The Canadian Journal of Statistics
Vol. 34, No. 4, 2006, Pages ???-???
La revue canadienne de statistique

Marginalized transition random effects models for multivariate longitudinal binary data

Ozlem ILK and Michael J. DANIELS

Key words and phrases: Bayesian hierarchical model; Hybrid MC.
MSC 2000: Primary ??; secondary ??.

Abstract: Generalized linear models with random effects and/or serial dependence are commonly used to analyze longitudinal data. However, interpretation and computation of marginal covariate effects can be difficult. Heagerty has proposed marginally specified logistic-normal models (1999) and marginalized transition models (2002) for longitudinal binary and categorical data in which the marginal mean is modeled explicitly in the presence of random effects and serial dependence, respectively. In this paper, we extend his work to handle multivariate longitudinal binary response data by proposing a framework consisting of a triple of regression models, which permits subject-specific inferences, while modeling the marginal mean response taking into account dependence across time via a Markov structure and across responses within a subject for a given time via random effects. Markov Chain Monte Carlo Methods, specifically Gibbs sampling with Hybrid steps, are used to sample from the posterior distribution of parameters. Relative and absolute model fit is assessed via DIC and posterior predictive checks, respectively. Methods are illustrated on data from Iowa Youth and Families Project (IYFP), which contains both missing responses and covariates.

Marginalized transition random effects models for multivariate longitudinal binary data

Résumé : Generalized linear models with random effects and/or serial dependence are commonly used to analyze longitudinal data. However, interpretation and computation of marginal covariate effects can be difficult. Heagerty has proposed marginally specified logistic-normal models (1999) and marginalized transition models (2002) for longitudinal binary and categorical data in which the marginal mean is modeled explicitly in the presence of random effects and serial dependence, respectively. In this paper, we extend his work to handle multivariate longitudinal binary response data by proposing a framework consisting of a triple of regression models, which permits subject-specific inferences, while modeling the marginal mean response taking into account dependence across time via a Markov structure and across responses within a subject for a given time via random effects. Markov Chain Monte Carlo Methods, specifically Gibbs sampling with Hybrid steps, are used to sample from the posterior distribution of parameters. Relative and absolute model fit is assessed via DIC and posterior predictive checks, respectively. Methods are illustrated on data from Iowa Youth and Families Project (IYFP), which contains both missing responses and covariates.

1. INTRODUCTION

Repeated measurements from the same subject observed over time are referred to as longitudinal data. The within-subject measurements over time are typically not independent. When more than one response is measured on each individual at each time (multivariate longitudinal data) a second correlation is introduced, correlation among responses. This paper proposes a model for multivariate binary data that directly models marginal covariate effects while taking these two types of dependence into account.

There is a wealth of literature on modeling longitudinal data. Some recent articles and texts include Verbeke & Molenberghs (2000), Pourahmadi & Daniels (2002), Diggle, Heagerty, Liang & Zeger (2002), and Singer & Willett (2003). Models for multivariate binary data are a currently developing area with still limited literature. Some of the models proposed for multivariate binary or categorical data use random effects models and/or latent variables to account for the multivariate structure of responses (Bandeen-Roche, Miglioretti, Zeger & Rathouz 1997; Legler & Ryan 1997). However, determination and computation of marginal covariate effects with these models can be difficult. Agresti (1997) proposed models for multivariate, longitudinal binary data that allow marginal inferences. However, these models only handle categorical covariates and do not exploit the ordering implicit in longitudinal data. Geys, Molenberghs & Ryan (1999) used pseudo-likelihood approaches for multivariate binary responses; the correlation structure in these models is not well-suited to longitudinal data and marginal covariate effects are difficult to compute. Ribaudo & Thompson (2002) built a three level model for multivariate longitudinal binary data in the context of quality of life data. They introduced dependence via random effects and directly modeled conditional (individual specific) covariate effects. Reboussin and Anthony (2001) constructed latent class marginal models for multivariate longitudinal binary data with estimation via estimating equations.

Generalized estimating equations (GEEs) are a semiparametric alternative (Liang & Zeger 1986) to full-likelihood based approaches. Missing at random (MAR) responses and/or covariates can be handled by re-weighting (Robins, Rotnitzky & Zhao 1994, 1995). However, such approaches are often inefficient compared to full likelihood based alternatives.

Our work here is most similar to that proposed by Fitzmaurice & Laird (1993), Heagerty (1999, 2002) and Miglioretti & Heagerty (2004) in that we propose a full likelihood based approach that directly models marginal (population level) covariate effects. We describe this previous work below. Fitzmaurice & Laird (1993) proposed a likelihood based method for multivariate binary data that directly models the marginal mean. They used a marginal logistic model to explain the mean structure and conditional log odds-ratios to model the time dependence. However, since the log odds-ratios considered were conditioned on all observed responses, and not just on the history (previous responses at a given time point), this model does not allow straightforward modeling of longitudinal correlation structures.

Heagerty (1999, 2002) introduced marginally specified logistic-normal models and marginalized transition models (MTM) for univariate longitudinal binary data. In both models, a marginal logistic regression was first specified for explaining the average response. The model was completed by introducing a conditional regression that allows for the longitudinal, within-subject, dependence, either via random effects or regressing on previous responses. These two regressions are tied together by appropriately constraining the 'intercept' parameters in the conditional regression model. Parameter estimation in both papers was handled by maximum likelihood and/or estimating equations.

Recently, Miglioretti & Heagerty (2004) developed marginalized multilevel models for binary data in the presence of time-varying covariates. The first level of this model was a marginal logistic regression model. The second level included a transition model to account for a temporal correlation within subjects plus a random effects term to incorporate correlation within a larger cluster.

In this paper, we build on Heagerty's work to handle **multivariate** longitudinal binary response data. We separate the random effects and transition structure used in Miglioretti & Heagerty (2004) by introducing a third level of the model. Estimation in this model is handled by MCMC methods, specifically Gibbs sampling with Hybrid steps (Neal 1996).

The use of a Bayesian approach to inference in these models provides several advantages. Under certain assumptions (usually MAR), uncertainty arising from missing data can be incorporated via Monte Carlo integration. For complex models like the ones proposed here, the Bayesian MCMC machinery provides a computational advantage over maximum likelihood approaches. For example, intermittent missing data can be addressed using a data augmentation step (Tanner & Wong 1987) while working with full conditional distributions for the parameters conditional on the complete set of data; intermittent missing data can complicate direct maximization of the likelihood and/or maximization via the EM algorithm. In addition, the dependence structure can be much more carefully modelled with potential efficiency gains over estimation done via estimating equations (that essentially treat the dependence as a nuisance). Finally, exact (up to MC error) approaches to inference can be done using the posterior distribution which avoids the need to use large sample theory for inference.

The Iowa Youth and Families Project (IYFP) is a longitudinal study of 451 Iowa families. The targets were 7-th graders, who along with their families were followed for more than eight years. The IYFP aimed to investigate the effect of negative life and economic events and characteristics of targets, such as their gender, on the emotional distress through the symptoms of anxiety, hostility and depression. The dataset involved many challenges: multivariate repeated responses, both time dependent and time independent covariates, and missing responses and covariates.

We introduce the model and approaches to handle the missing values and for model checking in Section 2. Section 3 describes the motivating dataset from Iowa Youth and Families Project (IYFP) in detail. In Section 4, results and comparisons with simpler approaches are presented.

2. MODEL

Multivariate longitudinal data introduces two kinds of correlation: a within-subject time dependence and a within-subject multivariate response dependence; the latter corresponding to the correlation between anxiety, hostility, and depression in the IYFP example. To take these two types of dependence into account, we propose a framework consisting of a triple of regression models: a marginal logistic regression to explain the mean response, a transition model to explain within-subject time dependence for each response, and a random effects model for the multivariate response structure at each time. In this section, we introduce the general model and provide details for the first order model. Further model features and model checking are also discussed.

2.1 General model: MTREM(p)

Let Y_{itj} be the jth response ($j = 1, \dots, J$) for the ith subject ($i = 1, \dots, n$) at time t ($t = 1, \dots, T$), and let X_{itj} be the corresponding set of covariates. Then, marginalized transition random effects models, MTREM(p) is defined as,

$$\text{logitP}(Y_{itj} = 1 | X_{i1j}, ..., X_{itj}) = X_{itj}\ \beta, \tag{1}$$

$$\text{logitP}(Y_{itj} = 1 | y_{i,t-1,j}, ..., y_{i,t-p,j}, X_{i1j}, \dots, X_{itj}) = \Delta_{itj} + \textstyle\sum_{m=1}^{p} \gamma_{itj,m} y_{i,t-m,j}, \tag{2}$$

$$\text{logitP}(Y_{itj} = 1 | y_{i,t-1,j}, ..., y_{i,t-p,j}, X_{i1j}, \dots, X_{itj}, b_{it}) = \Delta^*_{itj} + \lambda_j b_{it}, \tag{3}$$

where $t > p$, $b_{it} \sim N(0, \sigma_t^2)$, and $\lambda_1 = 1$ (for identifiability). We can re-write b_{it} as $b_{it} = \sigma_t z_i$, where $z_i \sim N(0, 1)$; this will be used later. The logit transformation is defined as the logarithm of the odds.

The first level given in (1) models the covariate effects on the population, i.e., marginal covariate effects. The marginal regression coefficients β contrast the log odds of success for different values of a covariate, X_{itj}, by averaging over individual variation, hence comparing subgroups. For instance, we can compare the log odds of observing depression in females versus males. Typically it is assumed that $P(Y_{itj} = 1 | X_{i1j}, \dots, X_{itj}) = P(Y_{itj} = 1 | X_{itj})$.

For each of the $j = 1, \ldots, J$ outcomes, level 2 given in (2) captures the temporal (longitudinal) dependence within the $t = 1, \ldots, T$ responses via a transition (Markov) model of order p. The term Δ_{itj} is a subject/time/response specific intercept that takes into account the nonlinear relationship between the logit of the conditional mean in (2) and the marginal mean in (1). This will become more transparent when we introduce the constraints on Δ_{itj} in the next subsection. The mth (of p) transition parameters, $\gamma_{itj,m}$, are specific to subject i, at time t, for response j. We model these parameters as $\gamma_{itj,m} = \alpha C_{itj,m}$ for $m = 1, ..., p$, where $C_{itj,m}$ is a vector of subject/time/response specific covariates that can differ by lag (order). This specification allows the transition parameters to be shared across responses and time and also allows them to differ by subject-specific covariates. With this formulation, we can, for instance, test whether last year's response for subjects having negative life events in the previous year has a differential effect on this year's distress. Equations (1) and (2) have the structure of J MTM's (one for each response) although we do allow dependence parameters and/or covariate effects to be shared across responses. In this specification, we assume the longitudinal correlation is induced by the direct effect of the previous year's responses for that particular response type.

Level 3, given in (3), models the correlation among the J responses at each time, conditional on the previous (in time) responses via random effects. The term Δ^*_{itj} is an intercept that connects (3) to (2); this will be clarified in the next subsection. An approximation to the form of the correlation among the J responses here can be obtained using a first order Taylor series (Goldstein & Rasbash 1996),

$$cor(y_{itj}, y_{itj'}) = \frac{\lambda_j \lambda_{j'} \sigma_t^2 \pi_{itj}[1-\pi_{itj}]\pi_{itj'}[1-\pi_{itj'}]}{\sqrt{(\lambda_j^2 \pi_{itj}^2 (1-\pi_{itj})^2 \sigma_t^2 + \pi_{itj}(1-\pi_{itj}))}\sqrt{(\lambda_{j'}^2 \pi_{itj'}^2 (1-\pi_{itj'})^2 \sigma_t^2 + \pi_{itj'}(1-\pi_{itj'}))}},$$

where $\pi_{itj} = \exp(\Delta^\star_{itj})/(1 + \exp(\Delta^\star_{itj}))$. The interpretations of the λ_j and σ_t^2 is most easily understood under the setting of $\pi_{itj} = \pi_{itj'}$ for all j and j' and $\pi_{itj} = \pi_{it'j}$ for all t and t', respectively. The parameter λ_j allows the dependence among the J responses at a given time to vary; if $\lambda_j = 1$ for $j = 2, \ldots, J$, then we are assuming equal correlation among the J responses, conditional on the previous (time) responses. This correlation is allowed to change over time by allowing the random effects variance, σ_t^2, to vary with time. If $\sigma_t^2 = \sigma^2$ for $t = 1, \ldots, T$, we are assuming this across response correlation is constant over time. This three-level specification also induces some cross-response temporal dependence; we will explore this by examining such correlations induced by the model in Section 4.

In the MTREM, we assume that the conditional mean of responses given the entire set of covariates is equal to the conditional mean given the covariate history. That is, at a time point t, $E(Y_{itj}|X_{iqj}, q = 1, ..., T) = E(Y_{itj}|X_{isj}, s \leq t)$. This assumption is reasonable for exogenous time-varying covariates but not necessarily for endogenous ones and is necessary for the validity of the constraint equations introduced in Section 2.2. We refer the reader to Miglioretti and Heagerty (2004) for an approach to deal with endogenous time varying-covariates in marginalized models when this assumption does not hold. We will revisit this issue in the analysis of the example in Section 4.3.

2.2 First order model: MTREM(1)

In this section, we provide some details specific to the MTREM(1). For higher order models ($p > 1$), see the Appendix.

Initial state model

Since there is no history data available for the initial state ($t = 1$), we cannot use the second level of the model which regresses on the responses at the previous time. We specify a simpler, alternative model for $t = 1$. For the distribution of response at first time point, we assume the following:

$\text{logitP}(Y_{i1j}=1|X_{i1j}) = X_{i1j}\beta^*$

$\text{logitP}(Y_{i1j}=1|X_{i1j}, b_{i1}) = \Delta^*_{i1j} + \lambda^*_j b_{i1}$

where $b_{i1} \sim N(0, \sigma_1^2)$, $b_{i1} = \sigma_1 z_i$, and $\lambda_1^* = 1$. This model for the initial state has a similar form to the marginalized logistic normal model introduced in Heagerty (1999), but here, the random effects are scaled by λ_j^* to account for the J differing responses. In addition, we do not require $\beta = \beta^\star$, as in longitudinal data, more variability is expected at baseline, and marginal covariate effects are often different than those at later times.

Constraints

The intercepts in the logistic regression on the conditional probabilities in (2) and (3), Δ_{itj} and $\Delta^\star_{itj}$, are determined by the other parameters. To determine Δ_{itj}, we solve the marginal constraint equation formed from the connection between (1) and (2) (we suppress the dependence on the parameters to simplify the expression of the constraints in the following). We use the relation between the marginal (1) and conditional (2) probabilities and the assumption given at the end of Section 2.1, $E(Y_{itj}|X_{iqj}, q=1,...,T) = E(Y_{itj}|X_{isj}, s \leq t)$, to obtain

$$P(Y_{itj}=1|X_{i1j},...,X_{itj}) = \sum_{y_{i,t-1,j}} P(Y_{itj}=1|y_{i,t-1,j}, X_{i1j},...,X_{itj}) P(y_{i,t-1,j}|X_{i1j},...,X_{i,t-1,j})$$

which can be re-written as

$$\frac{e^{X_{itj}\beta}}{1+e^{X_{itj}\beta}} = \sum_{y_{i,t-1,j}=0}^{1} \frac{e^{\Delta_{itj}+\gamma_{itj,1}y_{i,t-1,j}}}{1+e^{\Delta_{itj}+\gamma_{itj,1}y_{i,t-1,j}}} \frac{e^{y_{i,t-1,j}X_{i,t-1,j}\beta}}{1+e^{X_{i,t-1,j}\beta}}, \tag{4}$$

to solve for Δ_{itj}. For $t=2$, the second term in the sum on the RHS of (4) is replaced by $\frac{e^{y_{i1j}X_{i1j}\beta^*}}{1+e^{X_{i1j}\beta^*}}$ from the initial state model. Given Δ_{itj}, we can solve for Δ^*_{itj} by forming the following convolution equation from the connection between (2) and (3),

$$P(Y_{itj}=1|y_{i,t-1,j}, X_{i1j},...,X_{itj}) = \int P(Y_{itj}=1|y_{i,t-1,j}, X_{i1j},...,X_{itj}, b_{it}) dF(b_{it})$$

which can be re-written as

$$\frac{e^{\Delta_{itj}+\gamma_{itj,1}y_{i,t-1,j}}}{1+e^{\Delta_{itj}+\gamma_{itj,1}y_{i,t-1,j}}} = \int \frac{e^{\Delta^*_{itj}+\lambda_j\sigma_t z_i}}{1+e^{\Delta^*_{itj}+\lambda_j\sigma_t z_i}} \phi(z_i) dz_i. \tag{5}$$

To approximate this 1-dimensional integral, we use Gauss-Hermite quadrature. For $t=1$, we construct the convolution equation using the initial state model,

$$P(Y_{i1j}=1|X_{i1j}) = \int P(Y_{i1j}=1|X_{i1j}, b_{i1}) dF(b_{i1})$$

$$\frac{e^{X_{i1j}\beta^*}}{1+e^{X_{i1j}\beta^*}} = \int \frac{e^{\Delta^*_{i1j}+\lambda^*_j\sigma_1 z_i}}{1+e^{\Delta^*_{i1j}+\lambda^*_j\sigma_1 z_i}} \phi(z_i) dz_i$$

These equations are all solved using Newton-Raphson methods. See the Appendix on the web page (*www.stat.ufl.edu/ ~ mdaniels/research.html*) or Ilk (2004) for details.

Posterior sampling

We construct a Gibbs sampling algorithm to sample from the posterior distribution of $(b_{it}, \beta, \beta^*, \alpha, \lambda_j, \lambda^*_j, \sigma_t)$. The full conditional distributions for b_{it}, β, β^*, and α are sampled using Hybrid MC (Neal 1996). We chose the Hybrid algorithm over a standard random walk Metropolis-Hastings since the Hybrid uses information from the gradient of the log full conditional distributions to more efficiently move through the posterior space and since the gradients are available in closed form (given in the web appendix). The full conditionals of $(log(\sigma_t^2),\ \lambda_j)$ and $(log(\sigma_1^2), \lambda^*_j)$ are sampled using Hybrid MC after the random effects are integrated out to facilitate convergence (step 4. of the algorithm in Appendix **A**). The details of the entire algorithm are also given in Appendix **A**. For details on the prior distributions specified for the example, see Section 4.2.

Missing responses and covariates

We used data augmentation (Tanner & Wong 1987) to handle missing responses within the MCMC algorithm (for details, see Appendix **A** and **B**). To handle the missing covariates we extended some of the ideas in Ibrahim, Chen & Lipsitz (2002) to time-varying covariates. Their approach uses parametric models for the distribution of covariates by factoring the joint distribution into a sequence of one-dimensional conditional distributions to accommodate missing covariates that are continuous or categorical or both. See Appendix **B** for details on the covariates model and for the computational algorithm for data augmentation. In the rest of the manuscript, we will refer to the model for the missing covariates as the X model. We mention that intermittent missing data is computationally easy to handle under MAR in a data augmentation setup; intermittent missingness can create some computational problems within EM and Newton-Raphson maximization approaches (see, e.g., Pourahmadi, Daniels & Park, 2006).

2.3 Further model features

The joint model specified in (1)-(3), allows for gains in efficiency by being able to model all J responses together, potentially allowing covariate effects to be the same across responses as β is not indexed by j (response) in (1). However, allowing for all the covariate effects to differ by response can be accommodated through the design vector, X_{itj}, which can be specified such that $X_{itj}\beta \equiv X^{\star}_{it}\beta_j$. The specification of (1) also permits lagged values of covariates to be included in the model as necessary.

Similarly, in (2), allowing distinct transition parameters for each response j can be done through appropriate specification of the design vector $C_{itj,m}$ or by reparameterizing the model as $\alpha_j C^{\star}_{it,m}$.

The number of parameters in an MTREM(p) model is $[(p+1)*(J-1)] + [T] + [r3*(T-p)*p + r3*(p-1)] + [p*r1 + r2] + [n*T]$, where $r1$ is the number of parameters in marginal regression models, (1) for the initial state(s) model, $r2$ is the number of regression parameters in the MTREM(p) excluding the initial state(s) model regression parameters, and $r3$ is the number of columns in $C_{itj,m}$ in (2). The terms in square brackets correspond to the λ, σ, α, β and b, respectively.

Through our specification, we have accounted for a complex dependence structure and implemented an efficient sampling algorithm that at most requires evaluation of one-dimensional integrals. We assess the adequacy of this specification using the methods discussed in the next subsection.

2.4 Assessing model fit

2.4.1 Deviance Information Criterion

The Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & van der Linde 2002) can be used to choose among differing specifications of (2); for example, the order (p) or which covariates to include in the C matrix. We define the parameters, θ, as all the parameters in (1)-(3) and y_{obs} as the observed response data. Specifically, let θ be $(b_{it}, \beta, \beta^*, \alpha, \lambda_j, \lambda_j^*, \sigma_t)$. Define $Dev(\theta) = -2logLik(\theta|y_{obs})$, $\bar{\theta} = E[\theta|y_{obs}]$, and $\overline{Dev} = E[Dev(\theta)|y_{obs}]$. The DIC is formally defined as Dev$(\bar{\theta})$ + 2 p_D, where $p_D = \overline{Dev} - Dev(\bar{\theta})$. The MTREM(1) likelihood terms for subjects (i) with no missing responses takes the form,

$$\left[\prod_{t>1,j} \frac{\exp((\Delta^{\star}_{itj} + \lambda_j b_{it}) y_{itj})}{1 + \exp(\Delta^{\star}_{itj} + \lambda_j b_{it})}\right] \prod_{j} \frac{\exp((\Delta^{\star}_{i1j} + \lambda^*_j b_{i1}) y_{i1j})}{1 + \exp(\Delta^{\star}_{i1j} + \lambda^*_j b_{i1})}. \quad (6)$$

For individuals with missing responses, the likelihood will take a slightly different form. As an example, consider $T = 4$ and that the responses are observed at times 2 and 4, but missing at time 3. The likelihood term for this subject would replace the term in brackets in (6) with

$$\prod_{j} \sum_{y_{i3j}=0}^{1} [\frac{\exp((\Delta^{\star}_{i4j} + \lambda_j b_{i4}) y_{i4j})}{1 + \exp(\Delta^{\star}_{i4j} + \lambda_j b_{it})} \frac{\exp((\Delta^{\star}_{i3j} + \lambda_j b_{i3}) y_{i3j})}{1 + \exp(\Delta^{\star}_{i3j} + \lambda_j b_{i3})}] \prod_{j} \frac{\exp((\Delta^{\star}_{i2j} + \lambda_j b_{i2}) y_{i2j})}{1 + \exp(\Delta^{\star}_{i2j} + \lambda_j b_{i2})}. \quad (7)$$

Note that in (7), $\Delta^{\star}_{i4j}$ depends on y_{i3j}.

The missing covariates, which we did **not** include as "parameters", necessitates a further modification to the form of the DIC through the observed data likelihood. In particular, we approximated the integration over the missing covariates by sampling a values from the X model and replacing the likelihood above with $Lik(\theta) = \frac{1}{a}\sum_{x_{mis}} Lik(\theta|x_{mis})$; for details, see Appendix **B**.

2.4.2 Posterior Predictive Checks

Posterior predictive checks (Gelman, Carlin, Stern & Rubin 2003) can be used to check if the model had adequately captured the correlation structure of the responses. We conduct these checks using the following steps. For each iteration of the Gibbs sampler, we obtain a realization of θ and y_{mis} (from the data augmentation step). Given θ we draw a replicated full dataset, y_{rep} from the posterior predictive distribution conditional on θ; this involves sampling from independent Bernoulli distributions with probabilities a function of θ. We compute two sets of $q = 1, \ldots, Q$ statistics, $S_{rep,q}(\cdot)$ and $S_{obs,q}(\cdot)$ at each iteration. The first set are evaluated at y_{rep} and the second at the observed data y_{obs} supplemented with the current sampled values of y_{mis} from the data augmentation step (to obtain a complete dataset). The specific statistics, $S_{rep,q}(\cdot)$ and $S_{obs,q}(\cdot)$ used were the log odds ratio (LOR) evaluated for the $Q = \binom{TJ}{2}$ pairs $(y_{itj}, y_{it'j'})$ to assess the reasonableness of the association structure given in (2) and (3).

As a more parsimonious measure than the entire posterior predictive distributions, we also compute posterior predictive p-values, the proportion of the association measures based on replicated responses that were larger than or equal to the associations based on observed data (supplemented with the replicated missing responses), given as $\frac{number(S_q(y^l_{rep}) \geq S_q(y_{obs}, y^l_{mis}))}{L}$, where L is the total number of iterations, y^l_{rep} is the realization at iteration l, $l = 1, ..., L$, y^l_{mis} is the realization of the missing data from the data augmentation step, and $q = 1, \ldots, Q$ indexes the specific LOR. P-values less than 0.01 or larger than 0.99 were reported to be suspicious.

3. MOTIVATING EXAMPLE: IOWA YOUTH AND FAMILIES PROJECT (IYFP)

Briefly introduced in Section 1, the IYFP explores emotional distress of teenagers in Iowa families. This section provides more information on this example and includes details on missing data.

3.1 Data

The Iowa Youth and Families Project was a longitudinal study which began in 1989 to help understand the effect of economic hardship and social changes on family members and to help improve family life in Iowa during such changes. Economic stress, such as that experienced in rural parts of Midwest during 1980's, and negative life events were expected to be related to emotional distress.

Data used in this paper was part of 4-year follow up of 451 families from eight counties in north central Iowa. Targets were selected to be seventh graders, with an average age of about 13 years at the start of the project, with two married biological parents and with a sibling within four years of age.

Three response variables of interest, anxiety, hostility and depression, were measured using a symptom checklist. Responses were dichotomized as to whether the target was feeling at least one of the physical symptoms of distress. Some of the symptoms for distress included nervousness or shakiness, an urge to break things, or feeling low in energy.

Conger, Elder, Lorenz, Simons & Whitbeck (1994) stated that the farm crisis of 1980s had long term effects on these families in terms of relationships and individual emotional status. We try to account for this impact in the model by introducing random effects which induce dependence between the three responses during each year and represent an 'overall' measure of distress; see

Level 3, (3). We allow these random effects to vary over time to reflect potentially unmeasured factors which impact the components of distress for the targets. Conger *et al.* (1994) also concluded that economic hardship caused daily hassles and hence distress on parents, which in turn affected the well-being of their children through harsh parenting. The dependence in distress in the targets (children) over time is modelled via the direct effects of the previous year(s) response for each response individually; see Level 2, (2). We expect most of the correlation to result from the same response at the previous year (or two). However, we pointed out in Section 2 that there is some indirect cross-response temporal dependence induced by the model. There are alternative ways to characterize this dependence. We will discuss these in Section 5.

To help explain the emotional distress, information from targets and their parents were collected. From the targets, this included their gender and whether they experienced any negative life events (NLE) in the last 12 months, such as having a close friend move away. Household information included whether there were any negative economic events (NEE), such as changing jobs for a worse one, and whether they needed to have cutbacks (Cutbacks), such as taking on a second (part-time) job to help meet living expenses or changing residences to save money. Of interest was the effect of these factors on the emotional distress of the cohort (population) as measured through the anxiety, hostility, and depression outcomes. These questions are addressed by our model by directly modelling the marginal (i.e., population level) effects of these covariates, Level 1, (1). For more information about this project, see Elder & Conger (2000).

3.2 Missing responses and covariates

There were no missing data at baseline (1989). The percentage of individuals (i) with at least one year of missing responses or covariates was 15%, so a complete case analysis of this dataset would require the removal of 15% of subjects. There were both intermittent missingness and dropouts. Of all the measured covariates (see Table 1), three of the time-varying covariates, NLE, NEE, and cutbacks, were occasionally missing. The percentage of missingness in NLE, NEE, and cutbacks was 7.3%, 7.9%, and 8.6%, respectively, over all years.

Earlier work by Lorenz, Simons, Conger, Elder, Johnson & Chao (1997) reported that there were no differences in measured covariates or observed responses between those who dropped out and those who completed the study. In addition, discussions with the investigators who collected the data led us to believe that missingness was mostly related to relocation of subjects (such as due to a job change) and was not related to subjects being more/less distressed. Therefore, we assumed the missing responses were missing at random (MAR). However, with this full-likelihood based approach, we can also evaluate results under MNAR mechanisms using sensitivity analyses (Verbeke, Molenberghs, Thijs, Lesaffre & Kenward 2001; Daniels & Hogan 2000).

4. RESULTS

This section illustrates MTREM(1) and MTREM(2) on the IYFP data. Inferences using the MTREMs are also compared with inferences from simpler models.

4.1 Sampling algorithm details

Initial estimates for the sampling algorithm for the marginal regression coefficients in (1) were obtained by fitting independent logistic regression models in SAS. To assess convergence, we ran several chains with different starting values. All the chains converged to the same region of the parameter space.

To deal with missing covariates, we first ran the X model to impute 1,000 sets of covariates, and randomly picked five of these. Then, we ran five separate Gibbs sampling algorithms for the Y model (the Y model refers to the MTREM(p) given in Section 2, conditional on a full set of covariates), each for 2,100 iterations and discarded the first 100 in each chain as burn-in. The use of Hybrid MC facilitated fast convergence. Every fifth sample was retained to avoid autocorrelation problems.

4.2 Prior distributions

We specified diffuse proper priors for β, β^* and α; multivariate normal distributions with means of 0 and large variances $\sigma_\beta^2 I$, $\sigma_{\beta^*}^2 I$ and $\sigma_\alpha^2 I$ respectively. More informative priors were specified for λ and λ^*, normal distributions with mean 1 and variance 2. The priors were centered at the value 1 corresponding to 'equal correlation' among the responses at a given time, see (3). We chose the variance of 2 to reflect a (weak) prior belief towards equicorrelation among the J responses; however, the results were not very sensitive to the specification of the variance. The priors for σ_t^2 were proportional to $\frac{1}{(1+\sigma_t^2)^2}$, which places positive probability at $\sigma_t^2 = 0$ (no multivariate dependence, see (3)) and is on a similar scale to λ_j (the prior has a median of 1).

These were the specific priors chosen for our analysis of the IYFP data using the MTREM. Other choices of priors can be made and will not further complicate the sampling algorithm.

4.3 Models

Individual-level covariates, X_{itj} included in the marginal mean in (1) for all four models are given in Table 1. Experts were consulted to assist in choosing these covariates. To avoid multicollinearity, some of the variables collected were not included into model. For example, material needs, which is another economic pressure measure, was highly correlated with cutbacks, and it was excluded.

To check the exogeneity assumption for the time-varying covariates (negative life and economic events and cutbacks), we regressed them on the history of those covariates and previous responses, adjusting for baseline covariates (gender) and time. We found that none of the time varying covariates were predicted by responses; in particular, the confidence intervals for all the odds ratios covered one.

Ultimately, we fit and compared four models. Models I and II were MTREM(1)'s and Models III and IV were MTREM(2)'s. In Models I and III, the transition parameters were the same across responses (j), so $C_{itj,m} = C_{it,m}$ which implies $\gamma_{itj,m} = \gamma_{it,m}$. Models II and IV removed this restriction and included indicators of the type of response (j) in $C_{itj,m}$, effectively allowing a separate 'transition' (dependence) parameter for each type of response. Specifically, for Models I and III, $C_{it,m} = 1$, whereas, for Models II and IV, $C_{itj,m} = [1, Resp1_{itj}, Resp2_{itj}]$, where $m = 1$ for Model II and $m = 1, 2$ for Model IV.

4.4 Model Fit

The DIC values for the models I through IV were 3487, 3572, 3452, 3409 respectively. Model IV had the smallest DIC indicating the best fit among the four models. The corresponding p_D values are 618, 653, 596, 576 respectively. We point out that Models III and IV, which on the surface seem to have more parameters than Models I and II (see section 2.3), have fewer effective number of parameters, i.e., smaller p_D's. This reduction in the effective number of parameters in Models III and IV was most likely a by-product of adding the 2nd order dependence ($p = 2$) in (2) in our data example; this resulted in more explained variability and thus, fewer effective number of random effects.

To further examine the better fit of Model IV (vs. Model III), Model IV had important parameters not included in Model III. For example, the dependence parameters from regressing the responses in 1991 ($t = 3$) on the two previous years' responses, differed between whether the response was depression or anxiety; in Table 5, see $\alpha_{3,1;2}$ which has a posterior mean of 1.26 and a 95% credible interval of (.15,2.6), which excludes zero [find this notation defined in the caption of Table 5].

Posterior predictive checks for all four models were carried out by methods described in Section 2.4.2. Figures and tables are not shown due to space limitation. The p-values and the distributions of some of the the statistics for the posterior predictive checks indicated some problems in model fit, even for the best model, Model IV. In particular, the model appeared to have some trouble capturing the temporal correlation, especially with the hostility and anxiety responses at the later times. Four out of eighteen p-values calculated for associations in the second level of Model IV were outside (.01,.99). For example, the distribution of the log odds ratios for the probability of hostility in 1991 versus 1990 indicated that the observed (supplemented with the missing responses at each iteration) log odds ratios were consistently higher than the predicted ones under the model. However, even for the the most extreme discrepancies, as quantified by the posterior predictive p-values, the actual log odds ratios were never far apart in relative magnitude. We will discuss possible model expansions to account for this lack of fit in Section 5. The third level of the model captured the correlation structure of data adequately with p-values ranging from 0.104 to 0.603.

4.5 Parameter estimates and interpretations

Results from Model IV, the best fitting model, are given in Tables 2-5. Note that 95% interval estimates are not necessarily symmetric around the posterior mean as the (Bayesian) inference here does not rely on large-sample theory. In the next subsection, we report on inference on the quantities of interest as outlined in Section 3 using Model IV.

4.5.1 Inference on quantities of interest

At baseline (1989), the parameters corresponding to gender, negative life events, cutbacks, and negative economic events were not significant. However, the coefficients that allowed the probabilities for each of the three types of responses to differ were significant. Specifically, $\exp(\beta^*_{Resp1})$ had a posterior median of $\exp(-2.13) = 0.12$ with a 95% credible interval of $(0.03, 0.43)$, and $\exp(\beta^*_{Resp2})$ had a posterior median of $\exp(-1.65) = 0.19$ with a 95% credible interval of $(0.05, 0.71)$. These indicated that the probability of a subject being depressed was higher than having either anxiety or hostility.

In 1990, the effect of gender on distress was significant and differed among the three measures of distress. The odds of depression for females were $\exp(1.2) = 3.32$ times higher than males, with a 95% credible interval of $(1.7, 6.3)$. Moreover, females were $\exp(1.2-0.9) = 1.35$ times more likely to feel anxious, and $\exp(1.2 - 0.88) = 1.38$ times more likely to feel hostile compared to males. There was also a significantly lower probability of distress among individuals with no negative life events (posterior medians of $\exp(\tilde{\beta}_{NLE1})$ and $\exp(\tilde{\beta}_{NLE2})$ were 0.38 and 0.18, with credible intervals $(0.13, 1.01)$ and $(0.07, 0.43)$ respectively). The effects of negative economic events and cutbacks were not significant.

In the last two years (1991 and 1992), the parameter corresponding to gender ($\exp(\beta_{gender})$) had a posterior median of 2.69 and a 95% credible interval of $(1.68, 4.31)$ again indicating females were more likely to feel depressed. Similarly, females were 1.84 and 1.48 times more likely to feel anxious and hostile, respectively. The coefficients for negative life events, $\exp(\beta_{NLE1})$ (posterior median of 0.47 and 95% credible interval of $(0.21, 0.97)$) and $\exp(\beta_{NLE2})$ (posterior median of 0.33 and 95% credible interval of $(0.19, 0.53)$) indicated that distress was significantly lower for individuals with no negative life events. The effects of negative economic events and cutbacks were not significant.

4.5.2 General observations

Marginal covariate effects

The marginal covariate effect estimates ($\hat{\beta}$ and $\hat{\beta}^{*}$) under Model II were close to the estimates under Model I; the same was observed with Model III and Model IV estimates (note that only results for Model IV are shown in the tables; results for Models I, II and III are available from the authors upon request). That is, the parameter estimates for (1) appeared relatively insensitive to the change in the parameters in (2). Heagerty (2002) proved that β and α are orthogonal in MTM(1)'s, and hence marginal parameter estimates are consistent regardless of possible misspecification of second level model (although posterior standard deviations or credible intervals will be incorrect). Since the first two levels of MTREM(1) are essentially MTM(1)'s, we might expect similar robustness. This was also supported by work by Heagerty & Kurland (2001) who demonstrated via simulations that marginalized models result in smaller bias in estimates of marginal regression coefficients. Such robustness for MTREMs will be explored in more detail in future work.

Dependence

There was clearly strong, and differential temporal and cross-response dependence in both Levels 2 and 3 (see Table 5). We provide some examples below.

Level 2 dependence:

The parameters that capture the temporal dependence, α, changed slightly from one year to another, were significant at lags of two years, and were significantly different from zero overall. For example, for depression, the coefficient for regressing depression in 1992 ($t = 4$) on depression in 1990 ($t = 2$), ($\alpha_{4,2;1}$) had a posterior mean of 0.71 and a 95% credible interval of $(.36, 1.11)$ indicating strong temporal dependence even at lag two. In addition, the difference between the lag two transition parameter in 1991 for hostility versus depression ($\alpha_{3,1;3}$) was significant with a posterior mean of 1.27 and a credible interval of $(.12, 2.49)$, indicating stronger temporal correlation for hostility.

Level 3 dependence:

$\lambda^{\star}$ and $\tilde{\lambda}$ showed no evidence of the pairwise associations differing between anxiety, hostility, and depression (all credible intervals covered 1; see Tables 2-3). However, in 1991 and 1992, λ_3 (posterior mean of 2.1 and 95% credible interval of (1.2, 3.45)) indicated the association between hostility and depression was higher than the other pairwise associations (Table 5). In addition, the variability of the random intercepts at each time, σ_t^2, was smaller in 1991 and 1992 ($t = 3, 4$) than in 1989 and 1990 ($t = 1, 2$). This corresponds to a weaker association among the three responses at each time; to see this more clearly, refer back to the approximate form of the correlation, which is a function of λ_j and σ_t^2, in Section 2.1.

There was also some cross-response temporal association captured by the model (as discussed in Section 2.1). We computed posterior predictive p-values on these associations and found the model captured these associations well with only one p-value outside (.01,.99).

4.6 Comparison with MTM's

To assess the importance of modelling the multivariate dependence between the three responses, we also fit several MTM models (i.e., MTREM's without Level 3). In particular, we fit Model II and Model IV without Level 3. We also fit independent MTM(1) and MTM(2) for each of the three responses separately. The DIC of all these models suggested a much poorer fit; the DIC values ranged from 4149 for the Model II MTM to 4324 for independent MTM(1)'s. In addition, the posterior predictive p-values for these models were much more extreme than the results for the MTREM's even for the within response temporal log odds ratios.

The posterior means and standard deviations for the three independent MTMs (one for each of anxiety, hostility, and depression), along with the Model IV results, are given in Tables 2-5. We point out that there were identifiability problems for the initial state model for the MTM for depression with several covariates. As a result, the results of the initial state model for depression were not included in Table 2.

There were considerable differences in the posterior mean and standard deviations between the Model IV MTREM and the independent MTMs. For example, the posterior standard deviation for negative life events coefficients (NLE1 and NLE2) for 1989 (Table 2) were considerably smaller in the MTREM which pooled these coefficients across the three responses. Similar results were seen for the other coefficients and for the other years. In addition, there were considerable decreases in the posterior standard deviations of the dependence parameters in (2). For example, in Table 3, the coefficient for depression, α_{21} was 50% smaller in the MTREM versus the independent MTM for depression. Similar results were seen for the depression association parameters in Table 5.

These large differences in posterior standard deviations provides further justification for fitting the more complex MTREM beyond its ability to more accurately capture the dependence in this dataset as reflected in the DIC and the posterior predictive checks.

5. CONCLUSIONS AND DISCUSSION

We have proposed a full-likelihood based approach for multivariate longitudinal binary responses that directly models marginal means as function of covariates while accounting for longitudinal and multivariate correlation. Temporal (longitudinal) dependence was included via transition models and cluster (multivariate) correlation from the multivariate response at each time was captured via random effects. The model allowed for both baseline and exogenous time-varying covariates. Smaller estimates of variability were expected by modelling the binary longitudinal responses jointly and by allowing effects/parameters to be shared across responses. This can be seen by comparing the MTREM results to the independent MTM's.

Intermittent and dropout missing values in responses were handled by data augmentation and missing time-varying covariates were modeled using a parametric approach that accounted for the time-varying structure of the covariates.

Calculations and analyses in this paper were introduced for $p = 1$ and $p = 2$, i.e., first and second order Markov models. Extension to higher orders is possible. Another interesting extension would be to introduce temporal dependence in the random effects ($b_i \sim N(0, \Sigma)$, where $b_i = (b_{i1}, \ldots, b_{iT})^T$) and/or cross-temporal dependence between responses (e.g., extend Level 2 for $p = 1$ to $P(Y_{itj}|Y_{i,t-1,j}, Y_{i,t-1,j'})$). This might help to capture the temporal correlation not captured by the current model. Another possibility would be to remove (2) and the λ_j and introduce $J \times T$ vectors b_i in (3) with distribution $b_i \sim N(0, \Sigma_J \otimes \Sigma_T)$. However, the efficient sampling algorithm proposed here would require the evaluation of high-dimensional integrals. Such models are beyond the scope of this paper and left for future work both in terms of exact model specification and development of efficient algorithms to sample from the posterior distribution. A final extension/addition to the model would be to introduce a random intercept, b_{ij} in (2) to better account for long range temporal dependence within response (Schildcrout and Heagerty, 2005) or to assume the time-specific random intercept in (3), b_{it} is fixed over time, b_i.

The first author is in the process of preparing the complex code for fitting these models via a Fortran program to be posted at *www.stat.ufl.edu/∼mdaniels/research.html*.

APPENDIX

A: Details on sampling from posterior distribution of parameters in MTREM(1)

Computational Algorithm for p=1

Let θ be the vector of all parameters in the model, i.e. $(b_{it}, \beta, \beta^*, \alpha, \lambda_j, \lambda_j^*, \sigma_t)$. Also, let $\theta_{-\delta}$ be the vector of all parameters except δ. The following algorithm was used to obtain a sample from the posterior distribution of θ.

1. Assign starting values to θ.
2. Calculate Δ_{itj} from the marginal constraint equation (4) using Newton-Raphson.
3. Calculate Δ^*_{itj} from the convolution equations (5) using Newton-Raphson and Gauss-Hermite

Quadrature.

4. Sample from the full conditional for $(log(\sigma_t^2), \lambda_j, b_{it})$ using the following two steps:

a. Integrate out b_{it} by Gauss-Hermite Quadrature (1-dimensional integration), and sample $(log(\sigma_t^2), \lambda_j | \theta_{-(b,\sigma,\lambda)}, Y)$ by using Hybrid MC. Update Δ^*_{itj} for $t \geq 2$.

b. Sample from the full conditional of b_{it} using Hybrid MC.

5. Sample from the full conditional for $(log(\sigma_1^2), \lambda_j^*, b_{i1})$ using the following two steps:

a. Integrate out b_{i1} by Gauss-Hermite Quadrature (1-dimensional integration), and sample $(log(\sigma_1^2), \lambda_j^* | \theta_{-(b,\sigma,\lambda^*)}, Y)$ using Hybrid MC. Update Δ^*_{i1j}.

b. Sample from the full conditional of b_{i1} using Hybrid MC.

6. Sample from the full conditional of β using Hybrid MC. Update Δ_{itj} and Δ^*_{itj} for $t \geq 2$.
7. Sample from the full conditional of β^* using Hybrid MC. Update Δ_{i2j}, Δ^*_{i1j}, and Δ^*_{i2j}.
8. Sample from the full conditional of α using Hybrid MC. Update Δ_{itj} and Δ^*_{itj} for $t \geq 2$.
9. Repeat steps 4 through 8. Continue until convergence.

Steps 4-8 constituted the Gibbs sampling algorithm to sample from posterior distribution of θ. Convergence was assessed by running multiple chains and examining trace plots of a subset of the parameters.

The calculation of Δ_{itj} and Δ^*_{itj} as a function of the other parameters and the corresponding derivatives was quite involved and can be found at the following web page: *www.stat.ufl.edu/∼mdaniels/research.html* or in Ilk (2004). The full conditional distributions used in the above algorithm and the derivatives needed for the Hybrid MC algorithm can also be found there.

B: Details on missing responses and covariates in MTREM(1)

In the IYFP data, we had missingness in the following covariates: X_{itj3}, X_{itj4}, X_{itj5}, X_{itj6}, X_{itj7} (see Table 1). To deal with missing covariates, the likelihood was first partitioned as $P(Y|X,\theta)P(X_{mis}|X_{obs},\psi)$ and inferences for θ were based on

$$p(y|x_{obs},\theta) = \int p(y|X,\theta)p(X_{mis}|X_{obs},\psi)dX_{mis} \approx \frac{1}{a}\sum_{x_{mis}} p(y, x_{mis}|X_{obs},\theta)$$

where a values of X_{mis} were sampled from $p(x_{mis}|x_{obs},\psi)$. So, we are using a 'multiple imputation' approach to approximate this integral. In principle, there could be gains in efficiency by using the information in Y to inform about X, but this was not practical computationally (see Ilk, 2004).

Computational Algorithm for Data Augmentation:

As previously, let θ be the parameters in the Y model, $p(y|x,\theta)$ and ψ be the parameters related to joint distribution of covariates, $p(X_{mis}|X_{obs},\psi)$. The steps of the algorithm are:

1. Set k=0. Set starting values for $Y_{itj,mis}, X_{itj,mis}, \psi, \theta$.
2. Run the X model, $X_{itj,mis}|X_{itj,obs},\psi$, for 1000 iterations, and randomly select 5 sets of X matrices.

Specifically, repeat these two steps, a. and b., for 1000 iterations after burn-in:

a. Sample $X^{(k+1)}_{itj,mis}$ from its full conditional (see below for details).

b. Sample from the full conditional distribution of $\psi^{(k+1)}$ by Hybrid MC.

3. Sample $Y_{itj,mis}|X_{itj}, Y_{itj,obs}, \theta$ by the following steps:

a. Set $l = 2$.

b. Calculate Δ^*_{ilj} (which uses Y_{il-1j}) and $p_{ilj} = P(Y^{(k+1)}_{ilj,mis} = 1|Y_{imj}; m \neq l; \theta)$. If the missing cases are dropouts, then $p_{ilj} = \frac{e^{\Delta^*_{ilj}+\lambda_j b_{il}}}{1+e^{\Delta^*_{ilj}+\lambda_j b_{il}}}$. If they are intermittent missing, then p_{ilj} is

$$\left(\frac{e^{\Delta^*_{ilj}+\lambda_j b_{il}}}{1+e^{\Delta^*_{ilj}+\lambda_j b_{il}}}\frac{e^{y_{il+1j}(\Delta^*_{il+1j}+\lambda_j b_{il+1})}}{1+e^{\Delta^*_{il+1j}+\lambda_j b_{il+1}}}\right)/\left(\frac{e^{\Delta^*_{ilj}+\lambda_j b_{il}}}{1+e^{\Delta^*_{ilj}+\lambda_j b_{il}}}\frac{e^{y_{il+1j}(\Delta^*_{il+1j}+\lambda_j b_{il+1})}}{1+e^{\Delta^*_{il+1j}+\lambda_j b_{il+1}}}+\frac{1}{1+e^{\Delta^*_{ilj}+\lambda_j b_{il}}}\frac{e^{y_{il+1j}(\Delta^*_{il+1j}+\lambda_j b_{il+1})}}{1+e^{\Delta^*_{il+1j}+\lambda_j b_{il+1}}}\right).$$

c. Generate u from Uniform(0,1). If $u \leq p_{ilj}$, then set $Y_{ilj,mis}$=1, otherwise, set $Y_{ilj,mis} = 0$.

d. Set $l = l + 1$. Go back to step 3b. Repeat until responses at all time points are imputed.
4. Sample from the full conditional distribution of $\theta^{(k+1)}$. This is the Gibbs sampling step described in Section A.
5. Set $k = k + 1$. Go back to step 3. Continue until convergence.

Model for the missing exogenous time-varying covariates

At baseline, there were no cases with missing covariates. For t>1, Ibrahim *et al.*'s (2002) work was extended to include the previous time covariate information. We describe this next. For simplicity of notation, the indices i and j were suppressed. Let $X_t = (X_{t2}, X_{t3}, X_{t4}, X_{t5}, X_{t6}, X_{t7}, X_{t10}, X_{t11})$ and $X_{t-1} = (X_{t-1,3}, X_{t-1,4}, X_{t-1,5}, X_{t-1,6}, X_{t-1,7})$. Note that $X_{t-1,2}$, $X_{t-1,10}$ and $X_{t-1,11}$ are absent in the vector X_{t-1}. These covariates are either baseline covariates and/or indicators of time and thus implicitly are already included in X_t.

Under a Markov assumption, the joint distribution of covariates can be factored as

$$P(X_2, \ldots, X_T|\psi) = P(X_3|X_2, \psi)P(X_4|X_3, \psi) \cdots P(X_T|X_{T-1}, \psi)$$

Then, each component can be factored as

$$P(X_t|X_{t-1}, \psi) = P(X_{t3}, X_{t4}|X_{t2}, X_{t,10}, X_{t,11}, X_{t-1}, \psi_3)P(X_{t5}|X_{t2}, X_{t3}, X_{t4}, X_{t,10}, X_{t,11}, X_{t-1}, \psi_5)$$
$$*P(X_{t6}, X_{t7}|X_{t2}, X_{t3}, X_{t4}, X_{t5}, X_{t,10}, X_{t,11}, X_{t-1}, \psi_6) \qquad (8).$$

To sample from the joint distribution of X_{itj}, the following method was used. Since missing covariates are observed only on NLE (3 levels), NEE (2 levels), and cutbacks (3 levels), we have 3*2*3=18 possible sets for the missing covariate. We explicitly calculated the probability of each of these 18 possible combinations for each individual and sampled one of them; note, $X_{itj,mis} = X_{itj',mis}$ for all j and j'.

For the missing covariates with three 'ordered' levels (i.e. NLE and cutbacks), we used proportional odds models. For the missing covariates with two levels, we used logistic regression models. For the form of these regressions, see the web appendix. Priors for ψ were assumed to be independent and identically distributed normals with 0 mean and large variance.

The order of conditioning in (8) is arbitrary and may impact inferences. Ibrahim *et al.* (2002) examined the sensitivity to ordering and concluded that posterior inferences of ψ are generally quite robust to the ordering. For this analysis, there were five other possible orderings. Sensitivity of inferences to this ordering is left for future work.

C: Second order model, MTREM(2)

Some details on the initial states model ($t = 1, 2$) for the MTREM(2) are given here. We assumed a marginally specified logistic-normal model for $t = 1$, as we did for the MTREM(1), and an MTREM(1) for $t = 2$. Specifically, for $t = 1$, we assume

$$\text{logitP}(Y_{i1j} = 1|X_{i1j}) = X_{i1j}\beta^*$$
$$\text{logitP}(Y_{i1j} = 1|X_{i1j}, b_{i1}) = \Delta^*_{i1j} + \lambda^*_j b_{i1}$$

For $t = 2$, we specify the following model,

$$\text{logitP}(Y_{i2j} = 1|X_{i1j}, X_{i2j}) = X_{i2j}\widetilde{\beta}$$
$$\text{logitP}(Y_{i2j} = 1|y_{i1j}, X_{i1j}, X_{i2j}) = \Delta_{i2j} + \widetilde{\gamma}_{i2j,1}y_{i1j}$$
$$\text{logitP}(Y_{i2j} = 1|y_{i1j}, X_{i1j}, X_{i2j}, b_{i2}) = \Delta^*_{i2j} + \widetilde{\lambda}_j b_{i2}$$

λ^*_1 and $\widetilde{\lambda}_1$ are again set to be 1 for identifiability. Details for fitting these are given in the web appendix. Note that we allow the coefficients for these two component models, $\beta^\star$ and $\tilde{\beta}$ to differ from β for reasons similar to those given in the discussion of MTREM(1) in Section 2.2.

ACKNOWLEDGEMENTS

The data used in this paper was provided by Institute for Social and Behavioral Research, ISU Research Park, Ames. We would like to thank Becky Burzette, Ph.D. and Frederick O. Lorenz, Ph.D. for their help in data collection and clarifying some issues with data. Daniels was partially supported by NIH grants CA-85295 and HL-079457. This paper was based on a chapter in Ilk's PhD thesis at Iowa State University.

REFERENCES

A. Agresti (1997). A model for repeated measurements of a multivariate binary response. *Journal of the American Statistical Association*, 92, 315–321.

K. Bandeen-Roche, D. L. Miglioretti, S. L. Zeger, & P. J. Rathouz (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92, 1375-1386.

R. D. Conger, G. H. Elder, O. F. Lorenz, R. L. Simons, & L. B. Whitbeck (1994). *Families in Troubled Times: adapting to change in rural times.* Aldine De Gruyter, New York.

M. J. Daniels, & J. W. Hogan (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics*, 56, 1241–1248.

P. J. Diggle, P. Heagerty, K-Y. Liang, & S. L. Zeger (2002).*Analysis of Longitudinal Data, 2nd edition.* Oxford University Press, New York.

G. H. Elder, & R. Conger (2000). *Children of the Land.* The University of Chicago Press.

G. M. Fitzmaurice, & N. M. Laird (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80, 141–151.

A. Gelman, J. B. Carlin, H. S. Stern, & D. B. Rubin (2003). *Bayesian Data Analysis, 2nd edition.* Chapman and Hall Ltd. London; New York.

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

H. Goldstein & J. Rasbash (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series B*, 159, 505–513.

P. J. Heagerty (1999). Marginally Specified Logistic-normal Models for Longitudinal Binary Data. *Biometrics*, 55, 688–698.

P. J. Heagerty & B. F. Kurland (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88, 973–985.

P. J. Heagerty (2002). Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data. *Biometrics*, 58, 342–351.

J. G. Ibrahim, M-H. Chen , & S. R. Lipsitz (2002). Bayesian methods for generalized linear models with covariates missing at random. *The Canadian Journal of Statistics*, 30, 55–78.

O. Ilk (2004). *Exploratory multivariate longitudinal data analysis and models for multivariate longitudinal binary data.* PhD thesis, Iowa State University.

J. M. Legler, & L. M. Ryan (1997). Latent variable models for teratogenesis using multiple binary outcomes. *Journal of the American Statistical Association*, 92, 13–20.

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

F. O. Lorenz, R. L., Simons, R. D. Conger, G.H. Elder, C. Johnson & W. Chao (1997). Married and Recently Divorced Mothers' Stressful Events and Distress: Tracing Change Across Time. Journal of Marriage and the Family, 59, 219–232.

D. L. Miglioretti, & and P. J. Heagerty (2004). Marginal Modeling of Multilevel Binary Data with Time-Varying Covariates. *Biostatistics*, 5, 381–398.

R. M. Neal (1996). *Bayesian learning for neural networks.* Springer-Verlag, New York.

M. Pourahmadi, & and M. J. Daniels (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58, 225–231.

M. Pourahmadi, & M. J. Daniels, & T. Park (2006). Simultaneous Modelling of the Cholesky Decomposition of Several Covariance Matrices with Applications. *Journal of Multivariate Analysis* (in press).

B. A. Reboussin, & J. C. Anthony (2001). Latent class marginal regression models for longitudinal data: Modeling youthful drug involvement and its suspected influences. *Statistics in Medicine*, 20, 623–639.

H. Ribaudo, & S. G. Thompson (2002). The analysis of repeated multivariate binary quality of life data: A hierarchical model approach. *Statistical Methods in Medical Research*, 11, 69–83.

J. M. Robins, A. Rotnitzky, L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846–866.

J. M. Robins, A. Rotnitzky, L. P. Zhao (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106–121.

J. Schildcrout & P. J. Heagerty (2005). Marginalized models for moderate to long series of longitudinal binary response data. *Technical Report.*

J. D. Singer, & J. B. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, & A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583–616.

M. A. Tanner, & W. H. Wong (1987). The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association, 82, 528–540.

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

G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, & M. G. Kenward (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics*, 57, 7–14.

Received ???
Accepted ???

Ozlem ILK: `oilk@metu.edu.tr`
Department of Statistics, Middle East Technical University
Ankara
Turkey, 06531

Michael J. DANIELS: `mdaniels@stat.ufl.edu`
Department of Statistics, University of Florida
Gainesville, Florida
United States of America, 32611

Table 1: Individual- and household-level covariates included in (1). p refers to the order of the MTREM

Variable	Notation	Description
Gender	X_{itj2}	1= male, 2 = female
NLE	X_{itj3},X_{itj4}	whether subjects experienced any negative life events in the last 12 months. (X_{itj3} : 1 if they didn't have any negative events; 0 o.w., and X_{itj4} : 0 if they had lots of negative events; 1 o.w.)
NEE	X_{itj5}	whether the household had any negative economic events (X_{itj5} : 0 = no, 1= yes)
Cutbacks	X_{itj6},X_{itj7}	whether the household had any financial cutbacks (X_{itj6}=1 if they had no cutbacks; 0 o.w. and X_{itj7}=0 if they had more than 5 cutbacks last year; 1 o.w.)
Response	X_{itj8}, X_{itj9}	Resp1 (X_{itj8}=1 if response=anxiety; 0 o.w.), Resp2 (X_{itj9}=1 if response=hostility; 0 o.w.)
Time	X_{itj10}, X_{itj11}	For p=1, Time1 (X_{itj10}=1 if Year=1991; 0 if Year=1990,1992), Time2 (X_{itj11}=1 if Year=1992; 0 if Year=1990,1991) For p=2, Time (X_{itj10}=1 if Year=1992; 0 if Year=1991)
NLE*NEE	X_{itj12}, X_{itj13}	interaction effects of NLE1*NEE and NLE2*NEE
Gender*Response	X_{itj14}, X_{itj15}	interaction effects of gender*Resp1 and gender*Resp2
Response*Time	X_{itj16}, X_{itj17}, X_{itj18}, X_{itj19}	interaction effects of Resp1*Time1, Resp1*Time2, Resp2*Time1, and Resp2*Time2

Table 2: Posterior summaries for parameters in Model IV and independent MTMs for baseline (1989). The MTM for the depression response did not converge so the results are not reported in the table. As a result, the posterior means and standard deviations for Resp1 and Resp2 and Gender*Resp1, Gender*Resp2 correspond to the anxiety and hostility specific coefficients, not the differences between these coefficients and the depression ones.

	Model IV					Anx		Host		Depr	
				posterior		posterior		posterior		posterior	
Baseline	2.5%	50%	97.5%	mean	SD	mean	SD	mean	SD	mean	SD
Inter	1.64	2.03	2.47	2.04	0.2					—	—
Resp1	-3.51	-2.13	-0.85	-2.13	0.68	1.75	0.15				
Resp2	-2.91	-1.65	-0.34	-1.63	0.66			1.71	0.14		
Gender	-1.48	-0.61	0.14	-0.62	0.41					—	—
Gender*Resp1	-0.04	0.72	1.5	0.72	0.39	0.1	0.27				
Gender*Resp2	-0.37	0.41	1.15	0.4	0.38			-0.17	0.27		
NLE1	-3.07	-1.14	0.69	-1.13	0.97	-1.85	1.48				
								-0.78	1.41		
										—	—
NLE2	-1.18	-0.5	0.17	-0.49	0.36	-0.88	0.47				
								-0.15	0.43		
										—	—
NEE	-0.42	0.31	1.03	0.31	0.36	0.04	0.47				
								0.32	0.41		
										—	—
Cuts1	-1.02	-0.37	0.24	-0.38	0.31	-0.44	0.35				
								-0.3	0.39		
										—	—
Cuts2	-0.19	0.3	0.76	0.3	0.25	-0.05	0.31				
								0.54	0.3		
										—	—
NLE1*NEE	-1.36	0.81	2.79	0.78	1.08	0.88	1.61				
								1.2	1.62		
										—	—
NLE2*NEE	-1.2	-0.32	0.52	-0.32	0.46	0.01	0.58				
								-0.48	0.54		
										—	—
$log(\sigma_1^2)$	0.72	1.36	2.13	1.39	0.37						
λ_2^*	0.73	1.46	3.23	1.58	0.61						
λ_3^*	0.73	1.37	3.24	1.5	0.63						

Table 3: Posterior summaries for parameters in Model IV and the independent MTMs for 1990. We again introduce new notation for the α parameters; α_{21} corresponds to regressing the response at time 2 on the previous response for depression, $\alpha_{21}+\alpha_{22}$ for anxiety, and $\alpha_{21}+\alpha_{23}$ for hostility.

	Model IV					Anx		Host		Depr	
				posterior		posterior		posterior		posterior	
Time=2	2.5%	50%	97.5%	mean	SD	mean	SD	mean	SD	mean	SD
Inter	1.29	1.61	1.95	1.61	0.17					2.01	0.18
Resp1	-0.17	0.76	1.62	0.76	0.45	-0.5	0.24				
Resp2	-0.23	0.66	1.5	0.65	0.45			-0.68	0.22		
Gender	0.53	1.2	1.84	1.19	0.33					1.2	0.31
Gender*Resp1	-1.5	-0.9	-0.26	-0.9	0.32	-0.87	0.4				
Gender*Resp2	-1.48	-0.88	-0.24	-0.87	0.31			-0.78	0.39		
NLE1	-2.07	-0.97	0.01	-0.98	0.52	-1.33	0.76				
								-1.54	0.75		
										-0.33	0.94
NLE2	-2.69	-1.7	-0.85	-1.72	0.47	-2.8	0.87				
								-1.21	0.51		
										-1.13	0.61
NEE	-1.45	-0.35	0.61	-0.37	0.53	-1.4	0.91				
								-0.08	0.61		
										0.32	0.75
Cuts1	-0.76	-0.26	0.26	-0.26	0.26	0.04	0.32				
								-0.57	0.31		
										-0.33	0.38
Cuts2	-0.28	0.19	0.66	0.19	0.24	0.03	0.3				
								0.4	0.3		
										0.1	0.36
NLE1*NEE	-0.9	0.34	1.66	0.34	0.64	1.11	0.92				
								0.55	0.88		
										0.39	1.15
NLE2*NEE	0.04	1.04	2.2	1.08	0.55	2.03	0.94				
								0.73	0.66		
										0.28	0.8
α_{21}	0.56	0.97	1.4	0.98	0.22					1.02	0.46
α_{22}	-1.1	-0.19	0.67	-0.19	0.45	-0.08	0.55				
α_{23}	-0.57	0.32	1.25	0.33	0.46			0.61	0.55		
$log(\sigma_2^2)$	0.52	1.27	2.14	1.29	0.41						
$\widetilde{\lambda}_2$	0.54	1	1.92	1.06	0.35						
$\widetilde{\lambda}_3$	0.82	1.66	3.86	1.85	0.78						

Table 4: Posterior summaries for parameters in (1) for Model IV and the independent MTMs for 1991 and 1992.

	Model IV					Anx		Host		Depr	
				posterior		posterior		posterior		posterior	
Time> 2	2.5%	50%	97.5%	mean	SD	mean	SD	mean	SD	mean	SD
Inter	1.18	1.41	1.63	1.41	0.11					2.03	0.13
Resp1	-0.78	0.03	0.72	0.02	0.38	-0.08	0.16				
Resp2	-0.38	0.37	1.04	0.35	0.36			-0.08	0.16		
Gender	0.52	0.99	1.46	0.98	0.24					0.99	0.23
Gender*Resp1	-0.86	-0.38	0.17	-0.37	0.26	-0.37	0.3				
Gender*Resp2	-1.06	-0.6	-0.12	-0.59	0.24			-0.59	0.3		
Time	-0.21	0.21	0.68	0.22	0.23					0.2	0.19
Time*Resp1	-0.85	-0.44	-0.03	-0.45	0.21	-0.44	0.24				
Time*Resp2	-0.88	-0.45	-0.03	-0.46	0.22			-0.43	0.24		
NLE1	-1.56	-0.75	-0.03	-0.75	0.39	-0.69	0.51				
								-1.12	0.5		
										-0.48	0.55
NLE2	-1.64	-1.12	-0.64	-1.13	0.25	-0.91	0.32				
								-1.17	0.34		
										-1.76	0.5
NEE	-0.83	-0.32	0.24	-0.32	0.27	-0.26	0.34				
								-0.45	0.36		
										-0.5	0.56
Cuts1	-0.43	-0.08	0.32	-0.07	0.19	-0.05	0.23				
								0.06	0.25		
										-0.59	0.29
Cuts2	-0.21	0.09	0.41	0.10	0.16	-0.09	0.2				
								0.05	0.2		
										0.65	0.26
NLE1*NEE	-0.51	0.57	1.65	0.56	0.55	1	0.75				
								0.85	0.67		
										-0.73	0.76
NLE2*NEE	-0.07	0.55	1.14	0.54	0.31	0.48	0.38				
								0.53	0.4		
										1.16	0.6

Table 5: Posterior summaries for parameters in (2) and (3) for Model IV and the independent MTMs for 1991 and 1992. Again, we introduce some new notation for the α for ease of interpretation in this table; $\alpha_{t,t^\star;j}$ is the regression for the 2nd order Markov model from regressing the response at time t on $t^\star$. The coefficient with $j = 1$ corresponds to depression while the sum of the coefficients for $j = 1$ and $j = 2$ is for anxiety and the sum of $j = 1$ and $j = 3$ for hostility.

	Model IV					Anx		Host		Depr	
				posterior		posterior		posterior		posterior	
Time> 2	2.5%	50%	97.5%	mean	SD	mean	SD	mean	SD	mean	SD
$\alpha_{3,2;1}$	0.84	1.23	1.63	1.23	0.2					1.43	0.38
$\alpha_{3,1;1}$	-0.46	0.07	0.54	0.07	0.25					-0.3	0.62
$\alpha_{4,3;1}$	0.57	0.95	1.31	0.95	0.19					1.12	0.41
$\alpha_{4,2;1}$	0.36	0.71	1.11	0.71	0.19					0.88	0.4
$\alpha_{3,2;2}$	-0.84	-0.02	0.75	-0.04	0.4	-0.22	0.48				
$\alpha_{3,1;2}$	0.15	1.23	2.6	1.26	0.62	1.04	0.7				
$\alpha_{4,3;2}$	-0.35	0.48	1.29	0.48	0.42	0.32	0.5				
$\alpha_{4,2;2}$	-0.73	0.05	0.85	0.07	0.41	0.44	0.5				
$\alpha_{3,2;3}$	-0.28	0.55	1.36	0.54	0.41			0.36	0.49		
$\alpha_{3,1;3}$	0.12	1.24	2.49	1.27	0.61			0.91	0.71		
$\alpha_{4,3;3}$	-0.56	0.24	1.04	0.23	0.42			0.15	0.51		
$\alpha_{4,2;3}$	-1.56	-0.78	0.05	-0.77	0.42			-0.55	0.51		
$log(\sigma_3^2)$	-0.1	0.59	1.24	0.59	0.34						
$log(\sigma_4^2)$	-0.15	0.55	1.18	0.53	0.34						
λ_2	0.94	1.45	2.53	1.53	0.42						
λ_3	1.2	2	3.45	2.1	0.59						