Mixed-effects Additive Transformation Models with the R Package tramME

Bálint Tamási

Abstract

Regression models that accommodate correlated observations and potential nonlinear predictor-outcome relationships are fundamental in analyzing experimental and observational data. Unlike traditional parametric approaches, transformation models make weaker assumptions on the conditional response distribution, thus allowing for a more universal applicability to at least ordered univariate outcomes. This flexibility makes transformation models an attractive choice for modeling complex relationships in a wide range of domains. The R package **tramME** extends the transformation model framework with general random effect structures and penalized smooth terms to adapt to dependent data and nonlinear predictor-outcome relationships. This paper presents the statistical framework and implementation details of **tramME**, including its integration with other popular R packages for transformation modeling (**mlt**), mixed-effects (**lme4**) and additive models (**mgcv**). The package employs the efficient Template Model Builder framework (**TMB**) for fully parametric likelihood-based estimation and inference. Two illustrations demonstrate that **tramME** can readily model complex, dependent data structures under settings where the choice of the outcome distribution type is challenging.

Keywords: additive models, correlated data, mixed-effects models, penalization, R, regression, smoothing, transformation models.

1. Introduction

Regression models are among the most important tools in applied statistics. These models are used to analyze experimental results or data from observational settings when the scientific question pertains to the relationship between an outcome variable and a set of predictors. While linear regression is a widely-used and flexible model with a rich history, it is limited by specific assumptions that are often violated in real-world settings. These assumptions include conditional normality, conditional independence among observations in the sample, and a linear functional relationship between the predictors and the expectation of the outcome. To address these limitations, a range of extensions have been developed over the past decades, in part facilitated by the increasing availability of computing power.

Generalized linear models (GLM, Nelder and Wedderburn 1972) expand the set of available conditional distributions to the exponential family, allowing for specific nonlinear relationships between predictors and the conditional expectation of the outcome. Generalized additive models (GAM, Hastie and Tibshirani 1999) build on this by introducing general nonlinear covariate effects. To handle dependent data, the normal linear model, GLMs, and GAMs can be extended with random effects, which introduce additional structure to the random part of the regression model, leading to mixed models (GLMM, Stroup 2012 and GAMM, Wood 2017). More recently, distributional regression approaches, such as GAMLSS (Rigby and Stasinopoulos 2005), have been proposed to increase the flexibility of the parametric regression framework. These methodological developments in regression analysis are supported by a variety of software packages that implement the various extensions of the normal linear model, such as **nlme** (Pinheiro and Bates 2000) and **lme4** (Bates, Mächler, Bolker, and Walker 2015), which implement (generalized) linear and nonlinear mixed models. Additionally, the **glmmTMB** package (Brooks, Kristensen, van Benthem, Magnusson, Berg, Nielsen, Skaug, Maechler, and Bolker 2017) uses an efficient computational framework to fit many varieties and extensions of GLMMs. For modeling discrete ordinal responses, the likelihood-based approach provided by **ordinal** (Christensen 2023) package is a popular choice. For estimating additive models, the **mgcv** package (Wood 2017) is the most popular option, while **gamm4** (Wood and Scheipl 2020) combines the selection of smoothers in **mgcv** with the computational tools available in **lme4**. A Bayesian alternative for additive and mixed-effects modeling of a large variety of response types is provided by the **brms** package by Bürkner (2017). The **gamlss** package (Rigby and Stasinopoulos 2005) extends the additive modeling approach to multiple parameters of the conditional outcome distribution. The Bayesian estimation of these distributional regression models is implemented, for example, by the **bamlss** package (Umlauf, Klein, Simon, and Zeileis 2021).

Traditional *parametric* regression approaches are based on the assumption that the conditional response distribution can be fully described with a small number of parameters of a pre-specified distribution type. Modeling the relationship between the outcome and the predictors by writing these distributional parameters as (potentially nonlinear) functions of fixed and random effects not only enhances the interpretability of these models, but also requires simpler computational procedures for estimation and inference. However, selecting the correct distributional form can be challenging, and misspecifications can lead to inefficient or incorrect inferences (see, for example, Siegfried and Hothorn 2020, in the context of modeling count data, and Rutherford, Crowther, and Lambert 2015, in survival analysis). In survival analysis, which is traditionally treated as a separate branch of regression analysis (Klein and Moeschberger 2003), nonparametric and semiparametric approaches are currently much more prevalent, with the famous Cox proportional hazards model (Cox 1972) as their most important example. While these models do not rely on simple parametric distributional assumptions, their estimation is typically more involved, and the results could be harder to interpret.

This paper presents the **tramME** package (Tamási 2024) that implements a general mixed-effects additive transformation model for arbitrary outcome types. Transformation models do not rely on parametric distributional assumptions, but rather estimate the shape of the conditional outcome distribution with a flexible parameterization using a small number of parameters. In this sense, the model is “distribution-free” in that it can approximate any distribution at the reference level (i.e., a covariate configuration that leads to a zero value of the linear predictor) in a regression setting. The additive structure of the model class implemented in the package retains the interpretability of parametric regression models. Additionally, **tramME** can handle any, at least ordered, continuous or discrete univariate outcome and arbitrary random censoring and truncation. The correlation structure in the data and nonlinear predictor-outcome relationships are captured with potentially complex random effects and other penalized additive terms.

There are several R packages that follow a similar approach to **tramME** for regression modeling. The **rstpm2** package (Liu, Pawitan, and Clements 2017b, 2018) implements generalized survival models for potentially correlated time-to-event data. As we will see in Section 4.1, flexible parametric proportional hazards models for survival outcomes (Royston and Parmar 2002) constitute a special case of the linear transformation model framework of **tramME**. Several packages provide likelihood-based (**flexsurv**, Jackson 2016) or Bayesian (**rstanarm,** Brilleman, Elci, Novik, and Wolfe 2020) implementations for these. Manuguerra, Heller, and Ma (2020) developed the **ordinalCont** package for analyzing $(0,1)$ (or $[0,1]$, after rescaling) bounded visual analogue scale data with continuous ordinal regression, which can be regarded as another special case of the transformation model approach. A comparison between this package and **tramME** in an example analysis focusing only on the mixed-effects extension of linear transformation models can be found in Tamási and Hothorn (2021). A semiparametric approach for transformation modeling extends the regression model for discrete ordinal outcomes (Liu, Shepherd, Li, and Harrell 2017a) and uses the `orm()` function of the **rms** package (Harrell 2024). A systematic comparison between this approach, which estimates the transformation function with nonparametric maximum likelihood, and the fully parametric alternative represented by **tramME** can be found in Tian, Hothorn, Li, Harrell Jr., and Shepherd (2020). The **TransModel** package (Zhou, Zhang, and Lu 2022) provides semiparametric estimation of linear transformation models for censored survival data. Finally, the **tram** package by Hothorn, Barbanti, and Siegfried (2024) provides the function `mtram()` for marginally interpretable transformation models for clustered data (Barbanti and Hothorn 2024).

Although some of these packages may share functionalities with **tramME**, they have important differences in terms of their scope, approach to approximating transformations, applicability to grouped data, availability of nonlinear terms, and post-estimation procedures. **tramME** offers a conditional approach for distribution-free regression modeling of grouped data with potential nonlinear predictor-outcome relationships. It leverages the functionality from several popular R packages to formulate transformation models with complex random effects structures and a wide range of optional nonlinear additive terms. By using a fully parametric approach (as opposed to a more traditional semiparametric treatment, see Cheng, Wei, and Ying 1995; Zeng and Lin 2007; Lin, Zhou, and Zhou 2014; De Neve, Thas, and Gerds 2019, for examples) to approximating the outcome distribution, standard results from likelihood theory for inference can be readily used. The computations are done through a highly performant C++-based (Stroustrup 2013) framework for mixed model estimation (the package **TMB** by Kristensen, Nielsen, Berg, Skaug, and Bell 2016), which results in a fast and stable implementation of the model class (see the results in Appendix A for a comparison with the **glmmTMB** package). Moreover, the package offers a collection of methods for model evaluation and comparison, prediction, and simulation, beyond basic estimation and inference. Overall, **tramME** represents a flexible, powerful, and versatile tool for regression modeling, with a wide range of potential applications in various scientific fields.

The mixed model ecosystem of R (Bolker, Piaskowski, Tanaka, Alday, and Viechtbauer 2024), offers many options for specialized tasks and settings when working with dependent data. **tramME** fits into this ecosystem by providing a regression modeling framework for general (potentially censored or truncated) outcome types without the need to *a priory* commit to a specific distributional family. In this sense, **tramME** is a viable option for censored (specifically, but not exclusively, survival) outcomes, as we will see in Section 4.1, ordinalvalued responses (see Tamási and Hothorn 2021, for example applications), over-dispersed and zero(-one)-inflated data (see the example analyes in Tamási and Hothorn 2023), as well as heavy-tailed or skewed responses (see Section 4.2). The mixed effects additive framework provides several options for penalized smooth terms. Because the fitted outcome distribution can be flexibly evaluated on various scales, **tramME** can be seen as an alternative to quantile regression approaches, as the application in Section 4.2 will demonstrate.

This paper is structured as follows. Section 2 introduces mixed-effects additive transformation models, discussing the parameterization and estimation within a likelihood-based framework. Section 3 presents the **tramME** package and explains how it can be used to specify various transformation models. In Section 4, two example analyses are presented to demonstrate the functionality of **tramME**. Both cases require a distribution-free approach without committing to parametric outcome distributions *a priori*, and involve the use of random effects and other penalized additive terms due to the specifics of the data structure. Additional details and examples are given in the documentation and package vignettes on the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=tramME`.

2. Mixed-effects additive transformation models

Transformation models represent a very general class of regression models with many notable special cases and potential extensions. The basic idea of the approach can be traced back to the seminal work by Box and Cox (1964), which suggested using flexible transformations to cast the regression problem to a scale where a general outcome can be modeled in terms of a simple parametric distribution. The original Box-Cox regression relies on a one- or two-parameter family of functions to estimate the necessary transformation from the data together with other regression parameters. Over the following decades, various extensions of this approximation and estimation strategies were suggested (see, e.g., Doksum and Gasko 1990; Cheng *et al.* 1995; McLain and Ghosh 2013). Hothorn, Möst, and Bühlmann (2018) proposed a fully parametric approach for transformation models based on basis expansions of the outcome and maximum likelihood inference. The resulting framework is applicable for any, at least ordered, outcome types and incorporates a wide range of regression models. Hothorn (2020) describes the **mlt** package, the accompanying R implementation to this approach. The **tramME** package extends the functionality of **mlt** and its formula-based user interface **tram** (Hothorn *et al.* 2024) with random effects and nonlinear additive terms to handle general dependent data structures and potential nonlinear predictor-outcome relationships. Because the resulting model includes latent terms, the original likelihood-based approach needs an extension.

2.1. Model structure

For a univariate outcome Y from a sample space Ξ, mixed-effects additive transformation models parameterize the conditional cumulative distribution function (CDF) given the vector of predictors and random effects

$$F_Y(y \mid \boldsymbol{x}, \boldsymbol{\gamma}) = \mathsf{P}(Y \leq y \mid \boldsymbol{x}, \boldsymbol{\gamma}) = F_Z\left(h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta}) - \boldsymbol{x}_B^\top \boldsymbol{\beta} - \sum_{j=1}^{J} f_j(\boldsymbol{x}_{C_j}; \boldsymbol{\delta}_j) - \boldsymbol{z}^\top \boldsymbol{\gamma}\right), \quad (1)$$

where $\boldsymbol{x} = (\boldsymbol{x}_A^\top, \boldsymbol{x}_B^\top, \boldsymbol{x}_{C_1}^\top, \boldsymbol{x}_{C_2}^\top, \dots, \boldsymbol{x}_{C_J}^\top, \boldsymbol{z}^\top)^\top$ is a stacked vector comprising the covariates and the random effects design vector. The model consists of five main components:

1. $F_Z(\cdot)$ is a predefined inverse link function that does not contain unknown parameters. It maps the transformed outcomes to probabilities, i.e. $F_Z : \mathbb{R} \to [0, 1]$. F_Z is typically set to the CDF of a simple distribution with log-concave density f_Z.

2. $h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta})$ is the monotonic non-decreasing transformation function, parameterized with the vector $\boldsymbol{\vartheta}$. This function takes the same argument (y) as the distribution function of the outcome, and maps $h : \Xi \to \mathbb{R}$. It may be conditional on a set of covariates ($\boldsymbol{x}_A$), which contains stratification variables or represents other non-proportional (response-varying) effects. Otherwise, the term is just a function of the outcome variable ($h(y; \boldsymbol{\vartheta})$).

3. $\boldsymbol{x}_B^\top \boldsymbol{\beta}$ represents the set of parametric fixed-effects, which act as shift terms on the transformation scale.

4. The shift terms $f_j(\boldsymbol{x}_{C_j}; \boldsymbol{\delta}_j)$ introduce nonlinear effects of predictors or sets of predictors on the scale of the transformation function. To favor smoother solutions, these nonlinear terms can optionally be supplemented with various roughness penalties.

5. $\boldsymbol{z}^\top \boldsymbol{\gamma}$ represents the general random effects term with the design vector $\boldsymbol{z}$ and the vector of random effects, which is assumed to be multivariate normally distributed, $\boldsymbol{\gamma} \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ typically has a sparse block structure, which can be leveraged in the computations.

The choice of the inverse link function (F_Z) will affect the scale on which the elements of the linear predictor are interpreted (Hothorn *et al.* 2018). The typical choices for this function are the CDF of the standard logistic distribution ("expit" function, $F_{\mathrm{SL}}(z) = (1 + \exp(-z))^{-1}$), which leads to a proportional odds model, or the CDFs of the minimum extreme value distribution (the inverse of the complementary log-log function, $F_{\mathrm{MinEV}}(z) = 1 - \exp(-\exp(z))$) and maximum extreme value distribution (the inverse of the log-log link function, $F_{\mathrm{MaxEV}}(z) = \exp(\exp(-z))$), which result in proportional hazards and reverse-time proportional hazards models, respectively. Setting F_Z to the distribution function of the standard Gaussian distribution (Φ) produces effect estimates that are interpreted as differences in conditional means on a latent transformation scale.

Sometimes F_Z is referred to as the "error distribution" because it represents the distribution of an error term in the regression of the transformed outcome $h(Y) = \mu(\boldsymbol{x}) + \varepsilon_Z$, where $\varepsilon_Z \sim F_Z$ and $\mu(\boldsymbol{x})$ is the linear predictor. It is important to note that, although they play similar roles, the inverse link function of a transformation model cannot always be equated with inverse link function of a formally identical GLM. For example, the normal linear regression is a member of both the GLM (identity link, Gaussian family) and the transformation model class. As the first row of Table 1 shows, in the the transformation model family, this will correspond to a *probit link* ($F_Z = \Phi$) and a linear transformation function ($h(y; \boldsymbol{\vartheta}) = \vartheta_1 + \vartheta_2 y$). Nevertheless, in some special cases, the inverse links of the two model families are the same: Because F_Z always maps from the latent transformation scale to the probability scale, in transformation models for discrete ordinal outcomes (and, as a special case, binary responses), the inverse link function will be the same as in the corresponding GLM, because the latent-variable interpretation of this model is identical to the transformation approach. Table 1 of Hothorn *et al.* (2018) gives a more detailed overview about the connections of particular transformation models and members of the GLM family.

As it is shown in Equation 1, it is usually more natural to write the terms that constitute the linear predictor (Terms 3, 4 and 5) with negative signs. In this case, higher values of the linear predictor correspond to higher conditional expectations of the response. However, in some model variants, such as proportional hazards models discussed in Section 4.1, historically, a positive sign is more common because the scale of the transformation function has the interpretation of the (accumulated) risk of an event and the research question concerns how various factors affect this risk.

2.2. Parameterization

The baseline transformation in Equation 1 is parameterized with the help of suitable basis functions, $h(y; \boldsymbol{\vartheta}) = \boldsymbol{a}(y)^\top \boldsymbol{\vartheta}$, or, when interacting variables are present, the term is formulated with a tensor product basis $h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta}) = (\boldsymbol{a}(y)^\top \otimes \boldsymbol{b}(\boldsymbol{x}_A)^\top)\boldsymbol{\vartheta}$. The appropriate basis transformations $\boldsymbol{a}(y)$ are chosen depending on the outcome type (e.g., whether it is discrete or continuous) and the assumptions of the analysis, i.e., whether it is restricted to a specific functional form or treated as a general function (Hothorn *et al.* 2018).

For categorical outcomes with K levels, step functions $h(y; \boldsymbol{\vartheta}) = \vartheta_k$ for $k = 1, 2, \ldots, K-1$ are typically chosen. For continuous outcome types, the *Bernstein polynomial basis* (Farouki 2012) provides a convenient approach for smooth approximation of the transformation function:

$$\boldsymbol{a}_{\mathrm{Bs},K}(y)^\top \boldsymbol{\vartheta} = \frac{1}{K+1} \sum_{k=0}^{K} \vartheta_k \binom{K}{k} \tilde{y}^K (1-\tilde{y})^{K-k}, \tag{2}$$

where $\boldsymbol{a}_{\mathrm{Bs},K}(y)$ represents the Bernstein basis of order K and $\tilde{y}$ is the rescaled version of the argument y to the support $[0, 1]$. In **tramME**, the interval of the Bernstein approximation can be controlled with the `support` argument (see Section 4.1). Beyond this support, a linear extrapolation is applied. By restricting the elements of the parameter vector to $\vartheta_k \geq \vartheta_{k-1}$, the resulting approximation will be monotonic *non-decreasing* (to be applicable to continuous, discrete and discrete-continuous mixed-type outcomes). This can be ensured during the model fitting using a set of linear non-equality optimization constraints.

As a useful extension to count outcome types, Siegfried and Hothorn (2020) adapt the smooth parameterization of the transformation function by evaluating it only at discrete points $h(y; \boldsymbol{\vartheta}) = \boldsymbol{a}(\lfloor y \rfloor)^\top \boldsymbol{\vartheta}$. Furthermore, in some settings, simple functional forms, e.g., $h(y; \boldsymbol{\vartheta}) = \vartheta_1 + \vartheta_2 y$ or $h(y; \boldsymbol{\vartheta}) = \vartheta_1 + \vartheta_2 \log(y)$, may be assumed. These lead to well-known special cases of the model class, such as the normal, log-normal, Weibull etc. mixed-effects additive regression models (Hothorn *et al.* 2018).

The nonlinear shift terms of an additive transformation model are also expressed using basis functions, $f_j(\boldsymbol{x}_{C_j}; \boldsymbol{\delta}_j) = \boldsymbol{c}_j(\boldsymbol{x}_{C_j})^\top \boldsymbol{\delta}_j$. A wide range of options is available for $\boldsymbol{c}_j$, along with the corresponding penalty terms, to represent various smoothers for specific settings and tasks, e.g., thin plate splines, B-splines, P-splines, Markov random fields, Gaussian process smoothers and tensor product splines (Wood, Scheipl, and Faraway 2013). A detailed review of the most important alternatives is given in Wood (2017).

The smoothness penalties on the additive terms in **tramME** can be written in the quadratic form $\lambda_j \boldsymbol{\delta}_j^\top \boldsymbol{S}_j \boldsymbol{\delta}_j$, with the smoothing parameter λ_j, which controls how much penalty is imposed on the coefficients and is typically estimated from the data. The fixed penalty matrix $\boldsymbol{S}_j$ is determined by the exact form of the penalty (e.g., the order of differences between adjacent elements of $\boldsymbol{\delta}_j$). Additive terms with this penalty structure can typically be formulated as a set of fixed and random effects (see, Wood 2004, Appendix), and thus the mixed-effects additive transformation model can be reparameterized as a mixed-effects transformation model

$$F_Y(y \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) = F_Z\left(h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta}) - \tilde{\boldsymbol{x}}^\top \boldsymbol{\beta}^\star - \tilde{\boldsymbol{z}}^\top \boldsymbol{\gamma}^\star\right) \qquad \boldsymbol{\gamma}^\star \sim \mathcal{N}_{q^\star}(\mathbf{0}, \boldsymbol{\Sigma}^\star) \tag{3}$$

with a new set of fixed ($\boldsymbol{\beta}^\star$) and random-effects parameters ($\boldsymbol{\gamma}^\star$), and the corresponding design vectors, $\tilde{\boldsymbol{x}} = (\boldsymbol{x}_B^\top, \tilde{\boldsymbol{x}}_C^\top)^\top$ and $\tilde{\boldsymbol{z}} = (\boldsymbol{z}^\top, \tilde{\boldsymbol{z}}_C^\top)^\top$, respectively. The smoothing parameters of the penalized terms $(\lambda_1, \lambda_2, \dots)$ are estimated as transformed diagonal elements of the extended random effects covariance matrix $\boldsymbol{\Sigma}^\star$.

2.3. Estimation and inference

After transforming the original additive mixed-effects model to its mixed-effects representation, we can use standard fitting procedures for mixed models to estimate the parameters. Tamási and Hothorn (2021) present the maximum likelihood approach of **tramME** to efficiently estimate mixed models of the form of Equation 3 using the Template Model Builder (**TMB**, Kristensen *et al.* 2016). Given a dataset with observations indexed by $i = 1, 2, \dots, N$, the likelihood of the model parameters, i.e., the parameters of the transformation function, the fixed effects coefficients and the elements of the random effects covariance matrix, can be written in the form

$$\mathcal{L}(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star, \boldsymbol{\Sigma}^\star) = \int_{\boldsymbol{\gamma}^\star \in \mathbb{R}^{q^\star}} \prod_{i=1}^{N} \mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star) \phi(\boldsymbol{\gamma}^\star; \Sigma^\star) \, \mathsf{d}\boldsymbol{\gamma}^\star \tag{4}$$

where $\mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star)$ are the *conditional* likelihood contributions of the individual observations. As we will see shortly, these likelihood contributions can be directly expressed in terms of the conditional distribution function given in Equation 3. The observations are assumed to be independent given the length-$q^\star$ random effects vector, which is integrated out using the multivariate normal density function $\phi(\boldsymbol{\gamma}^\star; \Sigma^\star)$ with zero mean vector and $\boldsymbol{\Sigma}^\star$ covariance matrix. The marginal likelihood of the parameters is obtained by evaluating the integral in Equation 4 using numerical methods. **tramME** uses the Laplace approximation for this task because its implementation provided by the **TMB** package is fast and efficient, and it scales well to high-dimensional and complicated random effects structures.

Because transformation models directly parameterize the conditional distribution function of the response variable, it is simple to express likelihood contributions in terms of the CDF. As a result, **tramME** is able to handle not only "exactly observed" values of a continuous response, but also for discrete, randomly censored or truncated observations. The practical consequence of working with the outcome CDF is that these likelihood contributions can be evaluated very efficiently without having to rely on costly numerical integration steps to get the necessary log-probabilities (see, e.g., Brilleman *et al.* 2020, Section 3.2).

Let Y_i denote an observation of the sample $Y_1, Y_2, \dots, Y_N$. First, we derive the conditional likelihood corresponding to the interval-censored case, i.e., an imprecise observation that is only known up to an interval $Y_i \in (\underline{y}, \bar{y}] \subset \Xi$. Using the mixed-effects representation of

Equation 3, it can be simply expressed as the probability,

$$\begin{aligned}\mathsf{P}(Y_i \in (\underline{y}, \bar{y}] \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) &= F_Y(\bar{y} \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) - F_Y(\underline{y} \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) \\ &= \mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star).\end{aligned} \tag{5}$$

Left- and right-censoring are accounted for, as special cases of the expression above, by setting the second part of the difference to 0 and the first part to one, respectively (Klein and Moeschberger 2003, Chapter 3).

For discrete responses, with either finite ($\Xi = \{y_1, y_2, \dots, y_K\}$) or countably infinite sample spaces ($\Xi = \{y_1, y_2, y_3, \dots\}$), the likelihood takes on the same form as Equation 5,

$$\begin{aligned}\mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star) &= \mathsf{P}(Y_i = y_k \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) \\ &= F_Y(y_k \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) - F_Y(y_{k-1} \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star)\end{aligned}$$

with the edge cases $F_Y(y_0 \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) := 0$ and $F_Y(y_K \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) = 1$.

Although, in practice, exact observations of continuous random variables are impossible, (i.e., to some extent, all measurements are subject to rounding due to precision or detection limits), responses are frequently assumed to be exact for mathematical and computational convenience (Lindsey 1999, Section 5). The conditional likelihood contributions of these cases can be written as

$$\mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star) \approx f_Z\left(h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta}) - \tilde{\boldsymbol{x}}^\top \boldsymbol{\beta}^\star - \tilde{\boldsymbol{z}}^\top \boldsymbol{\gamma}^\star\right) h'(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta}),$$

where $f_Z(\cdot)$ is the density function that corresponds to the CDF $F_Z(\cdot)$ and $h'(\cdot)$ denotes the derivative of the transformation function with respect to its argument y. The expression above is an approximation of the likelihood for an observation on an interval $(y - \varepsilon, y + \varepsilon]$ where ε is the precision limit that tends to 0.

Truncation can be regarded as a form of sample selection, where an outcome is only observed if it falls within an interval $(y_l, y_r] \subset \Xi$. Delayed entry to studies (Klein and Moeschberger 2003, Chapter 3) and the presence of time-varying covariates are typical examples of how truncated observations arise in time-to-event analysis, but the problem can also occur in other contexts when the selection into the sample is dependent on the value of the outcome. The likelihood contribution of an interval-truncated observation, using the most general, interval-censored case, can be expressed as

$$\begin{aligned}\mathcal{L}_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}^\star \mid \boldsymbol{\gamma}^\star) &= \mathsf{P}(Y_i \in (\underline{y}, \bar{y}] \mid Y_i \in (y_l, y_r], \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) \\ &= \frac{\mathsf{P}(Y_i \in (\underline{y}, \bar{y}] \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star)}{\mathsf{P}(Y_i \in (y_l, y_r] \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star)} \\ &= \frac{F_Y(\bar{y} \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) - F_Y(\underline{y} \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star)}{F_Y(y_r \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star) - F_Y(y_l \mid \boldsymbol{x}_A, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \boldsymbol{\gamma}^\star)},\end{aligned}$$

with $y_l < \underline{y} < \bar{y} \le y_r$.

The likelihood-based inference in **tramME** follows the fully parametric approach by Hothorn *et al.* (2018), who provide asymptotic results for transformation models with general independent outcomes. Following Kristensen *et al.* (2016), **tramME** performs frequentist inference for penalized parameters (random effects and coefficients of the smooth terms) based on the joint covariance matrix estimate of the fixed and random parameters utilizing the generalized delta method. As Zheng and Cadigan (2021) show, this method approximates the mean squared error of the random effect predictions, which is appropriate for inference when the model assumes specific but unobserved random effects values that are reflected in the data. Wood (2017, Section 6.10) makes the case for an analogous method for the inference on penalized smooth terms based on the Bayesian argument put forward by Nychka (1988).

3. The tramME package

The R package **tramME** provides a modular, high-level implementation of likelihood-based estimation and inference in mixed-effects additive transformation models. It builds upon several well-known and well-tested packages to specify and fit flexible distribution-free regression models. Because mixed-effects additive transformation models represent an extension of the model class implemented in the **mlt** package (Hothorn 2020), **tramME** utilizes this package to set up the necessary inputs for the basis approximation of the baseline transformation function. Inputs required for the random effects and nonlinear additive terms are created with the help of the **lme4** (Bates *et al.* 2015) and **mgcv** packages (Wood 2017), respectively. In its computational core, **tramME** uses the **TMB** package (Kristensen *et al.* 2016) to numerically approximate the integral in Equation 4 for evaluating the marginal log-likelihood function and calculating its gradients efficiently through automatic differentiation.

The design of the package and the ways in which the user interacts with it reflect the modular nature of the regression model it implements. Figure 1 presents the flowchart of the steps of specifying and estimating a mixed-effects additive transformation model, distinguishing between the conceptual steps of model specification and estimation, the software components involved in these steps, as well as the main elements of the user interface.

The model components from Equation 1 are specified using a standard formula notation. The typical formula elements are demonstrated in the dummy example below:

$$\overbrace{\texttt{Surv(y, c)}}^{1}\ \underbrace{\texttt{| s1 + s2}}_{2} \sim \overbrace{\texttt{x1 + x2}}^{3} + \underbrace{\texttt{s(x3) + s(x4, by = f)}}_{4} + \overbrace{\texttt{(x1 | g1/g2)}}^{5}$$

1. The response can be a simple variable or a "`Surv`" object of the **survival** package (Therneau 2024) to define censored or truncated outcomes. **tramME** also provides a '`Resp`' class that extends this functionality (see an example of its usage in Section 4.1).
2. Interacting terms with the basis expansion of the response are used to introduce non-proportional (i.e., outcome-dependent) effects and to specify stratification variables (the elements of the vector $\boldsymbol{x}_\mathrm{A}$ in Equation 1). The bar operator on the left side of the formula follows the convention introduced by the **tram** package.
3. Parametric shift terms, i.e., the elements of the $\boldsymbol{x}_\mathrm{B}$ vector in Equation 1.
4. The nonlinear additive terms are specified using the same notation as in **mgcv**.
5. The notation to define the random effects structure follows **lme4**.

As special cases of penalized additive terms, random intercepts and slopes can be specified through the **mgcv** notation with `s(g1, bs = "re")` and `s(g1, x1, bs = "re")`, respec-

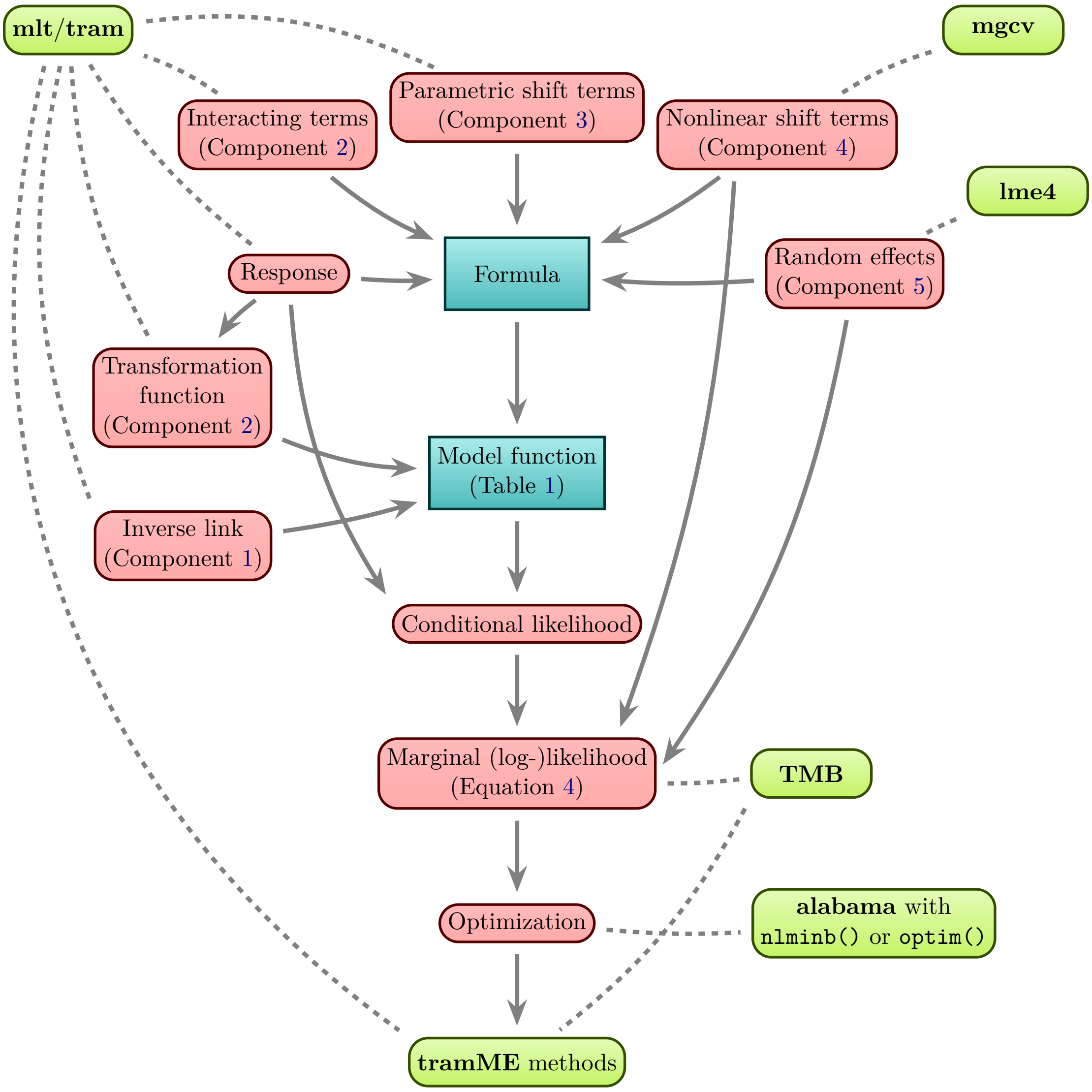

Figure 1: The flowchart of modeling with **tramME**. The red boxes show the elements of the statistical model (see Equation 1 for Model Components 1 to 5). The blue boxes indicate the main building blocks of the user interface, while the green boxes identify the key software components utilized by **tramME**. The grey arrows represent the process of specifying and estimating a mixed-effects transformation model. The dashed lines indicate which software components are used at specific steps of this process.

tively. However, it is recommended to use the **lme4** notation, which allows for a more compact representation of more complex random effects structures, such as nested and correlated random effects terms. Additionally, **tramME** internally distinguishes between "true" random effects (i.e., terms that capture characteristics of the sampling procedure) and penalized terms that only formally constitute random effects (Hodges 2014, Chapter 13) based on which notation is used to specify them and handles them differently in certain methods.

To specify the appropriate parameterization for the baseline transformation function and the

Function	Model name	F_Z	$\boldsymbol{a}(y)$
`LmME()`	Normal linear	Φ	Linear basis
`BoxCoxME()`	Non-normal (Box-Cox-type)	Φ	Bernstein basis
`ColrME()`	Continuous outcome logistic	F_{SL}	Bernstein basis
`CoxphME()`	Flexible parametric proportional-hazards	F_{MinEV}	Bernstein basis
`SurvregME()`	Parametric survival	Φ, F_{SL}, F_{MinEV}	Linear basis, log-linear basis
`PolrME()`	Discrete ordinal	Φ, F_{SL}, F_{MinEV}, F_{MaxEV}	Discrete basis
`LehmannME()`	Lehmann-alternative	F_{MaxEV}	Bernstein basis

Table 1: The model functions of **tramME** represent various combinations of the inverse link function (F_Z) and the basis functions used to approximate the baseline transformation ($\boldsymbol{a}(y)$). Φ stands for the CDF of the standard Gaussian distribution, $F_{\text{SL}} = \text{logit}^{-1}$, $F_{\text{MinEV}} = \text{cloglog}^{-1}$ and F_{MaxEV} are the standard logistic, minimum extreme value (Gompertz), and maximum extreme value (Gumbel) distribution functions, respectively.

type of inverse link function, the **tramME** package provides seven model functions (Table 1). These functions are defined as wrappers around the general `tramME()` function and represent specific special cases of the general mixed-effects additive transformation model (Equation 1) for particular outcome types and scales where proportionality is assumed. The four basis transformation options are the linear basis ($\boldsymbol{a}(y) = (1, y)^\top$), the log-linear basis ($\boldsymbol{a}(y) = (1, \log(y))^\top$), the Bernstein polynomial basis (see Equation 2), and the discrete basis ($\boldsymbol{a}(y) = (\mathbb{1}(y = 1), \mathbb{1}(y = 2), \dots, \mathbb{1}(y = K - 1))^\top$, where $\mathbb{1}(\cdot)$ denotes the indicator function) with optional interactions with basis transforms of some covariates. The **tramME** package currently offers four options for the inverse link function, including the standard Gaussian distribution function (Φ), the cumulative distribution function of the minimum extreme value distribution ($F_{\text{MinEV}}(z) = 1 - \exp(-\exp(z)) = \text{cloglog}^{-1}(z)$), the CDF of the maximum extreme value distribution ($F_{\text{MaxEV}}(z) = \exp(\exp(-z))$), and the standard logistic distribution function ($F_{\text{SL}}(z) = (1 + \exp(-z))^{-1} = \text{logit}^{-1}(z)$).

The conditional likelihood function is established by defining the model components through formula and model function choices, as well as the response type, which can be exactly observed continuous, discrete, censored, or truncated. The **TMB** package is used to numerically integrate over the vector of random effects (Equation 4), using the random effects densities implied by the penalty terms. The score functions are evaluated using automatic differentiation. To maintain the monotonicity of the fitted conditional CDFs, constrained optimization is performed with an augmented Lagrangian algorithm (Nocedal and Wright 2006, Chapter 17), using the **alabama** package (Varadhan 2023) and optimization routines implemented in `stats::nlminb()` and `stats::optim()`.

The "`tramME`" object fitted by **tramME** can be analyzed and presented using standard functionality provided by the package. Some of these methods are wrappers around functionality provided by the **mlt** and **TMB** packages. Inference on the parameters and model comparison is based on likelihood theory, and typical methods such as `coef()`, `vcov()`, `logLik()`, `residuals()` are available. Random effects and fitted additive terms can be evaluated in a similar way to other additive and mixed modeling packages. Because **tramME** models approximate the conditional outcome distribution, predictions on various scales (such as probability, density, odds, hazard) can be evaluated using the `predict()` method. Additionally, the `simulate()` function can be used to generate new samples from a fitted model by numerically inverting the conditional distribution function. Examples of the most important functionality are presented in Section 4 and the package vignettes (accessible with `vignette(package = "tramME")`).

4. Illustrations

This section presents two applications of the mixed-effects additive transformation modeling framework implemented in the **tramME** package. Both cases pose some technical challenges to alternative mixed and additive regression approaches, in terms of the outcome type, study design, and data or model structure that typically require computationally intensive bespoke models. As we will see, both regression problems can be handled within the transformation model framework of **tramME** in a relatively straightforward way. Although these examples do not aim to provide comprehensive analyses, their purpose is to showcase the range of functionalities provided by **tramME**.

For simpler illustrative examples and direct comparisons to alternative software implementations, where they are available, the reader is referred to Tamási and Hothorn (2021) and the package vignettes.

Because both example applications in this section involve complicated models and relatively large, complex data structures, we first slightly adjust the default optimization settings of the package to keep the calculations tractable. In the code below, the default tolerance level of the `nlminb` optimizer routine is increased from its default value of `1e-10`. Additionally, (mostly for demonstration purposes, as these changes will not have too much further effect on the results) we set the maximum numbers of iterations and evaluations to 300 and 400, respectively, which are the default numbers used in the **glmmTMB** package. Finally, we also activate the multi-core functionality of the **TMB** package.

```
R> library("tramME")
R> TMB::openmp(n = 4, DLL = "tramME")
R> opt <- optim_control(method = "nlminb",
+    iter.max = 300, eval.max = 400,
+    rel.tol = sqrt(.Machine$double.eps))
```

4.1. Child mortality in Nigeria

Kneib (2006a) presents an analysis of the childhood mortality data collected in the 2003 Nigeria Demographic and Health Survey, using a geoadditive hazard regression model. The original analysis investigated how various demographic, social and health-related factors of

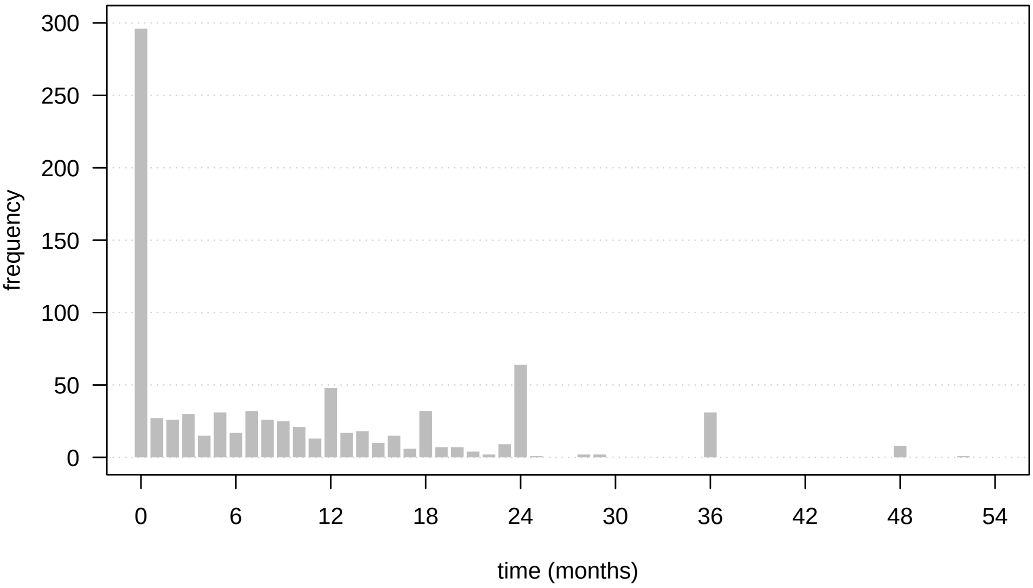

Figure 2: Frequencies of observed survival times in months in the Nigerian childhood mortality dataset.

the mothers and their children are associated with the mortality in children under five years of age. Given that the survey recorded the place of residence on a district level, the original analysis also incorporated the potential spatial variation of mortality. The dataset is available from `https://www.uni-goettingen.de/de/551628.html`, and additional details of the data and the original analysis are given in Kneib (2006a,b).

The survey collected survival times of children through retrospective interviews, resulting in exact event times being recorded only for the first two months due to memory effects. For other survival times, rounding to the nearest month was applied, leading to *interval-censored* outcomes that needed to be taken into account in the model estimation. Figure 2, which reproduces Figure 4 in Kneib (2006a), displays the monthly frequencies of observed times until death in the dataset. A concentration of reported events at months 12, 18, 24, 36, and 48 is clearly visible, while events between these time points are relatively infrequent, which is a common occurrence in retrospective interview data. To account for this heaping effect, following the analysis of Kneib (2006a), these observations were treated as interval-censored outcomes with 6- or 12-month-wide, non-overlapping, symmetric censoring intervals. In contrast to the event times, right-censoring times were exactly recorded as differences between interview and birth dates of the children.

One of the key questions of the original analysis was how breastfeeding affected child mortality. In many cases, the breastfeeding status changed over the course of the followup, and the time of cessation was recorded. This form of *time-dependent covariate* can be handled in survival analyses by augmenting the dataset and splitting up cases at cessation times into right-censored and *left-truncated* new data points. The original dataset of this example is already set up in this augmented format, otherwise the `tmerge()` function of the **survival** package could have been used to build the new dataset (Therneau, Crowson, and Atkinson 2024). The augmented dataset of this example consists of 11 280 observations. We can define the necessary interval-censored and left-truncated response object with the `Resp()` function of **tramME**. The original dataset encoded the exactly observed and right-censored events by setting the `censindinterval` variable to 1 and 0, respectively, while making the `heapingleft` and `heapingright` variables equal to the event or censoring times. The interval-censored observations are identified by differing `heapingleft` and `heapingright` boundary points and `censindinterval = 0`. Left-truncated observations, which occur when the breastfeeding indicator changes, are represented by the `trunctime` variable larger than 0. "`Resp`" objects are defined similarly to objects of the "`Surv`" class with `type = "interval2"`, i.e., exact observations correspond to cases where the left and right censoring boundaries are equal, and right censoring is encoded by setting `cright = Inf`. Additionally, `Resp()` enables the setting of truncation times for the data points.

```
R> nigeria$y <- with(nigeria, Resp(
+    tleft = ifelse(trunctime > 0, trunctime, NA),
+    cleft = heapingleft,
+    cright = ifelse(heapingleft == heapingright & censindinterval < 1,
+                    Inf, heapingright),
+    bounds = c(0, Inf))
+  )
R> nigeria$y[c(1, 2, 3, 13, 14, 21, 22)]

[1] {(   0,  30]}        {( 395, Inf]}        {(1215, Inf]| > 395}
[4] { 7}                 {(  30,  61]}        {(  61, Inf]}
[7] {(  91, 122]| >  61}
```

The code block above returns examples of the various possible outcome types. The first observation is left censored, i.e., we only know that the corresponding child died before the 30th day after birth. Observations two and three belong together and represent a case where the breastfeeding cessation occurred at day 395 and the child was 1215 days old at the end of the followup. Technically, this is achieved by defining a right-censored observation at day 395 and a right-censored observation at day 1215, which is truncated at the day when the breastfeeding status changed. The fourth data point shows a case where the child died on the seventh day after birth. The fifth example is interval-censored, where the event of death occurred between days 30 and 61. The two last data points, once again, represent a single case, where breastfeeding stopped at day 61, and the child died after the 91st but before the 122nd day after birth.

To investigate the relationship between child mortality and the various demographic and health factors, we estimate proportional hazards models of the general form,

$$F_T(t \mid \boldsymbol{x}) = 1 - \exp\left\{-\exp\left[h(t) + \sum_{j=1}^{J} f_j(x_j) + g(d) + \beta_{\texttt{bf}}\texttt{bf}(t) + \tilde{\boldsymbol{x}}^\top \boldsymbol{\beta}\right]\right\}. \tag{6}$$

This model specifies the conditional survival distribution using a set of (penalized) smooth functions of some continuous predictors $f_j(x_j)$, a term for capturing the unobserved, district-level spatial variability in the outcome $g(d)$, the fixed-effect of a time-dependent covariate (breastfeeding $\texttt{bf}(t)$, which takes on 1 until cessation and 0 afterwards) and finally some fixed-effects of time-fixed categorical covariates $\tilde{\boldsymbol{x}}^\top \boldsymbol{\beta}$. By setting $F_Z(z) = F_{\text{MinEV}}(z) = 1 -$

$\exp(-\exp(z))$, the transformation function $h(t)$ approximates the log-cumulative baseline hazard function and the parametric fixed-effects coefficients ($\beta_{\texttt{bf}}$, $\boldsymbol{\beta}$), as well as the penalized smooth terms ($f_j(x_j)$ and $g(d)$), are interpreted on the log-hazard scale.

Although the geoadditive proportional hazards model presented by Kneib (2006a) can be regarded as a special case of a mixed-effects additive transformation model (with $F_Z = F_{\mathrm{MinEV}}$ and a smooth parametric approximation of the cumulative baseline hazard), there are important differences between **tramME** and the approach presented in that article in terms of the parameterization of the smooth terms and the estimation strategy. Moreover, to keep our example application as simple as necessary, we only consider a subset of the potential covariates available in the original dataset. Despite the differences, it is worth noting that the results of this section are qualitatively very similar to those obtained by Kneib (2006a).

As an initial step, we investigate the (unconditional) spatial patterns in the survival data by estimating a parametric proportional hazards model with a single spatial term (i.e., keeping only the shift term $g(d)$ in the proportional hazards model in Equation 6) as a reduced rank Gaussian Markov random field smooth (Wood 2017, Section 5.8.1) defined on the set of districts. The list containing the neighborhood information (`nb`) is created from the adjacency matrix of the map object; the code to set up these inputs can be found in the supplementary material.

```
R> m_sp <- CoxphME(y ~ s(district, bs = "mrf", k = 20, xt = list(nb = nb)),
+    data = nigeria, control = opt, bounds = c(0, Inf), support = c(1, 1400),
+    log_first = TRUE, order = 8)
R> summary(m_sp)


Additive Parametric Cox Regression Model

        Formula:
y ~ s(district, bs = "mrf", k = 20, xt = list(nb = nb))

        Fitted to dataset nigeria

        Fixed effects parameters:
        =========================

No estimated shift coefficients.

        Smooth shift terms:
        ===================

             edf
s(district) 14.3

        Log-likelihood: -4817 (npar = 10)
```

Note that we explicitly set the bounds and the support of the outcome variable in the function call above. While `bounds` defines the theoretical boundaries of the response, `support` serves a more practical purpose: It defines the interval on which the flexible Bernstein polynomial approximation of the transformation function is performed (see Section 2.2); the function beyond this interval is assumed to be linear. Because, as Figure 2 shows, events beyond four years are very rare this seems a reasonable choice. In most analytical settings, **tramME** sets (through the **mlt** package) the `support` automatically using empirical quantiles, but in the current application the interval-censored and left-truncated response prevents the calculation of these. When the `log_first = TRUE` option is utilized for the smooth approximation of the transformation function, the Bernstein basis functions are computed using the natural logarithm of the response. This approach usually results in a significantly improved approximation for strictly positive right-skewed variables. The order of the Bernstein basis is set to 8 (the default is `order = 6`). For an in-depth discussion of practical considerations related to transformation modeling, including those that apply to **tramME**, refer to Hothorn (2020).

The point predictions of the spatial effects and the corresponding standard errors are returned by the `smooth_terms()` function.

```
R> sm_sp <- smooth_terms(m_sp)
R> head(sm_sp[[1]])

  district s(district)      se
1        1    -0.14201 0.17883
2        2    -0.25758 0.11050
3        3     0.24382 0.08732
4        4    -0.16909 0.15166
5        5     0.06247 0.11695
6        6     0.03603 0.14008
```

The left panel of Figure 3 presents the point estimates of the district-level spatial effects, which are interpreted on the log-hazard scale in our proportional hazards specification. To assess the size of the standard errors relative to the effect magnitudes, we calculate 95% confidence intervals and identify regions with significant positive or negative and insignificant spatial effects. The code used to generate these plots is omitted here but can be found in the replication material.

We now extend our proportional hazards specification with additional parametric and additive smooth terms. Additionally to the spatial term, we include the nonlinear shift effect of the mother's age and the number of the child in the birth order in the family. To enhance comparability with the results of Kneib (2006a), both of these effects are approximated with P-spline bases. The time-varying indicator of breastfeeding, the sex of the child, the educational level of the mother (at least primary education / no education), and the place of delivery (hospital / home) are included as parametric fixed-effects terms.

```
R> m_ph <- update(m_sp, . ~ . + breastfeed + sex + education + delivery +
+    s(ageatbirth, bs = "ps") + s(birthorder, bs = "ps"))
R> summary(m_ph)


Additive Parametric Cox Regression Model
```

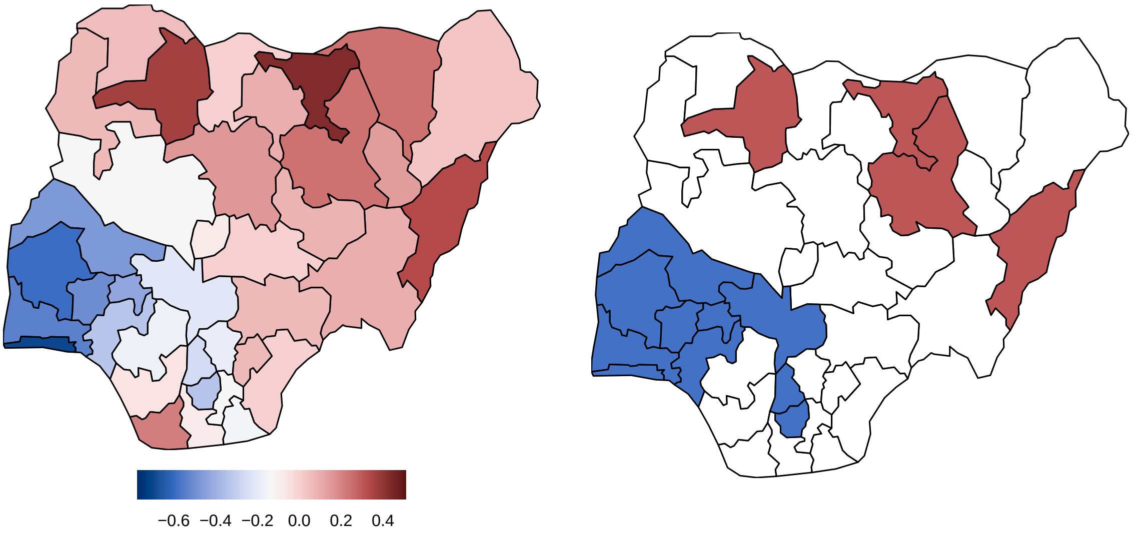

Figure 3: *Left panel*: Point estimates of the spatial effects on the log-hazard scale in the unconditional spatial model of the Nigerian child mortality data. *Right panel*: Districts with significantly negative (blue), positive (red) and insignificant (white) values based on the 95% Wald confidence intervals of the spatial effects.

```
        Formula:
y ~ s(district, bs = "mrf", k = 20, xt = list(nb = nb)) + breastfeed + 
    sex + education + delivery + s(ageatbirth, bs = "ps") + s(birthorder, 
    bs = "ps")

        Fitted to dataset nigeria

        Fixed effects parameters:
        =========================

                 Estimate Std. Error z value Pr(>|z|)    
breastfeedyes     -4.5128     0.1249  -36.14  < 2e-16 ***
sexmale            0.0729     0.0738    0.99  0.32309    
educationno       -0.3647     0.0954   -3.82  0.00013 ***
deliveryhospital  -0.4630     0.1071   -4.32  1.5e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Smooth shift terms:
        ===================

                edf
```

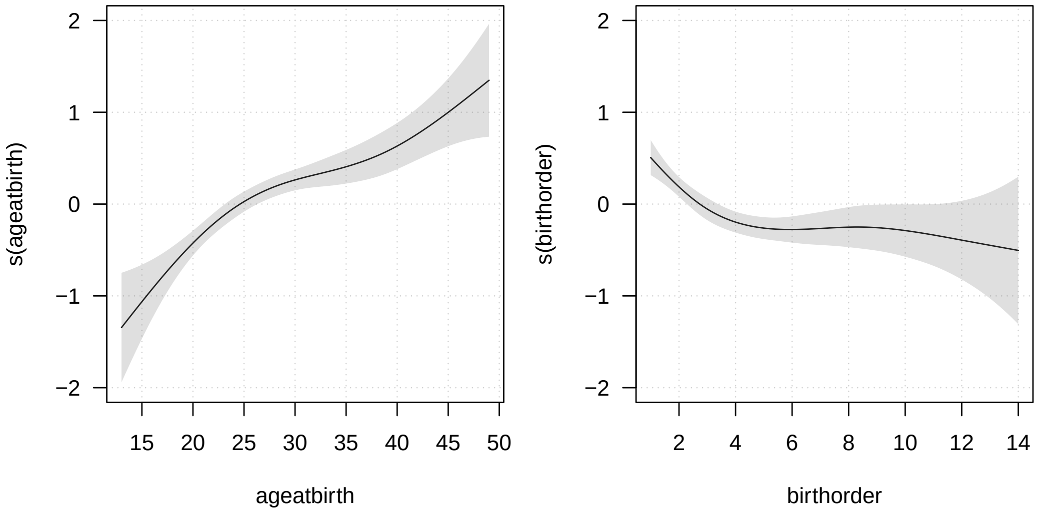

Figure 4: Penalized additive terms for the mother's age (*left panel*) and the number of the child in the birth order (*right panel*) in the proportional hazards model of child mortality in Nigeria. The solid lines represent the point estimates, while the shaded areas are the 95% pointwise confidence intervals.

```
s(ageatbirth)  4.34
s(birthorder)  4.15
s(district)   14.79

        Log-likelihood: -3181 (npar = 18)
```

The results suggest strong protective effect of breastfeeding on child mortality that corresponds to a hazard ratio of $\exp(\hat{\beta}_{\texttt{bf}}) = 0.011$.

The additive smooth terms for the mother's age (`ageatbirth`) and the number of the child in the birth order (`birthorder`) can be evaluated and plotted as

```
R> sm_ph <- smooth_terms(m_ph)
R> plot(sm_ph[2:3], panel.first = grid(), ylim = c(-2, 2))
```

According to Figure 4, both variables have significant nonlinear effects on the log-hazard scale based on their 95% confidence intervals. The corresponding effective degrees of freedom estimates can be extracted separately with the `edf_smooth()` method.

To aid model checking, **tramME** provides a `residuals()` method that implements a score residual corresponding to a constant term that is fixed at zero (see Tamási and Hothorn 2021, for the definition and its usage in a different regression setting). It can be shown that in a proportional hazards specification these score residuals are equal to the martingale (or Lagakos) residuals, which are useful tools of model diagnostics in regression analysis of survival data (Farrington 2000).

```
R> res <- residuals(m_ph)
```

Martingale residuals are frequently used to identify outliers in the data. The left panel of Figure 5 plots the residuals of the proportional hazards model against the observation indices. A clear outlier is marked with red circle. The observation in question is a child with the following characteristics

```
R> (outl <- model.frame(m_ph)[which.min(res), ])
```

```
              y breastfeed  sex education delivery district ageatbirth
1493 (1506, Inf]         no male  primary+     home        3         40
     birthorder
1493          9
```

The model-based estimate of survival probability beyond 1506 days for a boy with the same characteristics can be calculated with the `predict()` method:

```
R> predict(m_ph, type = "survivor", newdata = outl, q = 1506)
```

```
y             [,1]
  1.51e+03 0.01561
```

This very low probability explains why this observation is identified as an outlier in our model. To get a better picture of the conditional distribution of survival time, we can calculate the conditional quantiles for the same set of predictor values using `predict()` with `type = "quantile"`.

```
R> predict(m_ph, type = "quantile", newdata = outl, p = c(0.1, 0.5, 0.9))
```

```
prob      [,1]
    0.1  2.507
    0.5 14.089
    0.9 89.992
```

If we want to investigate further how the various predictors contribute to the negative prognosis of this outlier, we can decompose the predictions manually using the `model.matrix()` and `coef()` methods of the '`tramME`' object. First, we calculate the individual elements of the linear predictor (which is defined on the log-cumulative hazard scale in the case of '`CoxphME`') using the *complete* (with penalized and unpenalized elements) coefficient vector. We can then group and sum the individual elements to baseline transformation, additive smooth and fixed-effects terms, respectively, based on the parameter names. Finally, we transform the decomposed predictions to the cumulative hazard scale.

```
R> mm <- model.matrix(m_ph, data = outl)
R> b <- coef(m_ph, complete = TRUE)
R> Xm <- cbind(mm$Yr, mm$X, t(as.matrix(mm$Zt)))
R> vns <- sapply(names(b), function(n) all.vars(str2lang(n))[1])
R> lch <- sapply(split(Xm*b, vns), sum)
R> exp(lch)
```

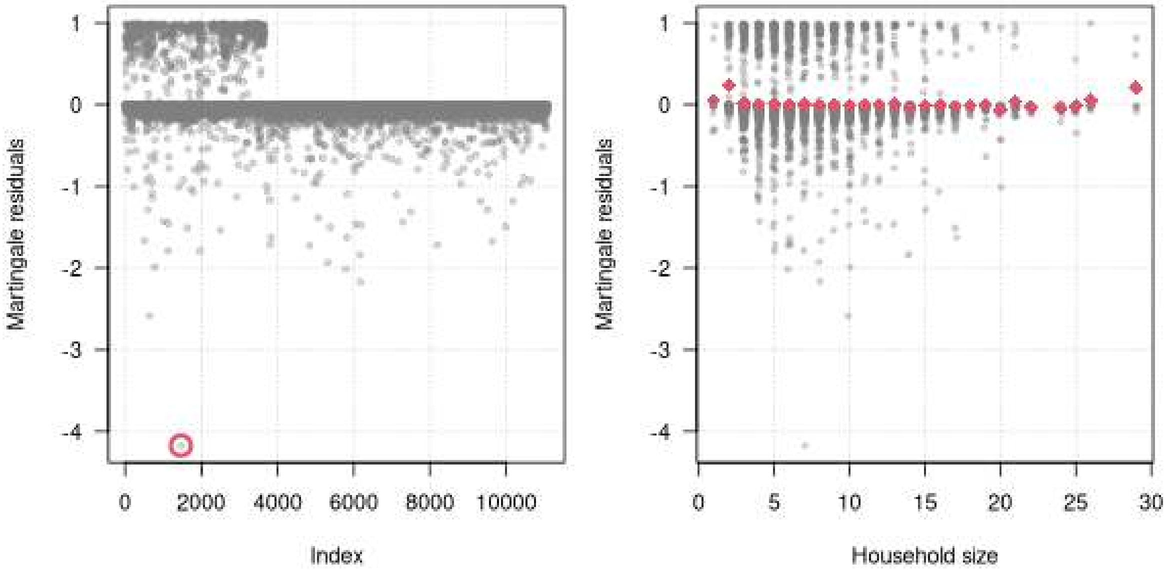

Figure 5: Residual plots against the observation index (*left panel*) and against the household size variable. In the left plot, the red circle shows an outlier observation. In the right-hand side plot, red diamonds show group means of the residuals.

```
     ageatbirth        birthorder    breastfeedyes deliveryhospital 
         1.8814            0.7734           1.0000           1.0000 
       district       educationno          sexmale                y 
         1.4569            1.0000           1.0756           1.8244 
```

The cumulative hazard corresponding to the reference level of the individual predictors is 1.82, which translates to a $\exp(-\Lambda(t \mid \boldsymbol{x})) = 0.16$ survival probability beyond 1506 days. Although being the ninth child of the family decreases the cumulative hazard by 23%, this is offset by the effect of the mother's age (88% increase) and the effect of the district (46% increase).

Score residuals can also be used to check whether variables that are not included in the model contain relevant information on the outcome. Plotting the residuals against a potential predictor, or testing formally the independence between the two could help to identify ways to extend our model. As an example, the right panel of Figure 5 shows the residuals from the additive proportional hazards specification against the household size variable. Because the martingale residuals are shown to have zero mean in large samples, marked departures from zero in the conditional mean structure is a sign for a relevant predictor. Although, based on the residual plot, the household size variable seems to be a relevant predictor for survival, we do not pursue this question further and extend our model in a different direction.

Our additive transformation model assumes proportional hazards, i.e., time-independence in the covariate effects. However, this assumption is often incompatible with the data in practice. In our analysis based on proportional hazards, the negative coefficient of the indicator variable for mothers' education (with levels of no education and at least primary education) suggests that children of mothers with no formal education have a lower risk of death. While we must be cautious about drawing causal conclusions based on our analysis of these observational data, this finding is somewhat unexpected. One explanation for this is that the effect of maternal education may depend on the survival time of the child. We can extend our regression model to allow for non-proportional hazards for the education indicator as

$$F_T(t \mid \boldsymbol{x}) = 1 - \exp\left\{ -\exp\left[h(t \mid \texttt{edu}) + \sum_{j=1}^{J} f_j(x_j) + g(d) + \beta_{\texttt{bf}}\texttt{bf}(t) + \tilde{\boldsymbol{x}}^\top \boldsymbol{\beta} \right] \right\},$$

where `edu` is the indicator of education level. The conditional transformation function can be further written as $h(t \mid \texttt{edu}) = h(t) + \beta_{\texttt{edu}}(t)\texttt{edu}$, which highlights the time dependence of the log-hazard ratio. In **tramME**, this model can be specified as

```
R> m_tv <- CoxphME(y | education  ~ breastfeed + sex + delivery +
+    s(district, bs = "mrf", k = 20, xt = list(nb = nb)) +
+    s(ageatbirth, bs = "ps") + s(birthorder, bs = "ps"),
+    data = nigeria, control = opt,
+    bounds = c(0, Inf), support = c(1, 1400),
+    log_first = TRUE, order = 8)
R> summary(m_tv)
```

```


Stratified Additive Parametric Cox Regression Model

	Formula:
y | education ~ breastfeed + sex + delivery + s(district, bs = "mrf",
    k = 20, xt = list(nb = nb)) + s(ageatbirth, bs = "ps") +
    s(birthorder, bs = "ps")

	Fitted to dataset nigeria

	Fixed effects parameters:
	=========================

                 Estimate Std. Error z value Pr(>|z|)
breastfeedyes     -4.5252     0.1251  -36.16  < 2e-16 ***
sexmale            0.0772     0.0738    1.05      0.3
deliveryhospital  -0.4222     0.1067   -3.96  7.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Smooth shift terms:
	===================

                edf
s(ageatbirth)  4.29
s(birthorder)  4.06
s(district)   15.08
```

```
Log-likelihood: -3156 (npar = 26)
```

To demonstrate non-proportional hazards in the effect of education, we plot the conditional baseline transformation functions, which correspond to the conditional log-cumulative hazards when we use `CoxphME()`, for two children who only differ in their mothers' education levels and their other predictors are set to the reference levels (with the `baseline_only = TRUE` argument of the `confband()` method). Additionally, we can evaluate the same functions on the probability scale and plot the survivor curves by setting `type = "survivor"`.

```
R> nd <- model.frame(m_tv)[rep(9, 2), ]
R> nd$education <- factor(c("no", "primary+"), levels = c("primary+", "no"),
+    labels = c("primary+", "no"))
R> cb_lch <- confband(m_tv, newdata = nd,
+    q = seq(1, 1000, length.out = 100), baseline_only = TRUE)
R> cb_sur <- confband(m_tv, newdata = nd, type = "survivor",
+    q = seq(1, 1000, length.out = 100), baseline_only = TRUE)
R> plot(cb_lch, single_plot = TRUE, lty = 1, lwd = 2,
+    panel.first = grid(), xlim = c(0, 1000), ylim = c(-2, 1),
+    ylab = "h(y | education)")
R> plot(cb_sur, single_plot = TRUE, lty = 1, lwd = 2,
+    panel.first = grid(), xlim = c(0, 1000), ylim = c(0, 1),
+    ylab = "Survivor")
```

As Figure 6 shows, the predicted curves cross both on the log-cumulative hazard and survivor scales, which is a sign of non-proportional hazards, i.e., the effect of the mother's education on the child's survival depends on time. However, the accompanying asymptotic confidence intervals overlap, indicating lack of strong evidence for these differences.

4.2. Burn victim recovery

In regression modeling of time-to-event outcomes, a common practice is to adopt a distribution-free, "semiparametric" approach. The widely-used Cox proportional-hazards model (Cox 1972) is a prime example for this. However, the distribution-free approach of transformation models can be beneficial in other cases where selecting a parametric conditional distribution for the outcome is challenging. The following example application illustrates such a scenario, highlighting the usefulness of the transformation models in diverse settings.

Spronk, Van Loey, Sewalt, Nieboer, Renneberg, Moi, Oster, Orwelius, Van Baar, Polinder, and the Quality of Life Study Group (2020) investigate the temporal patterns of recovery after burn injuries by analyzing data from a collection of 10 studies on burn patients. The pooled dataset consists of observations from different countries and represents a diverse population of patients in terms of demographics, burn etiology, severeness of injury and care received. The follow-up of recovery varied across studies from one year to several years and multiple measurements of the health-related quality-of-life (HRQoL) of the patients were taken, which permits a detailed longitudinal analysis of the recovery process. To harmonize the HRQoL outcome across studies, the EuroQoL 5-Dimensional measure (EQ-5D, Rabin and de Charro 2001) was calculated for all patients. The pooled dataset consists of 3744 observations of 1687

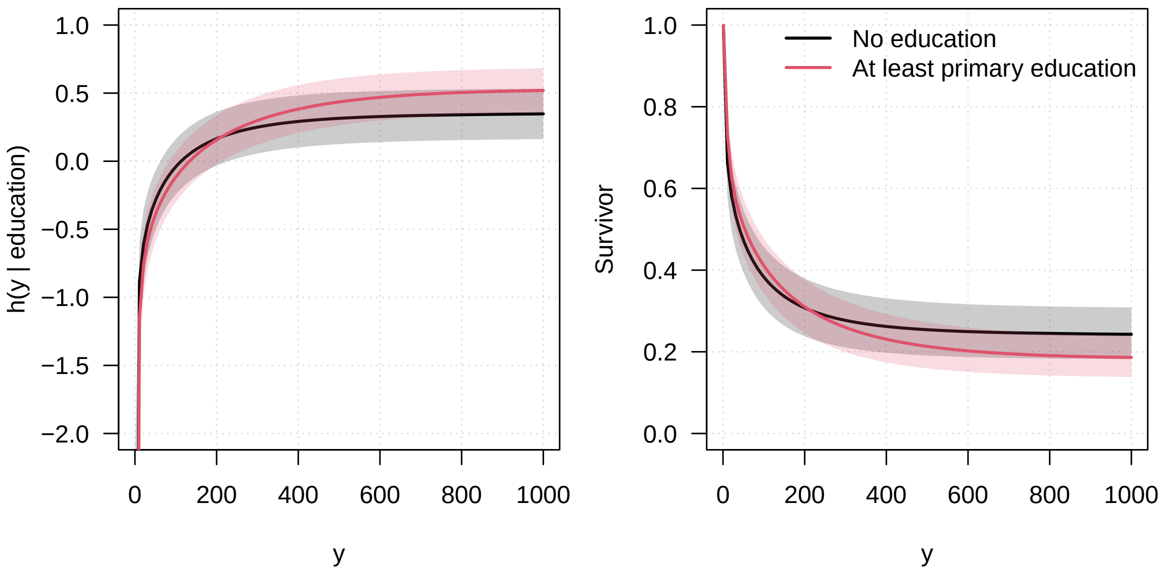

Figure 6: Conditional log-cumulative hazard and survivor curves for predictor configurations where only the maternal education level differs. The shaded areas indicate the 95% percent pointwise confidence intervals.

patients and is available as an online appendix of the original article at `https://doi.org/10.1371/journal.pone.0226653.s001`.

Mixed-effects (or hierarchical) models are popular tools in individual participant data (IPD) meta-analysis. Unlike in traditional aggregate-level meta analyses, the focus here is on combining results from multiple studies based on individual-level information (Riley, Tierney, and Stewart 2021). In *one-stage IPD meta-analysis*, the pooled data from various studies are analyzed using mixed-effects models that are able to account for correlations due to between-study heterogeneities. This task often requires very complicated multilevel models that can lead to computational challenges. Tamási, Crowther, Puhan, Steyerberg, and Hothorn (2022) discuss how mixed-effects transformation models of various complexities can be used to perform IPD meta-analysis.

In the following reanalysis of the individual-level data, we extend on the original regression analysis presented by Spronk *et al.* (2020) in three main directions using the additive mixed-effects transformation model methodology: First, we adopt a distribution-free approach of modeling the quality-of-life index, i.e., instead of assuming conditional normality for the outcome, we approximate the shape of its conditional distribution in a data-driven way. Secondly, to capture nonlinear patterns in the effects of the continuous predictors, on HRQoL (most importantly time since burn), we include these as additive smooth terms. GAMMs have recently been proposed to estimate nonlinear predictor-outcome relationships in meta-analysis of IPD (Belias, Rovers, Hoogland, Reitsma, Debray, and IntHout 2022). Finally, we use a more complicated random effects structure to allow for study-level departures from the mean recovery time patterns by including correlated random slopes. Including random slopes in IPD meta-analytic transformation models is discussed by Tamási *et al.* (2022).

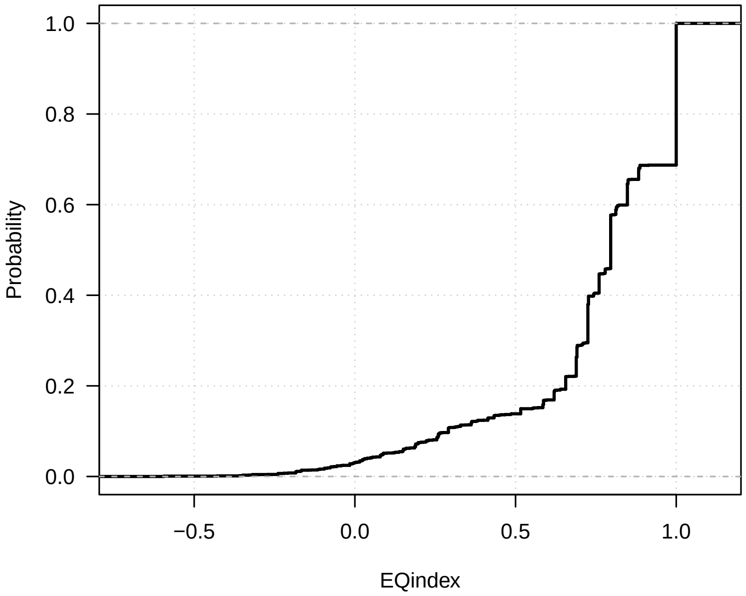

Figure 7: The empirical distribution of the quality-of-life index (EQ-5D) in the pooled dataset of burn patients.

In this example analysis, we focus on modeling the EQ-5D utility scores of the patients, which is a summary measure of HRQoL with a maximum value of 1, which corresponds to the sate of full health. A value of 0 of this index is associated with a health state equivalent to death, while values less than zero indicate a life quality that is "worse than death". The empirical distribution function of the EQ-5D index in the pooled dataset is shown in Figure 7. In addition to being bounded from above at 1, the plot highlights that some values, especially those close to the right boundary, are present many times in the sample. This is not surprising given that the index is calculated by combining five discrete scores, representing five dimensions of life quality, applying a complicated weighting scheme. In fact, about 31% of the response values are 1 in the sample. The mixed-effects regression analysis of the EQ-5D index in Spronk *et al.* (2020), similarly to typical analyses of quality-of-life scores, ignores these properties of the outcome and assumes conditional normality, as well as exactly observed response values.

As we saw in Section 2, the transformation model framework can accommodate responses with idiosyncratic properties (boundedness, discrete-continuous mixed-types etc.). By using a flexible approximation of the transformation function, the shape of the conditional outcome distribution can be estimated without *a priori* committing to a parametric distribution type. In these approximations, boundaries for the outcome distribution can also be introduced. Because continuous and discrete (as well as censored) observations are allowed in transformation models (see Section 2.3), even in the same dataset, rounding effects and discrete-continuous mixed-type outcomes can also be taken into account. (We already used this in the previous application in Section 4.1.) In our current example, **tramME** offers various strategies to handle the large number of ties in the response variable based on additional information available and the assumptions we are willing to make. We can ignore the whole problem and treat all responses as exactly observed values from a continuous distribution by using densities as likelihood contributions. As an alternative, we can ignore the ties that are lower than one but treat the HRQoL measurements that correspond to "full health" (EQ-5D index equals to 1) as discrete observations. The package vignette (Tamási and Hothorn 2023) presents an example where the a mixed-type discrete-continuous outcome is analyzed. If we know the values of the item scores and weights used for the calculation of the quality-of-life index, we can set the censoring intervals exactly and use them to define likelihood contributions for the discrete outcomes. Finally, we can opt for an empirical likelihood approach and use the discrete (multinomial) likelihood of the response (Owen 2001) $F_Y(y^{(k)} \mid \boldsymbol{x}, \boldsymbol{\gamma}) - F_Y(y^{(k-1)} \mid \boldsymbol{x}, \boldsymbol{\gamma})$, where $y^{(k)}$ is the outcome that corresponds to the kth observation in the ordered unique values $y^{(1)}, y^{(2)}, \ldots$ of the sample. In this example analysis we follow this latter approach and note that the other feasible strategies mentioned here would give qualitatively similar results.

To set up the response variable according to the empirical likelihood approach, we can define interval-censored observations for all unique response values in the dataset.

```
R> burn$EQi <- with(burn, as.Surv(R(EQindex, as.R.interval = TRUE)))
R> head(burn$EQi)
```

```
[1] [0.8150, 0.8176] [0.8834, 0.8872] [0.8500, 0.8517] [0.9139, 1.0000]
[5] [0.9139, 1.0000] [0.8150, 0.8176]
```

Using the discrete likelihood for the estimation does not necessarily mean that we must treat the outcome as a discrete variable in our statistical model. It is possible to choose a continuous parameterization of the outcome distribution and approximate the transformation function as a smooth function using continuous basis transformations, separate from defining the likelihood contributions. This approach combines elements from the discrete ordered regression framework proposed by Liu *et al.* (2017a) that uses a discrete likelihood and parameterization, and the smooth continuous approach represented by Lohse, Rohrmann, Faeh, and Hothorn (2017). Tian *et al.* (2020) provides a detailed empirical comparison between the two methods.

By using the empirical likelihood in conjunction with a continuous parameterization of the baseline transformation, we obtain a more concise representation of the conditional outcome distribution, requiring fewer parameters than the alternative discrete approach. As a result, the smooth parameterization provides computational advantages by reducing the dimensionality of the model, making it easier and faster to estimate. Moreover, since our aim in this example is to calculate distributional predictions by evaluating the conditional outcome distribution on various scales, a continuous parameterization that can interpolate between discrete data points is more consistent with the objectives of the analysis and the continuous nature of the phenomenon under investigation.

We estimate a continuous outcome logistic regression model (Lohse *et al.* 2017) of the HRQoL in the dataset of $i = 1, 2, \ldots, N$ studies of $j = 1, 2, \ldots, n_i$ patients in each, with $k = 1, 2, \ldots, T_{ij}$ HRQoL assessments Y_{ijk} at t_{ijk} time points. To analyze differences in recovery patterns, we also adjust for some important patient-specific information such as age, gender, the percentage of total body surface area burned (`TBSA`) and length of stay in hospital (`LOS`).

$$
\begin{aligned}
F_{Y_{ijk}}(y \mid \boldsymbol{x}_{ijk}, \boldsymbol{\gamma}_{ij}) = \operatorname{expit}\Big(&h(y) + f_1(t_{ijk}) + f_2(\texttt{TBSA}_{ij}) + f_3(\texttt{LOS}_{ij}) + f_4(\texttt{age}_{ij}) \\
&+ \beta\texttt{male}_{ij} + \gamma_{1,ij} + \gamma_{2,i} + \gamma_{3,i} t_{ijk}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\frac{F_{Y_{ijk}}(y \mid \boldsymbol{x}_{ijk}, \boldsymbol{\gamma}_{ij})}{1 - F_{Y_{ijk}}(y \mid \boldsymbol{x}_{ijk}, \boldsymbol{\gamma}_{ij})} = \exp\Big(&h(y) + f_1(t_{ijk}) + f_2(\texttt{TBSA}_{ij}) + f_3(\texttt{LOS}_{ij}) + f_4(\texttt{age}_{ij}) \\
&+ \beta\texttt{male}_{ij} + \gamma_{1,ij} + \gamma_{2,i} + \gamma_{3,i} t_{ijk}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{1,ij} &\sim \mathcal{N}(0, \tau_1^2) \\
\begin{pmatrix} \gamma_{2,i} \\ \gamma_{3,i} \end{pmatrix} &\sim \mathcal{N}_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_2^2 & \tau_3 \\ \tau_3 & \tau_4^2 \end{pmatrix} \right\},
\end{aligned}
$$

with $\boldsymbol{x}_{ijk} = (t_{ijk}, \texttt{TBSA}_{ij}, \texttt{LOS}_{ij}, \texttt{age}_{ij}, \texttt{male}_{ij})^\top$ and $\boldsymbol{\gamma}_{ij} = (\gamma_{1,ij}, \gamma_{2,i}, \gamma_{3,i})^\top$. In this mixed-effects additive transformation model, we approximate the transformation function of the HRQoL index with polynomials in Bernstein form, and the nonlinear shift terms of the numeric predictors with thin-plate splines (Wood 2017, Section 5.5.1). The patient-level residual heterogeneity is captured with random patient intercepts nested within the studies, while the study-level heterogeneity is modeled with random intercepts and slopes for time.

The model detailed above can be fitted in **tramME** with the function call

```
R> m_po <- ColrME(EQi ~ s(time) + s(TBSA) + s(LOS) + s(Age) + Gender +
+    (1 | Study:Patient) + (time | Study), data = burn, order = 8,
+    bounds = c(-Inf, 1), support = c(-0.1, 1), control = opt)
R> summary(m_po)
```

```

Additive Mixed-Effects Continuous Outcome Logistic Regression Model

        Formula:
EQi ~ s(time) + s(TBSA) + s(LOS) + s(Age) + Gender + (1 | Study:Patient) +
    (time | Study)

        Fitted to dataset burn

        Fixed effects parameters:
        =========================

           Estimate Std. Error z value Pr(>|z|)
Gendermale   -0.577      0.113   -5.11  3.2e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        Smooth shift terms:
```

```
        ===================

         edf
s(time) 8.84
s(TBSA) 1.00
s(LOS)  7.96
s(Age)  1.00

        Random effects:
        ===============

Grouping factor: Study:Patient (1687 levels)
Standard deviation:
(Intercept)
       1.49

Grouping factor: Study (10 levels)
Standard deviations:
(Intercept)        time
     1.2400      0.0295

Correlation:
     (Intercept)
time      -0.867


        Log-likelihood: -15223 (npar = 22)
```

According to the point estimate of the single parametric shift term, for any possible cutoff levels (i.e., assuming *proportional odds*) of the EQ-5D utility index, the odds of a smaller outcome is $\exp\left(\hat{\beta}\right) = 0.56$ times as high among males than females.

As previously, we can plot the fitted nonlinear terms for each numeric covariate.

```
R> plot(smooth_terms(m_po), panel.first = grid())
```

The nonlinear covariate effects shown in Figure 8, similarly to the parametric fixed effect estimate of gender, are interpreted on the log-odds scale. The general time pattern of recovery captured by the mixed-effects additive continuous outcome logistic regression is in line with the findings of Spronk *et al.* (2020): After a relatively large improvement within the first six months, the recovery becomes slower. Moreover, because the pooled dataset contains less observations with longer followup times, the uncertainty around the effect of time increases considerably after two years.

In their original analysis of the pooled burn patient dataset, Spronk *et al.* (2020) were interested in calculating expected, *population-level* recovery patterns for patients with various predictor configurations, and investigate how much time it takes for the HRQoL of these patients to return to some fixed reference level. Because the transformation model approach

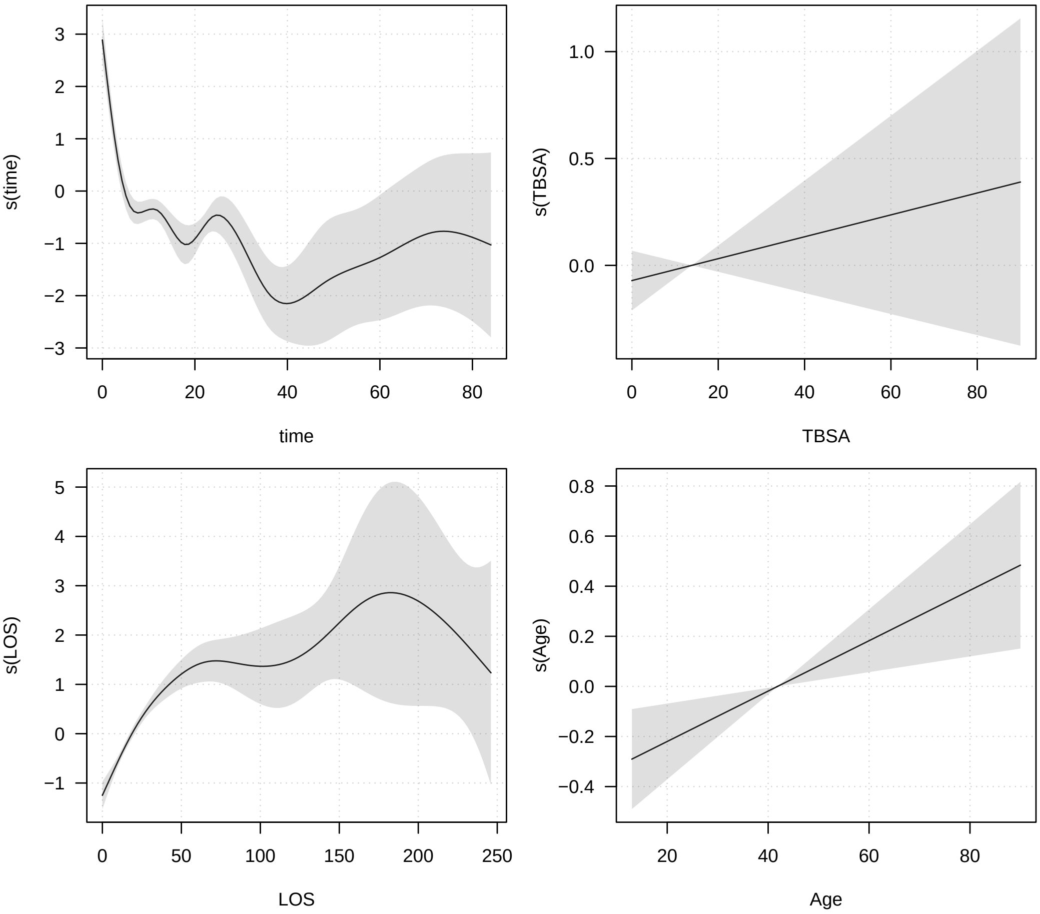

Figure 8: Penalized additive terms for time since burn (`time`, in months), total body surface area burned (`TBSA`, in percentage points) length of hospital stay (`LOS`, in days) and age (`Age`, in years) in the continuous outcome logistic regression model of burn patient quality-of-life. The solid lines represent the point estimates, while the shaded areas are the 95% pointwise confidence intervals.

approximates the whole conditional outcome distribution, it provides means to calculate other quantities beyond the conditional expectation, such as the *tail probability* of an event. In this example, we can calculate estimates of the probability that the quality-of-life of a patient, with a specific predictor configuration, exceeds some reference level at a given point in time.

The mixed-effects regression represents a conditional approach of modeling correlated data. This means that the models fitted with **tramME** represent conditional distributions for given studies and patients in our current analysis. If we want to calculate marginal, i.e. "population-level", distributions, we have to average over all subjects and studies in the population. Technically, this can be achieved by integrating out the random effect vector from the conditional distribution function

$$F_Y(y \mid \boldsymbol{x}) = \int_{\boldsymbol{\gamma}\in\mathbb{R}^q} F_Y(y \mid \boldsymbol{x}, \boldsymbol{\gamma})\phi(\boldsymbol{\gamma}; \boldsymbol{\Sigma})\,\mathsf{d}\boldsymbol{\gamma}.$$

In the general case, the integral does not have a closed form solution, and we have to resort to using numerical integration techniques.

To simplify the marginalization of the distributional predictions over the vector of random effects, we define a helper function.

```
R> mpredict <- function(obj, newdata,
+    scale = c("distribution", "density", "survivor"),
+    ndraws = 500, antithetic = FALSE, return_draws = FALSE,
+    ncpus = 1L, ...) {
+    scale <- match.arg(scale)
+    if (antithetic) ndraws <- floor(ndraws / 2)
+    re <- simulate(obj, type = "ranef", newdata = newdata, nsim = ndraws, ...)
+    if (antithetic) re <- c(re, lapply(re, `-`))
+    ndraws <- length(re)
+    FUN <- function(r) {
+      predict(obj, newdata = newdata, type = scale, ranef = r, ...)
+    }
+    if (ncpus > 1) pr <- parallel::mclapply(re, FUN, mc.cores = ncpus)
+    else pr <- lapply(re, FUN)
+    if (return_draws) return(pr)
+    rns <- rownames(pr[[1]])
+    cns <- colnames(pr[[1]])
+    pr <- unlist(pr)
+    d <- length(pr) / (nrow(newdata) * ndraws)
+    pr <- array(pr, dim = c(d, nrow(newdata), ndraws))
+    mp <- apply(pr, c(1, 2), mean)
+    mcv <- apply(pr, c(1, 2), var) / ndraws
+    if (antithetic) mcv <- mcv + apply(pr, c(1, 2), function(x) {
+      cov(head(x, ndraws/2), tail(x, ndraws/2)) / (2*ndraws)
+    })
+    rownames(mp) <- rownames(mcv) <- rns
+    colnames(mp) <- colnames(mcv) <- cns
+    attr(mp, "mc.se") <- sqrt(mcv)
+    mp
+  }
```

This function implements a simple Monte Carlo numerical integration scheme by combining `simulate()` and `predict()` methods of the '`tramME`' model class. It generates a sample from the fitted distribution of random effects, evaluates the conditional distributional predictions of the transformation model at these draws and calculates the mean of the results to evaluate the marginal distribution. To reduce the variability of these estimates, and thus decrease the required Monte Carlo iterations to achieve a given precision, the function also allows for antithetic variates (see, for example, Gentle 2003, Section 7.5.3). Although there are more efficient ways of numerically integrating over the random effects (an alternative is presented in Tamási and Hothorn 2021), the Monte Carlo integration is straightforward to implement using the available functionality of **tramME** and generalizes well to complex random effects structures.

The original analysis used a normal linear mixed model to calculate predictions for recovery patterns, which ignored the non-normal bounded nature of the outcome variable. Additionally, this approach treated the observations of EQ-5D index as exactly observed realizations from a continuous distribution, which is at odds with the large number of ties in the sample. To demonstrate the effect of these assumptions (i.e., conditional normality and exact observations) on the predicted patterns of recovery, we also estimate a mixed-effects additive transformation model with the same random effects and predictor structure but assuming conditional normal outcome distribution with the `LmME()` model function.

```
R> m_lm <- LmME(EQindex ~ s(time) + s(TBSA) + s(LOS) + s(Age) + Gender +
+    (1 | Study:Patient) + (time | Study), data = burn, control = opt)
```

Note that we use the original `EQindex` values and do not specify `bounds` and `support` in this model, neither do we set the polynomial order of the approximation of the baseline transformation. This is because we assume exact observations and estimate a normal additive mixed model, i.e., unbounded outcome, linear transformation function and standard Gaussian inverse link function.

In the calculations below, we are going to use the first subject in the dataset

```
R> model.frame(m_lm)[1, ]
```

```
  EQindex Gender time TBSA LOS Age Study Patient
1  0.8176   male   84 31.5  28  28     1     256
```

who is a 28-year-old male, with 31.5% of burned body surface area and 28 days of hospital stay. We evaluate the marginal probability of having a HRQoL index value larger than the 0.8 and 0.9 reference values, respectively. These tail probabilities are calculated at different time points between 0 and 36 months.

```
R> nd <- model.frame(m_lm)[rep(1, 100), ]
R> nd$time <- seq(0, 36, length.out = 100)
R> ncpus <- getOption("ncores", round(detectCores() / 2 - 1L))
R> pr_lm <- mpredict(m_lm, nd, scale = "survivor", q = c(0.8, 0.9),
+    ndraws = 500, antithetic = TRUE, seed = 100, ncpus = 1)
R> pr_po <- mpredict(m_po, nd, scale = "survivor", q = c(0.8, 0.9),
+    ndraws = 500, antithetic = TRUE, seed = 100, ncpus = 1)
```

The left panel of Figure 9 compares the resulting curves from the original mixed-effects additive continuous outcome logistic regression to the ones calculated from the normal additive mixed model.

To understand the source of the differences in the tail probabilities, the marginal distribution functions from the two models at 12 months are calculated, given the predictor configuration of the subject shown above.

```
R> nd2 <- model.frame(m_lm)[1, ]
R> nd2$time <- 12
R> cdf_lm <- mpredict(m_lm, nd2, scale = "distribution",
+    q = seq(-0.1, 1, length.out = 100),
```

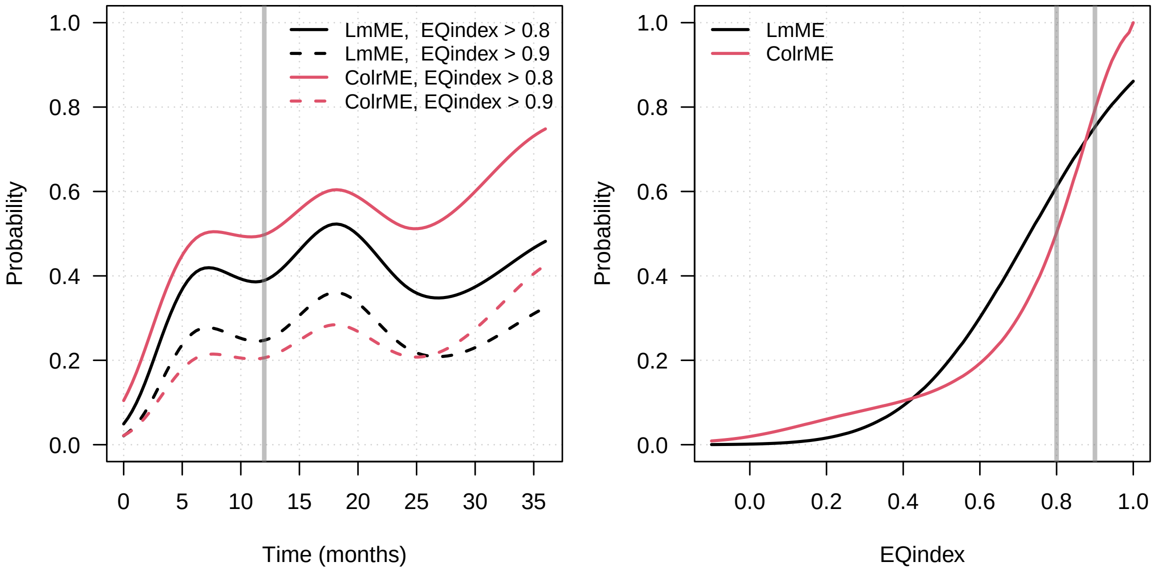

Figure 9: *Left panel:* Marginal tail probabilities of exceeding the EQ-5D reference levels of 0.8 and 0.9 calculated from the `LmME` and `ColrME` models, respectively. The grey line indicates month 12 (see right panel). *Right panel:* The marginal distributions of the HRQoL index at month 12, calculated from the `LmME` and `ColrME` models, respectively. The grey lines indicate the 0.8 and 0.9 reference levels.

```
+    ndraws = 500, antithetic = TRUE, seed = 100)
R> cdf_po <- mpredict(m_po, nd2, scale = "distribution",
+    q = seq(-0.1, 1, length.out = 100),
+    ndraws = 500, antithetic = TRUE, seed = 100)
```

The right-hand side plot in Figure 9 shows that the models produce markedly different approximations of the target distribution. The normal additive mixed model variant (`LmME`) adheres to parametric distributional assumptions, which is reflected in its approximation of the marginal outcome distribution. On the other hand, the transformation model approach (`ColrME`) estimates the shape of the conditional distribution from the data, resulting in a more flexible marginal approximation.

5. Summary and discussion

This article presented the R package **tramME**, which extends the transformation model family proposed by Hothorn *et al.* (2018) and implemented by Hothorn (2020) with random effects and nonlinear (optionally penalized) additive terms. Thanks to the general transformation model approach, **tramME** can be used to model any type (discrete, continuous, mixed-type, bounded etc.) of at least ordered univariate outcome. Additionally, as the example applications demonstrated, the model framework provides great flexibility in choosing the correct likelihood contributions that reflect best the properties of the individual observations and the assumptions of the analysis. As a result, the models implemented in the package can be adapted to settings that would otherwise require a lot of effort to implement (such as the proportional hazards regression for interval-censored data and time-dependent covariates in the example analysis of Section 4.1). Moreover, the modularity of the transformation model framework allows the user to define new models that are most appropriate for the analytic question at hand and may not be available from alternative implementations. The mixed-effects additive continuous outcome logistic regression for IPD meta-analysis in Section 4.2 is an example for such a model.

In **tramME**, the penalized smooth terms provide tools to take the temporal and spatial dependency of the data into account. Additionally, nonlinear covariate effects on the transformation scale can also be captured with these terms. With multiple nested or crossed random effects, grouped data can be efficiently accommodated in a regression setting. To allow for departures from implied assumptions of proportionality (e.g., proportional hazards or proportional odds), conditional transformation functions (i.e., stratification/outcome-dependent covariate effects) are also available in the package. These extensions make the transformation model framework a useful addition to the popular mixed and additive regression implementations.

Transformation models approximate the shape of outcome distribution flexibly from the data and thus many important traditional regression models are special cases of the model class. In fact, many of these models can be directly represented with specific parametric restrictions on the baseline transformation function (see Table 1). Because of this property, mixed-effects additive transformation models can be used to scrutinize distributional assumptions in GLMMs and GAMMs. The package vignette Tamási and Hothorn (2023) presents several examples for such comparisons.

tramME was designed for efficient estimation of interpretable yet flexible regression models that are able to handle dependent data structures, and for drawing valid inferences based on these models. While the extension of linear transformation models implemented by the package is able to adapt to any distribution at the reference level of the covariates, it typically restricts how different levels of the linear predictor relate to each other by including covariates as shift terms. Depending on the choice of the inverse link function, this restriction translates to proportionality assumptions on a specific scale (e.g., proportional hazards, proportional odds), which, in turn, lead to easily interpretable parameter estimates, and simplified inference. Of course, the assumed additive structure can be restrictive in some scenarios and for specific purposes. While **tramME** provides ways, to relax these assumptions by introducing conditional transformation functions $h(y \mid \boldsymbol{x}_A; \boldsymbol{\vartheta})$ (as we saw in Section 4.1), the resulting models are typically harder to fit and interpret. For this reason, if the goal is to generate highly accurate probabilistic predictions of an outcome given a set of covariates, **tramME** is not the most adequate tool. For these settings, several packages that implement proper distributional regression models are available (e.g., **gamlss**, Stasinopoulos and Rigby 2008, **brms**, Bürkner 2017, **bamlss**, Umlauf *et al.* 2021).

Because the underlying statistical theory is very general, there are many potential directions in which **tramME** could be extended in the future. New inferential procedures and methods to extract and represent results from fitted mixed-effects additive transformation models are continually added to the package. Simple model variants such as a mixed-effects additive extension of the count transformation model class by Siegfried and Hothorn (2020) are also among the plans. Siegfried, Kook, and Hothorn (2023) proposed a location-scale version of the linear transformation model, which allows a more flexible approximation of the outcome distribution. A rich and versatile alternative for the popular distributional regression models, such as GAMLSS, could be obtained by combining this modeling approach with the computation framework of **tramME**. Transformation models have recently been extended to multivariate regression problems by Klein, Hothorn, Barbanti, and Kneib (2022). Adapting the multivariate transformation approach to correlated data structures and introducing penalized smoothing would greatly improve the practical applicability of this model.

Acknowledgments

This work was financially supported by the UZH Postdoc Grant (FK-22-057). I thank Sandra Siegfried for her helpful comments and Torsten Hothorn for discussions about transformation models.

References

Barbanti L, Hothorn T (2024). “A Transformation Perspective on Marginal and Conditional Models.” *Biostatistics*, **25**(2), 402–428. `doi:10.1093/biostatistics/kxac048`.

Bates D, Mächler M, Bolker BM, Walker S (2015). “Fitting Linear Mixed-Effects Models Using **lme4**.” *Journal of Statistical Software*, **67**(1), 1–48. `doi:10.18637/jss.v067.i01`.

Belenky G, Wesensten NJ, Thorne DR, Thomas ML, Sing HC, Redmond DP, Russo MB, Balkin TJ (2003). “Patterns of Performance Degradation and Restoration During Sleep Restriction and Subsequent Recovery: A Sleep Dose-response Study.” *Journal of Sleep Research*, **12**(1), 1–12. `doi:10.1046/j.1365-2869.2003.00337.x`.

Belias M, Rovers MM, Hoogland J, Reitsma JB, Debray TPA, IntHout J (2022). “Predicting Personalised Absolute Treatment Effects in Individual Participant Data Meta-Analysis: An Introduction to Splines.” *Research Synthesis Methods*, **13**(2), 255–283. `doi:10.1002/jrsm.1546`.

Bolker B, Piaskowski J, Tanaka E, Alday P, Viechtbauer W (2024). “CRAN Task View: Mixed, Multilevel, and Hierarchical Models in R.” Version 2024-05-08, URL `http://CRAN.R-project.org/view=MixedModels`.

Box GEP, Cox DR (1964). “An Analysis of Transformations.” *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**(2), 211–252. `doi:10.1111/j.2517-6161.1964.tb00553.x`.

Brilleman SL, Elci EM, Novik JB, Wolfe R (2020). “Bayesian Survival Analysis Using the **rstanarm** R Package.” *arXiv 2002.09633*, arXiv.org E-Print Archive. `doi:10.48550/arXiv.2002.09633`.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). “**glmmTMB** Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling.” *The R Journal*, **9**(2), 378–400. `doi:10.32614/RJ-2017-066`.

Bürkner PC (2017). "**brms**: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software*, **80**(1), 1–28. `doi:10.18637/jss.v080.i01`.

Cheng SC, Wei LJ, Ying Z (1995). "Analysis of Transformation Models with Censored Data." *Biometrika*, **82**(4), 835–845. `doi:10.2307/2337348`.

Christensen RHB (2023). "**ordinal** — Regression Models for Ordinal Data." R package version 2023.12-4, URL `https://CRAN.R-project.org/package=ordinal`.

Cox DR (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**(2), 187–202. `doi:10.1111/j.2517-6161.1972.tb00899.x`.

De Neve J, Thas O, Gerds TA (2019). "Semiparametric Linear Transformation Models: Effect Measures, Estimators, and Applications." *Statistics in Medicine*, **38**(8), 1484–1501. `doi:10.1002/sim.8078`.

Doksum KA, Gasko M (1990). "On a Correspondence Between Models in Binary Regression Analysis and in Survival Analysis." *International Statistical Review / Revue Internationale de Statistique*, **58**(3), 243–252. `doi:10.2307/1403807`.

Farouki RT (2012). "The Bernstein Polynomial Basis: A Centennial Retrospective." *Computer Aided Geometric Design*, **29**(6), 379–419. `doi:10.1016/j.cagd.2012.03.001`.

Farrington CP (2000). "Residuals for Proportional Hazards Models with Interval-Censored Survival Data." *Biometrics*, **56**(2), 473–482. `doi:10.1111/j.0006-341X.2000.00473.x`.

Gentle JE (2003). *Random Number Generation and Monte Carlo Methods*. Statistics and Computing, 2nd edition. Springer-Verlag, New York, NY, U.S.A.

Harrell FE (2024). ***rms**: Regression Modeling Strategies*. `doi:10.32614/CRAN.package.rms`. R package version 6.8-1.

Hastie T, Tibshirani R (1999). *Generalized Additive Models*. Chapman & Hall/CRC, Boca Raton, FL, U.S.A.

Hodges JS (2014). *Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects*. CRC Press, Boca Raton, FL, U.S.A.

Hothorn T (2020). "Most Likely Transformations: The **mlt** Package." *Journal of Statistical Software*, **92**(1), 1–68. `doi:10.18637/jss.v092.i01`.

Hothorn T, Barbanti L, Siegfried S (2024). ***tram**: Transformation Models*. `doi:10.32614/CRAN.package.tram`. R package version 1.0-4.

Hothorn T, Möst L, Bühlmann P (2018). "Most Likely Transformations." *Scandinavian Journal of Statistics*, **45**(1), 110–134. `doi:10.1111/sjos.12291`.

Jackson C (2016). "**flexsurv**: A Platform for Parametric Survival Modeling in R." *Journal of Statistical Software*, **70**(8), 1–33. `doi:10.18637/jss.v070.i08`.

Klein JP, Moeschberger ML (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd edition. Springer-Verlag, New York, NY, U.S.A.

Klein N, Hothorn T, Barbanti L, Kneib T (2022). "Multivariate Conditional Transformation Models." *Scandinavian Journal of Statistics*, **49**(1), 116–142. `doi:10.1111/sjos.12501`.

Kneib T (2006a). "Mixed Model-based Inference in Geoadditive Hazard Regression for Interval-censored Survival Times." *Computational Statistics & Data Analysis*, **51**(2), 777–792. `doi:10.1016/j.csda.2006.06.019`.

Kneib T (2006b). *Mixed Model Based Inference in Structured Additive Regression.* Ph.D. thesis, Ludwig-Maximilians-Universität München. URL `https://edoc.ub.uni-muenchen.de/5011/`.

Kristensen K, Nielsen A, Berg CW, Skaug HJ, Bell BM (2016). "**TMB**: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software*, **70**(5), 1–21. `doi:10.18637/jss.v070.i05`.

Lin H, Zhou L, Zhou X (2014). "Semiparametric Regression Analysis of Longitudinal Skewed Data." *Scandinavian Journal of Statistics*, **41**(4), 1031–1050. `doi:10.1111/sjos.12080`.

Lindsey JK (1999). "Some Statistical Heresies." *Journal of the Royal Statistical Society. Series D (The Statistician)*, **48**(1), 1–40. `doi:10.1111/j.1467-9884.1999.tb09211.x`.

Liu Q, Shepherd BE, Li C, Harrell FE (2017a). "Modeling Continuous Response Variables Using Ordinal Regression." *Statistics in Medicine*, **36**(27), 4316–4335. `doi:10.1002/sim.7433`.

Liu XR, Pawitan Y, Clements MS (2017b). "Generalized Survival Models for Correlated Time-to-Event Data." *Statistics in Medicine*, **36**(29), 4743–4762. `doi:10.1002/sim.7451`.

Liu XR, Pawitan Y, Clements MS (2018). "Parametric and Penalized Generalized Survival Models." *Statistical Methods in Medical Research*, **27**(5), 1531–1546. `doi:10.1177/0962280216664760`.

Lohse T, Rohrmann S, Faeh D, Hothorn T (2017). "Continuous Outcome Logistic Regression for Analyzing Body Mass Index Distributions." *F1000Research*, **6**. `doi:10.12688/f1000research.12934.1`.

Manuguerra M, Heller GZ, Ma J (2020). "Continuous Ordinal Regression for Analysis of Visual Analogue Scales: The R Package **ordinalCont**." *Journal of Statistical Software*, **96**, 1–25. `doi:10.18637/jss.v096.i08`.

McLain AC, Ghosh SK (2013). "Efficient Sieve Maximum Likelihood Estimation of Time-Transformation Models." *Journal of Statistical Theory and Practice*, **7**(2), 285–303. `doi:10.1080/15598608.2013.772835`.

Nelder JA, Wedderburn RWM (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A (General)*, **135**(3), 370–384. `doi:10.2307/2344614`.

Nocedal J, Wright SJ (2006). *Numerical Optimization.* 2nd edition. Springer-Verlag, New York, NY, U.S.A.

Nychka D (1988). "Bayesian Confidence Intervals for Smoothing Splines." *Journal of the American Statistical Association*, **83**(404), 1134–1143. `doi:10.1080/01621459.1988.10478711`.

Owen AB (2001). *Empirical Likelihood*. 92. Chapman & Hall/CRC, Boca Raton, FL, U.S.A.

Pinheiro JC, Bates DM (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, NY, U.S.A. doi:10.1007/b98882.

Rabin R, de Charro F (2001). "EQ-5D: A Measure of Health Status from the EuroQol Group." *Annals of Medicine*, **33**(5), 337–343. doi:10.3109/07853890109002087.

Rigby RA, Stasinopoulos DM (2005). "Generalized Additive Models for Location, Scale and Shape." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**(3), 507–554. doi:10.1111/j.1467-9876.2005.00510.x.

Riley RD, Tierney JF, Stewart LA (eds.) (2021). *Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research*. Wiley. doi:10.1002/9781119333784.

Royston P, Parmar MKB (2002). "Flexible Parametric Proportional-hazards and Proportional-odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects." *Statistics in Medicine*, **21**(15), 2175–2197. doi:10.1002/sim.1203.

Rutherford MJ, Crowther MJ, Lambert PC (2015). "The Use of Restricted Cubic Splines to Approximate Complex Hazard Functions in the Analysis of Time-to-event Data: A Simulation Study." *Journal of Statistical Computation and Simulation*, **85**(4), 777–793. doi:10.1080/00949655.2013.845890.

Siegfried S, Hothorn T (2020). "Count Transformation Models." *Methods in Ecology and Evolution*, **11**(7), 818–827. doi:10.1111/2041-210X.13383.

Siegfried S, Kook L, Hothorn T (2023). "Distribution-Free Location-Scale Regression." *The American Statistician*, **77**(4), 345–356. doi:10.1080/00031305.2023.2203177.

Spronk I, Van Loey NEE, Sewalt C, Nieboer D, Renneberg B, Moi AL, Oster C, Orwelius L, Van Baar ME, Polinder S, the Quality of Life Study Group (2020). "Recovery of Health-Related Quality of Life after Burn Injuries: An Individual Participant Data Meta-Analysis." *PLOS ONE*, **15**(1), e0226653. doi:10.1371/journal.pone.0226653.

Stasinopoulos DM, Rigby RA (2008). "Generalized Additive Models for Location Scale and Shape (GAMLSS) in R." *Journal of Statistical Software*, **23**, 1–46. doi:10.18637/jss.v023.i07.

Stroup WW (2012). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Boca Raton, FL, U.S.A. doi:10.1201/b13151.

Stroustrup B (2013). *The C++ Programming Language*. 4th edition. Addison-Wesley.

Tamási B (2024). ***tramME**: Transformation Models with Mixed Effects*. doi:10.32614/CRAN.package.tramME. R package version 1.0.6.

Tamási B, Crowther M, Puhan MA, Steyerberg EW, Hothorn T (2022). "Individual Participant Data Meta-Analysis with Mixed-Effects Transformation Models." *Biostatistics*, **23**(4), 1083–1098. doi:10.1093/biostatistics/kxab045.

Tamási B, Hothorn T (2021). "**tramME**: Mixed-Effects Transformation Models Using Template Model Builder." *The R Journal*, **13**(1), 581–594. `doi:10.32614/RJ-2021-075`.

Tamási B, Hothorn T (2023). "Online Appendix: Flexible Regression for Correlated and Censored Ecological Data with Mixed-Effects Additive Transformation Models." R package **tramME** vignette, version 1.0.6, URL `https://CRAN.R-project.org/web/packages/tramME/vignettes/mixed-effects-additive-models.pdf`.

Therneau TM (2024). *A Package for Survival Analysis in R*. `doi:10.32614/CRAN.package.survival`. R package version 3.7-0.

Therneau TM, Crowson C, Atkinson E (2024). "Using Time Dependent Covariates and Time Dependent Coeffcients in the Cox Model." R package **survival** vignette, version 3.7-0, URL `https://CRAN.R-project.org/web/packages/survival/vignettes/timedep.pdf`.

Tian Y, Hothorn T, Li C, Harrell Jr FE, Shepherd BE (2020). "An Empirical Comparison of Two Novel Transformation Models." *Statistics in Medicine*, **39**(5), 562–576. `doi:10.1002/sim.8425`.

Umlauf N, Klein N, Simon T, Zeileis A (2021). "**bamlss**: A Lego Toolbox for Flexible Bayesian Regression (and Beyond)." *Journal of Statistical Software*, **100**(4), 1–53. `doi:10.18637/jss.v100.i04`.

Varadhan R (2023). ***alabama**: Constrained Nonlinear Optimization*. `doi:10.32614/CRAN.package.alabama`. R package version 2023.1.0.

Wood S, Scheipl F (2020). ***gamm4**: Generalized Additive Mixed Models using **mgcv** and **lme4***. `doi:10.32614/CRAN.package.gamm4`. R package version 0.2-6.

Wood SN (2004). "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models." *Journal of the American Statistical Association*, **99**(467), 673–686. `doi:10.1198/016214504000000980`.

Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2nd edition. CRC Press/Taylor & Francis Group, Boca Raton, FL, U.S.A.

Wood SN, Scheipl F, Faraway JJ (2013). "Straightforward Intermediate Rank Tensor Product Smoothing in Mixed Models." *Statistics and Computing*, **23**(3), 341–360. `doi:10.1007/s11222-012-9314-z`.

Zeng D, Lin DY (2007). "Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**(4), 507–564. `doi:10.1111/j.1369-7412.2007.00606.x`.

Zheng N, Cadigan N (2021). "Frequentist Delta-Variance Approximations with Mixed-Effects Models and TMB." *Computational Statistics & Data Analysis*, **160**, 107227. `doi:10.1016/j.csda.2021.107227`.

Zhou J, Zhang J, Lu W (2022). "**TransModel**: An R Package for Linear Transformation Model with Censored Data." *Journal of Statistical Software*, **101**, 1–12. `doi:10.18637/jss.v101.i09`.

A. Timing comparisons with glmmTMB

Because both **tramME** and **glmmTMB** (Brooks *et al.* 2017) rely on the same computational machinery (the **TMB** package by Kristensen *et al.* 2016), one might be interested in how the performance of the two packages compare in terms of computational speed.

In the following comparisons, we simulate datasets from a normal linear mixed-effects model with random slopes and intercepts that has a similar balanced structure as the famous `lme4::sleepstudy` data (Belenky, Wesensten, Thorne, Thomas, Sing, Redmond, Russo, and Balkin 2003). In the R function below, we set the number of groups (`K`) and number of observations (`N`) within groups separately.

```
R> sim_lmer <- function(N = 10, K = 20, b = c(251, 10), sd = c(26, 25, 6)) {
+    x <- runif(N * K, 0, 9)
+    y <- drop(cbind(1, x) %*% b) + rep(rnorm(K, sd = sd[2]), each = N) +
+      rep(rnorm(K, sd = sd[3]), each = N) * x + rnorm(N * K, sd = sd[1])
+    data.frame(x = x, y = y, g = factor(rep(seq(K), each = N)))
+  }
```

We fit three models to each simulated dataset: First, a GLMM with Gaussian family and identity link is estimated, using the **glmmTMB** package,

```
R> library("glmmTMB")
R> m1 <- glmmTMB(y ~ x + (x | g), data = df)
```

To directly compare the performance of the two packages, we estimate a mixed-effects transformation model with a linear transformation function and a probit link (`LmME`), which is equivalent to the Gaussian GLMM but uses a different parameterization (c.f. Tamási and Hothorn 2021). Finally, to evaluate the overhead of relaxing the parametric distributional assumption, we estimate a mixed-effects transformation model with flexible smooth transformation function and a probit link (`BoxCoxME`, see also Table 1), which nests the normal linear mixed model, but does not assume Gaussian errors. We use default settings for all model estimations.

```
R> library("tramME")
R> m2 <- LmME(y ~ x + (x | g), data = df)
R> m3 <- BoxCoxME(y ~ x + (x | g), data = df)
```

We run two sets of comparisons, first by changing the sizes of a fixed set of 20 groups, and next by changing the number of groups while keeping their sizes constant at 10. Figure 10 shows the estimation time distributions of 40 repetitions of each model and dataset combinations. The results suggest comparable estimation times with **glmmTMB** and **tramME**, and highlight that the computational price of distributional flexibility increases with the size of the sample.

Affiliation:

Bálint Tamási
Institut für Epidemiologie, Biostatistik und Prävention

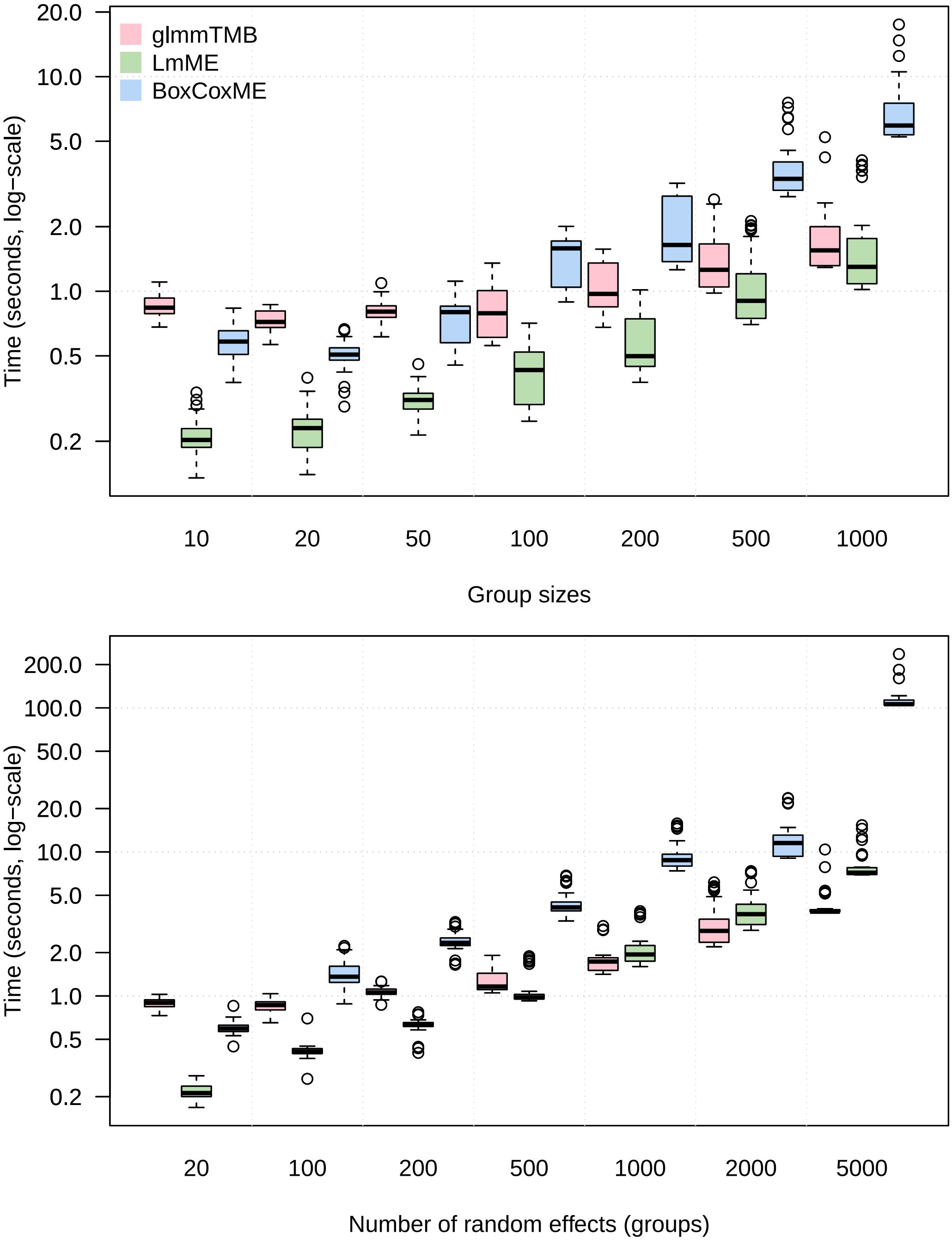

Figure 10: Time required to fit normal linear mixed-effects models (`glmmTMB`) and mixed-effects transformation models (`LmME` and `BoxCoxME`) to simulated datasets with various group sizes (*top panel*, keeping the number of groups at 20) and numbers of groups (*bottom panel*, keeping the group sizes at 10). Each model is fitted to each dataset 40 times and the distribution of estimation times are shown by the box plots.

Universität Zürich
Hirschengraben 84.
CH-8001 Zürich, Switzerland
E-mail: `Balint.Tamasi+tramME@gmail.com`