First-order marginalised transition random effects models with probit link function

Özgür Asar\textsuperscript{a,*}, Ozlem Ilk\textsuperscript{b}

\textsuperscript{a} CHICAS, Medical School, Lancaster University, UK
\textsuperscript{b} Department of Statistics, Middle East Technical University, Turkey

July 10, 2015

Abstract

Marginalised models, also known as marginally specified models, have recently become a popular tool for analysis of discrete longitudinal data. Despite being a novel statistical methodology, these models introduce complex constraint equations and model fitting algorithms. On the other hand, there is a lack of publicly available software to fit these models. In this paper, we propose a three-level marginalised model for analysis of multivariate longitudinal binary outcome. The implicit function theorem is introduced to approximately solve the marginal constraint equations explicitly. Probit link enables direct solutions to the convolution equations. Parameters are estimated by maximum likelihood via a Fisher-Scoring algorithm. A simulation study is conducted to examine the finite-sample properties of the estimator. We illustrate the model with an application to the data set from the Iowa Youth and Families Project. The R package \texttt{pnmtrem} is prepared to fit the model.

Keywords: correlated data, implicit differentiation, link functions, maximum likelihood estimation, subject-specific inference, statistical software.

1 Introduction

Longitudinal data comprise repeated measurements on the same subjects across time. Whilst data from the same subjects are typically dependent on each other, data from different subjects are typically independent. Often, multiple responses, e.g. multiple health outcomes or distress variables, from each subject are collected. These responses introduce two types of dependencies: 1) within-response (serial) dependence, and 2) multivariate response dependence at a given time point. To draw valid statistical inferences, both of these dependencies should be taken into account.

Conventional models for analysis of longitudinal data are marginal, transition and random effects models (Diggle et al., 2002). A recently popular method for discrete longitudinal data analysis is the framework of marginalised models, also known as marginally specified models. The framework typically combines the underlying features of the conventional models, and enables likelihood-based

\textsuperscript{*}Corresponding author: Özgür Asar, E-mail address: o.asar@lancaster.ac.uk, Tel.: +44 (0) 1524 593519.
inference for marginal mean parameters. Heagerty and Zeger (2000) define marginally specified models as a re-parameterised version of transition and/or random effects models in terms of the marginal mean and additional dependence parameters. Heagerty (1999, 2002), in his seminal papers, develops marginalised random effects and marginalised transition models, respectively. Both of these models are two-level logistic regression models. Whilst covariate effects are captured in the first levels, serial dependence is captured in the second levels via random effects and response history, respectively. Heagerty and Kurland (2001) show that marginal regression parameter estimates based on marginalised random effects models are less sensitive to dependence structure misspecification compared to those based on conventional random effects models. Heagerty (2002) and Lee and Mercante (2010) prove that parameters of the first and second levels of marginalised transition models are orthogonal. The marginalised modeling paradigm was primarily developed for binary data (Schildcrout and Heagerty, 2007; Ilk and Daniels, 2007; Lee et al., 2009; along with the aforementioned works of Heagerty). Later, it has been extended to ordinal (Caffo and Griswold, 2006; Lee and Daniels, 2007; Lee et al., 2013), count (Lee et al., 2011; Iddi and Molenberghs, 2012) and nominal data (Lee and Mercante, 2010). Amongst these works, Ilk and Daniels (2007) propose a three-level marginalised model for multivariate longitudinal binary data, called marginalised transition random effects model. With this model, whilst covariate effects are captured in the first level, serial and multivariate response dependencies are captured in the second and third levels via response history and random effects, respectively. In this paper, we extend marginalised transition random effects model in terms of link function, from logit to probit, and the parameter estimation methodology, from Bayesian methods (BM) to maximum likelihood (ML) estimation.

Probit and logit are popular link functions for modelling categorical data. These link functions are defined as the inverses of the distribution functions of the standard normal and the standard logistic distribution, respectively. They have similar behaviours in terms of placing probabilities. The only difference is at the extreme tails; logit places higher probabilities at the tails (Hedeker and Gibbons, 2006). Nonetheless, substantial and high quality data are needed to detect the difference (Doksum and Gakso, 1990, cited in Hedeker and Gibbons, 2006, pp. 153). Logit allows direct interpretation of the parameter estimates, as changes in (log) odds ratios. The interpretation is more challenging with probit. Nonetheless, (approximate) transitions between the parameter estimates based on these link functions is possible (Agresti, 2002; Griswold et al., 2013). For example, the JKB constant (Johnson et al., 1995, pp. 113-163, cited in Griswold et al., 2013) postulates the following: \( \beta_{\text{logit}} \approx \beta_{\text{probit}} \) where \( c = (15/16)(\pi/ \sqrt{3}) \approx 1.700437 \). One advantage of probit over logit is that it allows explicit form of the linkage between the levels of marginalised random effects models (Heagery and Zeger, 2000; Griswold et al., 2013; Caffo and Griswold, 2006). The use of probit link in multivariate modelling dates back to Ashford and Sowden (1970). Some recent examples on longitudinal mixed modelling are Hedeker and Gibbons (2006), Liu and Hedeker (2006), Varin and Czado (2010), amongst others.

Generalized estimating equations (GEE; Liang and Zeger, 1986) have been widely used to estimate the parameters of marginal models, especially for discrete outcome. Nonetheless, they might be inefficient because of being a semi-parametric method, compared to the full likelihood-based methods, e.g. ML and BM. BM are widely used in longitudinal data literature and have their own properties. Some distinguishing features of ML over BM are that parameter estimation requires less computational times, and related procedures are more automatized (Efron, 1986). In this paper, we consider ML for parameter estimation to avoid the computational burden.

Marginalised models with transition structures require solving marginal constraint equations (Heagerty, 2002; Schildcrout and Heagery, 2007; Ilk and Daniels, 2007; Lee and Mercante, 2010). Common literature for solving these equations has been built on optimisation methods, e.g. Newton-Raphson (N-R) algorithm. This might be computationally cumbersome and might yield convergence
problems. In this paper, we consider approximately explicit solutions of marginal constraint equations, and propose the use of the implicit function theorem for the first time in the scope of marginally specified models.

Publicly available software for analysis of multivariate longitudinal binary data is still rare. Available options include the SAS macro of Shelton et al. (2004), and the R (R Core Development Team, 2015) packages mmm (Asar and Ilk, 2013) and mmm2 (Asar and Ilk, 2014). In this study, we propose the R package pnmtrem for first-order marginalised transition random effects models with probit link. The package is available from the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/package= pnmtrem.

The paper is organized as follows. Whilst the general modelling framework is introduced in Section 2, first-order version is discussed in detail in Section 3. In Section 4, we discuss inference for the first-order model. Finite-sample behaviours of the estimator are investigated by a simulation study in Section 5. The first-order model is applied to a real data set in Section 6. Section 7 is a concluding discussion.

2 General framework

Let \( Y_{ij} \) denote the \( j \)th \((j = 1, \ldots, k)\) response of the \( i \)th \((i = 1, \ldots, n)\) subject at time \( t \) \((t = 1, \ldots, T)\). Also let \( X_{ij} \) denote the associated set of covariates, which might include time-varying and/or time-invariant covariates. The framework of the general model with inverse probit link is as follows:

\[
P_{itj}^{m} = P(Y_{itj} = 1 | X_{itj}) = \Phi(X_{itj}^T \beta),
\]

\[
P_{itj}^{d} = P(Y_{itj} = 1 | y_{i,t-1-j}, \ldots, y_{i,t-p-j}, X_{itj}) = \Phi(\lambda_{itj} + \sum_{m=1}^{p} \gamma_{itjm}y_{i,t-m,j}),
\]

\[
P_{itj}^{p} = P(Y_{itj} = 1 | y_{i,t-1-j}, \ldots, y_{i,t-p-j}, X_{itj}, b_{it}) = \Phi(\lambda_{itj}^{p} + \lambda_{itj}^{p}b_{it}),
\]

where \( \Phi(\cdot) \) is the distribution function of the standard normal.

In (1), the first level of the framework, \( \beta \) are marginal regression parameters. These parameters measure the relationship between covariates and responses, and allow comparing covariate subgroups, e.g. males vs. females, without conditioning on response history and/or random effects. The default setting assumes that intercepts and slopes are shared by different responses, i.e. we postulate \( \beta \) instead of \( \beta_{j} \). Nonetheless, one is able to specify different intercepts and slopes for multiple responses by including in \( X_{it} \) indicator variables for responses and interactions of these indicator variables with covariates, respectively. This specification provides model flexibility. We might gain in efficiencies considerably, e.g. when the relationships between covariates and multiple responses are not significantly different (Asar and Ilk, 2014). Another default setting is the assumption of accommodating only the relationship of responses with current covariates, i.e. \( P(Y_{itj} = 1|X_{itj}, \ldots, X_{it}) = P(Y_{itj} = 1|X_{itj}) \). Nonetheless, relationships with lagged covariates might be captured by including covariate history in \( X_{itj} \).

In (2), the second level of the framework, Markov model of order \( p \) is used to capture the serial dependence. Here, the \( m \)th transition parameters, \( \gamma_{itjm} \), can be written in terms of covariates, i.e. \( \gamma_{itjm} = \alpha_{itm}Z_{itjm} = \alpha_{itm}Z_{itj1,m} + \ldots + \alpha_{itm}Z_{itjl,m} \) for \( m = 1, \ldots, p \). \( \alpha_{itm} \) \((f = 1, \ldots, l)\) are time, covariate and order specific transition parameters. They capture the relationships between past and current responses. \( Z_{itj} \) have a form of design matrix with 1’s on the first column, and are typically a subset of \( X_{itj} \) with \( l \) covariates. The form of \( Z_{itj} \) permits flexibly specifying the association structures between past and current responses. For example, if one suspects that the lag-1 associations are different for males and females, then gender can be included in \( Z_{itj1} \). Similar to
the first level, \( \alpha_{e,m} \) are assumed to be shared across multiple responses. Response-specific transition parameters can be specified by including in \( Z_{tj} \), indicator variables for responses and interactions of these indicator variables with response history and other covariates.

In (3), the third level of the model, multivariate response dependence and individual variations are captured. \( b_{itj} \)’s are subject and time specific random effects coefficients that measure unobserved heterogeneity between subjects at time \( t \). We assume that \( b_{itj} \sim N(0, \sigma_{itj}^2) \). \( b_{itj} \) can be rewritten as \( b_{itj} = \sigma_j z_t \), where \( z_t \) is a standard normal random variable, which is useful in numerical integration. \( \lambda_j \)’s are response-specific parameters that scale \( b_{itj} \) with respect to the \( j \)th response, and accounts for multivariate response dependence. We set \( \lambda_1 \) to 1 for identifiability, and estimate \( \lambda_j \) for \( j = 2, \ldots \). Note that by specifying \( b_{itj} \)’s are time-varying, the model assumes that multivariate response dependencies might change across time.

\( \Delta_{itj} \)’s in (2) are subject, time and response specific intercepts. They take into account the (non-linear) relationship between marginal (\( P_{itj}^m \)) and transition probabilities (\( P_{itj}^r \)). Similarly, \( \Delta_{itj}^* \)’s in (3) are subject, time and response specific intercepts that account for the (non-linear) relationship between transition and random effects probabilities (\( P_{itj}^r \)).

We assume that conditional mean of responses given all covariates is equal to conditional mean for multivariate response dependence and individual variations. We set \( \lambda_1 = 1 \) to identify the parameters, and \( \lambda_j \) for \( j = 2, \ldots \) for identifiability. 

\begin{align*}
\Delta_{itj} & = \Phi(\gamma_{itj} + \gamma_{itj,1} y_{ijt-1,1}), \quad (5) \\
\Delta_{itj}^* & = \Phi(\lambda_j + \lambda_j b_{itj}), \quad (6)
\end{align*}

As before, \( b_{itj} \sim N(0, \sigma_{itj}^2) \) and \( b_{itj} = z_t \sigma_j, z_t \sim N(0,1) \); \( \lambda_1 = 1; \quad \gamma_{itj,1} = \alpha_{it1} Z_{it1,1} + \alpha_{it1,2} Z_{it1,2} + \ldots + \alpha_{itd} Z_{itd,1} \). Throughout, we call this framework as the \( \lambda \)-model.

\( \Delta_{itj} \)’s in (2) are subject, time and response specific intercepts. They take into account the (non-linear) relationship between marginal (\( P_{itj}^m \)) and transition probabilities (\( P_{itj}^r \)). Similarly, \( \Delta_{itj}^* \)’s in (3) are subject, time and response specific intercepts that account for the (non-linear) relationship between transition and random effects probabilities (\( P_{itj}^r \)).

We assume that conditional mean of responses given all covariates is equal to conditional mean for multivariate response dependence and individual variations. We set \( \lambda_1 = 1 \) to identify the parameters, and \( \lambda_j \) for \( j = 2, \ldots \) for identifiability. 

\begin{align*}
\Delta_{itj} & = \Phi(\gamma_{itj} + \gamma_{itj,1} y_{ijt-1,1}), \quad (5) \\
\Delta_{itj}^* & = \Phi(\lambda_j + \lambda_j b_{itj}), \quad (6)
\end{align*}

As before, \( b_{itj} \sim N(0, \sigma_{itj}^2) \) and \( b_{itj} = z_t \sigma_j, z_t \sim N(0,1) \); \( \lambda_1 = 1; \quad \gamma_{itj,1} = \alpha_{it1} Z_{it1,1} + \alpha_{it1,2} Z_{it1,2} + \ldots + \alpha_{itd} Z_{itd,1} \). Throughout, we call this framework as the \( \lambda \)-model.

\section{3 First-order model}

In this study, we focus on lag-1 dependence in (2). The framework for the first-order model becomes

\begin{align*}
P_{itj}^m & = P(Y_{itj} = 1|X_{itj}) = \Phi(X_{itj}\beta), \quad (4) \\
P_{itj}^r & = P(Y_{itj} = 1|y_{ijt-1,j}, X_{itj}) = \Phi(\Delta_{itj} + \gamma_{itj,1} y_{ijt-1,1}), \quad (5) \\
P_{itj}^r & = P(Y_{itj} = 1|y_{ijt-1,j}, X_{itj}, b_{itj}) = \Phi(\Delta_{itj}^* + \lambda_j b_{itj}). \quad (6)
\end{align*}

As before, \( b_{itj} \sim N(0, \sigma_{itj}^2) \) and \( b_{itj} = z_t \sigma_j, z_t \sim N(0,1) \); \( \lambda_1 = 1; \quad \gamma_{itj,1} = \alpha_{it1} Z_{it1,1} + \alpha_{it1,2} Z_{it1,2} + \ldots + \alpha_{itd} Z_{itd,1} \). Throughout, we call this framework as the \( \lambda \)-model.

\section{3.1 Linking levels of the \( t \geq 2 \) model}

Levels of the first-order model (4 - 6) are connected to each other for the model being a valid probabilistic model.
3.1.1 Linking first and second levels

Levels 1 (4) and 2 (5) are linked via the following marginal constraint equation,

\[ P(Y_{i,j} = 1|X_{i,j}) = \sum_{y_{i,j-1},y_{i,j-2}} P(Y_{i,j} = 1|y_{i,j-1},y_{i,j-2},X_{i,j})P(y_{i,j-1},y_{i,j-2}|X_{i,j-1},X_{i,j}), \]  

(9)

which is equivalent to

\[ \Phi(X_{i,j}|\beta) = \Phi(\Delta_{i,j})(1 - \Phi(X_{i,j-1}|\beta^*)) + \Phi(\Delta_{i,j} + \gamma_{i,j,1})\Phi(X_{i,j-1}|\beta^*), \]  

(10)

and

\[ \Phi(X_{i,j}|\beta) = \Phi(\Delta_{i,j})(1 - \Phi(X_{i,j-1}|\beta^*)) + \Phi(\Delta_{i,j} + \gamma_{i,j,1})\Phi(X_{i,j-1}|\beta^*). \]  

(11)

for \( t > 2 \) and \( t = 2 \), respectively. Hereafter, the discussion will be based on (11). We take the difference between (11) and (10) when necessary.

Since (11) does not permit explicitly writing \( \Delta_{i,j} \) in terms of \( \beta \) and \( \gamma_{i,j,1} \) (or \( \alpha_{t,1} \)), we use the implicit function theorem (IFT; Krantz and Parks, 2003) for an approximately explicit solution. Application of IFT is as follows.

Let \( F \) be a function of \( X_{i,j}, X_{i,j-1}, \beta, \Delta_{i,j}, \alpha_{t,1} \) and \( Z_{i,j,1} \) such that (by rewriting (11))

\[ F(X_{i,j}, X_{i,j-1}, \beta, \Delta_{i,j}, \alpha_{t,1}, Z_{i,j,1}) = \Phi(X_{i,j}|\beta) - \Phi(\Delta_{i,j})(1 - \Phi(X_{i,j-1}|\beta)) - \Phi(\Delta_{i,j} + \alpha_{t,1}Z_{i,j,1})\Phi(X_{i,j-1}|\beta) = 0. \]  

(12)

By IFT with first order implicit differentiation, i.e. first order approximation, \( \Delta_{i,j} \) can be obtained as

\[ \Delta_{i,j} = -\frac{\partial F}{\partial \beta}(\beta - x_0) - \frac{\partial F}{\partial \alpha_{t,1}}(\alpha_{t,1} - x_{t,10}), \]  

(13)

where

\[ \frac{\partial F}{\partial \beta} = X_{i,j}\Phi(X_{i,j}|\beta) + \Phi(\Delta_{i,j})(\Phi(X_{i,j-1}|\beta^*))X_{i,j-1,1} - \Phi(\Delta_{i,j} + \alpha_{t,1}Z_{i,j,1})\Phi(X_{i,j-1}|\beta)X_{i,j-1,1}, \]

\[ \frac{\partial F}{\partial \Delta_{i,j}} = -\phi(\Delta_{i,j})(1 - \Phi(X_{i,j-1}|\beta)) - \phi(\Delta_{i,j} + \alpha_{t,1}Z_{i,j,1})(\Phi(X_{i,j-1}|\beta)), \]

\[ \frac{\partial F}{\partial \alpha_{t,1}} = -\phi(\Delta_{i,j} + \alpha_{t,1}Z_{i,j,1})\Phi(X_{i,j-1}|\beta)Z_{i,j,1}. \]  

(14)

Here, \( \phi() \) is the density function of the standard normal, and \( \beta_0, \alpha_{t,10} \) and \( \Delta_{i,j,0} \) are fixed values around which IFT searches for solution. We set \( \beta_0 \) and \( \alpha_{t,10} \) to 0, since null hypotheses for \( \beta \) and \( \alpha_{t,1} \) are on equality of these parameters to 0. \( \Delta_{i,j,0} \) is obtained by solving (12) under \( \beta_0 \) and \( \alpha_{t,10} \) being 0. This yields \( \Delta_{i,j,0} = 0 \) for \( t > 2 \). N-R is used to obtain \( \Delta_{i,j,0} \). Based on our experience, this has very fast convergence due to the simple form of (12).

3.1.2 Linking second and third levels

Level 2 (5) and level 3 (6) are linked via the following convolution equation:

\[ P(Y_{i,j} = 1|y_{i,j-1,1},X_{i,j}) = \int P(Y_{i,j} = 1|y_{i,j-1,1},X_{i,j},b_{i,j})dF(b_{i,j}), \]  

(15)
which is equivalent to
\[
\Phi(\Delta_{ij} + \alpha_{\lambda_{i1}}Z_{it_{j1}Y_{t_{j}-1,j}}) = \int \Phi(\Delta_{ij}^* + \lambda_{j}b_{it}) f(b_{it}) db_{it}.
\] (16)

Following Griswold (2005), we can explicitly obtain \(\Delta_{ij}^*\) as
\[
\Delta_{ij}^* = \sqrt{1 + \lambda_{j}^2 \sigma_{i1}^2} (\Delta_{ij} + \alpha_{\lambda_{i1}}Z_{it_{j1}Y_{t_{j}-1,j}}).
\] (17)

Proof of (17) is given in Appendix A. \(\Delta_{ij}^*\) is now an explicit and deterministic function of \(\Delta_{ij}\) (hence, \(\Delta_{ij}^*\) is a function of \(X_{it_{j}}, X_{t_{j}-1,j}\) and \(\beta_{1}, \alpha_{\lambda_{i1}} Z_{it_{j1}}, Y_{t_{j}-1,j}, \lambda_{j}\) and \(\sigma_{i1}\).

3.2 Linking levels of the baseline model

First (7) and second (8) levels of the baseline model are linked via the following convolution equation:
\[
P(Y_{i_{1}j} = 1|X_{i_{1}j}) = \int P(Y_{i_{1}j} = 1|X_{i_{1j}}, b_{i}) dF(b_{i}).
\] (18)

\(\Delta_{ij}^*\) can be written as an explicit function of \(X_{i_{1j}}, \beta^*, \lambda^*_j\) and \(\sigma_{i1}\) such that
\[
\Delta_{ij}^* = \sqrt{1 + \lambda_{j}^2 \sigma_{i1}^2} X_{i_{1j}}\beta^*.
\] (19)

Proof of (19) can be easily adapted from the proof of (17).

4 Inference

4.1 Estimation

The likelihood of the first-order model is the product of the likelihood functions of the baseline and \(t \geq 2\) models. By re-writing the random effects coefficients as \(b_{i1} = \sigma_{i1}Z_{i}\) and \(b_{it} = \sigma_{it}Z_{i}\), it can be expressed as
\[
L(\theta|y) = L_1(\theta_1|y_1)L_2(\theta_2|y_2),
\] (20)

where
\[
L_1(\theta_1|y_1) = \prod_{i=1}^{N} \int \prod_{1}^{k} \left(P_{i_{1j}}^\prime \right)^{y_{i_{1j}}} \left(1 - P_{i_{1j}}^\prime \right)^{1-y_{i_{1j}}} \phi(z_{i}) dz_{i},
\] (21)
\[
L_2(\theta_2|y_2) = \prod_{i=1}^{N} \frac{1}{\prod_{t=2}^{T} \left(P_{i_{tj}}^\prime \right)^{y_{i_{tj}}} \left(1 - P_{i_{tj}}^\prime \right)^{1-y_{i_{tj}}} \phi(z_{i}) dz_{i}}.
\] (22)

Here, \(\theta = (\theta_1, \theta_2)\), where \(\theta_1 = (\beta^*, \lambda^*, \sigma_{i1}^2)\) with \(\lambda^* = (\lambda_{1}^*, \ldots, \lambda_{k}^*)\) and \(\theta_2 = (\beta, \alpha_{\lambda_{i1}}, \lambda, \sigma^2)\) with \(\lambda = (\lambda_{1}, \ldots, \lambda_{k})\) and \(\sigma^2 = (\sigma_{i1}^2, \ldots, \sigma_{it}^2)\), are parameters of the baseline and \(t \geq 2\) models, respectively; \(y_1\) and \(y_2\) are observed responses at baseline and \(t \geq 2\) time points, respectively. \(L_1(\theta_1|y_1)\) and \(L_2(\theta_2|y_2)\) are connected to each other via \(\beta^*\) at \(t = 2\) (see (10)). We model \(\log(\sigma_{i1})\), instead of \(\sigma_{i1}\) or \(\sigma_{i1}^2\), due to computational aspects. This transformation helps extending the parameter spaces from \([0, +\infty)\) to \((-\infty, +\infty)\). Estimates and standard errors regarding \(\sigma_{i1}\) or \(\sigma_{i1}^2\) can be easily obtained using the invariance property of ML estimates and delta method, respectively.
We need to use numerical methods to solve the integrals in (21) and (22), since there is no closed-form solutions. Since these integrals are one-dimensional, we use Gauss-Hermite quadrature with 20-points (Lesaffre and Spiessens, 2001; Agresti, 2002; McCulloch et al., 2008). Similarly, the closed-form solutions based on the first partial derivatives of the log-likelihood are not available. We use Fisher-Scoring (F-S) algorithm to obtain the parameter estimates iteratively. An advantage of F-S algorithm is that it only works with the first partial derivatives and does not require the second partial derivatives (Hedeker and Gibbons, 2006, pp. 162-165). Another advantage of the algorithm is that at convergence, inverse of the expected information matrix is a consistent estimator of the large sample variance-covariance matrix of the model parameters. With F-S algorithm, ML estimates are obtained iteratively as

\[
\hat{\theta}_s^{m+1} = \hat{\theta}_s^m + I(\hat{\theta}_s^m)^{-1} \frac{\partial \log(L_s(\hat{\theta}_s^m|y_s))}{\partial \hat{\theta}_s^m},
\]

where \( s = (1, 2); \ s = 1 \) corresponds to the baseline model and \( s = 2 \) corresponds to the \( t \geq 2 \) model; \( m \) represents the F-S step and \( I(\theta_s) \) is an empirical information matrix. \( \frac{\partial \log(L_s(\theta_s|y_s))}{\partial \theta_s} \) is the first partial derivative of the log-likelihood, calculations of which can be found in Appendix B. \( I(\theta_s) \) is calculated as

\[
I(\theta_1) = \sum_{j=1}^{N} \frac{1}{h(Y_{ij}|\theta_1)} \left( \frac{\partial h(Y_{ij}|\theta_1)}{\partial \theta_1} \right) T
\]

and

\[
I(\theta_2) = \sum_{j=1}^{N} \left( \sum_{t=2}^{T} \frac{1}{h(Y_{ij}|\theta_2)} \left( \frac{\partial h(Y_{ij}|\theta_2)}{\partial \theta_2} \right) \right) \left( \sum_{t=2}^{T} \frac{1}{h(Y_{ij}|\theta_2)} \left( \frac{\partial h(Y_{ij}|\theta_2)}{\partial \theta_2} \right) \right). \tag{25}
\]

Details of \( h(Y_{ij}|\theta_s) \) and \( \frac{\partial h(Y_{ij}|\theta_s)}{\partial \theta_s} \) for \( t = 1, \ldots, T \) can be found in Appendix B. Since \( \sigma_1 \) is time-specific and \( \lambda_t^s \) is response-specific for baseline, and \( \sigma_t \) and \( \alpha_{tj} \) are time-specific and \( \lambda_t \) is response-specific for \( t \geq 2 \), the calculations of \( I(\theta_1) \) and \( I(\theta_2) \) for these parameters are different compared to the calculations for \( \beta^* \) and \( \beta \). Details can be found in the online supplementary material.

### 4.2 Prediction

Predicting \( b_t = \sigma_t z_t \ (t = 1, \ldots, T) \) is equivalent to predicting \( z_t \). We obtain the predictions of \( z_t \)’s as the modes of log-conditional distributions of \( z_t \)’s given the data (Heagerty, 1999). This requires solving

\[
\left\{ \sum_{t=1}^{T} \sum_{j=1}^{k} \hat{\lambda}_j \hat{\sigma}_j \phi(\hat{d}_{ij})(Y_{ij} - \Phi(\hat{d}_{ij})) \Phi(\hat{d}_{ij}) (1 - \Phi(\hat{d}_{ij})) \right\} - z_t = 0, \tag{26}
\]

with respect to \( z_t \) usin N-R algorithm. Here, \( \hat{d}_{ij} = \hat{\Lambda}_{ij}^* + \lambda_t \hat{\sigma}_t z_i \). \( \hat{\Lambda}_{ij}^* \)’s are obtained as in (17) and (19) plugging-in the ML estimates of \( \theta_1 \) and \( \theta_2 \).

### 5 Simulation study

We conduct a Monte Carlo simulation study to examine the finite-sample behaviours of the marginal mean parameters. In each replications, we simulate data using the first-order model. The data sets
include bivariate binary responses, $Y_1$ and $Y_2$, and two associated covariates, $X_1$ and $X_2$, for 250 subjects with 4 follow-ups. We generate $X_1$ from Uniform(0, 1) as a time-independent variable. $X_2$ is taken as the response indicator variable. It takes 1 for $Y_1$, 0 for $Y_2$. We consider different sets of covariates for baseline and $t \geq 2$. Moreover, we consider varying relationships between the covariates and the responses, i.e. $\beta^t \neq \beta$. We specifically consider $\beta^t = (\beta_{0t}, \beta_{1t}) = (-0.5, 0.5)$ for $t = 1$, and $\beta = (\beta_0, \beta_1) = (-0.7, 0.7, 0.2)$ for $t = 2, 3, 4$. By the inclusion of response indicator as a covariate, we allow the responses to have different intercepts. Whilst the intercept is $\beta_0 + \beta_2 = -1 + 0.2 = -0.8$ for $Y_1$, it is $\beta_0 = -1$ for $Y_2$. The relationships between $X_1$ and $Y_1$, and $X_1$ and $Y_2$ are assumed to be the same, i.e. interaction between $X_1$ and $X_2$ is not included. In terms association structures, by keeping the marginal mean parameter setting same, we consider the following four cases:

**Case 1**

$$(\lambda_1^t, \lambda_2) = (0.9, 0.95), (\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (0.2, 0.25, 0.3, 0.35), (\alpha_{21,1}, \alpha_{31,1}, \alpha_{41,1}) = (0.3, 0.4, 0.5)$$

**Case 2**

$$(\lambda_1^t, \lambda_2) = (1.1, 1.15), (\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (0.2, 0.25, 0.3, 0.35), (\alpha_{21,1}, \alpha_{31,1}, \alpha_{41,1}) = (0.3, 0.4, 0.5)$$

**Case 3**

$$(\lambda_1^t, \lambda_2) = (0.9, 0.95), (\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (0.5, 0.55, 0.6, 0.65), (\alpha_{21,1}, \alpha_{31,1}, \alpha_{41,1}) = (0.3, 0.4, 0.5)$$

**Case 4**

$$(\lambda_1^t, \lambda_2) = (0.9, 0.95), (\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (0.2, 0.25, 0.3, 0.35), (\alpha_{21,1}, \alpha_{31,1}, \alpha_{41,1}) = (0.6, 0.7, 0.8)$$

with $\lambda_i^t$ and $\lambda_i$ being 1. The relationships between the lag-1 and current responses are assumed to be same for $Y_1$ and $Y_2$, i.e. $Z_{0j,t} = [1]$

Simulated data sets are analysed by the first-order model. The simulation procedure is replicated 500 times for each case. Analysis of a simulated data set (the last one) took 8.9 minutes on a PC with 4.00 GB RAM and 3.00 GHz processor. A simulated data set and the R script for data analysis are available in the user manual of the pmrtrim package.

Simulation results are displayed in Table 1. We report mean, percentage bias (Bias(%)), empirical standard deviations of the parameter estimates (SD), mean of the standard errors of the parameter estimates (meSE), and coverage probabilities of the corresponding 95% confidence intervals (CP). Parameters are approximately unbiased. Empirical standard deviations of the parameter estimates and the means of the standard error estimates are close to each other. Coverage probabilities are close to the nominal level of 0.95.

### 6 Example: Iowa Youth and Families Project data set

#### 6.1 Data

We apply the first-order model to the data set from the Iowa Youth and Families Project (IFYP; Elder and Conger, 2000; Ilk, 2008). The project was conducted to investigate long-term effects of the farm crisis that began in 1980’s in the U.S. 451 families from eight rural parts of the north central Iowa were selected. 7th graders with two alive and biological parents and a sibling within 4 years old were the target. The focus is on their well-being. The study was started in 1989, and conducted yearly until 1992. Then, it was conducted in 1994, 1995, 1997 and 1999. At each follow-up, both
Table 1: Simulation results based on 500 replications. For the details of the abbreviations, see the text.

<table>
<thead>
<tr>
<th>Case</th>
<th>SE</th>
<th>meSE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.50</td>
<td>0.12</td>
<td>0.95</td>
</tr>
<tr>
<td>Bias (%)</td>
<td>0.79</td>
<td>2.11</td>
<td>0.95</td>
</tr>
<tr>
<td>SE</td>
<td>0.11</td>
<td>0.19</td>
<td>0.95</td>
</tr>
<tr>
<td>meSE</td>
<td>0.12</td>
<td>0.20</td>
<td>0.95</td>
</tr>
<tr>
<td>CP</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

the parents and children were surveyed. At the beginning of the study, 48% of the 7th graders were male and their average age was 12.7 years.

Three main distress variables, anxiety, hostility and depression, were used to measure emotional statuses of the young people (Table 2). These variables were collected by a list of symptoms, e.g. including nervousness, shakiness, an urge to break things and feeling low in energy etc. The symptoms were then dichotomised (Ilk, 2008). The frequencies of the dichotomised distress variables are given in Table 3. The frequencies of depression were higher compared to those of anxiety and hostility, and the frequencies of the latter variables were close to each other. For instance, almost 93% of them reported at least one depression symptom at 1989, whilst the frequencies of anxiety and hostility were 83.2%. A set of explanatory variables, thought to be related with the distress variables, were also collected (Table 2). These variables include gender, degree of negative life event experiences of the young people, e.g. having a close friend moved away permanently, financial cutbacks, e.g. moving to a cheaper residence, and negative economical event experiences of their families, e.g. such as changing job for a worse one. Amongst the explanatory variables, whilst gender was time-invariant, the others were time-varying.

Transitional structure of our model requires equally-spaced data. Therefore, we analyse the first four follow-ups of the IYFP data set. Indicator variables for distress variables and time are considered as additional explanatory variables, and dummy variables are created for all the categorical covariates (Table 2). We coded the binary explanatory variables as 0 vs. 1 in our initial data analyses. However, the alternative coding of -1 vs. 1 is used due to convergence problems.
Table 2: Variable list of IYFP used in PNMTREM(1).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses</td>
<td>whether the young person had symptoms: (0 = \text{absence}, 1 = \text{presence})</td>
</tr>
<tr>
<td>anxiety</td>
<td>(0 = \text{absence}, 1 = \text{presence})</td>
</tr>
<tr>
<td>hostility</td>
<td>(0 = \text{absence}, 1 = \text{presence})</td>
</tr>
<tr>
<td>depression</td>
<td>(0 = \text{absence}, 1 = \text{presence})</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>gender of the young person: (-1 = \text{male}, 1 = \text{female})</td>
</tr>
<tr>
<td>NLE1</td>
<td>first indicator variable for negative life event experiences of young people: (1 = \text{some}, -1 = \text{none or many})</td>
</tr>
<tr>
<td>NLE2</td>
<td>second indicator variable for negative life event experiences of young people: (1 = \text{many}, -1 = \text{none or some})</td>
</tr>
<tr>
<td>NEE</td>
<td>whether the household had any negative economical event: (-1 = \text{no}, 1 = \text{yes})</td>
</tr>
<tr>
<td>cut1</td>
<td>first indicator variable for financial cutback experiences of the household: (1 = \text{between 1 and 5}, -1 = \text{none or more than 5})</td>
</tr>
<tr>
<td>cut2</td>
<td>second indicator variable for financial cutback experiences of the household: (1 = \text{more than 5}, -1 = \text{none or between 1 and 5})</td>
</tr>
<tr>
<td>resp1</td>
<td>first response indicator variable: (1 = \text{hostility}, -1 = \text{anxiety or depression})</td>
</tr>
<tr>
<td>resp2</td>
<td>second response indicator variable: (1 = \text{depression}, -1 = \text{hostility or anxiety})</td>
</tr>
<tr>
<td>time1</td>
<td>first indicator variable for follow-up time: (1 = 1991, -1 = 1990 \text{ or 1992})</td>
</tr>
<tr>
<td>time2</td>
<td>second indicator variable for follow-up time: (1 = 1992, -1 = 1990 \text{ or 1991})</td>
</tr>
</tbody>
</table>

Table 3: Frequency table of the distress variables across years.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>375 (83.2%)</td>
<td>347 (76.9%)</td>
<td>342 (75.8%)</td>
<td>327 (72.5%)</td>
</tr>
<tr>
<td>Hostility</td>
<td>375 (83.2%)</td>
<td>350 (77.6%)</td>
<td>342 (75.8%)</td>
<td>328 (72.7%)</td>
</tr>
<tr>
<td>Depression</td>
<td>418 (92.7%)</td>
<td>385 (85.4%)</td>
<td>378 (83.8%)</td>
<td>386 (85.6%)</td>
</tr>
</tbody>
</table>

6.2 Results

We specifically build two models. Whilst the set of explanatory variables are same, the models differ in terms of separating the lag-1 associations amongst the distress variables. Whilst the first model (Model 1 in Table 5) assumes these associations are shared across the responses, i.e. \(Z_{it1} = [1]\), the second model (Model 2 in Table 5) assumes that the associations are different for the distress variables, i.e. \(Z_{it1} = [1 \ \text{resp1} \ \text{resp2}]\). Results for baseline models are presented in Table 4. Note that the baseline results of Model 1 and Model 2 are same, since the specifications of the baseline parameter sets are same. Results for \(t \geq 2\) models are presented in Table 5.

We compare Model 1 and 2 by likelihood ratio test (LRT), since they are nested. Respective maximised log-likelihoods are \(-1236.78 (= -210.78 - 1026)\) and \(-1234.49 (= -210.78 - 1023.71)\). The LRT statistic is 4.58 (\(= -2 \times (-1026 - (-1023.71))\)), with a p-value of 0.60. This indicates that there is not enough evidence to conclude that Model 2 is a better model to analyse the IYFP data set compared to Model 1. Therefore, throughout the paper we only discuss the results of Model 1.

We check existence of multicollinearity problem by variance inflation factor. The largest value is 1.17 (results not shown here). This indicates that multicollinearity is not a problem for the analysis of the IYFP data set. We rely on the findings of Ilk and Daniels (2007) regarding the exogeneity of the time-varying covariates in the IFYP data set.
Table 4: Results for $t = 1989$. $H_0 : \lambda_{\text{hostility}}^* = 1$ and $H_0 : \lambda_{\text{depression}}^* = 1$; other parameters are tested for 0.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Est.</th>
<th>SE</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^*_0$</td>
<td>1.33</td>
<td>0.07</td>
<td>18.82</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta^*_\text{gender}$</td>
<td>-0.09</td>
<td>0.06</td>
<td>-1.41</td>
<td>0.16</td>
</tr>
<tr>
<td>$\beta^*_\text{NLE1}$</td>
<td>0.20</td>
<td>0.12</td>
<td>1.61</td>
<td>0.11</td>
</tr>
<tr>
<td>$\beta^*_\text{NLE2}$</td>
<td>0.41</td>
<td>0.12</td>
<td>3.27</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta^*_\text{NEE}$</td>
<td>0.03</td>
<td>0.05</td>
<td>0.72</td>
<td>0.47</td>
</tr>
<tr>
<td>$\beta^*_\text{cut1}$</td>
<td>0.08</td>
<td>0.07</td>
<td>1.15</td>
<td>0.25</td>
</tr>
<tr>
<td>$\beta^*_\text{cut2}$</td>
<td>-0.003</td>
<td>0.07</td>
<td>-0.04</td>
<td>0.97</td>
</tr>
<tr>
<td>$\beta^*_\text{resp1}$</td>
<td>-0.001</td>
<td>0.06</td>
<td>-0.02</td>
<td>0.99</td>
</tr>
<tr>
<td>$\beta^*_\text{resp2}$</td>
<td>0.29</td>
<td>0.07</td>
<td>4.17</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta^*_\text{gender+resp1}$</td>
<td>-0.04</td>
<td>0.06</td>
<td>-0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>$\beta^*_\text{gender+resp2}$</td>
<td>-0.08</td>
<td>0.07</td>
<td>-1.29</td>
<td>0.20</td>
</tr>
<tr>
<td>$\lambda_{\text{hostility}}$</td>
<td>1.10</td>
<td>0.79</td>
<td>0.12</td>
<td>0.91</td>
</tr>
<tr>
<td>$\lambda_{\text{depression}}^*$</td>
<td>1.04</td>
<td>0.71</td>
<td>0.05</td>
<td>0.96</td>
</tr>
<tr>
<td>log($\sigma_1^*$)</td>
<td>-0.41</td>
<td>0.41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3 Population-averaged results

At baseline (1989), only the intercept, one of the negative life event indicators (NLE2) and one of the response indicators (resp2) are significant. The estimate of intercept, $\hat{\beta}^*_0 = 1.33$, indicates that young people had high probability of distress at 1989. The estimate of the second response indicator variable, $\hat{\beta}^*_\text{resp2} = 0.29$, indicates that young people were more likely to report depression compared to anxiety and hostility. Insignificance of the first response indicator (p-value=0.99) indicates that reporting anxiety and hostility were equally likely. These findings are in agreement with the empirical frequencies (Table 3). Young people who had many negative life events were more likely to be distressed ($\hat{\beta}^*_\text{NLE2} = 0.41$). Pairwise correlations between anxiety, hostility and depression were not significantly different, p-values of $\lambda_{\text{hostility}}$ and $\lambda_{\text{depression}}^*$ were 0.91 and 0.96. The standard deviation estimate of the random effects distribution is 0.66 ($= \exp(-0.41)$), with a standard error of 0.27 ($= \sqrt{0.41^2 \times \exp(-0.41 \times 2)}$, by the delta method). The standard deviation is significantly different from 0, with a p-value of 0.007. Of note, we modified the p-value following Molenberghs and Verbeke (2007).

For 1990 – 1992, the intercept, gender, both negative life event indicators (NLE1, NLE2), negative economical events experience (NEE), one of the cutbacks indicators (cut1), one of the response indicators (resp2), one of the time indicators (time2) and the interaction between gender and second response indicator (gender * resp2) are significant. The estimate of the intercept ($\hat{\beta}_0 = 0.96$) indicates high probability of distress for 1990 – 1992, which tend to be higher compared to baseline, since $\hat{\beta}^*_0 > \hat{\beta}_0$. Females were more likely to report distress compared to males ($\hat{\beta}^*_\text{gender} = 0.18$). Furthermore, they were more likely to report depression ($\hat{\beta}^*_\text{gender+resp2} = 0.07$) compared to reporting anxiety or hostility. Note that gender is insignificant at 1989. This finding was also reported in Ge et al. (2001, cited in Ilk, 2008) and Ilk (2008). Experiencing many negative life events and any family-level negative economical events were associated with distress ($\hat{\beta}^*_\text{NLE1} = 0.14, \hat{\beta}^*_\text{NLE2} = 0.38$ and $\hat{\beta}^*_\text{NEE} = 0.08$). Reporting depression was more likely compared to reporting anxiety or hostility.
Table 5: Results for \( t \geq 1990 \). \( H_0: \lambda_{\text{hostility}} = 1 \) and \( H_0: \lambda_{\text{depression}} = 1 \); other parameters are tested for 0.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>0.96 0.05</td>
<td>0.96 0.05</td>
</tr>
<tr>
<td>( \beta_{\text{time}} )</td>
<td>0.18 0.03</td>
<td>0.18 0.03</td>
</tr>
<tr>
<td>( \beta_{\text{NEE}} )</td>
<td>0.14 0.04</td>
<td>0.14 0.05</td>
</tr>
<tr>
<td>( \beta_{\text{LE1}} )</td>
<td>0.38 0.05</td>
<td>0.38 0.05</td>
</tr>
<tr>
<td>( \beta_{\text{LE2}} )</td>
<td>0.08 0.03</td>
<td>0.08 0.03</td>
</tr>
<tr>
<td>( \beta_{\text{art}} )</td>
<td>0.06 0.03</td>
<td>0.07 0.03</td>
</tr>
<tr>
<td>( \beta_{\text{resp1}} )</td>
<td>0.02 0.03</td>
<td>0.02 0.03</td>
</tr>
<tr>
<td>( \beta_{\text{resp2}} )</td>
<td>0.01 0.04</td>
<td>0.01 0.04</td>
</tr>
<tr>
<td>( \beta_{\text{mel}} )</td>
<td>-0.07 0.04</td>
<td>-0.08 0.05</td>
</tr>
<tr>
<td>( \beta_{\text{time2}} )</td>
<td>-0.09 0.05</td>
<td>-0.09 0.05</td>
</tr>
<tr>
<td>( \beta_{\text{art} \times \text{resp1}} )</td>
<td>-0.01 0.03</td>
<td>-0.01 0.03</td>
</tr>
<tr>
<td>( \beta_{\text{art} \times \text{resp2}} )</td>
<td>0.07 0.04</td>
<td>0.07 0.04</td>
</tr>
<tr>
<td>( \beta_{\text{resp1} \times \text{time}} )</td>
<td>-0.002 0.03</td>
<td>-0.02 0.04</td>
</tr>
<tr>
<td>( \beta_{\text{resp2} \times \text{time}} )</td>
<td>0.004 0.04</td>
<td>0.003 0.04</td>
</tr>
<tr>
<td>( \beta_{\text{resp2} \times \text{time2}} )</td>
<td>-0.01 0.04</td>
<td>-0.01 0.04</td>
</tr>
<tr>
<td>( \alpha_{21,1} )</td>
<td>0.76 0.11</td>
<td>0.75 0.17</td>
</tr>
<tr>
<td>( \alpha_{22,1} )</td>
<td>0.11 0.16</td>
<td>0.70 0.48</td>
</tr>
<tr>
<td>( \alpha_{31,1} )</td>
<td>0.87 0.10</td>
<td>0.86 0.13</td>
</tr>
<tr>
<td>( \alpha_{32,1} )</td>
<td>0.08 0.11</td>
<td>0.74 0.46</td>
</tr>
<tr>
<td>( \alpha_{33,1} )</td>
<td>0.07 0.14</td>
<td>0.48 0.63</td>
</tr>
<tr>
<td>( \alpha_{41,1} )</td>
<td>0.90 0.10</td>
<td>0.86 0.12</td>
</tr>
<tr>
<td>( \alpha_{42,1} )</td>
<td>-0.04 0.12</td>
<td>-0.34 0.74</td>
</tr>
<tr>
<td>( \alpha_{43,1} )</td>
<td>0.12 0.13</td>
<td>0.93 0.35</td>
</tr>
<tr>
<td>( \lambda_{\text{hostility}} )</td>
<td>1.03 0.37</td>
<td>0.99 0.36</td>
</tr>
<tr>
<td>( \log(\sigma_1) )</td>
<td>1.21 0.49</td>
<td>1.18 0.49</td>
</tr>
<tr>
<td>( \log(\sigma_2) )</td>
<td>-0.48 0.25</td>
<td>-0.47 0.26</td>
</tr>
<tr>
<td>( \log(\sigma_3) )</td>
<td>-0.62 0.25</td>
<td>-0.59 0.26</td>
</tr>
<tr>
<td>( \log(\sigma_4) )</td>
<td>-0.62 0.26</td>
<td>-0.59 0.26</td>
</tr>
<tr>
<td>Max. loglk</td>
<td>-1026.00</td>
<td>-1023.71</td>
</tr>
</tbody>
</table>

(\( \hat{\beta}_{\text{resp2}} = 0.22 \)). On the other hand, reporting anxiety or hostility were not significantly different (p-value of \( \hat{\beta}_{\text{resp1}} = 0.78 \)). Whilst the distress levels were lower at 1992 compared to 1990 and 1991 (\( \hat{\lambda}_{\text{time2}} = -0.09 \)), the distress levels at 1990 and 1991 were not significantly different from each other (p-value of \( \hat{\beta}_{\text{resp1}} = 0.08 \)). The decrease in the distress probabilities at 1992 was not different for depression, anxiety and hostility; respective p-values for \( \hat{\beta}_{\text{resp1} \times \text{time2}} \) and \( \hat{\beta}_{\text{resp2} \times \text{time2}} \) are 0.92 and 0.25.

The marginal mean parameter estimates based on probit link can be interpreted in terms of odds-ratios, using the JKB constant; for details see Introduction. For instance, young people who experienced many negative life events were approximately 2.26 (= \( \exp(1.700437 	imes ((-1 \times 0.14 + 1 \times 0.38) - (1 \times 0.14 - 1 \times 0.38)) \)) times more likely to be distressed compared to those with some negative life events, and individuals in the latter group were 1.60 (= \( \exp(1.700437 	imes ((1 \times 0.14 - 1 \times 0.38) - (-1 \times 0.14 - 1 \times 0.38)) \)) times more likely to be distressed compared to those with no negative life events.

The transition parameter estimates are positive and significant: \( \hat{\alpha}_{21,1} = 0.76, \hat{\alpha}_{31,1} = 0.87, \hat{\alpha}_{41,1} = 0.90 \) with p-values < \( 1 \times 10^{-10} \). These indicate that that young people who were distressed at year before were more likely to be distressed at current year. As indicated by the baseline model, the
pairwise correlations between anxiety, hostility and depression were not significantly different; corresponding p-values were 0.94 and 0.68 for hostility and depression, respectively. The standard deviation estimates of the random effects distributions were 0.62 (= exp(−0.48)), 0.54 (= exp(−0.62)) and 0.54 (= exp(−0.62)) at 1990, 1991 and 1992, respectively. Respective standard errors were 0.16, 0.14 and 0.14, and all of these parameters were significant (p-values < 0.0001). These results indicate that the individual variations decreased through time (recall that \( \hat{\sigma}_1 = 0.66 \)) and close to each other at 1991 and 1992.

### 6.4 Subject-specific results

We calculate probabilities of reporting anxiety, hostility and depression for each individual at each year. We also calculate the marginal probabilities for comparison. These probabilities are plotted in Figure 1; only the results for depression are shown here due to page limits, others can be found in the online supplementary material. We label observed values by 0 and 1 according to absence and presence depression, respectively. Marginal probabilities range in a narrower interval compared to conditional probabilities. For instance, whilst the range for the marginal probabilities of being depressed at the period of 1990 - 1992 was (0.576, 0.971), it was (0.118, 0.999) for the conditional probabilities. This indicates that marginal probabilities are high even for young people who did not report depression. On the other hand, conditional probabilities leads to correct decisions. For instance, in Figure 1, the 0’s were associated with lower conditional probabilities. The associated box-plots reflect the location and scale of the marginal and conditional probabilities. Whereas the the conditional probabilities have a spread distribution with many outliers, the marginal probabilities have a stacked and narrow distribution.

Probabilities of a young person with ID=223 are presented in Table 6. This person was a female, with some negative life event experiences, no negative economical event experiences and cutbacks between 1 and 5, except in 1992 at which her family did not experience any cutbacks. She did not report distress at all. Predicted value of \( z_{223} \) is −2.45. This indicates that she was less likely to report distress compared to an average person, i.e. \( z_i = 0 \). For this person, the conditional probabilities lead to correct inferences compared to the marginal probabilities. For instance, at 1992, whereas the marginal probability of being anxious is 0.64, the conditional probability being anxious is 0.08. We also calculate conditional probabilities assuming that the person is an average person, i.e. setting \( z_{223} = 0 \). Related results are given under Conditional*. These probabilities are still subject, time and response specific, since \( \Delta_{ij} \) holds subject, time and response specific information. For instance, at 1992, probability of being anxious based on this method is 0.46.
Figure 1: Scatter and box plots of marginal vs. conditional probabilities for depression at 1989 (left panel) and 1990-1992 (right panel).
Table 6: Marginal and conditional probabilities for the young person with ID = 223.

<table>
<thead>
<tr>
<th>Time</th>
<th>Response</th>
<th>Gender</th>
<th>NLE</th>
<th>NEE</th>
<th>Cutbacks</th>
<th>Observed</th>
<th>Marginal</th>
<th>Conditional</th>
<th>Conditional*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>Anxiety</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.82</td>
<td>0.30</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.80</td>
<td>0.23</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.91</td>
<td>0.47</td>
<td>0.95</td>
</tr>
<tr>
<td>1990</td>
<td>Anxiety</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.78</td>
<td>0.09</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.78</td>
<td>0.09</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.90</td>
<td>0.14</td>
<td>0.77</td>
</tr>
<tr>
<td>1991</td>
<td>Anxiety</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.74</td>
<td>0.14</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.74</td>
<td>0.13</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>Betw. 1 &amp; 5</td>
<td>Absence</td>
<td>0.86</td>
<td>0.22</td>
<td>0.79</td>
</tr>
<tr>
<td>1992</td>
<td>Anxiety</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>None</td>
<td>Absence</td>
<td>0.64</td>
<td>0.08</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>None</td>
<td>Absence</td>
<td>0.65</td>
<td>0.08</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>Female</td>
<td>Some</td>
<td>No</td>
<td>None</td>
<td>Absence</td>
<td>0.85</td>
<td>0.19</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 7: Frequency table of the stayers. “All” stands for the subjects who reported the same answer for all the distress variables.

<table>
<thead>
<tr>
<th></th>
<th>Absence (0)</th>
<th>Presence (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>15 (3.3%)</td>
<td>215 (47.7%)</td>
</tr>
<tr>
<td>Hostility</td>
<td>9 (2%)</td>
<td>221 (49%)</td>
</tr>
<tr>
<td>Depression</td>
<td>2 (0.4%)</td>
<td>288 (63.9%)</td>
</tr>
<tr>
<td>All</td>
<td>2 (0.4%)</td>
<td>134 (29.7%)</td>
</tr>
</tbody>
</table>

6.5 Diagnostics

Longitudinal binary data sets almost surely include stayers, i.e. subjects who constantly report absence (0) or presence (1) of a binary variable at all time points, for which the subject with ID=223 is an example. The counts and percentages of the stayers in the IYFP data set are given in Table 7. For instance, 29.7% of the subjects reported 1 for all the three distress variables at all the time points. Marginal and conditional anxiety probabilities of the stayers in terms all the distress variables are summarised in Figure 2. Other results can be found the online supplementary material. Whilst the gray lines represent the subjects who always reported 1, the black lines represent the ones who always reported 0. Conditional probabilities are successful at correctly assigning the success probabilities for these subjects; higher probabilities for subjects reporting 1 and lower probabilities for those who reported 0. On the other hand, marginal probabilities are not able to distinguish these subjects.

We also calculate accuracy measures to summarise the predicted probabilities. We specifically use expected proportion of correct prediction (Herron, 1999) and area under the receiver operating characteristics curve (AUROC). Results (not shown here) show that conditional probabilities outperform the marginal probabilities. This difference is apparent especially in terms of AUROC. For instance, while the AUROC value for depression at 1990-1992 is 0.684 for marginal probabilities, it is 0.864 for the conditional probabilities.

7 Discussion and conclusion

In this paper, we have proposed a marginalised model for analysis of multivariate longitudinal binary data. It is an extension of the model proposed by Ilk and Daniels (2007). These authors use logit link, and estimate the parameters using BM, specifically Markov Chain Monte Carlo methods. Un-
automated Fortran codes are available from the personal website of Dr. Ilk. However, their procedure is computationally cumbersome and requires expertise in BM and Fortran. These aspects prohibit the routine use of the model. In this study, we replace logit link by probit, and use ML for parameter estimation. probit link enables us explicitly linking the second and third levels of the model, which is not possible with the logit link. On the other hand, parameter estimation with ML takes less time compared to BM. We propose the use of implicit function theorem to solve the marginal constraint equations directly. To the best of our knowledge, this application is proposed for the first time here for marginalised models. We have prepared the publicly available R package pnmtrem to fit the proposed model. Currently, the package provides a function for fitting the first-order model. The function considers both parameter estimation and random effects prediction. It has been tested under different conditions. For the details and usage, we refer the readers to the package manual.

We have conducted a simulation study to investigate the properties of the estimator under different scenarios. Results are satisfactory in terms of unbiasedness, efficiency and coverage. We have illustrated the first-order model with an application to the IYFP data set. Both population-averaged and subject-specific inferences have been illustrated. Our findings on the IYFP data analysis coincide with the findings of Ilk (2008). As a separate note, the IYFP data set is available upon request from the authors.

A natural extension of our work here would be fitting higher-order models. The variances of random effects could be modified by a subset of covariates, i.e. \( \log(\sigma_i) = M_tj \omega_t \) where \( M_tj \) is a set covariates and \( \omega_t \) are the associated parameters. Also, the random effects coefficients might be assumed to have a multivariate normal distribution, i.e. \( b_t \sim N(0, D) \) where \( D \) is a \( T \times T \) matrix. However, all of these extensions require intensive new derivations and implementations. Therefore, we leave them as future work.

Figure 2: Spaghetti plots of marginal (left panel) and conditional (right panel) anxiety probabilities for stayers in terms of all the distress variables. While gray lines represent subjects who reported 1, the black lines represent subjects who reported 0.
Appendices

A. Linking second and third levels of the $t \geq 2$ model

Whilst linking second and third levels of the $t \geq 2$ model, we claim the following

$$\int \Phi(\Delta_{ij}^* + \lambda_j \beta_{it}) f(b_{it}) db_{it} = \Phi \left( \frac{\Delta_{ij}^*}{\sqrt{1 + (\lambda_j \sigma_t)^2}} \right)$$

where $b_{it} \sim N(0, \sigma_t^2)$ and $b_{it} = z_i \sigma_t$, $z_i \sim N(0, 1)$. The related proof, which is modified from Griswold (2005), is given below.

Let $W_i \perp z_i$, where $W_i \sim N(0, 1)$, then,

$$W_i / (\lambda_j \sigma_t) \sim N(0, (\lambda_j \sigma_t)^{-2})$$

$$W_i / (\lambda_j \sigma_t) - z_i \sim N(0, 1 + (\lambda_j \sigma_t)^{-2})$$

$$\sqrt{1 + (\lambda_j \sigma_t)^{-2}} \sim N(0, 1)$$

and

$$\int \Phi(\Delta_{ij}^* + \lambda_j \beta_{it}) f(b_{it}) db_{it} = \int_{-\infty}^{\infty} \Phi(\Delta_{ij}^* + \lambda_j \sigma_t) \phi(z_i) dz_i$$

$$= \int_{-\infty}^{\infty} \Phi \left( \frac{W_i / (\lambda_j \sigma_t) - z_i}{\sqrt{1 + (\lambda_j \sigma_t)^{-2}}} \right) \phi(z_i) dz_i$$

$$= \int_{-\infty}^{\infty} \Phi \left( \frac{W_i / (\lambda_j \sigma_t) - z_i}{\sqrt{1 + (\lambda_j \sigma_t)^{-2}}} \right) \phi(z_i) dz_i$$

$$= \Phi \left( \frac{\Delta_{ij}^*}{\sqrt{1 + (\lambda_j \sigma_t)^2}} \right)$$

B.1 ML estimation of $\theta_1$

Maximizing the log-likelihood function of the baseline model, $L_1(\theta_1 | y_1)$, with respect to $\theta_1$ yields

$$\frac{\partial \log(L_1(\theta_1 | y_1))}{\partial \theta_1} \approx \sum_{i=1}^{N} \frac{1}{h(Y_i | \theta_1)} \frac{\partial h(Y_i | \theta_1)}{\partial \theta_1}$$

(27)

where

$$h(Y_i | \theta_1) \approx \sum_{q=1}^{20} w_q \exp \left[ \sum_{j=1}^{k} \left( Y_{ij} \log \left( \Phi(d_{ijq}) \right) + (1 - Y_{ij}) \log \left( 1 - \Phi(d_{ijq}) \right) \right) \right],$$

(28)

$$\frac{\partial h(Y_i | \theta_1)}{\partial \theta_1} \approx \sum_{q=1}^{20} w_q \left( \ell(Y_i | \theta_1) \sum_{j=1}^{k} \frac{\partial d_{ijq}}{\partial \theta_1} \phi(d_{ijq}) \left( \frac{Y_{ij} - \Phi(d_{ijq})}{(\Phi(d_{ijq}))^2} \right) \left( 1 - \Phi(d_{ijq}) \right) \right],$$

(29)

$$d_{ijq} = \sqrt{1 + \lambda_j^2 e^{2z_i} (X_{ij} \beta_j^*) + \lambda_j^2 e^{2z_i} \sqrt{2} z_q}.$$
Here, \( \log(\sigma_1) \) is equated to \( c_1 \) for simplicity of notation and \((z_q, w_q)\) for \( q = 1, \ldots, 20\) are Gauss-Hermite quadrature points and weights, respectively which are available in Abramowitz and Stegun (1972). The derivatives of \( d_{i,jq} \) with respect to \( \theta_1 = (\beta^*, \lambda^*, c_1) \) with \( \lambda^* = (\lambda_2^*, \ldots, \lambda_T^*) \) are given below.

\[
\frac{\partial d_{i,jq}}{\partial \beta^*} = \sqrt{1 + \lambda_2^* e^{2i} (X_{ij})} \\
\frac{\partial d_{i,jq}}{\partial \lambda_j^*} = (1 + \lambda_2^* e^{2i})^{-1/2} \lambda_j^* e^{2i} (X_{ij}) + e^{\epsilon} \sqrt{z_q} \\
\frac{\partial d_{i,jq}}{\partial c_1} = (1 + \lambda_2^* e^{2i})^{-1/2} \lambda_j^* e^{2i} (X_{ij}) + \lambda_j^* e^{\epsilon} \sqrt{z_q}
\]

**B.2 ML estimation of \( \theta_2 \)**

Similar to the baseline model, maximizing the log-likelihood function of the \( t \geq 2 \) model with respect to \( \theta_2 \) yields

\[
\frac{\partial \log(L_2(\theta_2|y_2))}{\partial \theta_2} = \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{1}{h(Y_i|\theta_2)} \frac{\partial h(Y_i|\theta_2)}{\partial \theta_2},
\]

where

\[
h(Y_i|\theta_2) = \sum_{q=1}^{20} w_q \exp \left\{ \sum_{j=1}^{k} \left( Y_{ij} \log \left( \Phi(d_{i,jq}) \right) + (1 - Y_{ij}) \log \left( 1 - \Phi(d_{i,jq}) \right) \right) \right\},
\]

\[
\frac{\partial h(Y_i|\theta_2)}{\partial \theta_2} \approx \sum_{q=1}^{20} w_q \left\{ \ell(Y_i|\theta_2) \left( \sum_{j=1}^{k} \frac{\partial d_{i,jq}}{\partial \theta_2} \Phi(d_{i,jq}) \left( \frac{Y_{ij} - \Phi(d_{i,jq})}{\Phi(d_{i,jq}) (1 - \Phi(d_{i,jq}))} \right) \right) \right\},
\]

\[
d_{i,jq} = \sqrt{1 + \lambda_2^* e^{2i} (\Delta_{ij} + \alpha_{c} \mathbf{Z}_{ij}(y_{ij-1})) + \lambda_j^* e^{\epsilon} \sqrt{z_q}}.
\]

Here, \( c_1 = \log(\sigma_1) \) for \( t \geq 2 \), and \((z_q, w_q)\) for \( q = 1, \ldots, 20 \) are Gauss-Hermite quadrature points and weights. Also note that explicit solution of \( \Delta_{ij} \) is given in (13). The derivatives of \( d_{i,jq} \) with respect to \( \theta_2 = (\beta, \alpha_{c}, \lambda, \epsilon) \) with \( \lambda = (\lambda_2^*, \ldots, \lambda_T^*) \) and \( \epsilon = (c_2, \ldots, c_T) \) are given below.

\[
\frac{\partial d_{i,jq}}{\partial \beta} = \sqrt{1 + \lambda_2^* e^{2i} (A_{ij})} \\
\frac{\partial d_{i,jq}}{\partial \alpha_{c}} = \sqrt{1 + \lambda_2^* e^{2i} (B_{ij} + \mathbf{Z}_{ij}(y_{ij-1}))} \\
\frac{\partial d_{i,jq}}{\partial \lambda_j} = (1 + \lambda_2^* e^{2i})^{-1/2} \lambda_j^* e^{2i} \left( (A_{ij} \beta_0 + B_{ij} \alpha_{c,10}) + A_{ij} \beta + \alpha_{c,1} (B_{ij} + \mathbf{Z}_{ij}(y_{ij-1})) \right) + e^{\epsilon} \sqrt{z_q} \\
\frac{\partial d_{i,jq}}{\partial \epsilon} = (1 + \lambda_2^* e^{2i})^{-1/2} \lambda_j^* e^{2i} \left( (A_{ij} \beta_0 + B_{ij} \alpha_{c,10}) + A_{ij} \beta + \alpha_{c,1} (B_{ij} + \mathbf{Z}_{ij}(y_{ij-1})) \right) + \lambda_j^* e^{\epsilon} \sqrt{z_q}
\]

where

\[
A_{ij} = \frac{\partial F}{\partial \beta}(\beta_{0}, \alpha_{c,10}, A_{ij,0}), \quad B_{ij} = \frac{\partial F}{\partial \alpha_{c}}(\beta_{0}, \alpha_{c,10}, A_{ij,0})
\]
References


