

Estimated estimating equations: Semiparametric inference for clustered/longitudinal data

Jeng-Min Chiou

Academia Sinica, Taipei, Taiwan

and Hans-Georg Müller †

University of California, Davis, USA

Summary. We introduce a flexible marginal modelling approach for statistical inference for clustered/longitudinal data under minimal assumptions. This estimated estimating equations (EEE) approach is semiparametric and the proposed models are fitted by quasi-likelihood regression, where the unknown marginal means are a function of the fixed-effects linear predictor with unknown smooth link, and variance-covariance is an unknown smooth function of the marginal means. We propose to estimate the nonparametric link and variance-covariance functions via smoothing methods, while the regression parameters are obtained via the estimated estimating equations. These are score equations that contain nonparametric function estimates. The proposed EEE approach is motivated by its flexibility and easy implementation. Moreover, if data follow a generalized linear mixed model (GLMM), with either specified or unspecified distribution of random effects and link function, the proposed model emerges as the corresponding marginal (population-average) version and can be used to obtain inference for the fixed effects in the underlying GLMM, without the need to specify any other components of this GLMM. Among marginal models, the EEE approach provides a flexible alternative to modelling with generalized estimating equations (GEE). Applications of EEE include diagnostics and link selection. The asymptotic distribution of the proposed estimators for the model parameters is derived, enabling statistical inference. Practical illustrations include Poisson modelling of repeated epileptic seizure counts and simulations for clustered binomial responses.

Keywords: Diagnostics; Generalized estimating equations; Generalized linear mixed model; Link selection; Marginal model; Quasi-likelihood; Repeated measurements; Semiparametric regression; Smoothing; Variance-covariance function.

1. Introduction

Generalized linear mixed models (GLMMs) extend the framework of linear mixed models and of generalized linear models (GLMs), by allowing for non-Gaussian

†*Address for Correspondence:* Hans-Georg Müller, Department of Statistics, University of California, One Shield Avenue, Davis, CA 95616 U.S.A.

E-mail address: mueller@wald.ucdavis.edu

data, nonlinear link functions and the inclusion of random effects and of correlated errors. They have become a favored tool for the modelling of clustered and longitudinal data, in particular, for repeated or correlated non-Gaussian data, such as binomial or Poisson type responses that are commonly encountered in longitudinal studies. Due to the wide range of applications of GLMMs, these models have received substantial attention during the last decade and are available in the major software packages.

The computational burden associated with high dimensional numerical integration has limited past studies of GLMMs to the case of simplified models (e.g., random intercept models), to tractable random effects distributions (e.g., the Gaussian and conjugate distributions such as the beta-binomial and negative binomial models), or to conditional inference for the regression coefficients, conditioning on the random effects (Zeger and Karim, 1991). A variety of novel approaches have been proposed to overcome the computational difficulties, with the goal to improve inference and estimation procedures for the fixed effects in GLMMs. These include Gibbs sampling (Zeger and Karim, 1991), penalized quasi-likelihood and marginal quasi-likelihood (Breslow and Clayton, 1993), pseudo-likelihood based on approximate marginal models (Wolfinger and O'Connell, 1993), and maximum likelihood with Monte Carlo versions of EM, Newton-Raphson and simulated maximum likelihood algorithms (McCulloch, 1997), among many others (Jiang, 1998). These approaches typically require Gaussian distribution assumptions for the random effects. Methods for non-normal random effects are less common and limited to specialized cases (Magder and Zeger, 1996; Lee and Nelder, 1996, 2001; Gamerman, 1997). Approaches that drop distributional assumptions on the random effects are due to Aitkin (1999) and Follmann and Lambert (1989).

An alternative approach to modelling clustered/longitudinal data are marginal or population-average models, with the generalized estimating equations (GEE) and its variants providing prominent examples. The relationship between subject-specific models such as GLMMs and population-average models has been investigated among others by Zeger, Liang and Albert (1988) and more recently by Heagerty (1999) and Heagerty and Zeger (2000). The latter authors proposed a marginally specified GLMM by introducing an adjustment function in the fixed-effects linear predictors to connect the GLMM to a marginal model of GEE type, postulating the same link function for both GLMM and marginal models, and using numerical integration and Newton-Raphson iteration to maximize the likelihood. However, as we demonstrate below, the link function stipulated in a GLMM generally does not coincide with the link function of the corresponding marginal model, due to the influence of the random effects.

Heagerty and Kurland (2001) showed that under misspecified random effects models the conventional GLMM may be subject to substantial bias, while their marginally specified GLMM is less susceptible. This bias issue has also been discussed by Crowder (1995, 2001), Chaganty (1997) and Sutradhar and Das

(1999). On the other hand, efficiency of the estimated regression coefficients in marginal models based on estimating equations relies on correct specification of the covariance or correlation structure. In this regard, Thall and Vail (1990) proposed a family of covariance models for longitudinal data, Pourahmadi (1999, 2000) provided joint mean-covariance models and studied the asymptotics of the maximum likelihood estimators, while Chiou (2003) studied nonparametric covariance estimation for GEE.

The estimated estimating equations (EEE) method proposed in this article is a marginal or population-average approach based on semiparametric quasi-likelihood regression, where the unknown marginal means are a function of the fixed-effects linear predictor with unknown link, and variance-covariance is an unknown function of the marginal means. The EEE approach allows for statistical inference for clustered/longitudinal data under minimal assumptions. We propose to estimate the nonparametric link and variance-covariance functions via nonparametric smoothing methods while the regression parameters are obtained via estimated estimating equations, which are score equations that contain nonparametric function estimates.

Besides its intrinsic flexibility within the class of marginal models, another distinguishing property of the EEE approach is its relationship to a GLMM. Assuming the longitudinal/clustered data follow a GLMM with either known or unknown components (distribution of random effects and link function), the corresponding marginal model then corresponds to the EEE specifications, as we shall demonstrate below. Even if the underlying GLMM is fully parametric with known random effect distributions and link functions, the corresponding marginal model has unknown and unconventional link and covariance structure and in general does not conform to the specifications of a GEE. In this sense, the GEE approach and an underlying GLMM lead an uneasy co-existence. In contrast, a subject-specific GLMM and the EEE approach for its corresponding population-average marginal model are simultaneously applicable to the same data, as the underlying assumptions are compatible. Inference for the fixed-effects parameter in GLMM is then possible via semiparametric quasi-likelihood in the marginal EEE, for which the usual components of marginal models (link function, correlation structure) are chosen data-adaptively. This has the advantage that the numerical difficulties in fitting GLMM can be avoided and no specification needs to be made in regard to the link function or the nature of the random effects of the underlying GLMM.

As we demonstrate in Section 4 below, the proposed EEE approach can be obtained as a marginal version of an underlying GLMM. Alternatively, EEE can be introduced as a flexible extension of GEE without any reference to an underlying GLMM; this is the approach we take in Section 2. Specific advantages of the proposed EEE approach are: (1) It is a conceptually simple marginal approach; (2) Specification of the link function is not needed; (3) Specification of the correlation structure of the data is not needed, as determination of a smooth

underlying covariance function is part of the model fitting; (4) The asymptotic distributions of the parameter estimates can be derived under remarkably weak assumptions, enabling asymptotic inference; (5) It serves as a tool for checking results of prior analyses of clustered/longitudinal (for example with GEE or GLMM) and is useful for diagnostics, link function selection and model building; (6) In cases where EEE is considered for the situation of an underlying GLMM, specification of the random effects of this GLMM is not required, and neither the distribution family nor its parameters or the link function need to be assumed known; (7) Under the premise of (6), we may obtain inference for the fixed effects parameter of the underlying GLMM via EEE.

This article is organized as follows. The model and the assumptions of the proposed semiparametric EEE approach are introduced in Section 2, including nonparametric variance-covariance estimates and quasi-likelihood estimation of the regression parameters as well as the estimation algorithm. Section 3 provides the asymptotic results necessary for inference. Connections between EEE and GLMM are discussed in Section 4, including illustrating examples of the relationship between the corresponding link functions. Practical performance of the proposed methods through an application to a clinical study on epileptic seizures and a simulation experiment are discussed in Section 5. Concluding remarks are presented in Section 6 and a brief description of smoothing methods and proofs can be found in the Appendix.

2. Estimated estimating equations

2.1. Semiparametric marginal modelling

To introduce the proposed semiparametric marginal model for clustered/longitudinal data, denote by $\{(Y_{ij}, X_{ij})_{1 \leq i \leq n, 1 \leq j \leq m_i}\}$ the j th repeated observation for the i th subject or experimental unit, where Y_{ij} is the outcome variable associated with the vector of explanatory variables X_{ij} in \mathbb{R}^p . The proposed semiparametric marginal model consists of a systematic part, the linear predictor, and two nonparametric components, the link and variance-covariance functions. The following conditions (M1)-(M3) are required.

(M1) There exists an unknown but smooth and invertible link function $h(\cdot)$ such that

$$E(Y_{ij}|X_{ij} = \mathbf{x}) = h(\mathbf{x}^\top \boldsymbol{\alpha}_0), \quad (1)$$

where $\boldsymbol{\alpha}_0$ in \mathbb{R}^p is a vector of regression coefficients, satisfying the identifiability constraint $\|\boldsymbol{\alpha}_0\| = 1$.

We note that a constraint on $\boldsymbol{\alpha}_0$ is required for identifiability because the link function $h(\cdot)$ is unknown and needs to be estimated. We note that $h(\cdot)$ corresponds to the so-called inverse link function in the conventional GLM (McCullagh and Nelder, 1989). Inference for the regression coefficients $\boldsymbol{\alpha}_0$ is of primary

interest; the interpretation of the regression coefficients will depend on the link function.

(M2) There exists an unknown and smooth variance function $V(\cdot)$ such that

$$\text{Var}(Y_{ij}|X_{ij} = \mathbf{x}) = V\{h(\mathbf{x}^\top \boldsymbol{\alpha}_0)\} . \quad (2)$$

Under (M1) and (M2), the marginal model is a quasi-likelihood type regression model where the variance function is a function of the mean only. The functional relationship between the variance and the mean is smooth, and the shape of both link and variance functions is assumed unknown.

When dealing with clustered/longitudinal data, modelling dependencies is a central issue. Let Y_{i1} and Y_{i2} be any two dependent observations from the i th cluster or subject with the corresponding covariates X_{i1} and X_{i2} . We assume that the covariance function for observations within the same cluster or subject is a smooth function of the means.

(M3) There exists a covariance function $\Gamma(\cdot, \cdot)$ that is an unknown smooth surface such that if Y_{i1} and Y_{i2} are outcome variables from the same cluster or subject, associated with the predictor vectors X_{i1} and X_{i2} , respectively, then

$$\text{Cov}(Y_{i1}, Y_{i2} | X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2) = \Gamma\{h(\mathbf{x}_1^\top \boldsymbol{\alpha}_0), h(\mathbf{x}_2^\top \boldsymbol{\alpha}_0)\} . \quad (3)$$

Observations from different clusters or subjects are assumed to be independent.

While the assumption that the variance function is a function of the means is also made in GEE type marginal models as considered by Liang and Zeger (1986), the handling of dependencies for observations within the same cluster or subject is different. The correlation of the repeated clustered measurements is considered through a common “working” correlation matrix in GEE. Including the “working” correlation, if chosen correctly, can improve efficiency but is not necessarily an essential feature of GEE, see, e.g., Fitzmaurice (1995) and Crowder (1995, 2001). In contrast, the covariance structure in our proposal is modelled nonparametrically as a function of the means and therefore is an essential component that is part of the model fitting.

2.2. Estimation of the model components

We use column vectors $\mathbf{y}_i = \text{vec}(y_{ij})$ and shorthand notation for column vectors such as $h(\boldsymbol{\alpha}^\top \mathbf{x}_i) = \text{vec}(h(\boldsymbol{\alpha}^\top \mathbf{x}_{ij}))$. The regression parameter vector $\boldsymbol{\alpha}$ is obtained by solving the estimating or quasi-score equations,

$$\sum_{i=1}^n D_i^\top \Omega_i^{-1} \{\mathbf{y}_i - \boldsymbol{\mu}_i\} = 0 . \quad (4)$$

Here, $D_i = (\partial \boldsymbol{\mu}_i / \partial \boldsymbol{\alpha}^\top)$ is a $m_i \times p$ matrix of full rank, $\boldsymbol{\mu}_i = \text{vec}(\mu_{ij})$ with $\mu_{ij} = h(\boldsymbol{\alpha}^\top \mathbf{x}_{ij})$, and V_i^{-1} is the inverse of the variance-covariance matrix for the observations that belong to the i th cluster or subject. The estimating equations for EEE contain nonparametric function estimates for link and variance-covariance functions.

Initial estimates for the regression parameter $\boldsymbol{\alpha}$ can be obtained, for example, by sliced inverse regression (see Li, 1991), ignoring the correlation structure within clusters. Given current estimates for regression coefficients $\hat{\boldsymbol{\alpha}}$, nonparametric link function $\hat{h}(\cdot)$ and variance-covariance functions $\hat{V}(\cdot)$ and $\hat{\Gamma}(\cdot, \cdot)$, updated regression parameter estimates $\tilde{\boldsymbol{\alpha}}$ are obtained by solving the estimated estimating equations,

$$\sum_{i=1}^n \hat{D}_i^\top \hat{\Omega}_i^{-1} (\mathbf{y}_i - \hat{h}(\boldsymbol{\alpha}^\top \mathbf{x}_i)) = 0, \quad (5)$$

where $\hat{D}_i = \partial \hat{h}(\boldsymbol{\alpha}^\top \mathbf{x}_i) / \partial \boldsymbol{\alpha}^\top$, evaluated at $\hat{\boldsymbol{\alpha}}$, and the elements of the covariance matrix $\hat{\Omega}_i$ are $\hat{V}(\hat{\mu}_{ij})$, $1 \leq j \leq m_i$, for the diagonal elements, and $\hat{\Gamma}(\hat{\mu}_{ij}, \hat{\mu}_{i\ell})$, $1 \leq j, \ell \leq m_i$, $j \neq \ell$, for the non-diagonal elements. An additional normalization step required to satisfy the identifiability constraint leads to the updated estimate of the parameter vector $\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} / \|\tilde{\boldsymbol{\alpha}}\|$.

Smoothing methods are used to update the marginal link function $h(\cdot)$ as well as its first derivative $h^{(1)}(\cdot)$, which are both needed to define the estimating equations. We use local polynomial fitting for smoothing and nonparametric estimation of derivatives, denoting the corresponding smoothers by the generic notation $S_L^{(\nu)}$ (see Appendix A); other smoothers (kernels, splines etc.) could also be used. Given current regression parameter estimates $\hat{\boldsymbol{\alpha}}$ and corresponding linear predictors $\hat{\eta}_{ij} = \hat{\boldsymbol{\alpha}}^\top \mathbf{x}_{ij}$, the marginal link function $h(\cdot)$ and its first derivative $h^{(1)}(\cdot)$ are updated via a smoothing step,

$$\hat{h}^{(\nu)}(\eta; \hat{\boldsymbol{\alpha}}) = S_L^{(\nu)} \{ \eta, b_\nu; (\hat{\eta}_{ij}, y_{ij})_{1 \leq i \leq n; 1 \leq j \leq m_i} \}, \quad \nu = 0, 1, \quad (6)$$

which is short-hand for smoothing the scatterplot $\{(\hat{\eta}_{ij}, y_{ij})_{1 \leq i \leq n; 1 \leq j \leq m_i}\}$ and evaluating the smooth at the point η . This smoothing step requires the choice of smoothing parameters b_ν , to be discussed later.

Estimates for the nonparametric variance and covariance functions, $V(\cdot)$ in (M2) and $\Gamma(\cdot, \cdot)$ in (M3), are obtained in two additional smoothing steps. To facilitate the notation, let $e_{ij\ell}$ be the ‘‘raw’’ variance-covariance estimates, $e_{ij\ell} = (y_{ij} - \hat{\mu}_{ij})(y_{i\ell} - \hat{\mu}_{i\ell})$, for $1 \leq i \leq n$ and $1 \leq j, \ell \leq m_i$, where $\hat{\mu}_{ij} = \hat{h}(\hat{\boldsymbol{\alpha}}^\top \mathbf{x}_{ij})$, and $\hat{h}(\cdot)$, $\hat{\boldsymbol{\alpha}}$ are current estimates. The variance function is estimated by smoothing the squared residuals,

$$\hat{V}\{u; \hat{h}(\cdot), \hat{\boldsymbol{\alpha}}\} = S_L^{(0)} \{ u, b; (\hat{\mu}_{ij}, e_{ijj})_{1 \leq i \leq n; 1 \leq j \leq m_i} \}, \quad (7)$$

and the covariance function by two-dimensional smoothing of the cross-products of the residuals,

$$\hat{\Gamma}\{u_1, u_2; \hat{h}(\cdot), \hat{\boldsymbol{\alpha}}\} = S_L^{(0)}[(u_1, u_2), (b_1, b_2); \{(\hat{\mu}_{ij}, \hat{\mu}_{i\ell}), e_{ij\ell}\}_{1 \leq i \leq n; 1 \leq j, \ell \leq m_i; j \neq \ell}]. \quad (8)$$

Note that the elements $\{e_{ij\ell}\}_{j=\ell}$ are used in the one-dimensional smoothing step for estimating the variance function only, but not in the two-dimensional step for estimating the covariance function. This is because extra variation along the diagonal may be induced by measurement errors. Such errors may lead to a ridge along the diagonal of the variance-covariance surface (Staniswalis and Lee, 1998), and including these diagonal elements can cause bias in the covariance function estimate, due to non-differentiability of the surface along the diagonal.

The inverse of the variance-covariance matrix plays an important role in the EEE approach. For computational reasons, the matrix $\boldsymbol{\Omega}$ is decomposed into the diagonal variance matrix \mathbf{V} and a non-diagonal covariance matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Omega} = \mathbf{V} + \boldsymbol{\Gamma}$. Here, the diagonal matrix \mathbf{V} consists of the elements $V(\mu_{ij})$, and the matrix $\boldsymbol{\Gamma}$ of the non-diagonal elements $\Gamma(\mu_{ij}, \mu_{i\ell})$, $j \neq \ell$, with 0's as diagonal elements. Furthermore, we may express $\boldsymbol{\Gamma}$ by a spectral decomposition such that $\boldsymbol{\Gamma} = \mathbf{U}\mathbf{U}^\top$, where $\mathbf{U} = \boldsymbol{\Theta}\boldsymbol{\Lambda}^{1/2}$ with the eigenvectors $\{\boldsymbol{\rho}_i\}$ of $\boldsymbol{\Gamma}$ as the columns of the orthogonal matrix $\boldsymbol{\Theta}$ and the corresponding eigenvalues $\{\lambda_i\}$ as the elements of the diagonal matrix $\boldsymbol{\Lambda}$. Using the Sherman-Morrison-Woodbery formula (see e.g. Golub and Van Loan, 1996), the inverse of the variance-covariance matrix $\boldsymbol{\Omega}$ can thus be obtained by

$$\boldsymbol{\Omega}^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{U}^\top\mathbf{V}^{-1}\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{V}^{-1}, \quad (9)$$

assuming that $(\mathbf{I} + \mathbf{U}^\top\mathbf{V}^{-1}\mathbf{U})$ is nonsingular.

In practice, the estimated variance-covariance matrix $\boldsymbol{\Omega}$ and also $\boldsymbol{\Gamma}$ may be near-singular. In this case, we project the matrix $\boldsymbol{\Gamma}$ to an eigenspace consisting of the eigenvectors with corresponding positive eigenvalues (possibly exceeding a given threshold). Accordingly, the dimensions of the orthogonal matrix $\boldsymbol{\Theta}$ and the diagonal matrix $\boldsymbol{\Lambda}$ are reduced to $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\Lambda}}$ which are the corresponding matrices with positive eigenvalues only, with $\hat{\mathbf{U}} = \hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Lambda}}^{1/2}$ and $\hat{\boldsymbol{\Gamma}} = \hat{\mathbf{U}}\hat{\mathbf{U}}^\top$. Together with the estimated variance $\hat{\mathbf{V}}$, we obtain $\check{\boldsymbol{\Omega}}^{-1} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\hat{\mathbf{U}}(\mathbf{I} + \hat{\mathbf{U}}^\top\hat{\mathbf{V}}^{-1}\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^\top\hat{\mathbf{V}}^{-1}$. For better approximation, we add a nuisance scale parameter ϕ such that $\hat{\boldsymbol{\Omega}} = \hat{\phi}\check{\boldsymbol{\Omega}}$ and $\hat{\phi} = \chi^2(\mathbf{y}; \hat{\boldsymbol{\mu}}, \check{\boldsymbol{\Omega}})/(\sum_{i=1}^n m_i - p)$, with

$$\chi^2\{\mathbf{y}; \hat{\boldsymbol{\mu}}(b_1), \check{\boldsymbol{\Omega}}(\hat{\boldsymbol{\alpha}}_0; b_2)\} = \sum_{i=1}^n (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^\top \check{\boldsymbol{\Omega}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i). \quad (10)$$

Here $\check{\boldsymbol{\Omega}}_i^{-1}$ is a submatrix of $\hat{\boldsymbol{\Omega}}^{-1}$ of size m_i that corresponds to the data from the i -th cluster or subject, and $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \dots, \hat{\mu}_{im_i})^\top$.

An automatic bandwidth selection method is required for the smoothing steps. Cross-validation is one of the possible selection methods but is computationally expensive and has a tendency towards undersmoothing. Instead, we

employ an adaptive bandwidth selection method that is based on the generalized Pearson chi-square statistic χ^2 , leading to

$$\hat{b}_{opt} = \arg \min_{b_1, b_2} |\chi^2\{\mathbf{y}; \hat{\boldsymbol{\mu}}(b_1), \check{\boldsymbol{\Omega}}(\hat{\boldsymbol{\alpha}}; b_2)\} - (\sum_{i=1}^n m_i - p)|.$$

The entire estimation procedure can be implemented iteratively until some convergence criterion is met. In practice, as shown in the simulation results, a one-step estimation procedure gives satisfactory results when for example the initial estimates of the regression coefficients are obtained by sliced inverse regression (SIR; Li, 1991).

3. Asymptotic inference

Interest in the asymptotic results focuses on the asymptotic properties of the estimated regression parameters which correspond to the coefficients for the fixed effects linear predictors. Often a hypothesis of interest, for example a treatment effect, can be formulated in terms of these parameters. Using the marginal EEE approach, the key to obtain the asymptotic normality of the estimated regression parameters lies in uniform consistency properties of the estimated link and variance-covariance functions.

For the estimated link function $\hat{h}(\cdot)$ (6), uniform convergence can be obtained in a manner similar to the proof of Theorem 1 of Chiou and Müller (1999). Under appropriate regularity conditions on the design density of the linear predictor and the smoother, link function estimates $\hat{h}^{(\nu)}(\cdot)$ (6) satisfy

$$\sup_{\eta \in I_\eta} |\hat{h}^{(\nu)}(\eta) - h^{(\nu)}(\eta)| = o_p(1), \quad \nu = 0, 1, \quad (11)$$

where I_η is a compact interval based on the domain of the linear predictor η , which is assumed to be compact.

Consistency results for variance-covariance surface estimation hold under mild regularity conditions. For the estimated variance function $\hat{V}(\cdot)$ (7), the proof is similar to Theorem 2 of Chiou and Müller (1999). Analogous techniques can be applied to derive uniform consistency with modifications due to the dependency of data within clusters. The consistency result for the estimated covariance function $\hat{\Gamma}(\cdot, \cdot)$ (8) involves two-dimensional smoothing (Ruppert and Wand, 1994). Again, under regularity conditions on the design density, uniform consistency properties can be obtained, leading to

$$\sup_{u \in I_\mu} |\hat{V}(u) - V(u)| = o_p(1), \quad (12)$$

$$\sup_{u_1, u_2 \in I_\mu} |\hat{\Gamma}(u_1, u_2) - \Gamma(u_1, u_2)| = o_p(1), \quad (13)$$

where I_μ is a compact interval based on the domain of the means μ and their design density in (7) and (8).

Because of the identifiability constraint $\|\alpha_0\| = 1$, the p -dimensional vector of the regression parameters is actually of rank $p - 1$. Let $\hat{\alpha}' = (\hat{\alpha}_1, \dots, \hat{\alpha}_{p-1})^\top$, where $\hat{\alpha}_i = \tilde{\alpha}_i / \|\tilde{\alpha}\|$ and $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_p)$ are the solutions of the estimating equations (5) without implementing the normalization step. The derivation of the asymptotic covariance of the parameter vector will be based on the normalized version $\hat{\alpha}'$. To facilitate the notation, we define $f_i : \mathbb{R}^p \rightarrow \mathbb{R}$, $1 \leq i \leq p - 1$, such that $f_i(\mathbf{u}) = u_i / \|\mathbf{u}\|$, where $\mathbf{u} = (u_1, \dots, u_p)^\top$, and $f : \mathbb{R}^p \rightarrow \mathbb{R}^{p-1}$ such that $f(\mathbf{u}) = \{f_1(\mathbf{u}), \dots, f_{p-1}(\mathbf{u})\}^\top$. Accordingly, we have $\hat{\alpha}_i = f_i(\tilde{\alpha})$ and $\hat{\alpha}' = f(\tilde{\alpha})$, and further define a $(p - 1) \times p$ matrix $(Df)(\mathbf{u})$ such that

$$(Df)(\mathbf{u}) = \left\{ \frac{\partial f_i(\mathbf{u})}{\partial u_j} \right\}_{1 \leq i \leq p-1, 1 \leq j \leq p}. \quad (14)$$

The basic idea is to obtain the asymptotic covariance of $\hat{\alpha}'$ by adjusting the asymptotic covariance of $\tilde{\alpha}$ via the multivariate delta method.

Theorem 1. In the semiparametric marginal model (M1)-(M3), under appropriate regularity conditions, the following event holds with probability $1 - \delta$, for a given arbitrarily small $\delta > 0$: There exists a solution $\hat{\alpha}$ of the estimated estimating equation (5) that satisfies $\|\hat{\alpha}\|=1$ and as $n \rightarrow \infty$,

$$\sqrt{n} \{f(\hat{\alpha}) - f(\alpha_0)\} \xrightarrow{d} N_{p-1}[\mathbf{0}, \{(Df)(\alpha_0)\} \Sigma^{-1} \{(Df)(\alpha_0)\}^\top], \quad (15)$$

where $\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n D_i^\top \Omega_i^{-1} D_i$ and $D_i = \{\partial h(\alpha^\top \mathbf{x}_i) / \partial \alpha^\top\}|_{\alpha_0}$.

A sketch of the proof is given in Appendix B. The asymptotic covariance on the r.h.s. of (15) can be consistently estimated by substituting estimates obtained in the last step of iteration (E1)-(E3) for the unknown quantities. This leads to the approximation

$$f(\hat{\alpha}) \sim N_{p-1} \left[f(\hat{\alpha}), \frac{1}{n} \{(Df)(\hat{\alpha})\} \hat{\Sigma}^{-1} \{(Df)(\hat{\alpha})\}^\top \right], \quad (16)$$

where $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \hat{D}_i^\top \hat{\Omega}_i^{-1} \hat{D}_i$ and $\hat{D}_i = \{\partial h(\alpha^\top \mathbf{x}_i) / \partial \alpha^\top\}|_{\hat{\alpha}}$, which provides the basis for inference in the finite sample situation.

4. Relationship of estimated estimating equations to generalized linear mixed models

4.1. The marginal version of a generalized linear mixed model

We begin with a brief description of conditionally specified GLMMs for clustered/longitudinal data. Let $\{(Y_{ij}, X_{ij}, Z_{ij})_{1 \leq i \leq n, 1 \leq j \leq m_i}\}$ denote the j th repeated observation for the i th subject or experimental unit, where Y_{ij} are as in

Section 2.1, and the associated vectors of explanatory variables are X_{ij} in \mathbb{R}^p for the fixed and Z_{ij} in \mathbb{R}^q for the random effects, where both X_{ij} and Z_{ij} are assumed to be random vectors. The GLMM is defined conditionally as

$$E(Y_{ij}|\mathbf{b}_i, X_{ij}, Z_{ij}) = g(X_{ij}^\top \boldsymbol{\beta}_0 + Z_{ij}^\top \mathbf{b}_i), \quad (17)$$

where $\boldsymbol{\beta}_0$ in \mathbb{R}^p and \mathbf{b}_i in \mathbb{R}^q are vectors of fixed and random regression parameters, respectively, and $g(\cdot)$ is the link function. The link function $g(\cdot)$ is assumed to be known in conventional versions of GLM and GLMM. For our considerations it is enough to assume that the function $g(\cdot)$ is smooth (infinitely often differentiable) but unknown otherwise. The distribution of the random effects \mathbf{b}_i may also be assumed unknown, with an unknown probability density function $f_{\mathbf{b}}(\cdot)$, with the property that all moments of the distribution $F_{\mathbf{b}}(\cdot)$ exist.

A Taylor expansion of $g(X_{ij}^\top \boldsymbol{\beta}_0 + Z_{ij}^\top \mathbf{b}_i)$ around the argument $X_{ij}^\top \boldsymbol{\beta}_0$, using $E(Y_{ij}|X_{ij} = \mathbf{x}) = E\{E(Y_{ij}|X_{ij} = \mathbf{x}, Z_{ij}, \mathbf{b}_i)\} = \int E\{E(Y_{ij}|X_{ij} = \mathbf{x}, Z_{ij}, \mathbf{b}_i = \mathbf{b})\} f_{\mathbf{b}}(\mathbf{b}) d\mathbf{b}$, yields

$$E(Y_{ij}|X_{ij} = \mathbf{x}) = g(\mathbf{x}^\top \boldsymbol{\beta}_0) + \sum_{k=1}^{\infty} \gamma_{0k} g^{(k)}(\mathbf{x}^\top \boldsymbol{\beta}_0), \quad (18)$$

where $\gamma_{0k} = E\{\gamma_k(Z_{ij})\}$, $\gamma_k(\mathbf{z}) = \int (\mathbf{z}^\top \mathbf{b})^k f_{\mathbf{b}}(\mathbf{b}) d\mathbf{b} / k!$ and $g^{(k)}(\cdot)$ is the k -th derivative of $g(\cdot)$. If the sum on the right-hand-side of (18) exists for all possible values of $u = \mathbf{x}^\top \boldsymbol{\beta}_0$, then a new link function $h(\cdot)$ is determined by $h(u) = g(u) + \sum_{k=1}^{\infty} \gamma_{0k} g^{(k)}(u)$, where we assume h to be invertible. Therefore, $E(Y_{ij}|X_{ij} = \mathbf{x}) = h(\mathbf{x}^\top \boldsymbol{\beta}_0)$, i.e. the mean function in the marginal version of the GLMM satisfies (1) in (M1), with $\boldsymbol{\alpha}_0 = \boldsymbol{\beta}_0$. This identity together with the following considerations implies that inference for $\boldsymbol{\alpha}_0$ in the population-average model obtained with EEE, as discussed in Section 3, simultaneously provides inference for the fixed effects parameter $\boldsymbol{\beta}_0$ in a subject-specific GLMM of the form (17).

As a consequence of $\boldsymbol{\alpha}_0 = \boldsymbol{\beta}_0$ and another Taylor expansion we obtain, under additional regularity conditions, that there exists a function $s(\cdot)$ such that for all possible values of $u = \mathbf{x}^\top \boldsymbol{\alpha}_0$

$$E(Y_{ij}^2|X_{ij} = \mathbf{x}) = s(\mathbf{x}^\top \boldsymbol{\alpha}_0).$$

Defining the function $v = s \circ h^{-1}$, we obtain

$$E(Y_{ij}^2|X_{ij} = \mathbf{x}) = v\{h(\mathbf{x}^\top \boldsymbol{\alpha}_0)\}.$$

The function $V(u) = v(u) - u^2$ then plays the role of a variance function in the corresponding marginal model. We find that the variance function of the marginal version of the GLMM satisfies (2) in (M2), with the nonparametric link function v .

For clustered/longitudinal data, the random effects included in the GLMM play a crucial role to account for the dependencies within clusters. Let Y_{i1} and Y_{i2} be any two dependent observations as in (M3) with the corresponding vector of covariates (X_{i1}, Z_{i1}) and (X_{i2}, Z_{i2}) , and a subject-specific random effect \mathbf{b}_i . Assume further that, conditional on the random effect \mathbf{b}_i , Y_{i1} and Y_{i2} are independent. By yet another Taylor expansion, using again $\boldsymbol{\alpha}_0 = \boldsymbol{\beta}_0$ and $E(Y_{i1}Y_{i2}|X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2) = E\{E(Y_{i1}Y_{i2}|X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2, Z_{i1}, Z_{i2})\} = \int E\{E(Y_{i1}Y_{i2} | \mathbf{b}_i = \mathbf{b}, X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2, Z_{i1}, Z_{i2})\}f_{\mathbf{b}}(\mathbf{b})d\mathbf{b}$, we find

$$\begin{aligned} E(Y_{i1}Y_{i2} | X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2) &= E\left\{\int \sum_{\ell=0}^{\infty} \sum_{j+k=\ell} \frac{g^{(j)}(\mathbf{x}_1^\top \boldsymbol{\alpha}_0) g^{(k)}(\mathbf{x}_2^\top \boldsymbol{\alpha}_0)}{j! k!} \right. \\ &\quad \left. \times (Z_{i1}^\top \mathbf{b})^j (Z_{i2}^\top \mathbf{b})^k f_{\mathbf{b}}(\mathbf{b})d\mathbf{b}\right\} \\ &= \sum_{\ell=0}^{\infty} \sum_{j+k=\ell} \zeta_{0jk} \binom{\ell}{j} g^{(j)}(\mathbf{x}_1^\top \boldsymbol{\alpha}_0) g^{(k)}(\mathbf{x}_2^\top \boldsymbol{\alpha}_0) \\ &=: r(\mathbf{x}_1^\top \boldsymbol{\beta}_0, \mathbf{x}_2^\top \boldsymbol{\alpha}_0), \end{aligned}$$

where $\zeta_{0jk} = E\{\zeta_{jk}(Z_{i1}, Z_{i2})\}$ and $\zeta_{jk}(\mathbf{z}_1, \mathbf{z}_2) = \int (\mathbf{z}_1^\top \mathbf{b})^j (\mathbf{z}_2^\top \mathbf{b})^k f_{\mathbf{b}}(\mathbf{b})d\mathbf{b}$, and $r(\cdot, \cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function that we assume to be well-defined. Again, under the assumption of the existence of the summands on the right-hand-side above and defining the function $\rho(u_1, u_2) = r\{h^{-1}(u_1), h^{-1}(u_2)\}$ for all possible values of $u_1 = \mathbf{x}_1^\top \boldsymbol{\alpha}_0$ and $u_2 = \mathbf{x}_2^\top \boldsymbol{\alpha}_0$ so that $r(u_1, u_2) = \rho\{h(u_1), h(u_2)\}$, we have

$$E(Y_{i1}Y_{i2}|X_{i1} = \mathbf{x}_1, X_{i2} = \mathbf{x}_2) = \rho\{h(\mathbf{x}_1^\top \boldsymbol{\alpha}_0), h(\mathbf{x}_2^\top \boldsymbol{\alpha}_0)\}. \quad (19)$$

Defining $\Gamma(u_1, u_2) = \rho(u_1, u_2) - u_1 u_2$, these derivations imply that the marginal version of the GLMM satisfies model assumption (3) regarding the covariance structure in (M3).

We conclude that under some mild regularity conditions, the marginal version of GLMM is a quasi-likelihood model with unknown link and variance-covariance functions. The variance function is a function of the mean as in the conventional GLM, while the dependence of the repeated measurements within the same subject or cluster is described by a variance-covariance structure that also is a function of the means. These features of the marginal version of a GLMM are seen to be the defining features of EEE. This close connection between marginal versions of GLMM and EEE provides additional motivation for the EEE approach and shows that assumptions (M1)-(M3) are reasonable and not only sufficient for asymptotic inference, but also necessary for a subject-specific interpretation of EEE.

4.2. Conditional and marginal link functions

We illustrate the relationship between marginal and conditional versions of a GLMM for a simplified situation where the random effects b_i are one-dimensional,

assuming that the distribution of $Y_{ij}|b_i$ is independent Bernoulli with random effects b_i that are identically and independently normally distributed with mean 0 and variance σ^2 . The corresponding conditional version of the GLMM is

$$E(Y_{ij} | b_i, X_{ij} = \mathbf{x}) = g(\mathbf{x}^\top \boldsymbol{\beta}_0 + b_i).$$

Following the development in the derivation of (M1) for the marginal version of this GLMM in Section 4.1, we obtain the marginal link function $h(\cdot)$,

$$h(u) = g(u) + \sum_{k=1}^{\infty} \gamma_k g^{(k)}(u) = g(u) + \sum_{k=1}^{\infty} h_k(u), \quad (20)$$

defining the adjustment functions $h_k(u) = \gamma_k g^{(k)}(u)$, $k \geq 1$, where $\gamma_k = \int b^k f_b(b) db / k!$.

Setting $g_s(u) = \{e^u / (1 + e^u)\}^s$, we find $h_{2k-1}(u) = 0$ for $k = 1, 2, \dots$, and $h_2(u) = \{g_1(u) - 3g_2(u) + 2g_3(u)\} / 2!$, $h_4(u) = 3\{g_1(u) - 15g_2(u) + 50g_3(u) - 60g_4(u) + 24g_5(u)\} / 4!$, and so on, according to (18). We note $Eb = 0$ implies that $\gamma_1 = 0$ and $h_1(\cdot) = 0$. The non-zero adjustment functions h_2 and h_4 are displayed in Fig. 1B. Variance and covariance are obtained as in (2) and (3). Under the conditional Bernoulli assumption, we find $Var(Y_{ij} | X_{ij} = \mathbf{x}) = V\{h(\mathbf{x}^\top \boldsymbol{\alpha}_0)\} = h(\mathbf{x}^\top \boldsymbol{\alpha}_0) - h^2(\mathbf{x}^\top \boldsymbol{\alpha}_0)$. The explicit form of the function $\rho(\cdot, \cdot)$ in (19) may be complicated, but it is not needed since the function $\Gamma(\cdot, \cdot)$ in (3) can be directly estimated from the data, and the same applies for the link function $h(\cdot)$.

In the following, we compare conditional link function $g(\cdot)$ and marginal link function $h(\cdot)$ through several examples, where we assume that the predictors X_{ij} are uniformly distributed in $(-1, 1)$. In these examples, the data are assumed to have 50 clusters, each with five repeated measurements in the conditional GLMM.

Logit link and normal random effects. The random effects b_i are assumed to follow the standard normal distribution, and conditional on the random effects b_i , the response variables Y_{ij} are Bernoulli with conditional means $\mu_{ij}^C = g(\eta_{ij}^C) = e^{\eta_{ij}^C} / (1 + e^{\eta_{ij}^C})$, where $\eta_{ij}^C = \mathbf{x}_{ij}^\top \boldsymbol{\alpha}_0 + b_i$. The relationship between the logit link $g(\cdot)$ and the marginal link $h(\cdot)$ is as in (20). The nonparametric link $h(\cdot)$, obtained by using the four adjustment functions h_i (obtained as above, see Fig. 1B), in this case is similar to but nevertheless noticeably different from the conditional link g (Fig. 1A).

Log link and normal mixture random effects. The random effects b_i follow a mixture of two normal distributions $0.5N(1, 0.5^2) + 0.5N(-1, 1.5^2)$. Conditional on the random effects b_i , the responses Y_{ij} are Poisson with conditional means $\mu_{ij}^C = g(\eta_{ij}^C) = e^{\eta_{ij}^C}$. The resulting $h_k(\cdot)$, for $k = 1, \dots, 4$, in (20) are $h_1(u) = h_3(u) = 0$, $h_2(u) = 5e^u / 2!$, and $h_4(u) = 3.75e^u / 4!$. We approximate the resulting marginal link function $h(\cdot)$ by using the first four adjustment functions (Fig. 1C), leading to clear differences between marginal link $h(\cdot)$ and conditional link $g(\cdot)$. The non-zero adjustment functions h_2 and h_4 are displayed in Fig. 1D.

Log link and Gamma random effects. The random effects b_i follow a Gamma distribution with density $f_b(b) = (rb)^{\alpha-1}re^{-rb}/\Gamma(b)$, $b \geq 0$, with shape parameter $\alpha = 1.5$ and rate parameter $r = 1$. The response variables Y_{ij} are normally distributed with conditional means $\mu_{ij}^C = g(\eta_{ij}^C) = \log(\eta_{ij}^C)$. The first four adjustment functions $h_k(\cdot)$ are found to be $h_1(u) = 0$, $h_2(u) = -.75/u^2$, $h_3(u) = 1/u^3$, and $h_4(u) = -3.9375/u^4$. The comparison of conditional and marginal link functions is in Fig. 1E, and the adjustment functions are illustrated in Fig. 1F. The differences between the two link functions are quite striking for this model.

Identity link. If we assume that the link function is the identity link, we have $g^{(1)}(\cdot) = 1$ and $g^{(k)}(\cdot) = 0$ for $k \geq 2$. According to (20), since $Eb = 0$, $h_k(\cdot) = 0$ for all k . Therefore, for the identity link function, the conditional and marginal link functions coincide.

Insert Figure 1 around here.

5. Illustrations of estimated estimating equations

5.1. Epileptic seizure count data

This data set stems from a progabide trial of 59 epileptics (Leppik et al., 1987) and has become a reference data set to illustrate methods for longitudinal data (Thall and Vail, 1990; Breslow and Clayton, 1993; Diggle, Liang and Zeger, 2002; Lee and Nelder, 1996; Jowaheer and Sutradhar, 2002). Patients in this study were randomized to receive either the new drug or a placebo. The number of seizure counts was recorded in four consecutive 2-week intervals with baseline counts recorded in the preceding 8-week period. Age in years for each patient is also available. We follow previous suggestions that the observations of one patient should be deleted, leaving 58 patients in the analysis. We first consider the predictor variables as defined in Thall and Vail (1990), which include *Base* as the logarithm of the baseline seizure counts divided by 4, *Age* as the logarithm of the age in years when entering the study, *Trt* as the indicator variable with 1 for treatment and 0 for placebo, *Visit₄* as the indicator variable with 1 for the fourth visit and 0 at other visits, and *TrtBase* as the interaction term of *Trt* and *Base*.

Insert Table 1 around here.

The parameter estimates are compared in Table 1 for four different approaches. These include the proposed EEE approach, with unknown random effects, link and variance-covariance functions (Model *EEE₁*), GLMMs based on

Poisson variance function and log link with a random intercept (Model $GLMM_1$), and the conventional GEE approach (Model GEE_1), assuming an overdispersed Poisson variance function with log link and unstructured working correlation structure. In addition, the comparison includes the best model (Model $T\&V$) among a class of models specifically adapted to these data and proposed by Thall and Vail (1990). This model features a specified parametric variance, $\mu_t + (\alpha_0 + \alpha_t)\mu_t^2$, and covariance functions, $\alpha_0\mu_t\mu_u$, where μ_t is the mean at time t , and α_0 and α_t are free model parameters.

We note that Model EEE_1 does not include an intercept term since the identifiability constraint requires that the norm of the estimated parameter vector is one (see (M1)). Models $GLMM_1$ and GEE_1 were fitted by PROC NLMIXED and PROC GENMOD, respectively, in SAS (SAS Institute, 2000). The four fitted models differ in the significance attached to the explanatory variables and in the sign of the regression coefficient for age. Only the proposed Model EEE_1 and Model $T\&V$ attach significance to the treatment effect. In all models baseline seizure count has a positive significant effect, in contrast to the interaction effect of treatment and baseline that is not significant in any model. The significance of the effect of the fourth visit varies among the four models. While *Age* has a negative coefficient in Model EEE_1 without intercept, it is positive in the other models which all include a negative intercept in the linear predictor. This discrepancy could be due to an effect of including versus not including an intercept term, lack of fit of the parametric link functions, a nonlinear *Age* effect that is not well reflected in the linear predictor, or possible interaction effects between age and other predictors.

To further investigate the nature of the age effect, we divide age into three categories by introducing indicator variables: Age_1 for age bracket [25, 35) and Age_2 for those with age ≥ 35 , default being age < 25 . We also explore interaction effects between age and treatment by setting $TrtAge_1 = Trt \cdot Age_1$ for the interaction term between Trt and Age_1 and $TrtAge_2 = Trt \cdot Age_2$ for that between Trt and Age_2 . We drop the term $TrtBase$ because of its insignificance. The models which include these interactions are referred to as EEE_2 , GEE_2 and $GLMM_2$, and the corresponding parameter estimates are listed in Table 2.

Insert Table 2 around here.

Overall, the inference obtained for Model EEE_2 appears to be reasonable. Age is seen to interact with the treatment effect. Compared with the previous analysis, the model fits for $GLMM_2$ and GEE_2 now attach significance to the treatment effect, although it is less significant than in Model EEE_2 . The EEE approach shows a highly nonlinear interaction of treatment and age, which explains the negative age parameter seen in the previous analysis: For the untreated group, increasing age has an increasingly negative (seizure suppressing)

effect; for the treated group, the treatment works best for the youngest age group, and least for the middle age group. The tendency of these interactions is the same for the two parametric models as for EEE, and while they do not reach significance in these models, it appears that the inclusion of these interactions (and perhaps allowing nonlinear age effects through the indicator variables) is instrumental in reaching the significance level for the treatment effect.

Insert Figure 2 around here.

The nonparametric link and variance-covariance function estimates for this situation are displayed in Fig. 2. The marginal link function estimate (solid curve) in Fig. 2 (upper panel) reveals clear deviations from the inverse of the log link function, $exp(\hat{\eta})$ (dashed curve). The log link function is the canonical link for Poisson responses under the framework of the GLM, and is being used as link for the parametric models. The ridge line rising along the diagonal in Fig. 2 (lower panel) is the estimated variance function. It indicates additional variation for the response variable along the diagonal, and essentially supports a quasi-Poisson variance structure in which the variance is proportional to the mean.

The nonparametric nature of the link and variance-covariance functions in EEE enhances the flexibility for modelling clustered data. The interpretation of the parameter vector estimate obtained in EEE is intrinsically tied to the link function. The nonparametric link function estimate in Fig. 2 (upper panel) is monotone increasing and suggests possible parametric link functions. It is quite different from the parametric log link function, that has been used in previous analyses. The nonparametric link function estimate is useful to gauge the change in the response under changing covariate levels.

5.2. Model diagnostics and interpretation

We have shown in Section 4 that the marginal link function obtained by integrating over the random effects of a GLMM generally differs from the conditional link function. The estimated nonparametric link function, that is an integral part of the EEE approach, may however serve as a diagnostic tool for commonly specified marginal link functions. An example is provided in Fig. 2, where the shape of the estimated link function clearly differs from the inverse of a log link function and instead motivates a two-phase linear link.

Implementing this suggestion, the unknown link function $h(\cdot)$ for the marginal mean in (M1) is modelled by a two-phase linear function of the linear predictor $\eta = \mathbf{x}^\top \boldsymbol{\alpha}_0$, leading to the model

$$E(Y | \mathbf{x}) = \mu = \beta_0 + \beta_1 \eta + \beta_2 (\eta - \eta_c)_+,$$

where $\{\beta_i; i = 0, 1, 2\}$ and η_c are parameters that determine the link function. Here $(\eta - \eta_c)_+ = (\eta - \eta_c)$ if $\eta \geq \eta_c$, and $(\eta - \eta_c)_+ = 0$ otherwise. The regression coefficients β_i and the parameter η_c could be estimated via quasi-likelihood or a profile likelihood. In the following, we adopt a simple alternative, obtaining estimates of these parameters by minimizing the L_2 distance between the two-phase linear link function and the nonparametric link from the EEE approach.

Insert Figure 3 around here.

The L_2 distance between the parametric and the nonparametric links is minimized at the change point location $\hat{\eta}_c = 0.72$, as shown in the upper panel of Fig. 3, and the corresponding minimizing coefficient values are (1.460, 5.003, 12.161) for $(\beta_0, \beta_1, \beta_2)$. As the lower panel of Fig. 3 demonstrates, the nonparametric link that was obtained in the EEE analysis is closely approximated by this two-phase linear link function, which then can be used as link for EEE or other marginal models. Choosing this link, the comparisons between EEE (Model EEE_3 , noting that since the link is fixed, an intercept term may be included) with the conventional GEE with unstructured working correlation (Model GEE_3) are in Table 3.

Insert Table 3 around here.

Models EEE_3 and GEE_3 are seen to be very close in terms of estimated regression coefficients, attaching significance to almost the same effects, with very similar estimates for treatment effect, baseline, fourth visit effect and treatment-age1 interaction. Comparing the results for Models EEE_2 and EEE_3 , the significant effects are the same, except for the interaction of treatment and the higher age group. Regarding Models GEE_2 and GEE_3 , the differences in both significant effects and parameter estimates are large. This is likely a consequence of the differences in the link functions that are used, the canonical log link in Model GEE_2 and the two-phase linear link in Model GEE_3 where the latter resulted from the empirical link analysis afforded through the EEE approach.

It is often of interest to evaluate the effect on the response when changing/increasing a predictor by one unit, while keeping other covariates fixed. For both GEE and EEE, this effect depends on the chosen link function. For some parametric links such as log or logit links, relatively straightforward interpretations exist, while for most nonlinear and nonparametric links, the predicted effect of changes in the level of one covariate will depend on the levels of other covariates. For example, fixing baseline seizure count at the median level 22, the predicted values of seizure counts for the fourth visit can be calculated for the

three age groups under placebo and treatment for the three models as shown in Table 4. The methods agree reasonably well with each other.

Insert Table 4 around here.

This analysis suggests that the true link is close to the two-phase linear shape, and that Models GEE_3 and EEE_3 reflect the nature of the complex relationship between the predictors and the response reasonably well, indicating that the treatment effect is modified by age group. In this example, the diagnostic features of the EEE approach lead us to the conclusion that the log link that is canonical for Poisson data may not be the most suitable link. Additional discussion on the interpretation of a nonparametric link function can be found in Klimov, Delecroix, and Simar (2002).

5.3. Simulation study for binary response data

The simulation design assumes an underlying GLMM for binary responses with logit link and a random intercept. Two distributions for the random intercept b are considered: a “centered” Gamma distribution, obtained by subtracting the mean from a Gamma distribution with shape parameter 10 and rate parameter 1, and a Gaussian mixture of $Normal(-1, 10)$ and $Normal(1, 10)$ with equal probabilities. We assume three explanatory variables x_1 , x_2 and x_3 , where $x_{1ij} = i/n + j/m$, x_{2ij} is assumed to be uniformly distributed in $(-1, 1)$ and x_{3ij} is either 0 or 1 with probability $\frac{1}{2}$. The underlying regression coefficients are assumed to be $\alpha_1 = 0.8$, $\alpha_2 = -0.6$ and $\alpha_3 = 0$ for x_1 , x_2 and x_3 , respectively. The fixed effects parameter vector of interest in the underlying GLMM (17) is $\beta_0 = (\alpha_0, \alpha_1, \alpha_2)^\top$, and according to the derivations in Section 4.1, we may use the identity $\beta_0 = \alpha_0$, where α_0 is the parameter vector of the EEE (1). Accordingly, we obtain inference for this fixed effects parameter in the GLMM and the equivalent parameter in EEE and compare the results in this simulation that we obtain for this parameter via the GLMM and EEE approaches.

Each simulated data set consists of 50 individuals with 5 repeated observations each. We note that the “stock” GLMM that we apply in this analysis is supplied with the correct link function but operates under the erroneous assumption that the random effects are Gaussian.

Insert Tables 5 and 6 around here.

The results are shown in Table 5 for the Gamma random effects and in Table 6 for the Gaussian mixture, based on 200 simulation runs each. The proposed

EEE approach clearly outperforms GLMM in terms of the overall mean squared error (MSE) and the empirical power (rejection rate) for $H_0 : \alpha_i = 0$, $i = 1, 2, 3$, based on testing at the 5% significance level; for the first two parameters, empirical power (rejection rates) should be as large as possible, while for the third parameter it should be at 5%, since there the null hypothesis holds.

The proposed EEE approaches have the highest empirical rejection rates for α_1 and α_2 , indicating that EEE has more power to detect significant covariate effects than GLMM in these settings. On the other hand, the rejection rates of EEE for α_3 implies close tracking of the nominal level. The only case where GLMM tracks the target rejection rate better is for α_3 in the Gaussian mixture case. Power of the GLMM to detect nonzero parameters is uniformly low, and lengths of 95% confidence intervals are substantially larger for GLMM throughout. Overall, this simulation confirms the theoretical considerations from which we concluded that inference for the fixed effects parameter in GLMM is feasible via EEE. The practical finite sample performance of EEE in this simulation setting is seen to be quite competitive, even though this approach does not rely on assumptions regarding random effects or link function.

6. Concluding remarks

The proposed EEE approach provides semiparametric estimation and inference for the marginal modelling of clustered/longitudinal data, and works with unknown link and variance-covariance functions. This approach is flexible, conceptually simple and easy to implement. Neither does it require high-dimensional integration as GLMM, nor does it rely on pre-specified link function, working correlation or specified nature of random effects as GEE or GLMM do. The estimation procedure can be implemented in an iterative algorithm that alternates between nonparametric updating steps for link and variance-covariance functions and parametric updating steps by solving a current estimated estimating equation. The asymptotic normality property of the estimated regression parameters is retained.

One of the interesting findings is that under mild assumptions, the marginal version of a given GLMM with smooth unknown link function and unknown nature of the random effects falls into the framework of EEE, and moreover that the parameter vector in the EEE approach is identical to the fixed effects parameter vector of an underlying GLMM. This demonstrates the compatibility between the subject-specific GLMMs and the population-average EEE approach and, for statistical practice, allows the user to obtain inference for the fixed effects parameter in a GLMM through the corresponding parameter in EEE. Our simulations confirm this, and in this sense, EEE, although corresponding to a marginal approach, carries some subject-specific interpretation.

The proposed marginal model can be regarded as an extension of the semi-parametric quasi-likelihood regression (SPQR) approaches developed in Chiou

and Müller (1998, 1999) to the case of clustered/longitudinal data. The correlation of clustered/longitudinal data is reflected by the marginal nonparametric covariance function which is a key component of the proposed EEE approach. Estimation of the corresponding covariance surface as a function of the means is easy to implement through two-dimensional smoothing. This device allows for fairly arbitrary dependence structures which do not need to be pre-specified. Link, variance and covariance function estimates provide additional information about the marginal relationships in the data, noting that these are the only relationships that can be checked against actual observations. In that sense, the proposed approach is fully data-driven and uses only features of the data that can be empirically checked.

For these reasons, the proposed model provides valuable diagnostics for link function selection, and more generally model checking for any model fitted by one of the standard approaches. Beyond a convenient implementation of the marginal version of GLMMs it is also a flexible alternative to the GEE family of models, and is a data-exploratory tool. It is noteworthy that previous analyses of the epilepsy data that were discussed in Section 5.1 have led to different conclusions, depending on which among various standard approaches was adopted. Our detailed analysis, employing the diagnostic features of the EEE approach, is pointing to misspecified model components as a possible cause of these differences.

We conclude that the proposed EEE approach serves two main functions: It guides appropriate extension and modification of more standard parametric approaches, and also may be used to obtain inference for the fixed effects in GLMMs. The proposed semiparametric marginal modelling approach through estimated estimating equations therefore provides a flexible and “model-robust” tool for analysis and inference of clustered/longitudinal data.

Acknowledgements

We wish to thank the Associate Editor and Editor for helpful comments. Special thanks are due to a referee for detailed scrutiny and many clarifications. The reviewer’s suggestions led to numerous improvements in the paper. This research was supported in part by NSF grants DMS02-04869 and DMS03-54448, and a grant from Academia Sinica 94B001-1.

Appendix A: Smoothing methods

We introduce generic notation for the smoothers and a brief description of the smoothing procedures. We first consider the single covariate case. Let $\{(U_i, V_i)_{1 \leq i \leq k}\}$ be scatterplot data, where k is the number of data, with underlying regression function $g(u) = E(V|U = u)$. The nonparametric regression function estimate or smoother for $g^{(\nu)}(u)$, the ν th derivative of $g(u)$, is denoted

by $\hat{g}^{(\nu)}(u) = S^{(\nu)}\{u, b; (U_i, V_i)_{1 \leq i \leq k}\}$, where b is a smoothing parameter. The smoothers can be implemented as kernel estimators, splines or local polynomial fits.

We choose the local polynomial smoother, denoted by $S_L^{(\nu)}$, and obtained by fitting polynomials by the weighted least squares method to the data in local windows (Fan and Gijbels, 1996). The smoothed estimates are obtained by minimizing a weighted least squares criterion,

$$\min_{\{a_j\}} \sum_{i=1}^n K \left\{ \frac{u - U_i}{b} \right\} \left\{ V_i - a_0 - \sum_{j=1}^q a_j (u - U_i)^j \right\}^2,$$

and setting

$$S_L^{(\nu)}\{u, b, (U_i, V_i)_{1 \leq i \leq k}\} = \nu! \hat{a}_\nu. \quad (21)$$

This smoother has nice features and includes automatic bias adjustment near endpoints. Note that the degree of polynomial q is usually chosen such that $q - \nu$ is odd; we choose $q = 1$ for $\nu = 0$ and $q = 2$ for $\nu = 1$. For the kernel function, we use the Epanechnikov kernel $K(x) = (1 - x^2)1_{\{\|x\| \leq 1\}}$, which is the optimal kernel in this situation (Müller, 1987). Other common choices for the kernel function include the Gaussian kernel, $K(x) = \exp(-x^2/2)$ or the bi-quadratic kernel, $K(x) = (1 - x^2)^2 1_{\{\|x\| \leq 1\}}$.

For smoothing the covariance function, a multivariate smoothing method is needed. The regression function is then $g(u_1, u_2) = E(V|U_1 = u_1, U_2 = u_2)$ for the case of two predictors. In analogy to the one-dimensional case, we use local weighted least squares, fitting local planes to the data. Given scatterplot data with bivariate predictors $\{(U_{i1}, U_{i2}, V_i)_{i=1, \dots, k}\}$, the smoother S_L is denoted as

$$\hat{g}(u_1, u_2) = S_L\{(u_1, u_2), (b_1, b_2); (U_{i1}, U_{i2}, V_i)_{1 \leq i \leq k}\} = \hat{a}_0, \quad (22)$$

where $(\hat{a}_0, \hat{a}_1, \hat{a}_2)$ are the minimizers of the local weighted sum of squares

$$\sum_{i=1}^n K \left(\frac{u_1 - U_{i1}}{b_1}, \frac{u_2 - U_{i2}}{b_2} \right) \{V_i - a_0 + a_1(U_{i1} - u_1) + a_2(U_{i2} - u_2)\}^2.$$

Here, $K(\cdot, \cdot) \geq 0$ is a two-dimensional kernel function, such as a two-dimensional analogue of the Epanechnikov weight function, $K(u, v) = \{1 - (u^2 + v^2)^{1/2}\} 1_{\{u^2 + v^2 \leq 1\}}$, and (b_1, b_2) is a pair of bandwidths. This corresponds to the local fitting of a weighted least squares plane within the window, evaluated at the midpoint of this window.

Appendix B: Sketch of proof of Theorem 1

We first define the quasi-score function for the i th cluster,

$$U_i\{\boldsymbol{\alpha}; (D_i, \Omega_i), \boldsymbol{\alpha}_0\} = D_i^\top \Omega_i^{-1} \{\mathbf{y}_i - h(\boldsymbol{\alpha}^\top \mathbf{x}_i)\},$$

where α_0 is the vector of true regression coefficients, $D_i = \partial h(\alpha^\top \mathbf{x}_i) / \partial \alpha^\top$ is evaluated at α_0 , and Ω_i is a square matrix of size n_i with the (k, ℓ) -th non-diagonal element $\Gamma(\mu_{ik}, \mu_{i\ell})$ and the k th diagonal element $\Omega(\mu_{ik})$. The generalized estimating equations for the parameters of the regression coefficients α can then be written as

$$0 = U\{\alpha; (D_i, \Omega_i)_{i=1, \dots, n}, \alpha_0\} = \sum_{i=1}^n U_i\{\alpha; (D_i, \Omega_i), \alpha_0\}. \quad (23)$$

As discussed in Section 9.3.1 of McCullagh and Nelder (1989), under appropriate conditions the limiting distribution of $\sqrt{n}(\hat{\alpha}_0 - \alpha_0)$, where $\hat{\alpha}_0$ is the solution to (23), is asymptotically Normally distributed with mean $\mathbf{0}$ and covariance Σ^{-1} as in (15).

In the semiparametric marginal model, the link and variance-covariance functions are unknown and are estimated nonparametrically in an iterative manner. We define a set \mathcal{A}_n to be a neighborhood in \mathbb{R}^p around the true value of the regression parameter α_0 such that

$$\mathcal{A}_n = \{\alpha : \sqrt{n} \|\alpha - \alpha_0\| = c, \text{ for some } c, 0 < c < \infty\}.$$

The unknown link and variance-covariance functions are replaced with the corresponding nonparametric estimates $\hat{h}(\cdot)$ (6), respectively and $\hat{V}(\cdot)$ (7) and $\hat{\Gamma}(\cdot, \cdot)$ (8), evaluated at $\bar{\alpha}$, $\bar{\alpha} \in \mathcal{A}_n$. Then the “nonparametric” generalized estimating equations are defined by letting

$$\begin{aligned} U_i\{\alpha; (\hat{D}_i, \hat{\Omega}_i), \bar{\alpha}\} &= \hat{D}_i^\top \hat{\Omega}_i^{-1} \{\mathbf{y}_i - \hat{h}(\alpha^\top \mathbf{x}_i)\} \\ &= \sum_{j,k=1}^{m_i} \mathbf{x}_{ij} \hat{h}^{(1)}(\bar{\alpha}^\top \mathbf{x}_i) \omega_{ijk} \{y_{ik} - \hat{h}(\alpha^\top \mathbf{x}_{ik})\}, \end{aligned}$$

where ω_{ijk} is the (j, k) element of $\hat{\Omega}_i^{-1}$, and the nonparametric generalized estimating equations for the regression coefficients are then

$$U\{\alpha; (\hat{D}_i, \hat{\Omega}_i)_{i=1, \dots, n}, \bar{\alpha}\} = \sum_{i=1}^n U_i\{\alpha; (\hat{D}_i, \hat{\Omega}_i), \bar{\alpha}\}. \quad (24)$$

The quasi-information matrix $\mathbf{I}(\alpha_0)$ and its expected value $\mathbf{i}(\alpha_0)$ are obtained by taking the derivative of $U\{\alpha; (D_i, \Omega_i)_{i=1, \dots, n}, \alpha_0\}$ with respect to α such that

$$\mathbf{i}(\alpha_0) = \sum_{i=1}^n D_i^\top \Omega_i^{-1} D_i = \sum_{i=1}^n \sum_{j,k=1}^m \mathbf{x}_{ij} \mathbf{x}_{ij}^\top h^{(1)}(\alpha_0^\top \mathbf{x}_{ij}) h^{(1)}(\alpha_0^\top \mathbf{x}_{ik}) \omega_{ijk},$$

$$\mathbf{I}(\alpha_0) = \mathbf{i}(\alpha_0) + \sum_{i=1}^n \sum_{j,k=1}^m \mathbf{x}_{ij} \mathbf{x}_{ij}^\top \{h^{(2)}(\alpha_0^\top \mathbf{x}_{ij}) \omega_{ijk} + h^{(1)}(\alpha_0^\top \mathbf{x}_{ij}) \omega'_{ijk}\} (y_{ik} - \mu_{ik}),$$

where $h^{(2)}(\cdot)$ is the second derivative of the marginal link function $h(\cdot)$ and ω'_{ijk} is the (j, k) -th element of the first derivative of Ω_i^{-1} .

Similar to the estimating equations defined above, we further define $\mathbf{I}^*(\bar{\alpha})$ by replacing the link and variance-covariance functions with the corresponding nonparametric estimates in $\mathbf{I}(\alpha_0)$. The proof of the theorem hinges on the following results: $|\mathbf{U}\{\alpha; (\hat{D}_i, \hat{\Omega}_i)_{i=1, \dots, n}, \bar{\alpha}\} - \mathbf{U}\{\alpha; (D_i, \Omega_i)_{i=1, \dots, n}, \alpha_0\}| = o_p(\sqrt{n})$, $\sup_{\bar{\alpha} \in \mathcal{A}_n} |\mathbf{I}^*(\bar{\alpha}) - \mathbf{I}(\bar{\alpha})| = o_p(n)$, and $\sup_{\bar{\alpha} \in \mathcal{A}_n} |\mathbf{I}(\bar{\alpha}) - \mathbf{i}(\alpha_0)| = O_p(\sqrt{n})$. The detailed proofs of the above results are similar to those given in Chiou and Müller (1999) and are omitted here. Based on these results and using a Taylor series expansion, it can be shown that $\sqrt{n}(\tilde{\alpha} - \hat{\alpha}_0) = o_p(1)$, where $\hat{\alpha}_0$ is the solution of the estimating equations (23) with correctly specified link and variance-covariance functions, whereas $\tilde{\alpha}$ is the solution to the estimating equations (5) without normalization. Coupling this with the asymptotic normality results of McCullagh and Nelder (1989) above, we find $\sqrt{n}(\tilde{\alpha} - \alpha_0) \xrightarrow{d} N(0, \Sigma^{-1})$. The asymptotic covariance estimate of $f(\hat{\alpha})$ in (15) is obtained accordingly by a multivariate delta method. Further details can be found in the proof of Theorem 3 in Chiou and Müller (2004).

References

- Aitkin, M. (1999) A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 117-128.
- Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9-25.
- Chaganty, N. R. (1997) An alternative approach to the analysis of longitudinal data via generalized estimating equations. *J. Statist. Planng. Infer.*, **63**, 39-54.
- Chiou, J.-M. (2003) Nonparametric covariance estimation in generalized estimating equations for longitudinal data. Manuscript.
- Chiou, J.-M. and Müller, H.-G. (1998) Quasi-likelihood regression with unknown link and variance functions. *J. Am. Statist. Ass.*, **93**, 1376-1387.
- Chiou, J.-M. and Müller, H.-G. (1999) Nonparametric quasi-likelihood. *Ann. Statist.*, **27**, 34-64.
- Chiou, J.-M. and Müller, H.-G. (2004) Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scand. J. Statist.*, **31**, 367-386.
- Climov, D., Delecroix, M. and Simar, L. (2002) Semiparametric estimation in single index Poisson regression: a practical approach. *Statist. Med.*, **29**, 1047-1070.
- Crowder, M. (1995) On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82** 407-410.

- Crowder, M. (2001) On repeated measures analysis with misspecified covariance structure. *J. R. Statist. Soc. B*, **63**, 55-62.
- Diggle, P. J., Heagerty, P. J., Liang, K-Y. and Zeger, S. L. (2002) Analysis of Longitudinal Data. Oxford University Press.
- Fan, J. and Gijbels, I. (1996) Local Polynomial Modeling and Its Applications. London: Chapman and Hall.
- Fitzmaurice, G. M. (1995) A caveat concerning independence estimating equations with multiple multivariate binary data. *Biometrics*, **51**, 309-317.
- Follman, D. and Lambert, D. (1989) Generalizing logistic regression by nonparametric mixing. *J. Am. Statist. Ass.*, **84**, 295-300.
- Gamerman, D. (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statist. Comput.*, **7**, 57-68.
- Golub, G. H., Van Loan, C. F. (1996) Matrix Computations. Johns Hopkins University Press.
- Heagerty, P. J. (1999) Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688-698.
- Heagerty, P. J. and Kurland, B. F. (2001) Misspecified maximum likelihood estimation and generalized linear mixed models. *Biometrika*, **88**, 973-985.
- Heagerty, P. J. and Zeger, S. L. (2000) Marginalized multilevel models and likelihood inference. *Statist. Science*, **15**, 1-26.
- Jiang, J. (1998) Consistent estimators in generalized linear mixed models. *J. Am. Statist. Ass.*, **93**, 720-729.
- Jowaheer, V. and Sutradhar, B. C. (2002) Analysing longitudinal count data with overdispersion. *Biometrika*, **89**, 389-399.
- Koch, G. G., Carr, G. J., Amara, I. A. and Uryniak, T. J. (1990) Categorical data analysis, in *Statistical Methodology in the Pharmaceutical Sciences*, ed. D. A. Berry. New York: Marcel Dekker Inc., 391-475.
- Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models. *J. R. Statist. Soc. B*, **58**, 619-678.
- Lee, Y. and Nelder, J. A. (2001) Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987-1006.
- Leppik, L. E., Dreifuss, F. E., Porter, R., Bowman, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J. Stackman, J. Graves, N., Sutula, T., Welty, T. Vickery, J., Brundage, R., Gates, J., Gummit, R. J. and Gutierrez, A. (1987) A controlled study of progabide in partial seizures: Methodology and results. *Neurology*, **37**, 963-968.
- Li, K.-C. (1991) Sliced inverse regression for dimension reduction. *J. Am. Statist. Ass.*, **86**, 316-327.
- Liang, K-Y and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.
- Magder, L. S. and Zeger, S. L. (1996) A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J. Am. Statist. Ass.*, **91**,

- 1141-1151.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. London: Chapman and Hall.
- McCulloch, C. E. (1997) Maximum likelihood algorithms for generalized linear models. *J. Am. Statist. Ass.*, **92**, 162-170.
- Müller, H.-G. (1987) Weighted local regression and kernel methods for nonparametric curve fitting. *J. Am. Statist. Ass.*, **82**, 231-238.
- Pourahmadi, M. (1999) Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, **86**, 677-690.
- Pourahmadi, M. (2000) Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, **87**, 425-435.
- Ruppert, D. and Wand, M. P. (1994) Multivariate weighted least squares regression. *Ann. Statist.*, **22**, 1346-1370.
- SAS Institute (2000) *SAS/STAT User's Guide*, Version 8. SAS Institute Inc., Cary, N.C.
- Staniswalis, J. G. and Lee, J. J. (1998) Nonparametric regression analysis of longitudinal data. *J. Am. Statist. Ass.*, **93**, 1403-1418.
- Sutradhar, B. C. and Das, K. (1999) On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika*, **86**, 459-465.
- Thall, P. F. and Vail, S. C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657-671.
- Wolfinger, R., and O'Connell, M. (1993) Generalized linear mixed models: A pseudo-likelihood approach. *J. Statist. Comput. Simul.*, **48**, 233-243.
- Zeger, S. L., Liang, K. Y. and Albert, P. S. (1988) Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.
- Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects: A Gibbs sampling approach. *J. Am. Statist. Ass.*, **82**, 231-238.

Table 1. Estimated regression parameters (with standard errors in parentheses) for the epilepsy data, comparing Models EEE_1 (the proposed EEE approach), $GLMM_1$ (GLMM with a random intercept), GEE_1 (GEE with unstructured working correlation matrix), $T\&V$ (best model of Thall and Vail, 1990). Estimates that are significant at 5% level are marked with ‘*’

Variable	EEE_1	$GLMM_1$	GEE_1	$T\&V$
<i>Intercept</i>		-1.297 (.160)*	-2.417 (.861)*	-1.350 (.904)
<i>Base</i>	.801 (.042)*	.885 (.129)*	.923 (.091)*	.877 (.103)*
<i>Age</i>	-.484 (.031)*	.471 (.341)	.817 (.247)*	.531 (.266)*
<i>Trt</i>	-.339 (.125)*	-.697 (.417)	-.559 (.390)	-.958 (.390)*
<i>Visit₄</i>	-.084 (.069)	-.148 (.059)*	-.147 (.111)	-.159 (.072)*
<i>TrtBase</i>	.044 (.061)	.193 (.217)	-.144 (.173)	.352 (.196)

Table 2. Estimated regression parameters (with standard errors in parentheses) for the epilepsy data, comparing Models EEE_2 (the proposed EEE approach), $GLMM_2$ (GLMM with a random intercept), GEE_2 (GEE with unstructured working correlation matrix)

Variable	EEE_2	$GLMM_2$	GEE_2
<i>Intercept</i>		.032 (.269)	.083 (.197)
<i>Base</i>	.576 (.053)*	.983 (.101)*	.963 (.083)*
<i>Age₁</i>	-.057 (.053)	.058 (.237)	.118 (.145)
<i>Age₂</i>	-.170 (.068)*	.195 (.286)	.381 (.147)*
<i>Trt</i>	-.598 (.030)*	-.577 (.260)*	-.452 (.167)*
<i>Visit₄</i>	-.186 (.076)*	-.150 (.059)*	-.147 (.114)
<i>TrtAge₁</i>	.395 (.058)*	.519 (.329)	.460 (.272)
<i>TrtAge₂</i>	.296 (.094)*	-.051 (.398)	-.179 (.314)

Table 3. Estimated regression parameters (with standard errors in parentheses) for the epilepsy data, using the two-phase linear link function and comparing Models EEE_3 (the proposed EEE approach) and GEE_3 (GEE with unstructured working correlation matrix)

Variable	EEE_3	GEE_3
<i>Intercept</i>	-.303 (.036)*	-.283 (.056)*
<i>Base</i>	.608 (.018)*	.578 (.037)*
<i>Age₁</i>	.158 (.036)*	.180 (.052)*
<i>Age₂</i>	.076 (.044)	.175 (.064)*
<i>Trt</i>	-.344 (.039)*	-.322 (.053)*
<i>Visit₄</i>	-.062 (.026)*	-.064 (.024)*
<i>TrtAge₁</i>	.186 (.048)*	.187 (.080)*
<i>TrtAge₂</i>	.059 (.055)	-.022 (.085)

Table 4. Predicted seizure counts for Models EEE_2 , EEE_3 , GEE_3 at the fourth visit and baseline seizure count 22 (median), for placebo and treatment, differentiated by age group

Age	EEE_2			EEE_3			GEE_3		
	Treat	Placebo	Diff	Treat	Placebo	Diff	Treat	Placebo	Diff
[0, 25)	2.412	7.232	-4.820	3.095	4.816	-1.720	3.043	4.656	-1.613
[25, 35)	4.717	6.617	-1.900	4.816	6.924	-2.108	4.877	6.760	-1.883
[35, ∞)	3.137	5.467	-2.330	3.770	5.522	-1.753	3.807	6.671	-2.864

Table 5. Simulation results for the estimated regression coefficients with the random intercept following a centered $Gamma(10, 1)$ distribution

Parameter	Method	Bias	Stdev	MSE	Rej. rate (5% level)	95% c.i. half-length
α_1	EEE	0.0285	0.0879	0.0085	100%	0.2708
	GLMM	-0.1092	0.3905	0.1644	43%	0.7851
α_2	EEE	0.1281	0.1751	0.0471	62%	0.4229
	GLMM	0.0208	0.3403	0.1162	35%	0.7411
α_3	EEE	0.1045	0.2057	0.0533	5%	0.5627
	GLMM	0.0244	0.3582	0.1289	1%	0.8056

Table 6. Simulation results for the estimated regression coefficients with the random intercept following Gaussian mixture $Normal(-1, 10)$ and $Normal(1, 10)$ with equal probabilities

Parameter	Method	Bias	SE	MSE	Rej. rate (5% level)	95% c.i. half-length
α_1	EEE	0.0196	0.1080	0.0120	99%	0.2637
	GLMM	0.0655	0.4666	0.2203	49%	0.8985
α_2	EEE	0.1133	0.1510	0.0354	64%	0.4094
	GLMM	0.0199	0.4117	0.1699	25%	0.7957
α_3	EEE	0.0910	0.2231	0.0580	9%	0.5493
	GLMM	0.0354	0.4655	0.2179	5%	0.8757

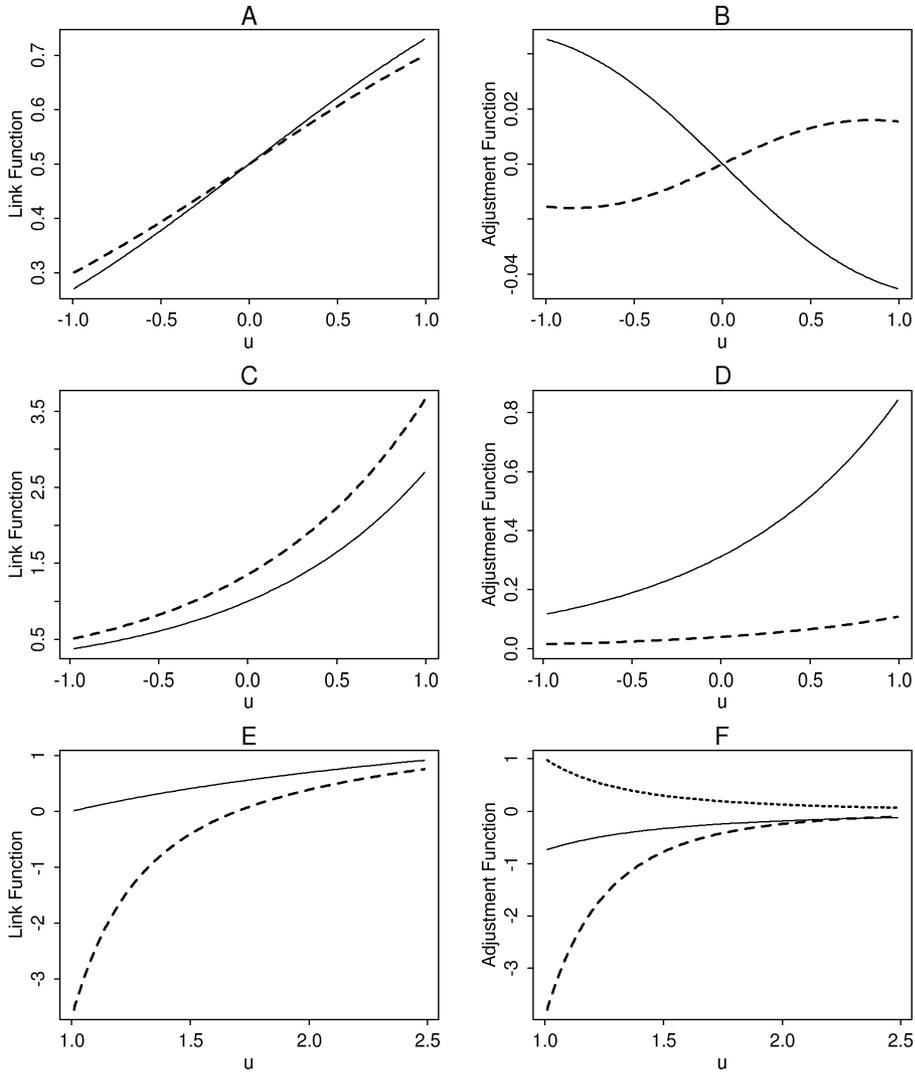


Fig. 1. Comparisons of conditional and marginal link functions for the Logit-Normal (A and B), the Poisson-Mixture (C and D) and the Normal-Gamma (E and F) models. Left-hand panels display conditional link functions g (dashed) and marginal link functions h (solid), and right-hand panels display adjustment functions h_2 (solid), h_3 (dotted where applicable) and h_4 (dashed).

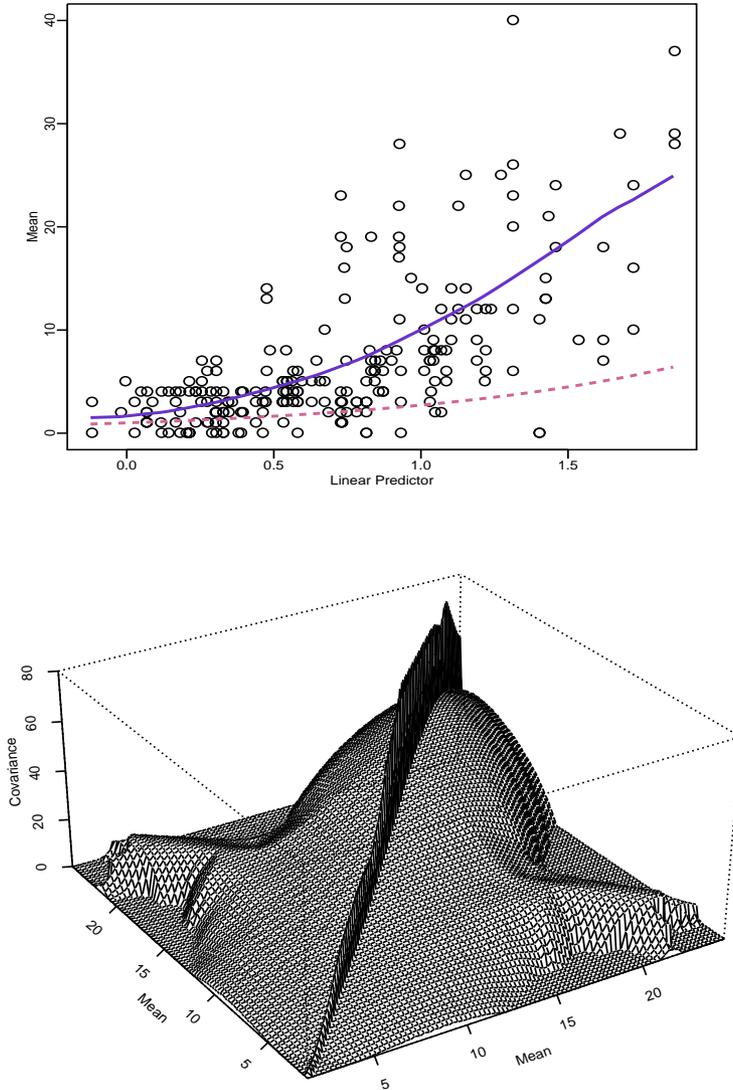


Fig. 2. Estimated link (upper panel) and variance-covariance (lower panel) functions of the proposed EEE approach for epileptic seizure count data. The dashed curve on the upper panel is an exponential function of the estimated linear predictor.

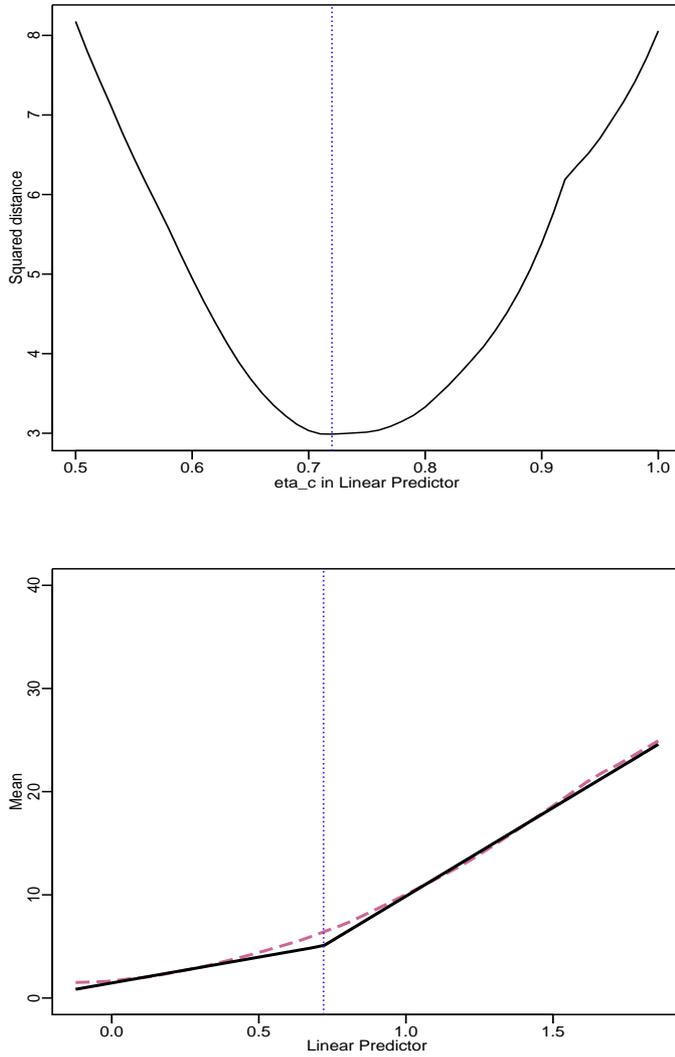


Fig. 3. Upper panel: Squared distance as a function of change point in linear predictor based on the two-phase regression function. Lower panel: Estimated two-phase regression function with the optimal change point at 0.72 (solid), superimposed on the nonparametric link (dashed).