the former one, the variance of the data can exceed or undervalue the predefined variance in the model, referred as over- and underdispersion, respectively.

To account for both, Molenberghs et al. (2007, 2010) proposed the so-called combined modeling framework through the inclusion of two distinguish sets of random effects within the exponential family. Maximum likelihood with partial marginalization is routinely applied to obtain inferences, implying subject-specific interpretations on their parameters even though this is not always preferred. A study may, for example, aim to investigate a treatment effect on the population rather than the treatment effect on a particular subject. Efendi et al. (2014) constructed a marginalized version of the combined model for time-to-event outcomes to obtain population-averaged interpretations for the fixed effects, while still maintaining a subject-specific interpretation for the random effects parameters.

By using two case studies originating from an observational demographic (sociological) and a randomized, double-blinded, parallel group multi-center (clinical) study, existing models were initially considered in the analyses. Boundary problems in the estimation procedure were detected in the sociological study, and linked with the possible presence of underdispersion within the data. Alternative univariate and hierachical models for counts have been proposed that easily allows the fit of both under- and overdispersion, and/or longitudinal/hierarchical discrete data. An extended family of discrete finite mixtures was constructed, and applied to the sociological study. Significant improvements on several goodness-of-fit quantities like AIC and BIC were observed, underscoring the practical relevance of this extension. An extended discrete Weibull model is considered to analyze both studies. Random effects were entered in the model to address the hierarchical/longitudinal structures of the data. Results show significant improvements in goodness-of-fit measures in comparison with traditional models, while remaining simplicity in their interpretation.

To conclude, a second-order moment-based estimation approach is proposed that enables the marginal estimation of negative values for the random effects parameters originating from Poisson mixed-effects models, therefore lifting boundary issues in classical estimation approaches such as partial marginalization.

Scientific Acknowledgement, Conflict of Interest and Personal Contribution

Scientific Acknowledgments

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. The research leading to these results has also received funding from KU Leuven GOA project: "New approaches to the social dynamics of long term fertility change". We thank Mr. R. Bijl, a Flemish genealogist and member of the V.V.F.-Dendermonde (Vlaamse Vereniging voor Familiekunde, Dendermonde), for granting us access to his demographic data on Moerzeke.

Conflict of Interest

There is no conflict of interest.

Personal Contribution

Chapter 2: Highlight the presence of underdispersion through a GLMM and CM framework for time-to-event outcomes; study boundary issues in a conventional hierarchical estimation approach; implementation of the CM framework and corresponding submodels in SAS; the analysis and interpretation of the results for the CM and related submodels.
Chapter 3: Overview of the literature and introduction; derivation of closed-form expressions for lower and upper boundaries of the mean and variance function; exploration of dispersion, zero-inflation and heavy-tailed regions through several indices; formulation of the extension with random effects; the analysis of the

epilepsy and Moerzeke datasets, alongside the interpretation of the results; exploration of the deviance surfaces with a large simulated set of data; formulation of the conclusion; implementation of all considered models and large simulated set of data in SAS and R, respectively.
Chapter 4: Literature study of existing work in FMMs with possible negative weights; proposal of the discrete FMM family with negative weights; mathematical and statistical derivations and explorations within the mixture of two Poissons; implementation of all models in SAS; the analysis and interpretation of the results for all models in the paper; formulation of the concluding remarks.
Chapter 5: Overview of existing approximation methods for fitting GLMMs; exploration of negative variance components in the Poisson-normal approach; reporting derivations of the GEE2 approach; implementation of the proposed GEE2 in SAS for the PN case; the analysis and interpretation of the results for the Moerzeke dataset.

Bibliography

Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhyā: The Indian Journal of Statistics, Series B*, 53(2):233–243.

Bartholomew, D. J. (1969). Sufficient conditions for a mixture of exponentials to be a probability density function. *The Annals of Mathematical Statistics*, 40(6):2183–2188.

Boatwright, P., Borle, S., Kadane, J. B., Minka, T. P., and Shmueli, G. (2006). Conjugate analysis of the conway-maxwell-poisson distribution. *Bayesian Analysis*, 1(2):363–374.

Booth, J. G., Casella, G., Friedl, H., and Hobert, J. P. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, 3(3):179–191.

Breslow, N. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association*, 85(410):565–571.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 33(1):38–44.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25.

Broyden, C. G. (1967). Quasi-Newton methods and their application to function minimisation. *Mathematics of Computation*, 21(99):368–381.

Burger, D. A., Schall, R., Ferreira, J. T., and Chen, D.-G. (2020). A robust Bayesian mixed effects approach for zero inflated and highly skewed longitudinal count data emanating from the zero inflated discrete Weibull distribution. *Statistics in Medicine*, 39(9):1275–1291.

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of applied econometrics*, 1(1):29–53.

Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions–A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, 2(1):1–30.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12:132–136.

Cournil, A., Legay, J. M., and Schächter, F. (2000). Evidence of sex-linked effects on the inheritance of human longevity: A population-based study in the Valserine valley (French jura), 18-20th centuries. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1447):1021–1025.

Cox, D. R. (1962). *Renewal theory*. Methuen Ltd., London.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, 70(1):269–274.

Cox, D. R. and Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.

Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 42(4):909–917.

De Beule, O. (1962). *Moerzeke. Religieus-sociale monografie van een plattelandsgemeente*. Erasmus, Ledeberg.

De Ridder, J. (1986). *Moerzeke 1710-1796, een historisch-demografische analyse van een plattelandsparochie in Oost-Vlaanderen*. Belgisch centrum voor landelijke geschiedenis, Leuven.

Dean, C. B. (1991). Estimating equations for mixed Poisson models. In *Estimating Functions (ed. V. P. Godambe)*, pages 35–46. Oxford University Press, Oxford.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–38.

Duchateau, L. and Janssen, P. (2007). *The frailty model*. Springer Science & Business Media.

Efendi, A., Molenberghs, G., and Iddi, S. (2014). A marginalized combined gamma frailty and normal random-effects model for repeated, overdispersed, time-to-event outcomes. *Communications in Statistics–Theory and Methods*, 43(22):4806–4828.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721.

Engel, B. and Keen, A. (1994). A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, 48(1):1–22.

Englehardt, J. D. and Li, R. (2011). The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis: An International Journal*, 31(3):370–381.

Everitt, B. S. and Hand, D. J. (1981). *Mixtures of discrete distributions. In: Finite Mixture Distributions. Monographs on Applied Probability and Statistics.* Springer, Dordrecht.

Famoye, F. (1993). Restricted generalized Poisson regression model. *Communications in Statistics – Theory and Methods*, 22(5):1335–1354.

Faught, E., Wilder, B. J., Ramsay, R. E., Reife, R. A., Kramer, L. D., Pledger, G. W., and Karim, R. M. (1996). Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. Topiramate yd study group. *Neurology*, 46(6):1684–1690.

Fears, T. R., Benichou, J., and Gail, M. H. (1996). A reminder of the fallibility of the Wald statistic. *The American Statistician*, 50(3):226–227.

Felgueiras, M., Martins, J., and Santos, R. (2012). Pseudo-convex mixtures. In *AIP Conference Proceedings (Vol. 1479, No. 1)*, pages 1125–1128. American Institute of Physics.

Gavrilov, L. A. and Gavrilova, N. S. (1991). *Biology of life span: A quantitative approach*. Taylor & Francis, Harwood.

Gavrilov, L. A. and Gavrilova, N. S. (2001). Biodemographic study of familial determinants of human longevity. *Population: An English Selection*, 13(1):197–221.

Gavrilova, N. S. and Gavrilov, L. A. (2001). When does human longevity start? Demarcation of the boundaries for human longevity. *Journal of Anti-Aging Medicine*, 4(2):115–124.

Gavrilova, N. S., Gavrilov, L. A., Evdokushkina, G. N., and Semyonova, V. G. (1998a). Mechanisms of familial transmission of human longevity: Comparison of maternal and paternal contributions into offspring lifespan. In *PAA Annual Meeting (Chicago)*, page 142.

Gavrilova, N. S., Gavrilov, L. A., Evdokushkina, G. N., and Semyonova, V. G. (1999). Demarcation of the boundaries for human longevity. Abstract. *The Gerontologist*, 39(1):146–147.

Gavrilova, N. S., Gavrilov, L. A., Evdokushkina, G. N., Semyonova, V. G., Gavrilova, A. L., Evdokushkina, N. N., Kushnareva, Y. E., Kroutko, V. N., and Andreyev, A. Y. (1998b). Evolution, mutations, and human longevity: European royal and noble families. *Human Biology*, 70(4):799–804.

Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283.

Geys, H., Molenberghs, G., and Ryan, L. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, 94(44):734–745.

Geys, H., Molenberghs, G., and Ryan, L. M. (1997). Pseudo-likelihood inference for clustered binary data. *Communications in Statistics–Theory and Methods*, 26(11):2743–2767.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78(1):45–51.

Griswold, M. E. and Zeger, S. L. (2004). On marginalized multilevel models and their computation. *Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 99*.

Gudmundsson, H., Gudbjartsson, D. F., Kong, A., Gudbjartsson, H., Frigge, M., Gulcher, J. R., and Stefánsson, K. (2000). Inheritance of human longevity in Iceland. *European Journal of Human Genetics*, 8(10):743–749.

Haselimashhadi, H., Vinciotti, V., and Yu, K. (2018). A novel Bayesian regression model for counts with an application to health data. *Journal of Applied Statistics*, 45(6):1085–1105.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, 55(3):688–698.

Heagerty, P. J. and Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Statistical Science*, 15(1):1–26.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Hinde, J. and Demétrio, C. G. B. (1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27(2):151–170.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396.

Huang, T., Peng, H., and Zhang, K. (2017). Model selection for Gaussian mixture models. *Statistica Sinica*, 27(1):147–169.

Iddi, S. and Molenberghs, G. (2012). A joint marginalized multilevel model for longitudinal outcomes. *Journal of Applied Statistics*, 39(11):2413–2430.

Iddi, S. and Molenberghs, G. (2013). A marginalized model for zero-inflated, overdispersed and correlated count data. *Electronic Journal of Applied Statistical Analysis*, 6(2):149–165.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate discrete distributions (Vol. 444)*. John Wiley & Sons.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(2):127–145.

Kassahun, W., Neyens, T., Molenberghs, G., Faes, C., and Verbeke, G. (2015). A joint model for hierarchical continuous and zero-inflated overdispersed count data. *Journal of Statistical Computation and Simulation*, 85(3):552–571.

Klakattawi, H. S., Vinciotti, V., and Yu, K. (2018). A simple and adaptive dispersion regression model for count data. *Entropy*, 20(2):142.

Knopp, K. (1990). *Theory and application of infinite series*. Courier Corporation.

Kokonendji, C. C., Dossou-Gbété, S., and Demétrio, C. G. B. (2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *SORT–Statistics and Operations Research Transactions*, 28(2):201–214.

Korpelainen, H. (1999). Genetic maternal effects on human life span through the inheritance of mitochondrial DNA. *Human Heredity*, 49(4):183–185.

Kulasekera, K. B. (1994). Approximate MLE's of the parameters of a discrete Weibull distribution with type I censored data. *Microelectronics Reliability*, 34(7):1185–1188.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15(3):209–225.

le Cessie, S. and Van Houwelingen, J. C. (1994). Logistic regression for correlated binary data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):95–108.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2018). *Generalized linear models with random effects: Unified analysis via H-likelihood (Vol. 153)*. CRC Press.

Lescrauwaet, A.-K., Debergh, H., Vincx, M., and Mees, J. (2010). Fishing in the past: Historical data on sea fisheries landings in Belgium. *Marine Policy*, 34(6):1279–1289.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1):3–24.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4):817–827.

Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Ribeiro Jr, E. E., Demétrio, C. G., and Hinde, J. (2019). A Weibull-count approach for handling under-and overdispersed longitudinal/clustered data structures. *Statistical Modelling*, 19(5):569–589.

Mardia, K. V. (1970). *Families of bivariate distributions*. Griffin's statistical monographs & courses. Griffin.

Matthijs, K., Van de Putte, B., and Vlietinck, R. (2002). The inheritance of longevity in a Flemish village (18th-20th century). *European Journal of Population*, 18(1):59–81.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall/CRC, London.

McLachlan, G. J. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, 89(426):633–644.

Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer, New York.

Molenberghs, G. and Verbeke, G. (2011). A note on a hierarchical interpretation for negative variance components. *Statistical Modelling*, 11(5):389–408.

Molenberghs, G., Verbeke, G., and Demétrio, C. G. B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, 13(4):513–531.

Molenberghs, G., Verbeke, G., Demétrio, C. G. B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugated random effects. *Statistical Science*, 25(3):325–347.

Molenberghs, G., Verbeke, G., Efendi, A., Braekers, R., and Demétrio, C. G. B. (2015). A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *Statistical Methods in Medical Research*, 24(4):434–452.

Morais, A. L. and Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions. *Computational Statistics & Data Analysis*, 55(3):1410–1425.

Myers, R. H., Montgomery, D. C., Vining, G. G., and Robinson, T. J. (2012). *Generalized linear models: With applications in engineering and the sciences (Vol. 791)*. John Wiley & Sons.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5):300–301.

Naylor, J. C. and Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3):214–225.

Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete Weibull distribution. *SORT–Statistics and Operations Research Transactions*, 39(1):127–146.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8(4):343–366.

O'Hagan, A., Murphy, T. B., and Gormley, I. C. (2012). Computational aspects of fitting mixture models via the expectation-maximization algorithm. *Computational Statistics & Data Analysis*, 56(12):3843–3864.

Oliveira, I. R. C., Molenberghs, G., Verbeke, G., Demétrio, C. G. B., and Dias, C. T. S. (2017). Negative variance components for non-negative hierarchical data with correlation, over-, and/or underdispersion. *Journal of Applied Statistics*, 44(6):1047–1063.

Palmgren, J. (1989). Regression models for bivariate binary responses. *UW Biostatistics Working Paper Series. Working Paper 101*.

Paul, S. R. and Plackett, R. L. (1978). Inference sensitivity for Poisson mixtures. *Biometrika*, 65(3):591–602.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60(310):516–522.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44(4):1033–1048.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, 47(3):825–839.

Pryseley, A., Tchonlafi, C., Verbeke, G., and Molenberghs, G. (2011). Estimating negative variance components from Gaussian and non-Gaussian data: A mixed models approach. *Computational Statistics & Data Analysis*, 55(2):1071–1085.

Puig, P. and Valero, J. (2006). Count data distributions: Some characterizations with applications. *Journal of the American Statistical Association*, 101(473):332–340.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21.

Rao, J. N. K. and Scott, A. J. (1987). On simple adjustments to Chi-square tests with sample survey data. *The Annals of Statistics*, 15(1):385–397.

Renard, D., Molenberghs, G., and Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis*, 44(4):649–667.

Roberts, G., Rao, J. N. K., and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74(1):1–12.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77(3):485–497.

Roy, D. (2003). The discrete normal distribution. *Communications in Statistics – Theory and Methods*, 32(10):1871–1883.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53(2):255–260.

Salaris, L., Tedesco, N., and Poulain, M. (2013). Familial transmission of human longevity: A population-based study in an inland village of Sardinia (Italy), 1850-2010. *Vienna Yearbook of Population Research*, 11:325–349.

Santos, R., Felgueiras, M., and Martins, J. P. (2016). Pseudo-convex mixtures generated by shape-extended stable distributions for extremes. *Journal of Statistical Theory and Practice*, 10(2):357–374.

Sato, H., Ikota, M., Sugimoto, A., and Masuda, H. (1999). A new defect distribution metrology with a consistent discrete exponential formula and its applications. *IEEE Transactions on Semiconductor Manufacturing*, 12(4):409–418.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–727.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879–885.

Sellers, K. F. and Raim, A. (2016). A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, 99:68–80.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 4(2):943–961.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142.

Sur, P., Shmueli, G., Bose, S., and Dubey, P. (2015). Modeling bimodal discrete data using Conway-Maxwell-Poisson mixture models. *Journal of Business & Economic Statistics*, 33(3):352–365.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46(3):657–671.

Tibaldi, F., Barbosa, F. T., and Molenberghs, G. (2004). Modelling associations between time-to-event responses in pilot cancer clinical trials using a Plackett-Dale model. *Statistics in Medicine*, 23(14):2173–2186.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Tuerlinckx, F., Rijmen, F., Verbeke, G., and De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, 59(2):225–255.

Van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., Beekman, M., Smith, K. R., and Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature Communications*, 10(1):1–12.

Vandresse, M. (2013). A household projection model for Belgium based on individual household membership rates. *Measuring Uncertainty in Population Forecasts: A New Approach.*, page 257.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New York.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, 59(2):254–262.

Warne, R. T. (2020). *Statistics for the social sciences: A general linear model approach*. Cambridge University Press.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61(3):439–447.

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of business & economic statistics*, 13(4):467–474.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48(3-4):233–243.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; A gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86.

Zeviani, W. M., Ribeiro Jr, P. J., Bonat, W. H., Shimakura, S. E., and Muniz, J. A. (2014). The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*, 41(12):2616–2626.

Zhang, B. and Zhang, C. (2005). Finite mixture models with negative components. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 31–41. Springer, Berlin, Heidelberg.

Zhang, M.-H. and Cheng, Q.-S. (2004). Determine the number of components in a mixture model by the extended KS test. *Pattern Recognition Letters*, 25(2):211–216.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77(3):642–648.

Zhu, R. and Joe, H. (2009). Modelling heavy-tailed count data using a generalised Poisson–inverse Gaussian family. *Statistics & Probability Letters*, 79(15):1695–1703.

Zou, Y., Geedipally, S. R., and Lord, D. (2013). Evaluating the double Poisson generalized linear model. *Accident Analysis & Prevention*, 59:497–505.

Supplementary Material for Chapter 2

S.1 Pseudo-likelihood Estimation for the Plackett-Dale Model

The pseudo-likelihood estimator $\widehat{\mathbf{\Phi}}$ is defined as the maximizer of Eq. (2.1). Consistency and asymptotic normality results (le Cessie and Van Houwelingen, 1994; Arnold and Strauss, 1991; Geys et al., 1999) imply that $\widehat{\mathbf{\Phi}}$ converges in probability to the true parameter value $\mathbf{\Phi}_0$, and $\sqrt{N}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0)$ converges in distribution to $N_p(\mathbf{0}, J(\mathbf{\Phi}_0)^{-1}K(\mathbf{\Phi}_0)J(\mathbf{\Phi}_0)^{-1})$ with $J(\mathbf{\Phi})$ a $p \times p$ matrix with rl-elements

$$J_{r\ell} \quad = \quad -\sum_{(s,t)\in S} E_{\mathbf{\Phi}}\left(\frac{\partial^2 \ln f_{T_s,T_t}(t_{is},t_{it})}{\partial\phi_r\partial\phi_\ell}\right), \tag{S.1}$$

and $K(\mathbf{\Phi})$ a $p \times p$ matrix with rl-elements

$$K_{r\ell} \quad = \quad \sum_{(s,t)\in S} E_{\mathbf{\Phi}}\left(\frac{\partial \ln f_{T_s}(t_{is},t_{it})}{\partial\phi_r}\cdot\frac{\partial \ln f_{T_t}(t_{is},t_{it})}{\partial\phi_\ell}\right). \tag{S.2}$$

Due to the presence of $\boldsymbol{\theta}$, the PD model allows us to estimate and interpret the strength of the association between a pair of survival times $\mathbf{T}_r$ and $\mathbf{T}_\ell$, through global cross ratios (the θ parameters in the model). The *global cross-ratio* $\theta_{r\ell}(\mathbf{t}_r, \mathbf{t}_\ell)$ can depend on time as well as on covariates. In the standard definition of Plackett, this is presumed constant for a given covariate combination. Generally, the cross-ratio is defined as:

$$\theta_{r\ell}(\mathbf{t}_r,\mathbf{t}_\ell) = \frac{F_{T_r,T_\ell}\cdot[1 - F_{T_r} - F_{T_\ell} + F_{T_r,T_\ell}]}{[F_{T_r} - F_{T_r,T_\ell}]\cdot[F_{T_\ell} - F_{T_r,T_\ell}]}. \tag{S.3}$$

$\theta = 1$ corresponds to independence and ranges from 0 to $+\infty$. The log odds ratio is often utilized to avoid range restrictions and for which, due to symmetry, standard errors are more readily interpretable. All plots will be based on the log odds ratio. Nevertheless, it is often easier to work with a transformation of θ such as Spearman's ρ or Kendall's τ. There exists a relationship between Kendall's τ and θ for any copula $C(\mathbf{t}_r, \mathbf{t}_\ell, \theta)$ (Genest and MacKay, 1986):

$$\tau(\theta) = 4 \int_0^1 \int_0^1 C_{T_r,T_\ell}(\mathbf{t}_r, \mathbf{t}_\ell, \theta) C_{T_r,T_\ell}(d\mathbf{t}_r, d\mathbf{t}_\ell, \theta) - 1. \tag{S.4}$$

This relationship is independent of the marginal distributions and only depends on the copula function C_{T_r,T_ℓ} (Schweizer and Wolff, 1981). Estimates and confidence intervals, using the delta method, are accordingly easily obtained. There is no closed-form for Kendall's τ in the PD case and an estimate has to be obtained directly from Eq. (S.4). We have developed a SAS IML 9.14 macro to this effect. Nelsen (2007) provided more details on copulas and the relationships between association measures. All macros pertaining to the model and test statistics are available from the first author upon request.

The relationship between Spearman's ρ and θ is (Tibaldi et al., 2004):

$$\rho(\theta) = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \cdot \ln \theta}{(\theta - 1)^2}. \tag{S.5}$$

An estimate follows from $\widehat{\rho} = \rho(\widehat{\theta})$, with variance estimated using the delta method. Figure 1 graphically displays the relationships between these three quantities (θ, ρ, and τ). The (ρ, τ) plot shows an almost linear relationship. Depending on the context, one can choose any of these three quantities to study association. By comparing Eq. (S.4) and Eq. (S.5), we observe that the computation of τ is more complex given that it involves numerical integration. In contrast, ρ is very easy to obtain by plugging the estimated value of θ into Eq. (S.5).

S.2 Test Statistics for the Plackett-Dale Model

This section summarizes work done by Geys et al. (1997).

The essential difference is that the pseudo-likelihood versions of the Wald, score, or likelihood ratio statistics asymptotically do not follow a χ^2 distribution. These authors established appropriate asymptotic results. We will present the tests sufficiently generally, so that they apply to any subvector $\boldsymbol{\phi}$ of the parameter vector. For example, $\theta_{r\ell} = 1$ corresponds to independence between T_{ir} and $T_{i\ell}$ and can be written as $\mathrm{H}_0 : \theta_{r\ell} - 1 = 0, 1 \le r < \ell \le 3$.

Assume we are interested in a hypothesis of the type $\mathrm{H}_0 : \boldsymbol{\varphi} = \boldsymbol{\varphi}_0$ where $\boldsymbol{\varphi}$ denotes a q-dimensional subvector of the p-dimensional vector of regression parameters $\boldsymbol{\Phi}$ and write $\boldsymbol{\Phi} = (\boldsymbol{\varphi}', \boldsymbol{\zeta}')'$. The tests will be discussed in turn.

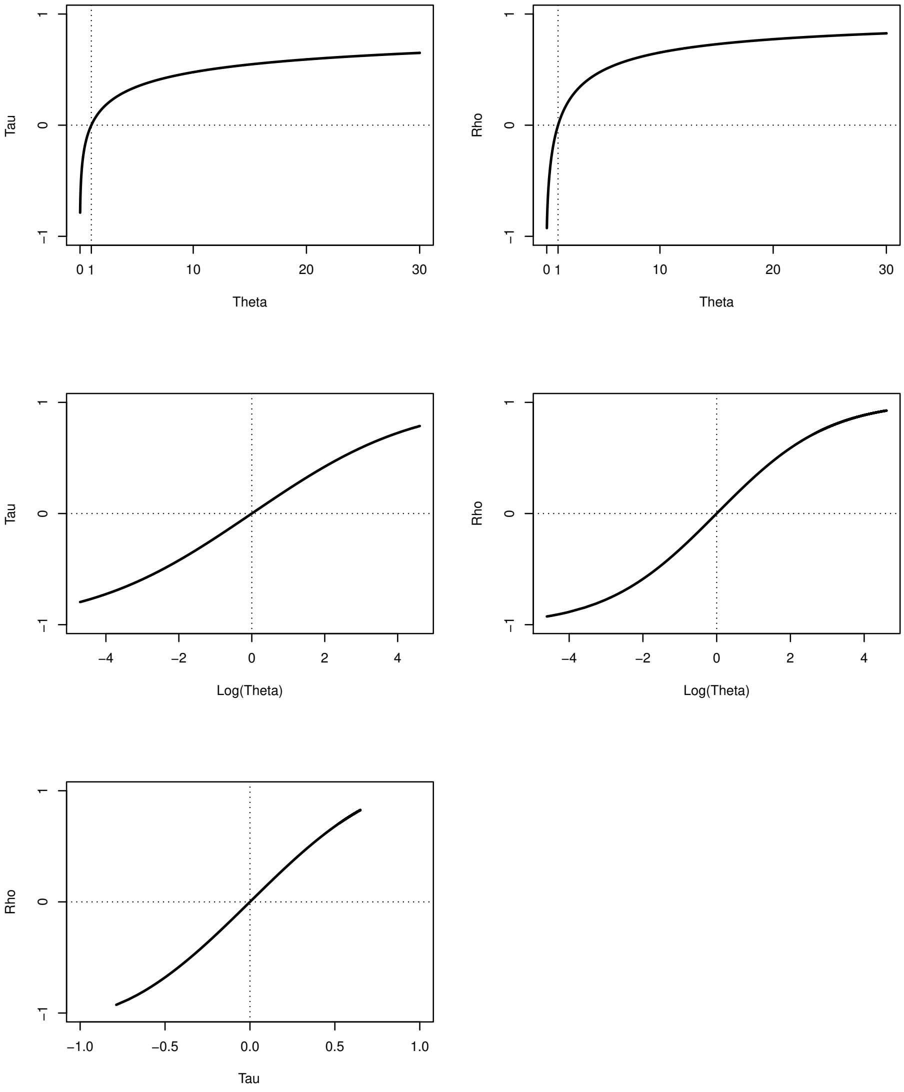


Figure 1: Relationship between θ, $\log(\theta)$, τ, and ρ plotted in pairs.

S.2.1 Wald Test Statistics

The asymptotic normality properties of the pseudo-likelihood estimators will be utilized to construct this test. We use the next result:

$$W^* = N(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)'\Sigma^{-1}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) \sim \chi^2_q. \tag{S.6}$$

$\Sigma_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ denotes the $q \times q$ submatrix of $\Sigma = J^{-1}KJ$. The matrices J and K were defined according to Eq. (S.1) and Eq. (S.2). The matrix Σ can be estimated by using the pseudo-likelihood estimate $\widehat{\boldsymbol{\Phi}}$. The Wald statistic is very easy to obtain and the more convenient one in cases where model fitting is very time consuming. Fears et al. (1996) proved that it is highly sensitive to changes in parameterization. For example, we can see, through the delta method, that the value of the Wald statistic obtained for the hypothesis H$_0$: $\boldsymbol{\varphi} = \mathbf{0}$ doubles the one corresponding to H$_0$: $\boldsymbol{\varphi}^2 = \mathbf{0}$, where $\boldsymbol{\varphi}^2$ refer to the element-wise squared terms of $\boldsymbol{\varphi}$. When interest lies in $[\theta^2]^\alpha$ with $\alpha > 0$, for example, the Wald test statistics is α^{-1} times the Wald statistic for θ^2 (Fears et al., 1996). Corresponding p-values also depend on the choice of α.

In this study, the fact that individual association parameters will be tested implies $\varphi_0 = \theta_{r\ell}$ and $q = 1$. $W^* \sim \chi^2_1$ and the normal distribution on the square root can be used to produce p-values.

S.2.2 Pseudo-score Test Statistics

This test is constructed by fitting the null model and it has the advantage over the Wald test that it is invariant to reparameterization. Let us call $\boldsymbol{U}(\boldsymbol{\Phi})$ the pseudo-score vector, specifically the derivative of the log of the pseudo-likelihood; and $\boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\Phi})$ the q-dimensional subvector. An empirically corrected (e.c.) version of this pseudo-score can be defined as

$$S^*(e.c.) = \frac{1}{N}\left[\boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_0, \hat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0))' J^{\boldsymbol{\varphi}\boldsymbol{\varphi}} \Sigma^{-1}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} J^{\boldsymbol{\varphi}\boldsymbol{\varphi}} \boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_0, \hat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0))\right], \tag{S.7}$$

where $\hat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0)$ presents the maximum pseudo-likelihood estimator of $\boldsymbol{\zeta}$ for $\boldsymbol{\varphi}$ equal to $\boldsymbol{\varphi}_0$, $J^{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ denotes the $q \times q$ submatrix of the inverse of J, and $J^{\boldsymbol{\varphi}\boldsymbol{\varphi}}\Sigma^{-1}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}J^{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ is evaluated under the null hypothesis. Under mild regularity conditions, $S^*(e.c.) \sim \chi^2_q$. Rotnitzky and Jewell (1990) observed computational problems in the framework of generalized estimating equations. An alternative model-based variant was suggested:

$$S^*(m.b.) = \frac{1}{N}\left[\boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_0, \hat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0))' J^{\boldsymbol{\varphi}\boldsymbol{\varphi}} \boldsymbol{U}_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}_0, \hat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0))\right]. \tag{S.8}$$

Its asymptotic distribution under H_0 is given by $\sum_{j=1}^{q} \eta_j \chi^2_{1(j)}$ where $\chi^2_{1(j)}$ are all independent random variables with χ^2_1 distribution and $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_q$ are the eigenvalues of $(J^{\boldsymbol{\varphi}\boldsymbol{\varphi}})^{-1}\Sigma_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ under H_0.

To simplify calculations and to have a χ^2_q distribution, an adjusted pseudo-score statistic is proposed, similar to Rotnitzky and Jewell (1990):

$$S^*_a(m.b.) = S^*(m.b.)/\bar{\eta}, \quad \bar{\eta} = \sum_{j=1}^{q} \eta_j/q. \tag{S.9}$$

Rao and Scott (1987) and Roberts et al. (1987) have proposed several adjustments. One interesting feature of all tests is that, in the maximum likelihood context, all eigenvalues are equal to one and therefore all three statistics coincide. In our scalar case, $S^*(m.b.) = S^*_a(m.b.)$ holds because $q = 1$.

S.2.3 Pseudo-likelihood Ratio Test Statistics

Another proposal for testing H_0 is based on likelihood ratio ideas:

$$G^{*2} = 2[pl(\widehat{\boldsymbol{\Phi}}) - pl(\boldsymbol{\varphi}_0, \widehat{\boldsymbol{\zeta}}(\boldsymbol{\varphi}_0))] \tag{S.10}$$

and is termed pseudo-likelihood ratio test statistic. The asymptotic distribution of G^* can be written as $\sum_{j=1}^{q} \eta_j \chi^2_{1(j)}$, with $\chi^2_{1(j)}$ independently distributed according to χ^2_1 and $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_r$ the eigenvalues of $(J^{\boldsymbol{\varphi}\boldsymbol{\varphi}})^{-1}\Sigma_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$ under H_0, similar as before.

Similarly, Geys et al. (1997) defined a modified pseudo-score statistic:

$$G^{*2}_a = G^{*2}/\bar{\eta}, \tag{S.11}$$

that can be approximated by χ^2_q. It can be proven that the formulation of G^{*2} is an approximation to a Wald test. The pseudo-likelihood ratio test needs more time to compute, since the model ought to be fitted twice, for the reduced and full models. Furthermore, it is generally known from pseudo-likelihood theory that the Wald test embraces the lowest power. From a practical perspective, however, preference is often given to the Wald test. Implementation of all test statistics is carried out within the SAS IML procedure.

S.3 SAS Code for Weibull-gamma-normal Model

```
/****************************
SOFTWARE: SAS 9.4.
OBJECTIVE: Analyzing Moerzeke data with the Weibull-gamma-normal
           framework;
DATASET: Moerzeke data, containing information about 457 families;
VARIABLE DESCRIPTION:
- ID: Family ID;
- FamilyMember: Family member indicator, i.e., F = father,
        M = mother, C = first born child;
- Sexnum: Binary indicator of the gender of first born child, i.e.,
        1 = boy, 0 = girl;
- y: Discretized life expectancy of a household member;
AUTHOR: M. Luyts (L-Biostat);
****************************/
libname m 'C:\Users\u0106491\Desktop\Moerzeke data';

/* Weibull model, via NLMIXED */
proc nlmixed data=MoerzekeFinal_trunc tech=newrap maxit=1000;
  parms Beta_11=-0.113 Beta_12=-0.113 Beta_13=-0.113 Beta_21=-1.067
        Beta_22=-0.899 Beta_23=-3.800 rho1=4.799 rho2=5.770
        rho3=3.002 L1=-1.51 L2=-1.69 L3=0.457;
  if FamilyMember='M' then rho=rho1;
  else if FamilyMember='F' then rho=rho2;
  else rho=rho3;
  if FamilyMember='M' then lambda=exp(L1);
  else if FamilyMember='F' then lambda=exp(L2);
  else lambda=exp(L3);
  if FamilyMember='M' then eta=Beta_11*SexNum + Beta_21*YearOfBirth;
  else if FamilyMember='F' then eta=Beta_12*SexNum + Beta_22
        *YearOfBirth;
  else eta=Beta_13*SexNum + Beta_23*YearOfBirth;
  expeta=exp(eta);
  if FamilyMember='M' then ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)
        +eta-lambda*(Age_t**rho)*expeta;
  else if FamilyMember='F' then ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)
        +eta-lambda*(Age_t**rho)*expeta;
  else ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)+eta-lambda*(Age_t**rho)
```

```
            *expeta;
  model Age_t ~ general(ll);
  estimate "lambda1" exp(L1);
  estimate "lambda2" exp(L2);
  estimate "lambda2" exp(L3);
run;

/* Weibull-gamma model, via NLMIXED */
proc nlmixed data=MoerzekeFinal_trunc tech=newrap maxit=1000;
  parms Beta_11=-0.113 Beta_12=-0.113 Beta_13=-0.113 Beta_21=-1.067
            Beta_22=-0.899 Beta_23=-3.800 rho1=4.799 rho2=5.770
            rho3=3.002 L1=0.21 L2=0.185 L3=1.58 alpha=459;
  lambda1=exp(L1);
  lambda2=exp(L2);
  lambda3=exp(L3);
  if FamilyMember='M' then eta=Beta_11*SexNum + Beta_21*YearOfBirth;
  else if FamilyMember='F' then eta=Beta_12*SexNum + Beta_22*YearOfBirth;
  else eta=Beta_13*SexNum + Beta_23*YearOfBirth;
  expeta=exp(eta);
  if FamilyMember='M' then ll=log(lambda1)+log(rho1)+(alpha+1)*log(alpha)
            +(rho1-1)*log(Age_t)+eta-(alpha+1)*log(lambda1*(Age_t**rho1)
            *expeta+alpha);
  else if FamilyMember='F' then ll=log(lambda2)+log(rho2)+(alpha+1)
            *log(alpha)+(rho2-1)*log(Age_t)+eta-(alpha+1)*log(lambda2
            *(Age_t**rho2)*expeta+alpha);
  else ll=log(lambda3)+log(rho3)+(alpha+1)*log(alpha)+(rho3-1)*log(Age_t)
            +eta-(alpha+1)*log(lambda3*(Age_t**rho3)*expeta+alpha);
  model Age_t ~ general(ll);
  estimate "lambda1" exp(L1);
  estimate "lambda2" exp(L2);
  estimate "lambda2" exp(L3);
run;

/* Weibull-normal model, via NLMIXED */
proc nlmixed data=MoerzekeFinal_trunc qpoints=50;
  parms Beta_11=-0.113 Beta_12=-0.113 Beta_13=-0.113 Beta_21=-1.067
            Beta_22=-0.899 Beta_23=-3.800 rho1=4.799 rho2=5.770
            rho3=3.002 L1=0.21 L2=0.185 L3=1.58 sigma=1;
  if FamilyMember='M' then rho=rho1;
  else if FamilyMember='F' then rho=rho2;
  else rho=rho3;
```

```
  if FamilyMember='M' then lambda=exp(L1);
  else if FamilyMember='F' then lambda=exp(L2);
  else lambda=exp(L3);
  if FamilyMember='M' then eta=Beta_11*SexNum + Beta_21*YearOfBirth
          + b1;
  else if FamilyMember='F' then eta=Beta_12*SexNum + Beta_22*YearOfBirth
          + b1;
  else eta=Beta_13*SexNum + Beta_23*YearOfBirth + b1;
  expeta=exp(eta);
  if FamilyMember='M' then ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)+eta
          -lambda*(Age_t**rho)*expeta;
  else if FamilyMember='F' then ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)
          +eta-lambda*(Age_t**rho)*expeta;
  else ll=log(lambda)+log(rho)+(rho-1)*log(Age_t)+eta-lambda*(Age_t**rho)
          *expeta;
  model Age_t ~ general(ll);
  random b1 ~ normal(0, sigma**2) subject=ID;
  estimate "lambda1" exp(L1);
  estimate "lambda2" exp(L2);
  estimate "lambda2" exp(L3);
run;

/* Weibull-gamma-normal model, via NLMIXED*/
proc nlmixed data=MoerzekeFinal_trunc tech=quanew qpoints=50 maxit=1000;
  parms Beta_11=-0.113 Beta_12=-0.113 Beta_13=-0.113 Beta_21=-1.067
          Beta_22=-0.899 Beta_23=-3.800 rho1=4.799 rho2=5.770
          rho3=3.002 L1=0.21 L2=0.185 L3=1.58 alpha=459 sigma=1;
  if FamilyMember='M' then rho=rho1;
  else if FamilyMember='F' then rho=rho2;
  else rho=rho3;
  if FamilyMember='M' then lambda=exp(L1);
  else if FamilyMember='F' then lambda=exp(L2);
  else lambda=exp(L3);
  if FamilyMember='M' then eta=Beta_11*SexNum + Beta_21*YearOfBirth
          + b1;
  else if FamilyMember='F' then eta=Beta_12*SexNum + Beta_22*YearOfBirth
          + b1;
  else eta=Beta_13*SexNum + Beta_23*YearOfBirth + b1;
  expeta=exp(eta);
  if FamilyMember='M' then ll=log(lambda)+log(rho)+(alpha+1)*log(alpha)
          +(rho-1)*log(Age_t)+eta-(alpha+1)*log(lambda*(Age_t**rho)
```

```
          *expeta+alpha);
  else if FamilyMember='F' then ll=log(lambda)+log(rho)+(alpha+1)
          *log(alpha)+(rho-1)*log(Age_t)+eta-(alpha+1)*log(lambda
          *(Age_t**rho)*expeta+alpha);
  else ll=log(lambda)+log(rho)+(alpha+1)*log(alpha)+(rho-1)*log(Age_t)+eta
          -(alpha+1)*log(lambda*(Age_t**rho)*expeta+alpha);
  model Age_t ~ general(ll);
  random b1 ~ normal(0, sigma**2) subject=ID;
  estimate "lambda1" exp(L1);
  estimate "lambda2" exp(L2);
  estimate "lambda2" exp(L3);
run;

/* Marginal Weibull-gamma-normal model, via NLMIXED*/
proc nlmixed data=MoerzekeFinal_trunc tech=quanew qpoints=50 maxit=1000;
  bounds alpha>0;
  parms Beta_11=-0.113 Beta_12=-0.113 Beta_13=-0.113 Beta_21=-1.067
          Beta_22=-0.899 Beta_23=-3.800 rho1=4.799 rho2=5.770
          rho3=3.002 L1=0.21 L2=0.185 L3=1.58 alpha=459 sigma=1;
  if FamilyMember='M' then rho=rho1;
  else if FamilyMember='F' then rho=rho2;
  else rho=rho3;
  if FamilyMember='M' then lambda=exp(L1);
  else if FamilyMember='F' then lambda=exp(L2);
  else lambda=exp(L3);
  if FamilyMember='M' then eta=Beta_11*SexNum + Beta_21*YearOfBirth
          + b1 - sigma/2;
  else if FamilyMember='F' then eta=Beta_12*SexNum + Beta_22*YearOfBirth
          + b1 - sigma/2;
  else eta=Beta_13*SexNum + Beta_23*YearOfBirth + b1 - sigma/2;
  expeta=exp(eta);
  if FamilyMember='M' then ll=log(lambda)+log(rho)+(alpha+1)*log(alpha)
          +(rho-1)*log(Age_t)+eta-(alpha+1)*log(lambda*(Age_t**rho)
          *expeta+alpha);
  else if FamilyMember='F' then ll=log(lambda)+log(rho)+(alpha+1)
          *log(alpha)+(rho-1)*log(Age_t)+eta-(alpha+1)*log(lambda
          *(Age_t**rho)*expeta+alpha);
  else ll=log(lambda)+log(rho)+(alpha+1)*log(alpha)+(rho-1)*log(Age_t)
          +eta-(alpha+1)*log(lambda*(Age_t**rho)*expeta+alpha);
  model Age_t ~ general(ll);
  random b1 ~ normal(0, sigma**2) subject=ID;
```

```
  estimate "lambda1" exp(L1);
  estimate "lambda2" exp(L2);
  estimate "lambda2" exp(L3);
run;
```

Supplementary Material for Chapter 3

S.4 General Overview of Over- and/or Underdispersed Count Models

Due to the restricted mean-variance relationship of the Poisson log-linear GLM models (i.e., $\mathsf{E}(Y_i) = \mathsf{Var}(Y_i) \equiv \lambda_i$, where $Y_i \in \mathbb{N}$, $i = 1, \ldots, n$, is Poisson distributed with parameter $\lambda_i \in \mathbb{R}_{\geq 0}$, the set of positive real numbers, and with $\ln(\lambda_i) = \mathbf{x}_i^{'} \cdot \boldsymbol{\beta}$, where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ denote a p-dimensional vector of covariates for ith observation and the associated parameter vector, respectively), various extensions have been proposed in the literature (Breslow, 1984, Lawless, 1987, Hinde and Demétrio, 1998) - and in what follows, some popular models will be examined in more detail.

S.4.1 Quasi-Poisson Model

A straightforward modification, in the context of the exponential family, is to allow the dispersion (scale) parameter, denoted by δ, to not be restricted to 1. This leads us to $\mathsf{Var}(Y_i) = \delta \cdot \mathsf{E}(Y_i)$, where $\delta > 1$ and $\delta < 1$ indicates over- and underdispersion, respectively. This results in the so-called quasi-Poisson model (Wedderburn, 1974), where the point estimates of $\boldsymbol{\beta}$ are identical to those of the standard Poisson model, but standard errors are scaled by $\sqrt{\delta}$ resulting in possible differences in inferences on covariates compared to the standard Poisson model.

S.4.2 Negative Binomial Model

Another elegant way to provide flexibility is through a two-stage model. A popular approach in this context is to assume that $Y_i \mid \lambda_i \sim \mathsf{Poi}(\lambda_i)$ and that the parameter λ_i is itself a random variable with mean μ_i and variance σ_i^2. By using standard results on iterated expectations we have:

$$\mathsf{E}(Y_i) = \mu_i, \qquad \mathsf{Var}(Y_i) = \mu_i + \sigma_i^2. \tag{S.12}$$

A popular specific distributional choice is $\lambda_i \sim \text{Gamma}(\alpha, \alpha^{-1})$, for reasons of identiability (Duchateau and Janssen, 2007), leading to the negative binomial (NB) model. Choosing the Gamma distribution has the advantage of (1) satisfying the mean's scale for count outcomes and (2) obtaining closed-forms for the marginal mean and variance, and even for the entire marginal distribution (Molenberghs et al., 2007). The corresponding (marginal) probability mass function, mean and variance of the model are equal to

$$\begin{aligned} P(Y_i = y_i \mid \mathbf{x}_i) &= \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y}, \\ \text{with } \lambda_i = \text{e}^{\mathbf{x}_i' \cdot \beta}, &\qquad \text{E}(Y_i) = \lambda_i, \qquad \text{Var}(Y_i) = \lambda_i + \alpha \cdot \lambda_i^2. \end{aligned} \tag{S.13}$$

respectively (Lawless, 1987, Cameron and Trivedi, 1986). We should note that, from a hierarchical/conditional viewpoint, only overdispersion can be examined (since for a valid Gamma distribution $\alpha > 0$).

S.4.3 Conway-Maxwell-Poisson Model

The Conway-Maxwell-Poisson (COM) model, first introduced by Conway and Maxwell (1962), is suitable for analyzing count data that exhibit either over- or underdispersion. Even though its existence has been known for several decades, most research on this model was done during the last decade. Shmueli et al. (2005), for example, investigated the statistical properties of the COM distribution. While in a Bayesian context, Boatwright et al. (2006) developed the conjugate distributions for the parameters of the COM distribution. The probability mass function of the model can be expressed as

$$\begin{aligned} P(Y_i = y_i \mid \mathbf{x}_i) &= \frac{1}{Z(\lambda_i, \tau)} \cdot \frac{\lambda_i^{y_i}}{(y_i!)^{\tau}}, \\ \text{with } \lambda_i = \text{e}^{\mathbf{x}_i' \cdot \boldsymbol{\beta}}, &\qquad Z(\lambda_i, \tau) = \sum_{n=0}^{+\infty} \frac{\lambda_i^n}{(n!)^{\tau}}. \end{aligned} \tag{S.14}$$

The domain of admissible parameters for which the probability mass function above defines a probability distribution is $(\lambda_i, \tau) > 0$, and $0 < \lambda_i < 1, \tau = 0$. Some well-known discrete data models result from this. When τ equals 1, it reduces to the standard Poisson model. When $\tau \to +\infty$, the COM model approaches a Bernoulli model with success parameter $\pi_i = \frac{\lambda_i}{1+\lambda_i}$. While if $\tau = 0$ and $\lambda_i < 1$, the geometric model with success probability $1 - \lambda_i$ is obtained. In terms of dispersion, specific focus is put on the mean and variance functions and the nature of the different dispersion regions can be found in Appendix S.8.2.

The mean and variance can be approximated by

$$\mathsf{E}(Y_i) = \lambda_i \frac{\partial \mathsf{log} Z(\lambda_i, \tau)}{\partial \lambda_i} \approx \lambda_i^{1/\tau} - \frac{\tau - 1}{2 \cdot \tau}, \;\; \mathsf{Var}(Y_i) = \frac{\partial \mathsf{E}(Y_i)}{\partial \mathsf{log} \lambda_i} \approx \frac{1}{\tau} \lambda_i^{1/\tau}. \quad \text{(S.15)}$$

S.4.4 Double Poisson Model

The double Poisson (DP) model, based on the double-exponential family of Efron (1986), has hardly been investigated and applied since its first introduction three decades ago. Winkelmann (2008) and Hilbe (2011) indicated that the normalizing constant is the bottleneck in applying the DP by showing that fitted models with its normalizing constant approximated by Efron's original method are not exact. For these and other reasons, different approximations have been proposed in the literature. A full discussion can be found in Zou et al. (2013).

The probability mass function of the DP model can be written as

$$\begin{aligned} P(Y_i = y_i \mid \mathbf{x}_i) &= K(\lambda_i, \phi) \cdot \phi^{1/2} \cdot \mathsf{e}^{-\phi \cdot \lambda_i} \cdot \frac{\mathsf{e}^{-y_i} \cdot y_i^{y_i}}{y_i!} \cdot \left(\frac{\mathsf{e} \cdot \lambda_i}{y_i} \right)^{\phi \cdot y_i}, \\ \text{with } \lambda_i = \mathsf{e}^{\mathbf{x}_i' \cdot \boldsymbol{\beta}}, \quad & \frac{1}{K(\lambda_i, \phi)} \approx 1 + \frac{1 - \phi}{12 \cdot \phi \cdot \lambda_i} \cdot \left(1 + \frac{1}{\phi \cdot \lambda_i} \right). \end{aligned} \quad \text{(S.16)}$$

where $K(\lambda_i, \phi)$ is the normalizing constant that is often close to 1. The corresponding mean and variance can be approximated by

$$\mathsf{E}(Y_i) \approx \lambda_i, \qquad \mathsf{Var}(Y_i) \approx \frac{\lambda_i}{\phi}. \quad \text{(S.17)}$$

Thus, the DP model allows for both overdispersion ($\phi < 1$) and underdispersion ($\phi > 1$). While for $\phi = 1$, the standard Poisson model results.

S.5 Proof of Dispersion for Discrete Exponential Case

Theorem 2. *Let Y_i, $i = 1, \ldots, n$, be (type 1) Discrete exponential distributed, with $\lambda > 0$. Then, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$).*

Proof. To prove it, three situations are examined:

1. $\underline{\mathrm{E}(Y_i) > \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1 - \mathrm{e}^{-\lambda})} > \frac{\mathrm{e}^{-\lambda}}{(1 - \mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} < 0 \qquad \Rightarrow \perp$$

2. $\underline{\mathrm{E}(Y_i) = \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})} = \frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} = 0 \qquad \Rightarrow \lambda \to +\infty$$

3. $\underline{\mathrm{E}(Y_i) < \mathrm{Var}(Y_i)?}$

$$\frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})} < \frac{\mathrm{e}^{-\lambda}}{(1-\mathrm{e}^{-\lambda})^2} \Leftrightarrow \mathrm{e}^{-\lambda} > 0 \qquad \Rightarrow \surd$$

Thus, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$)!

□

S.6 Proof of Mean and Variance Convergence for Discrete Weibull Case

Lemma 3 (d'Alembert's ratio test). *Let $\sum_{n=0}^{+\infty} a_n$ be an infinite serie, and consider*

$$L = \lim_{n\to+\infty} \left| \frac{a_{n+1}}{a_n} \right| .$$

1. *If $L < 1$, then the series converges absolutely;*
2. *If $L > 1$, then the series diverges;*
3. *If $L = 1$ or the limit fails to exist, then the test is inconclusive.*

Lemma 4 (Raabe–Duhamel's test). *Let $a_n > 0$ $(\forall n)$. Define*

$$b_n = n \cdot \left(\frac{a_n}{a_{n+1}} - 1 \right) .$$

If $L = \lim_{n\to+\infty} b_n$ exists, there are three possibilities:

1. *If $L > 1$, then the series $\sum_{n=0}^{+\infty} a_n$ converges;*
2. *If $L < 1$, then the series $\sum_{n=0}^{+\infty} a_n$ diverges;*
3. *If $L = 1$, then the test is inconclusive.*

Theorem 5. *Let Y_i, $i = 1, \ldots, n$, be (type 1) DW distributed. Then, it can* *be shown that*

$$\mathrm{E}(Y_i)\left(=\mu=\sum_{n=1}^{+\infty} q^{n^\rho}\right)<+\infty,$$
$$\mathrm{Var}(Y_i)\left(=2\cdot\sum_{n=1}^{+\infty} n\cdot q^{n^\rho}-\mu-\mu^2\right)<+\infty.$$

Proof. A trivial proof can be conducted for $\rho \geq 1$, since $\sum_{n=0}^{+\infty} q^{n^\rho} \leq \sum_{n=0}^{+\infty} q^n = (1-q)^{-1}$, $\sum_{n=0}^{+\infty} n\cdot q^{n^\rho} \leq \sum_{n=0}^{+\infty} n\cdot q^n$, and the series $\sum_{n=0}^{+\infty} n\cdot q^n$ converges. Indeed, using Lemma 3 with $a_n = n\cdot q^n$, it can easily be shown that $L<1$. For $\rho<1$, Lemma 4 can be used, where $a_n = n\cdot[q^{n^\rho}-q^{(n+1)^\rho}]$ and $a_n = n^2\cdot[q^{n^\rho}-q^{(n+1)^\rho}]$ are proper choices for $\mathrm{E}(Y_i)$ and $\mathrm{Var}(Y_i)$, respectively. □

Additionally, based on the integral test (Knopp, 1990) and assuming $q=e^{-\lambda}$, the following lower and upper boundaries can be obtained for the mean and variance expression ($\forall t \in 1,2,\ldots$):

$$\begin{aligned}
\mathsf{E}(Y_i) &\in \left[\sum_{n=1}^{t} q^{n^\rho} + \int_{t+1}^{+\infty} q^{n^\rho}\cdot dn; \sum_{n=1}^{t} q^{n^\rho} + \int_{t}^{+\infty} q^{n^\rho}\cdot dn\right] \\
&\in \left[\sum_{n=1}^{t} e^{-\lambda\cdot n^\rho} + \frac{1}{\rho\cdot\lambda^{1/\rho}}\cdot\Gamma[1/\rho;\lambda\cdot(t+1)^\rho];\right. \\
&\qquad \left.\sum_{n=1}^{t} e^{-\lambda\cdot n^\rho} + \frac{1}{\rho\cdot\lambda^{1/\rho}}\cdot\Gamma(1/\rho;\lambda\cdot t^\rho)\right],
\end{aligned}$$

$$\begin{aligned}
\mathsf{Var}(Y_i) \in \Bigg[& 2\sum_{n=1}^{t} nq^{n^\rho} - \sum_{n=1}^{t} q^{n^\rho} - \left(\sum_{n=1}^{t} q^{n^\rho}\right)^2 - \frac{2}{\rho\cdot\lambda^{1/\rho}}\Gamma(1/\rho;\lambda\cdot t^\rho)\sum_{n=1}^{t} q^{n^\rho} \\
& -\frac{1}{\rho^2\cdot\lambda^{2/\rho}}\left[\Gamma(1/\rho;\lambda\cdot t^\rho)\right]^2 - \frac{1}{\rho\cdot\lambda^{1/\rho}}\Gamma(1/\rho;\lambda\cdot t^\rho) + \frac{2}{\rho\cdot\lambda^{1/\rho}} \\
& \cdot\Gamma[2/\rho;\lambda\cdot(t+1)^\rho]; 2\sum_{n=1}^{t} nq^{n^\rho} - \sum_{n=1}^{t} q^{n^\rho} - \left(\sum_{n=1}^{t} q^{n^\rho}\right)^2 - \frac{2}{\rho\cdot\lambda^{1/\rho}} \\
& \cdot\Gamma[1/\rho;\lambda\cdot(t+1)^\rho]\cdot\sum_{n=1}^{t} q^{n^\rho} - \frac{1}{\rho^2\cdot\lambda^{2/\rho}}\cdot\{\Gamma[1/\rho;\lambda\cdot(t+1)^\rho]\}^2 \\
& -\frac{1}{\rho\cdot\lambda^{1/\rho}}\cdot\Gamma[1/\rho;\lambda\cdot(t+1)^\rho] + \frac{2}{\rho\cdot\lambda^{1/\rho}}\cdot\Gamma(2/\rho;\lambda\cdot t^\rho)\Bigg].
\end{aligned}$$

S.7 SAS Code for Hierarchical Discrete Exponential and Discrete Weibull Model

```
/****************************
SOFTWARE: SAS 9.4.
OBJECTIVE: Analyzing Moerzeke data with the DE and DW approach;
DATASET: Moerzeke data, containing information about 457 families;
VARIABLE DESCRIPTION:
- ID: Family ID;
- FamilyMember: Family member indicator, i.e., F = father,
          M = mother, C = first born child;
- Sexnum: Binary indicator of the gender of first born child, i.e.,
          1 = boy, 0 = girl;
- y: Discretized life expectancy of a household member;
AUTHOR: M. Luyts (L-Biostat);
****************************/
libname m 'C:\Users\u0106491\Desktop\Moerzeke data';

/* Model from exponential case, via NLMIXED */
proc nlmixed data=m.MoerzekeFinal;
  parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
          beta1X=-0.018 beta1XX=-0.018 sigma=1;
  if FamilyMember='F' then eta=beta0XX + beta1XX*SexNum + u;
  else if FamilyMember='M' then eta=beta0X + beta1X*SexNum + u;
  else eta=beta0 + beta1*SexNum + u;
  expeta=exp(eta);
  ll=eta*y - log(expeta + 1)*y + log(1 - (expeta/(expeta+1)));
  model y ~ general(ll);
  random u ~ normal(0, exp(sigma)**2) subject=id;
  estimate 'random intercept' exp(sigma);
run;

/* Model from Weibull case, via NLMIXED */
proc nlmixed data=m.MoerzekeFinal;
  parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
          beta1X=-0.018 beta1XX=-0.018 sigma=1 rho=1;
  if FamilyMember='F' then eta=beta0XX + beta1XX*SexNum + u;
  else if FamilyMember='M' then eta=beta0X + beta1X*SexNum + u;
  else eta=beta0 + beta1*SexNum + u;
  lambda=log(exp(eta)+1) - log(exp(eta));
```

```
   if y=0 then prob=1 - exp(-1*lambda);
   else prob=exp(-1*lambda*(y**rho))-exp(-1*lambda*((y+1)**rho));
   ll=log(prob);
   model y ~ general(ll);
   random u ~ normal(0, exp(sigma)**2) subject=id;
   estimate 'random intercept' exp(sigma);
run;
```

S.8 Characteristic Indices for the Negative Binomial, Conway-Maxwell-Poisson, and Double Poisson Model

S.8.1 Negative Binomial Distribution

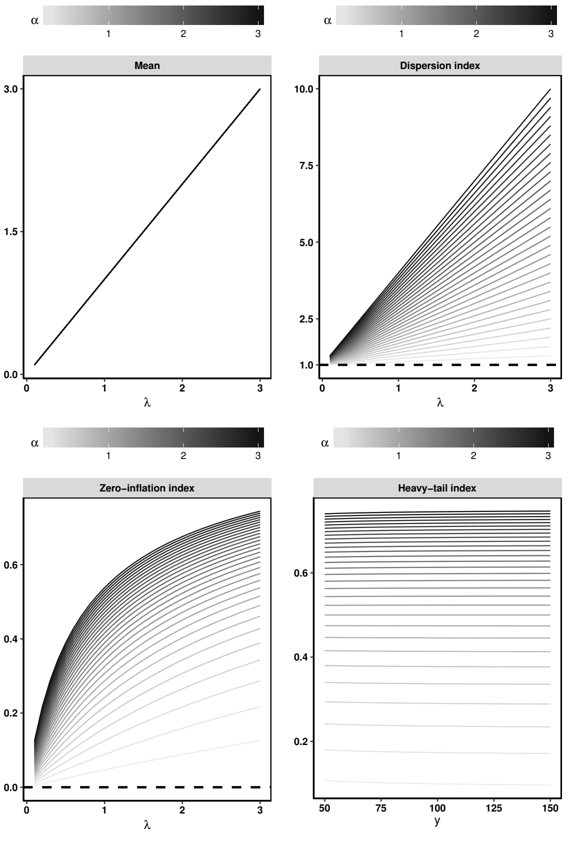


Figure 2: *Characteristic indices of the negative binomial distribution related to the Poisson distribution.*

S.8.2 Conway-Maxwell-Poisson Distribution

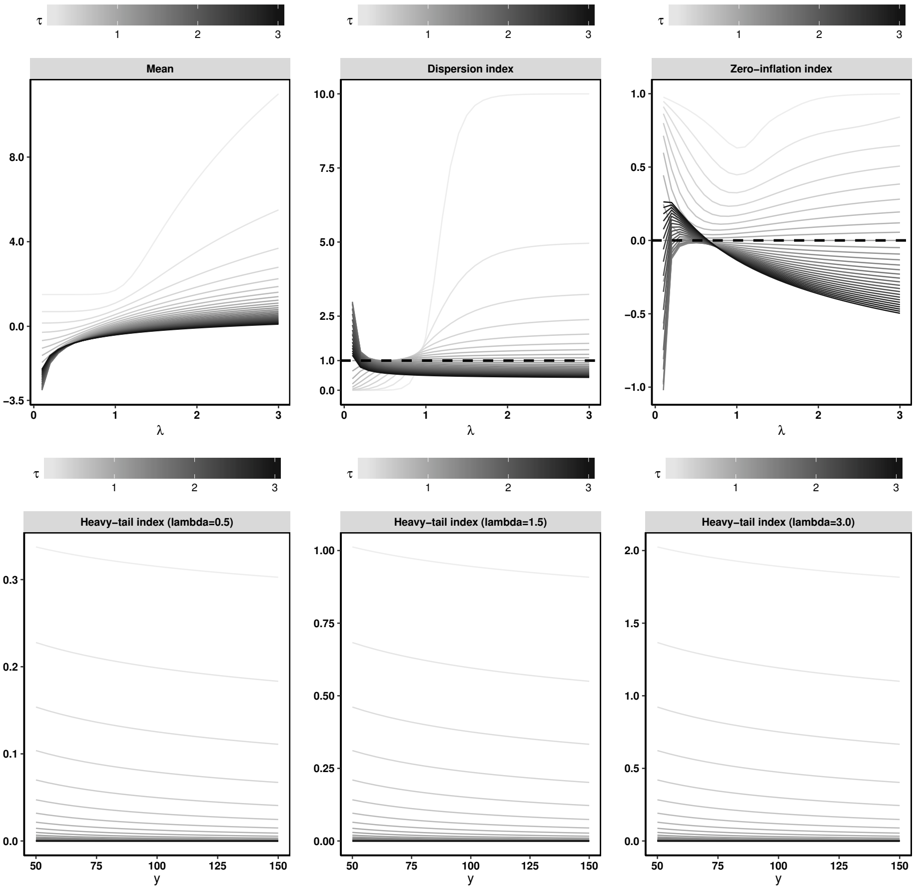


Figure 3: *Characteristic indices of Conway-Maxwell-Poisson distribution related to the Poisson distribution.*

S.8.3 Double Poisson Distribution

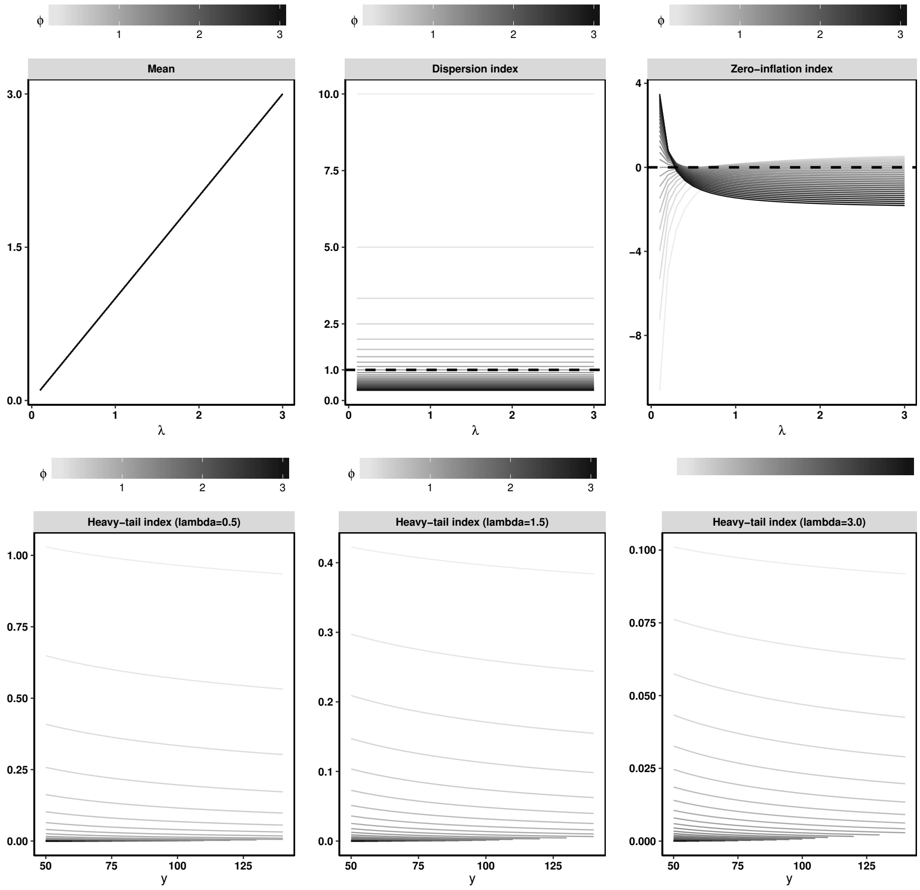


Figure 4: *Characteristic indices of double Poisson distribution related to the Poisson distribution.*

Supplementary Material for Chapter 4

S.9 Derivation of the Boundaries of π_1

Consider a mixture of 2 Poissons, i.e., Eq. (4.4). By utilizing constraint (4.3) into this mixture, we get:

$$\pi_1 \cdot \left[\mathrm{e}^{-\lambda_1} \lambda_1^y - \mathrm{e}^{-\lambda_2} \lambda_2^y \right] \geq -\mathrm{e}^{-\lambda_2} \lambda_2^y, \qquad \forall y \in \mathbb{N}. \tag{S.18}$$

If $\mathrm{e}^{-\lambda_1} \lambda_1^y - \mathrm{e}^{-\lambda_2} \lambda_2^y \in \mathbb{R}_+$:

$$\pi_1 \geq \frac{-\mathrm{e}^{-\lambda_2} \lambda_2^y}{\mathrm{e}^{-\lambda_1} \lambda_1^y - \mathrm{e}^{-\lambda_2} \lambda_2^y}, \qquad \forall y \in \mathbb{N}. \tag{S.19}$$

If $\mathrm{e}^{-\lambda_1} \lambda_1^y - \mathrm{e}^{-\lambda_2} \lambda_2^y \in \mathbb{R}_-$:

$$\pi_1 \leq \frac{-\mathrm{e}^{-\lambda_2} \lambda_2^y}{\mathrm{e}^{-\lambda_1} \lambda_1^y - \mathrm{e}^{-\lambda_2} \lambda_2^y}, \qquad \forall y \in \mathbb{N}. \tag{S.20}$$

Thus, π_1 is bounded by $[R_1, R_2]$, where

$$R_1 = \max_{\forall y \in \mathbb{N}} \left\{ \frac{\mathrm{e}^{-\lambda_2} \lambda_2^y}{\mathrm{e}^{-\lambda_2} \lambda_2^y - \mathrm{e}^{-\lambda_1} \lambda_1^y} \mid \mathrm{e}^{-\lambda_2} \lambda_2^y - \mathrm{e}^{-\lambda_1} \lambda_1^y \in \mathbb{R}_- \right\}, \tag{S.21}$$

$$R_2 = \min_{\forall y \in \mathbb{N}} \left\{ \frac{\mathrm{e}^{-\lambda_2} \lambda_2^y}{\underbrace{\mathrm{e}^{-\lambda_2} \lambda_2^y - \mathrm{e}^{-\lambda_1} \lambda_1^y}_{(X)}} \mid \mathrm{e}^{-\lambda_2} \lambda_2^y - \mathrm{e}^{-\lambda_1} \lambda_1^y \in \mathbb{R}_+ \right\}. \tag{S.22}$$

To obtain closed-form expressions for R_1 and R_2, the first derivate of (X) w.r.t. y will be calculated:

$$\begin{aligned}
\frac{\partial}{\partial y}\left[(X)\right] &= \frac{\overbrace{\frac{\partial}{\partial y}\left(\mathrm{e}^{-\lambda_2}\lambda_2^y\right)}^{(\star)}\cdot\left(\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y\right)-\left(\mathrm{e}^{-\lambda_2}\lambda_2^y\right)}{\left(\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y\right)^2} \\
&\quad \frac{\cdot\overbrace{\frac{\partial}{\partial y}\left(\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y\right)}^{(\star\star)}}{} \\
&\quad (\star)\frac{\partial}{\partial y}\left(\mathrm{e}^{-\lambda_2}\lambda_2^y\right)=\ln(\lambda_2)\mathrm{e}^{-\lambda_2}\lambda_2^y \\
&\quad (\star\star)\frac{\partial}{\partial y}\left(\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y\right)=\ln(\lambda_2)\mathrm{e}^{-\lambda_2}\lambda_2^y-\ln(\lambda_1)\mathrm{e}^{-\lambda_1}\lambda_1^y \\
&= \frac{\left[\ln(\lambda_1)-\ln(\lambda_2)\right]\cdot\left(\mathrm{e}^{-\lambda_2}\lambda_2^y\mathrm{e}^{-\lambda_1}\lambda_1^y\right)}{\left(\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y\right)^2}. \qquad \text{(S.23)}
\end{aligned}$$

Thus, the first derivative of (X) is positive and negative for $\ln(\lambda_1) > \ln(\lambda_2)$ and $\ln(\lambda_1) < \ln(\lambda_2)$, respectively.

This implies different scenarios:

Table S.1: *Different scenarios.*

	↗ $\ln(\lambda_1) > \ln(\lambda_2)$	↘ $\ln(\lambda_1) < \ln(\lambda_2)$
$[R_1]$ $\mathrm{e}^{-\lambda_2}\lambda_2^y - \mathrm{e}^{-\lambda_1}\lambda_1^y \in \mathbb{R}_-$	(1)	(3)
$[R_2]$ $\mathrm{e}^{-\lambda_2}\lambda_2^y - \mathrm{e}^{-\lambda_1}\lambda_1^y \in \mathbb{R}_+$	(2)	(4)

1. $\frac{\mathrm{e}^{-\lambda_2}\lambda_2^y}{\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y} \in \mathbb{R}_-$ because $\mathrm{e}^{-\lambda_2}\lambda_2^y \in \mathbb{R}_+$. Thus,

$$\begin{aligned}
R_1 &= \lim_{y\to\infty}\frac{\mathrm{e}^{-\lambda_2}\lambda_2^y}{\mathrm{e}^{-\lambda_2}\lambda_2^y-\mathrm{e}^{-\lambda_1}\lambda_1^y} \\
&= -\mathrm{e}^{-\lambda_2}\cdot\lim_{y\to\infty}\frac{\left(\frac{\lambda_2}{\lambda_1}\right)^y}{\mathrm{e}^{-\lambda_1}-\mathrm{e}^{-\lambda_2}\left(\frac{\lambda_2}{\lambda_1}\right)^y} \\
&= -\mathrm{e}^{-\lambda_2}\cdot\frac{0}{\mathrm{e}^{-\lambda_1}}=0. \qquad \text{(S.24)}
\end{aligned}$$

2. $\frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y} \in \mathbb{R}_+$ because $e^{-\lambda_2}\lambda_2^y \in \mathbb{R}_+$. Thus,

$$\begin{aligned} R_2 &= \left.\frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y}\right|_{y=0} \\ &= \frac{e^{-\lambda_2}}{e^{-\lambda_2} - e^{-\lambda_1}}. \end{aligned} \tag{S.25}$$

3. $\frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y} \in \mathbb{R}_-$ because $e^{-\lambda_2}\lambda_2^y \in \mathbb{R}_+$. Thus,

$$\begin{aligned} R_1 &= \left.\frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y}\right|_{y=0} \\ &= \frac{e^{-\lambda_2}}{e^{-\lambda_2} - e^{-\lambda_1}}. \end{aligned} \tag{S.26}$$

4. $\frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y} \in \mathbb{R}_+$ because $e^{-\lambda_2}\lambda_2^y \in \mathbb{R}_+$. Thus,

$$\begin{aligned} R_2 &= \lim_{y\to\infty} \frac{e^{-\lambda_2}\lambda_2^y}{e^{-\lambda_2}\lambda_2^y - e^{-\lambda_1}\lambda_1^y} \\ &= e^{-\lambda_2} \cdot \lim_{y\to\infty} \frac{1}{e^{-\lambda_2} - e^{-\lambda_1}\left(\frac{\lambda_1}{\lambda_2}\right)^y} \\ &= e^{-\lambda_2} \cdot \frac{1}{e^{-\lambda_2}} = 1. \end{aligned} \tag{S.27}$$

S.10 Proof that $[0,1] \subset [R1, R2]$ $(\forall \lambda_1, \lambda_2; \lambda_1 \neq \lambda_2)$

Theorem 6. *Given boundaries (4.7) for* π_1. *Then,* $[0,1] \subset [R1, R2]$ *(*$\forall \lambda_1, \lambda_2; \lambda_1 \neq \lambda_2$*).*

Proof. To prove this, one need to show that (1) $\frac{e^{-\lambda_2}}{e^{-\lambda_2}-e^{-\lambda_1}} > 1$ if $\lambda_1 > \lambda_2$ and (2) $\frac{e^{-\lambda_2}}{e^{-\lambda_2}-e^{-\lambda_1}} < 0$ if $\lambda_1 < \lambda_2$. Since the e^{-x} is a monotonic decreasing function bounded between 0 and 1 for all $x \in \mathbb{R}_+$, $\lambda_1, \lambda_2 \in \mathbb{R}_+$:

1. $\underline{\lambda_1 > \lambda_2 :}$

$$e^{-\lambda_2} - e^{-\lambda_1} < e^{-\lambda_2} \Leftrightarrow \frac{e^{-\lambda_2}}{e^{-\lambda_2} - e^{-\lambda_1}} > \frac{e^{-\lambda_2}}{e^{-\lambda_2}} = 1 \tag{S.28}$$

2. $\underline{\lambda_1 < \lambda_2 :}$

$$e^{-\lambda_1} - e^{-\lambda_2} > 0 \Leftrightarrow \frac{e^{-\lambda_2}}{e^{-\lambda_2} - e^{-\lambda_1}} = \frac{-e^{-\lambda_2}}{e^{-\lambda_1} - e^{-\lambda_2}} < 0 \tag{S.29}$$

□

S.11 Fitted Univariate & Finite Mixture Models

S.11.1 Fitted Univariate Models

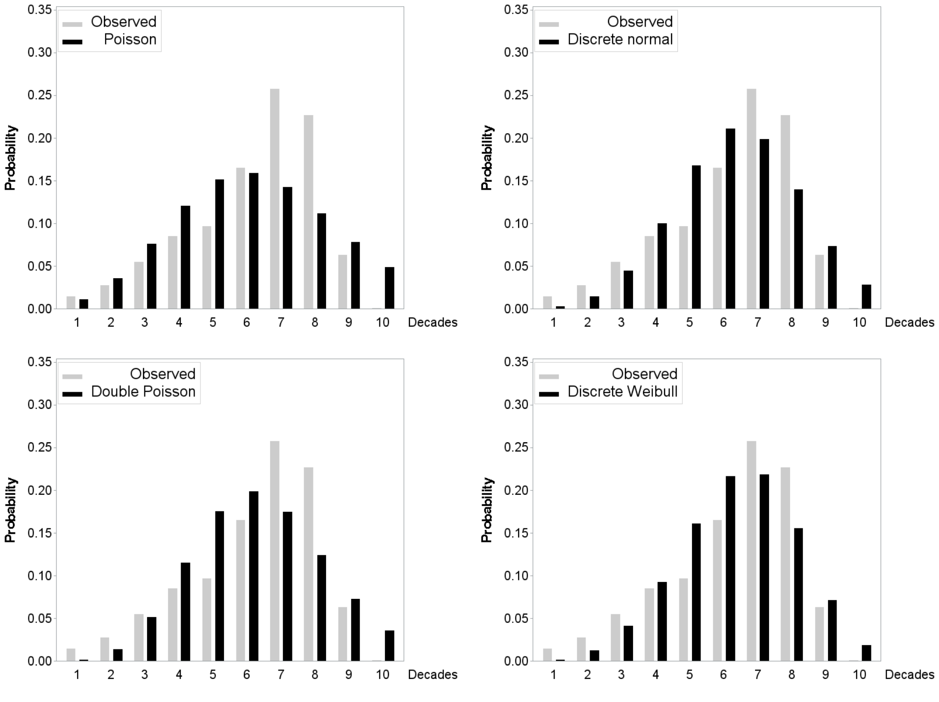


Figure 5: *Moerzeke data. Bar charts of fitted univariate models.*

S.11.2 Fitted Finite Mixture Models with Similar Components

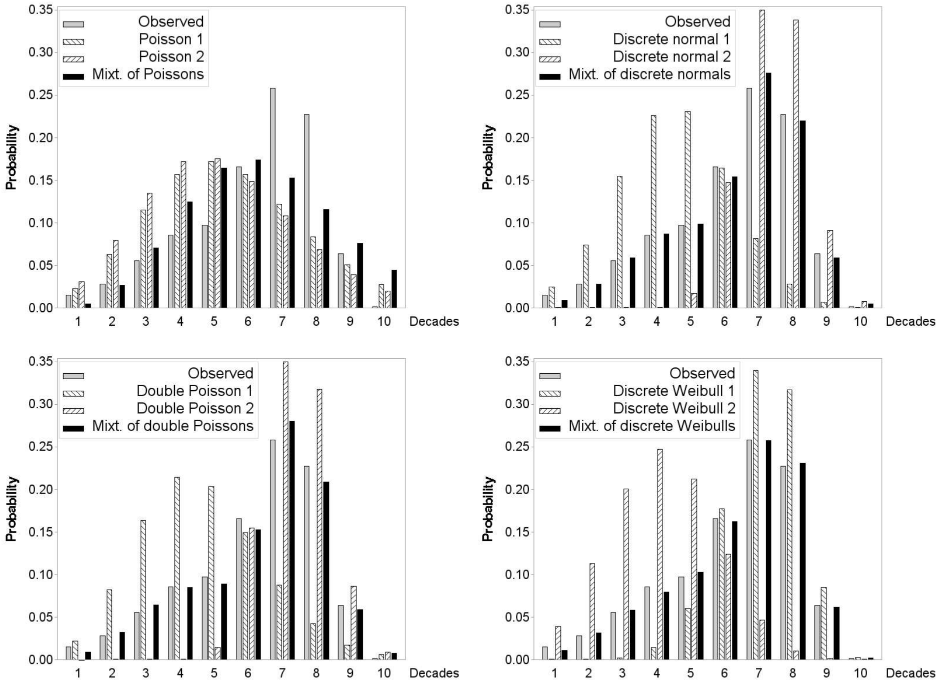


Figure 6: *Moerzeke data. Bar charts of fitted finite mixture models with similar components.*

S.11.3 Fitted Finite Mixture Models with Different Components

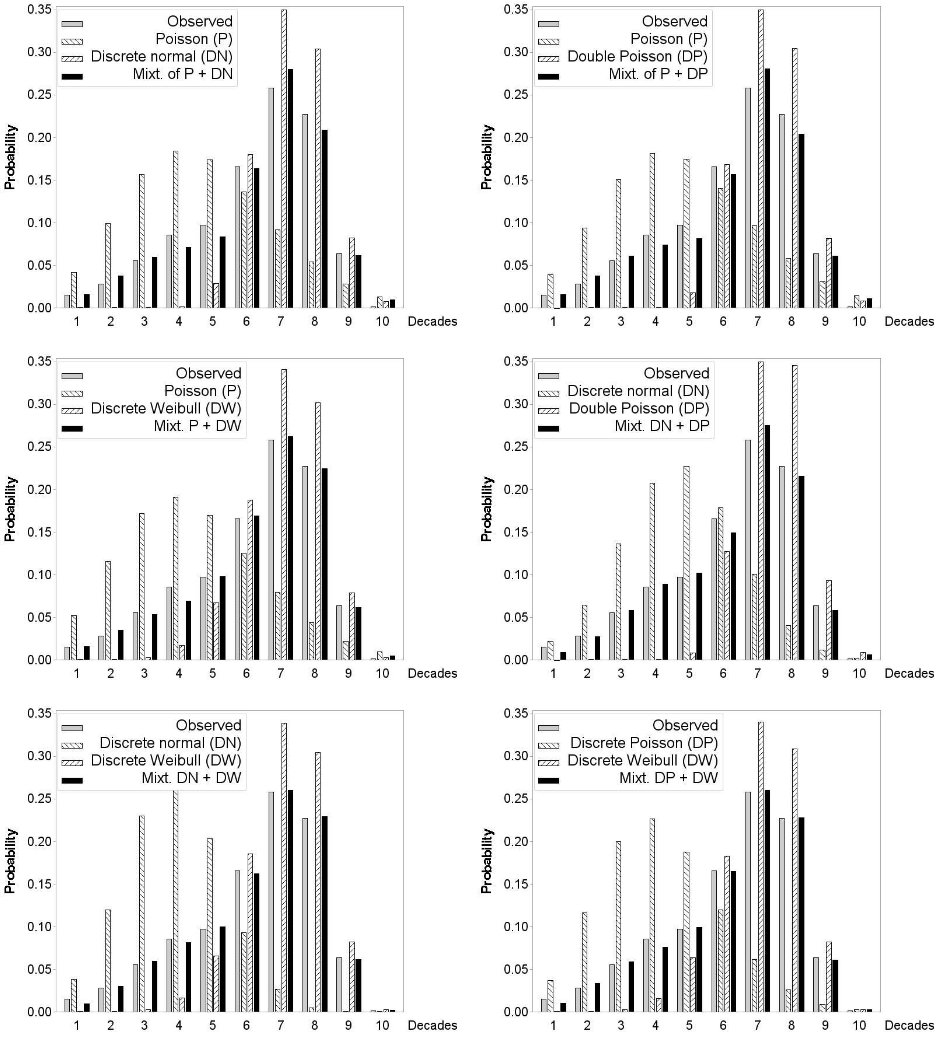


Figure 7: *Moerzeke data. Bar charts of fitted finite mixture models with different components.*

Supplementary Material for Chapter 5

S.12 Marginal Moments (up to the fourth order) of the Poisson-normal Model

S.12.1 Non-central Moments

$$\mathsf{E}\left(Y_{ij}\right) = \mathsf{E}\left[\mathsf{E}\left(Y_{ij} \mid \mathbf{b}_i\right)\right] = \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} \equiv \mu_{ij}, \tag{S.30}$$

$$\mathsf{E}\left(Y_{ij}^2\right) = \mathsf{E}\left[\mathsf{E}\left(Y_{ij}^2 \mid \mathbf{b}_i\right)\right] = \mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}, \tag{S.31}$$

$$\mathsf{E}\left(Y_{ij}^3\right) = \mathsf{E}\left[\mathsf{E}\left(Y_{ij}^3 \mid \mathbf{b}_i\right)\right] = \mathsf{e}^{3\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{9}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 3\mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}, \tag{S.32}$$

$$\begin{aligned} \mathsf{E}\left(Y_{ij}^4\right) = \mathsf{E}\left[\mathsf{E}\left(Y_{ij}^4 \mid \mathbf{b}_i\right)\right] &= \mathsf{e}^{4\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+8\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 6\mathsf{e}^{3\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{9}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 7\mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} \\ &\quad + \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}. \end{aligned} \tag{S.33}$$

S.12.2 Central Moments

$$\mathsf{E}\left(Y_{ij} - \mu_{ij}\right) = 0, \tag{S.34}$$

$$\begin{aligned} \mathsf{E}\left[\left(Y_{ij} - \mu_{ij}\right)^2\right] &= \mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - \mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} \\ &= \mu_{ij} + \mu_{ij}^2\left(\mathsf{e}^{\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 1\right) \end{aligned} \tag{S.35}$$

$$\begin{aligned} \mathsf{E}\left[\left(Y_{ij} - \mu_{ij}\right)^3\right] &= \mathsf{e}^{3\beta}\left(\mathsf{e}^{\frac{9}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 3\mathsf{e}^{\frac{5}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 2\mathsf{e}^{\frac{3}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}\right) \\ &\quad + 3\mathsf{e}^{2\beta}(\mathsf{e}^{2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - \mathsf{e}^{\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}) + \mathsf{e}^{\beta+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}, \end{aligned} \tag{S.36}$$

$$\begin{aligned} \mathsf{E}\left[\left(Y_{ij} - \mu_{ij}\right)^4\right] &= \mathsf{e}^{4\mathbf{x}_{ij}^{'}\boldsymbol{\beta}}\left(\mathsf{e}^{8\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 4\mathsf{e}^{5\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 6\mathsf{e}^{3\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 3\mathsf{e}^{2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}\right) \\ &\quad + \mathsf{e}^{3\mathbf{x}_{ij}^{'}\boldsymbol{\beta}}\left(6\mathsf{e}^{\frac{9}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 12\mathsf{e}^{\frac{5}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} + 6\mathsf{e}^{\frac{3}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}\right) \\ &\quad + \mathsf{e}^{2\mathbf{x}_{ij}^{'}\boldsymbol{\beta}}\left(7\mathsf{e}^{2\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}} - 4\mathsf{e}^{\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}\right) + \mathsf{e}^{\mathbf{x}_{ij}^{'}\boldsymbol{\beta}+\frac{1}{2}\mathbf{z}_{ij}^{'}D\mathbf{z}_{ij}}. \end{aligned} \tag{S.37}$$

S.13 Second-order Estimating Equations for the Poisson Mixed Approach with Random Intercept

The score equations for the Poisson mixed approach of Section 5.3.1 are expressed as follows:

$$\begin{aligned}\mathbf{Q}(\Theta) &= \sum_i \mathbf{P}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{f}_i \\ &= \sum_i \begin{bmatrix}\frac{\partial\boldsymbol{\mu}_i}{\partial\beta} & \frac{\partial\boldsymbol{\mu}_i}{\partial d}\\ \frac{\partial\boldsymbol{\eta}_i}{\partial\beta} & \frac{\partial\boldsymbol{\eta}_i}{\partial d}\end{bmatrix}' \begin{bmatrix}\mathsf{Var}(\mathbf{Y}_i) & \mathsf{Cov}(\mathbf{Y}_i,\mathbf{S}_i)\\ \mathsf{Cov}(\mathbf{Y}_i,\mathbf{S}_i) & \mathsf{Var}(\mathbf{S}_i)\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{Y}_i-\boldsymbol{\mu}_i\\ \mathbf{S}_i-\boldsymbol{\eta}_i\end{bmatrix} \\ &= \mathbf{0},\end{aligned} \tag{S.38}$$

where $\Theta = (\beta, d)'$. Furthermore, the model-based and empirically corrected standard errors of the parameter estimates are derived as the square root of the diagonal entries of

$$U^* = \left[\sum_i \mathbf{P}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{P}_i\right]^{-1}, \tag{S.39}$$

$$U^{**} = U^*\left(\sum_i \mathbf{P}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{f}_i\mathbf{f}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{P}_i\right)U^*, \tag{S.40}$$

respectively.

S.13.1 Components of the Second-order Estimating Equations

Let $B = \mathsf{e}^{\beta}$ and $\Delta = \mathsf{e}^{\frac{1}{2}d}$. Components of the score Eq. (S.38) are expressed as follows:

$$\mu_{ij} = B\Delta, \tag{S.41}$$

$$\eta_{ij} = B^2\Delta^4 + B\Delta, \tag{S.42}$$

$$\frac{\partial\mu_{ij}}{\partial\beta} = B\Delta, \tag{S.43}$$

$$\frac{\partial\mu_{ij}}{\partial d} = \frac{1}{2}B\Delta, \tag{S.44}$$

$$\frac{\partial\eta_{ij}}{\partial\beta} = B\Delta + 2B^2\Delta^4, \tag{S.45}$$

$$\begin{aligned}\frac{\partial \eta_{ij}}{\partial d} &= \frac{1}{2}B\Delta + 2B^2\Delta^4, && \text{(S.46)}\\ \mathsf{Var}(Y_{ij}) &= B^2\Delta^2(\Delta^2-1)+B\Delta, && \text{(S.47)}\\ \mathsf{Cov}(Y_{ij},S_{ij}) &= B^2\Delta^4(\Delta^4-1)+B\Delta(3\Delta^2-1)+1, && \text{(S.48)}\\ \mathsf{Var}(S_{ij}) &= B^4\Delta^8(\Delta^8-1)+B^3\Delta^5(6\Delta^4-2)+B^2\Delta^2(7\Delta^2-1)+B\Delta. && \text{(S.49)}\end{aligned}$$

S.13.2 Solution of the Second-order Estimating Equations

By inserting Eq. (S.41)–(S.49) into the proposed score Eq. (S.38), the following expression can be found:

$$\begin{aligned}\mathbf{Q}(\Theta) &= \frac{NB^2\Delta^2}{\det(\mathbf{\Sigma}_i)}\cdot\begin{bmatrix}K_{11} & K_{12}\\ K_{21} & K_{22}\end{bmatrix}\cdot\begin{bmatrix}\underbrace{\frac{1}{N}\sum_i\sum_j Y_{ij}}_{k_1} - B\Delta\\ \underbrace{\frac{1}{N}\sum_i\sum_j Y_{ij}^2}_{k_2} - (B^2\Delta^4+B\Delta)\end{bmatrix}\\ &= \mathbf{0}, && \text{(S.50)}\end{aligned}$$

with

$$\begin{aligned}K_{11} &= B^3\Delta^7(\Delta^4-1)^2+B^2\Delta^4(5\Delta^2-1)(\Delta^2-1)+2B\Delta^3, && \text{(S.51)}\\ K_{12} &= -B^2\Delta^4(\Delta^2-1)^2, && \text{(S.52)}\\ K_{21} &= \frac{1}{2}[B^3\Delta^7(\Delta^4-1)(\Delta^4-3)+B^2\Delta^4(5\Delta^4-12\Delta^2+3)], && \text{(S.53)}\\ K_{22} &= \frac{1}{2}[-B^2\Delta^4(\Delta^2-3)(\Delta^2-1)+2B\Delta^3]. && \text{(S.54)}\end{aligned}$$

Solving these equations results in the following equations:

$$\begin{cases}B^2 &= \frac{k_1^4}{k_2-k_1}\\ \Delta^2 &= \frac{k_2-k_1}{k_1^2}\end{cases} \Rightarrow \begin{cases}\widehat{\beta} &= 2\mathsf{ln}(k_1)-\frac{1}{2}\mathsf{ln}(k_2-k_1)\\ \widehat{d} &= \mathsf{ln}(k_2-k_1)-2\mathsf{ln}(k_1)\end{cases}. \qquad \text{(S.55)}$$

Thus, to capture the estimates of β of d, non-central sample first- and second-order moments are needed in this approach, i.e., k_1 and k_2, respectively.

S.13.3 Model-based and Sandwich Standard Errors of the Second-order Estimating Equations

To obtain the model-based and sandwich standard errors, a PROC IML macro has been written, where the input values are the given non-central first, second, third and fourth moment of the data, i.e., k_1, k_2, k_3, and k_4, respectively, and the output derived consists of (1) the parameter estimates obtained from the score equation, (2) model-based and sandwich standard errors for these estimates.

```
/****************************
SOFTWARE: SAS 9.4.
OBJECTIVE: Develop a SAS IML macro to obtain the
           parameter estimates, model-based and sandwich
           standard errors of GEE2;
INPUT:
- N: Sample size;
- k1: Non-central first-order sample moment;
- k2: Non-central second-order sample moment;
- k3: Non-central third-order sample moment;
- k4: Non-central fourth-order sample moment;
OUTPUT: List of parameters, consisting of
(1) Estimate of beta (beta_hat);
(2) Estimate of d (d_hat);
(3) Model-based standard error of beta (se_mb_beta);
(4) Sandwich standard error of beta (se_sandwich_beta);
(3) Model-based standard error of d (se_mb_d);
(4) Sandwich standard error of d (se_sandwich_d);
AUTHOR: M. Luyts (L-Biostat);
****************************/
%macro score_PN(N, k1, k2, k3, k4);
  PROC IML ;
    /*Parameter estimates*/
    beta_hat=2*log(&k1)-0.5*log(&k2-&k1);
    d_hat=log(&k2-&k1)-2*log(&k1);
    /*Model-based standard errors*/
    B=exp(beta_hat);
    delta=exp(0.5*(d_hat));
    A={1 0.5, 1 0.5};
    C={0 0, 1 1};
    D=(B*delta*A)+(2*(B**2)*(delta**4)*C);
    Sigma11=(B*delta*((delta**2)-1))+1;
```

```
    Sigma12=((B**2)*(delta**4)*((delta**4)-1))+(B*delta
          *((3*(delta**2))-1))+1;
    Sigma22=((B**3)*(delta**7)*((delta**8)-1))+((B**2)
          *(delta**4)*((6*(delta**4))-2))+(B*delta
          *((7*(delta**2))-1))+1;
    V={0 0, 0 0};
    V[1,1]=Sigma11;
    V[1,2]=Sigma12;
    V[2,1]=Sigma12;
    V[2,2]=Sigma22;
    Sigma=B*delta*V;
    Q=&N#t(D)*inv(Sigma)*D;
    Ustar=inv(Q);
    se_mb_beta=sqrt(Ustar[1,1]);
    se_mb_d=sqrt(Ustar[2,2]);
    /*Sandwich standard errors*/
    L11=&k2-(2*B*delta*&k1)+((B**2)*(delta**2));
    L12=&k3-(B*delta*&k2)-((((B**2)*(delta**4))
          +(B*delta))*&k1)+((((B**2)*(delta**4))
          +(B*delta))*B*delta);
    L22=&k4-(2*(((B**2)*(delta**4))+(B*delta))*&k2)
          +((((B**2)*(delta**4))+(B*delta))**2);
    L={0 0, 0 0};
    L[1,1]=L11;
    L[1,2]=L12;
    L[2,1]=L12;
    L[2,2]=L22;
    K=&N#L;
    Ustarstar=Ustar*t(D)*inv(Sigma)*K*inv(Sigma)*D*Ustar;
    se_sandwich_beta=sqrt(Ustarstar[1,1]);
    se_sandwich_d=sqrt(Ustarstar[2,2]);
    print beta_hat d_hat se_mb_beta se_sandwich_beta
          se_mb_d se_sandwich_d;
  quit;
%mend;
```