Essays on the Modeling of Binary Longitudinal Data

with Time-dependent Covariates

by

Elsa Aimara Vazquez Arreola

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2020 by the
Graduate Supervisory Committee:

Jeffrey R Wilson, Chair
Paul R Hahn
Ming-Hung Kao
Mark Reiser
Yi Zheng

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Longitudinal studies contain correlated data due to the repeated measurements on the same subject. The changing values of the time-dependent covariates and their association with the outcomes presents another source of correlation. Most methods used to analyze longitudinal data average the effects of time-dependent covariates on outcomes over time and provide a single regression coefficient per time-dependent covariate. This denies researchers the opportunity to follow the changing impact of time-dependent covariates on the outcomes. This dissertation addresses such issue through the use of partitioned regression coefficients in three different papers.

In the first paper, an alternative approach to the partitioned Generalized Method of Moments logistic regression model for longitudinal binary outcomes is presented. This method relies on Bayes estimators and is utilized when the partitioned Generalized Method of Moments model provides numerically unstable estimates of the regression coefficients. It is used to model obesity status in the Add Health study and cognitive impairment diagnosis in the National Alzheimer's Coordination Center database.

The second paper develops a model that allows the joint modeling of two or more binary outcomes that provide an overall measure of a subject's trait over time. The simultaneous modelling of all outcomes provides a complete picture of the overall measure of interest. This approach accounts for the correlation among and between the outcomes across time and the changing effects of time-dependent covariates on the outcomes. The model is used to analyze four outcomes measuring overall the quality of life in the Chinese Longitudinal Healthy Longevity Study.

The third paper presents an approach that allows for estimation of cross-sectional and lagged effects of the covariates on the outcome as well as the feedback of the response on future covariates. This is done in two-parts, in part-1, the effects of time-dependent covariates on the outcomes are estimated, then, in part-2, the outcome influences on future values of the covariates are measured. These model parameters are obtained through a Generalized Method of Moments procedure that uses valid moment conditions between the outcome and the covariates. Child morbidity in the Philippines and obesity status in the Add Health data are analyzed.

ACKNOWLEDGMENTS

I want to thank my parents, my siblings, my advisor, my aunt, my friends, my professors, and every person who has been part of my life for their support.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Longitudinal studies with binary outcomes are conducted across a wide range of scientific fields. Examples include longitudinal clinical trials that seek to understand why heart rate changes from normal to abnormal and national longitudinal surveys that attempt to understand the changing habits of the population with time. In some longitudinal studies, the focus is placed on a single outcome variable. In others, multiple outcomes that provide a joint measure of a subject's trait over time are of interest. This is the case when abnormal heart rate, abnormal blood pressure and abnormal heart wall thickness are used to measure heart function over time. All longitudinal studies collect data on subjects or units that are observed and measured over time. Datasets coming from longitudinal studies usually contain time-independent covariates, so their values don't change over time, such as race. These datasets also contain time-dependent covariates with values that might vary from one time-point to another, like weight. The changing values of time-dependent covariates change their impact on the outcomes; for example, weight gain might increase the risk for high blood pressure. There are also situations where there might be feedback of importance from the outcome to the covariate, such that the outcome influences future values of the time-dependent covariates. A situation where feedback is present happens when depression results in cognitive impairment, and having cognitive impairment makes patients more depressed.

The main purpose of longitudinal studies is to identify these changing associations and feedback processes from covariates to outcomes and vice versa across time.

1.1 Generalized Linear Models

Generalized Linear Models (GLM) are used when estimating regression coefficients for independent binary, count, or continuous outcomes that are assumed to follow a distribution from an exponential family (McCullagh & Nelder, 1989). Let y_i be the observed outcome of interest for subject i and $\boldsymbol{x_i}$ be the vector of covariates, under the exponential family framework the density or mass function of y_i is expressed as:

$$f(y_i|\boldsymbol{x_i}) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\varphi)} + c(y_i, \varphi)\right\}$$

where θ_i is the location parameter related to the mean, φ is the dispersion parameter related to the variance with $a_i(.)$, $b(.)$ and $c(.)$ known functions. For all types of outcomes, the first two moments are $\mu_i = E(y_i|\boldsymbol{x_i}) = b'(\theta_i)$ and $v_i = var(y_i|\boldsymbol{x_i}) = b''(\theta_i)a_i(\varphi)$. All generalized linear models have three components, the random component which is given by the vector of responses $\boldsymbol{y}$, the systematic component $\eta_i = \boldsymbol{x_i'\beta}$ which is based on the covariates, and the link component that matches θ_i to the systematic component through a monotone differentiable function $g(.)$ known as the link function such that $\theta_i = g(\eta_i)$.

The logistic regression model for binary outcomes is a generalized linear model with logit link such that:

$$\text{logit}\big(P(Y_i = 1|\boldsymbol{x}_i)\big) = \boldsymbol{x_i'\beta}$$

1.2 Generalized Estimating Equations

The Generalized Estimating Equations (GEE) approach provides an extension to Generalized Linear Models (GLM) for analyzing longitudinal data (K.-Y. Liang & Zeger, 1986). This approach provides population-averaged regression coefficients that are estimated while accounting for the correlation among outcomes from the same subject. This approach leads to more efficient parameter estimates for the regression coefficients.

Suppose that you observe N subjects across T time-points and that repeated observations coming from different subjects are independent, while observations from the same subject might be correlated. Let $\boldsymbol{y_i} = (y_{i1}, y_{i2}, \dots, y_{iT})'$ be the vector of observed outcomes for subject i $(i = 1, \dots, N)$ and $\boldsymbol{X_i} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots, \boldsymbol{x}_{iT})'$ its corresponding $T \times p$ matrix of covariates. Under the GEE approach, the marginal expectation of y for subject i at time t $(t = 1, 2, \dots, T)$, μ_{it}, is modeled assuming that $E(y_{it}|\boldsymbol{x}_{it}) = g^{-1}(\boldsymbol{x}'_{it}\boldsymbol{\beta})$ and $v_{it} = var(y_{it}) = v(\mu_{it})\varphi$ such that the variance is a function of the mean. Then for each subject i each with measurements at T different time points, we have a vector of means $\boldsymbol{\mu_i} = (\mu_{i1}, \ \mu_{i2}, \dots, \mu_{iT})$ and a diagonal matrix $\boldsymbol{A_i} = \text{diag}\{v(\mu_{i1}), v(\mu_{i2}), \dots, v(\mu_{iT})\}$. This method assumes that there exists a working correlation matrix $\boldsymbol{R}(\boldsymbol{\alpha})$ that appropriately describes the correlation among outcomes coming from the same subject and might depend on a vector of unknown parameters $\boldsymbol{\alpha}$ of length S. Let $\boldsymbol{V}_i(\boldsymbol{\alpha})$, the variance-covariance matrix for $\boldsymbol{y}_i$ related to $\boldsymbol{R}(\boldsymbol{\alpha})$ be $\boldsymbol{V_i}(\boldsymbol{\alpha}) = \boldsymbol{A_i^{1/2}}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{A_i^{1/2}}$. Zeger and Liang (Zeger & Liang, 1986) showed that the vector of regression coefficients $\boldsymbol{\beta}$ can estimated by solving the Generalized Estimating Equations (GEE)

$$\boldsymbol{g}_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}$$

where $\boldsymbol{D}_i' = \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\beta}}$. They proved that under mild regularity conditions, the estimator $\widehat{\boldsymbol{\beta}}$ derived from the GEEs above converges in probability to $\boldsymbol{\beta}$, the true parameter of regression coefficients, and is asymptotically normal. Thus, as $N \to \infty$

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to N(\boldsymbol{0}, \boldsymbol{\Gamma})$$

Where $\boldsymbol{\Gamma} = \lim_{N\to\infty} \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D}_i'\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}_i(\boldsymbol{\beta})\boldsymbol{g}_i(\boldsymbol{\beta})'\right)\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D}_i'\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\right)^{-1}$

1.3 Generalized Method of Moments for Longitudinal Data

Generalized Method of Moments (GMM) estimators are an extension to the method of moments. Suppose that you want to estimate a parameter $\boldsymbol{\beta}$ of p regression coefficients from a longitudinal dataset with N subjects. Assume that there exists a function $\boldsymbol{g}_i(\boldsymbol{\beta})$**,** the estimating equations (EE), from $\mathbb{R}^p \to \mathbb{R}^q$ that is continuous and differentiable with respect to $\boldsymbol{\beta}$ and that $E\left(\boldsymbol{g}_i(\boldsymbol{\beta})\right)$ exists for all i and $\boldsymbol{\beta}$. Then the population moment conditions are $E\left(\boldsymbol{g}_i(\boldsymbol{\beta})\right) = \boldsymbol{0}$ with sample moment conditions $\boldsymbol{g}_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}_i(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{D}_i'\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$. Let $\boldsymbol{W}_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}_N(\boldsymbol{\beta})\boldsymbol{g}_N(\boldsymbol{\beta})'$ be a weighting matrix that converges to a positive definite matrix W as $N \to \infty$. Hansen (Hansen, 1982) showed that $\widehat{\boldsymbol{\beta}} = argmin\left(\boldsymbol{g}_N(\boldsymbol{\beta})'\boldsymbol{W}_N^{-1}\boldsymbol{g}_N(\boldsymbol{\beta})\right)$ provides a consistent estimator of $\boldsymbol{\beta}$ for which the asymptotic distribution is normal.

Qu, Linsday, and Li (Qu et al., 2000) developed a GMM estimator for the regression coefficients $\boldsymbol{\beta}$ based on GEE that avoided the direct estimation of the vector $\boldsymbol{\alpha}$

that characterizes $\boldsymbol{R}(\boldsymbol{\alpha})$. They argued that the inverse of the working correlation matrix $\boldsymbol{R}^{-\mathbf{1}}(\boldsymbol{\alpha})$ can be represented by a linear combination of S matrices such that

$$\boldsymbol{R}^{-1}(\boldsymbol{\alpha}) = \sum_{s=1}^{S} \alpha_s M_s$$

where $\alpha_1, \alpha_2, \ldots, \alpha_S$ are unknown constants and $M_1, M_2, \ldots, M_S$ are a set of known basis matrices. Then, they expanded the GEEs, $\boldsymbol{g}_N(\boldsymbol{\beta})$, to

$$\boldsymbol{g}_N(\boldsymbol{\beta}) = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} \boldsymbol{D}_i' \boldsymbol{A}_i^{1/2} \boldsymbol{M}_1 \boldsymbol{A}_i^{1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \sum_{i=1}^{N} \boldsymbol{D}_i' \boldsymbol{A}_i^{1/2} \boldsymbol{M}_S \boldsymbol{A}_i^{1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \end{pmatrix}$$

and estimated $\boldsymbol{\beta}$ using the GMM estimator $\widehat{\boldsymbol{\beta}} = argmin\left(\boldsymbol{g}_{\boldsymbol{N}}(\boldsymbol{\beta})' \boldsymbol{W}_{\boldsymbol{N}}^{-\mathbf{1}} \boldsymbol{g}_{\boldsymbol{N}}(\boldsymbol{\beta})\right)$ where $\boldsymbol{W}_{\boldsymbol{N}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}_{\boldsymbol{N}}(\boldsymbol{\beta}) \boldsymbol{g}_{\boldsymbol{N}}(\boldsymbol{\beta})'$. This approach lets them combine the estimating equations optimally.

1.4 Bayesian Estimation

In Bayesian statistics, inferences for parameters of interest $(\boldsymbol{\theta})$ are summarized through random draws from the posterior distribution of such parameters, $p(\boldsymbol{\theta}|\boldsymbol{y})$. This posterior distribution depends on the likelihood function of the data $p(\boldsymbol{y}|\boldsymbol{\theta})$ and the prior distribution of the parameters, $p(\boldsymbol{\theta})$, such that

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

The goal of Bayesian computation is to obtain sets of independent draws $\boldsymbol{\theta}^{\boldsymbol{t}}$ with $t = 1, \ldots, T$, from the posterior distribution, with enough draws T so that any function of the parameters, $g(\boldsymbol{\theta})$, can be estimated with reasonable accuracy (Gelman et.al, 2014). Bayesian computations are usually done using Markov Chain Monte Carlo (MCMC) methods such as Metropolis, Metropolis-Hastings, and Hamiltonian Monte Carlo sampling algorithms.

Markov Chain Monte Carlo methods are used to generate draws from a distribution that approximates the posterior, $p(\boldsymbol{\theta}|\boldsymbol{y})$ when such distribution can be evaluated but not easily sampled from (Givens & Hoeting, 2013). All these methods are based on Markov Chains, which are sequences of vectors composed of random variables $\{\boldsymbol{\theta}^{(t)}\}$, $t = 0,1,2, \ldots$, where the next observed values $\boldsymbol{\theta}^{(t+1)}$ are only dependent on the present values $\boldsymbol{\theta}^{(t)}$. When using MCMC methods, we build Markov Chains that start at some point, $\boldsymbol{\theta}^{(0)}$, and then for each iteration t ($t = 1,2, \ldots, T$), we draw values $\boldsymbol{\theta}^{(t)}$ from a transition or proposal distribution $J\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}\right)$ which depends on the previous values $\boldsymbol{\theta}^{(t-1)}$. The proposal distribution, $J\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}\right)$, must be constructed to guarantee that the Markov Chains converge to a unique stationary distribution that is the posterior, $p(\boldsymbol{\theta}|\boldsymbol{y})$. According to Gelman, et.al. (Gelman et al., 2014), the key aspect in MCMC methodology is to create Markov processes whose stationary distribution is $p(\boldsymbol{\theta}|\boldsymbol{y})$. Thus, for a sufficiently large number of simulations, the chain corresponding to such Markov process will have a marginal distribution that approximates the posterior (Givens & Hoeting, 2013). We now describe the steps followed in Metropolis, Metropolis-Hastings, and Hamiltonian Monte Carlo sampling algorithms.

1.4.1 Metropolis Algorithm

The metropolis sampling algorithm is mostly used when there are no conjugate priors for parameters of interest. It proceeds by sampling a new proposed value $\boldsymbol{\theta}^*$ that is close to the previous value $\boldsymbol{\theta}^{(t-1)}$ using a symmetric proposal distribution. A proposal distribution $J\,(.\,|.)$ is symmetric if $J(\boldsymbol{\theta}_b|\boldsymbol{\theta}_a) = J(\boldsymbol{\theta}_a|\boldsymbol{\theta}_b)$ meaning that the probability of going from $\boldsymbol{\theta}_a$ to $\boldsymbol{\theta}_b$ is the same as the probability of going from $\boldsymbol{\theta}_b$ to $\boldsymbol{\theta}_a$.

The Metropolis algorithm with T random draws starts by sampling a starting random value, $\boldsymbol{\theta}^{(0)}$, from the proposal distribution such that $p\left(\boldsymbol{\theta}^{(0)}|\boldsymbol{y}\right) > 0$. Then, according to Hoff (Hoff, 2009) at each iteration t $(t = 1, \dots, T)$, the following process is conducted:

1. Sample $\boldsymbol{\theta}^* \sim J\left(\boldsymbol{\theta}^*\middle|\boldsymbol{\theta}^{(t-1)}\right)$, where $\boldsymbol{\theta}^*$ is a new proposed value that the vector of parameters might take at iteration t
2. Compute acceptance ratio

$$r = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})}{p(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{y})} \qquad (1)$$

   Where $p(\boldsymbol{\theta}^*|\boldsymbol{y})$ represents the posterior probability of the proposed value $\boldsymbol{\theta}^*$ and $p\left(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{y}\right)$ is the posterior probability of the parameter value at the previous iteration, $\boldsymbol{\theta}^{(t-1)}$.

3. Sample $u \sim Uniform(0,1)$

   If $u < r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$

   If $u > r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$

For step 3, if the posterior probability of the proposed value $\boldsymbol{\theta}^*$ is higher than that of the previous value $\boldsymbol{\theta}^{(t-1)}$, then $r > 1$ and the proposed value will certainly be part of our sample of values. However, if $r < 1$ there will be times when we will take the proposed value $\boldsymbol{\theta}^*$ as our new value $\boldsymbol{\theta}^{(t)}$ and others when we will keep the previous value such that $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$. Thus, there is the possibility that we will have the same value for the parameter for several consecutive iterations.

1.4.2 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is also mostly used when there are no conjugate prior distributions and follows the same procedure for building Markov Chains as the Metropolis algorithm. However, this method is different from the Metropolis algorithm in that the proposal distributions are usually not symmetric. Each iteration t $(t = 1, \dots, T)$ consists of the following steps:

1. Sample $\boldsymbol{\theta}^* \sim J\left(\boldsymbol{\theta} \middle| \boldsymbol{\theta}^{(t-1)}\right)$ where $\boldsymbol{\theta}^*$ is the new value proposed for the parameter of interest and $J\left(\boldsymbol{\theta} \middle| \boldsymbol{\theta}^{(t-1)}\right)$ is not necessarily symmetric
2. Compute acceptance ratio

$$r = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})}{p(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{y})} \times \frac{J\left(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*\right)}{J(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})} \qquad (2)$$

   Where $p(\boldsymbol{\theta}^*|\boldsymbol{y})$ is the posterior probability of the proposed value $\boldsymbol{\theta}^*$ given the data, $p\left(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{y}\right)$ is the posterior probability of the parameter's previous value given the data, $J\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}\right)$ is the probability of going from $\boldsymbol{\theta}^{(t-1)}$ to $\boldsymbol{\theta}^*$ under the proposal distribution and $J\left(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*\right)$ represents the probability of going from $\boldsymbol{\theta}^*$ to $\boldsymbol{\theta}^{(t-1)}$ also under the proposal.
3. Sample $u \sim Uniform(0,1)$

   If $u < r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$

If $u > r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$

The Metropolis algorithm is a special case of the Metropolis-Hastings algorithm. Since the proposal distribution is symmetric and thus $J(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*) = J(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$, then in the Metropolis sampler the factor $\frac{J(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}{J(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})} = 1$ which reduces the ratio in (2) to the ratio in (1).

As explained before, in Metropolis-based algorithms it might be the case that the proposed values $\boldsymbol{\theta}^*$ are rejected in more than one consecutive draw and that we keep the same value $\boldsymbol{\theta}^{(t)}$ for several consecutive iterations, causing autocorrelation among the simulated draws. This is due to the random walk behavior of the Metropolis-based algorithms and can make such sampling algorithms inefficient, especially when estimating high dimension parameter vectors. As such, there are situations in which more efficient algorithms are needed.

1.4.3 Hamiltonian Monte Carlo Algorithm

Gelman et.al. (2014) define the Hamiltonian Monte Carlo algorithm as a generalization of the metropolis algorithm that subdues its random walk behavior and allows HMC to move faster in the parameter space by including momentum variables, resulting in faster mixing and convergence, especially for high dimensional parameter vectors. This algorithm is most used when there are several parameters to be estimated and the Metropolis algorithm moves slowly through the target distribution. It depends on a vector of auxiliary variables $\boldsymbol{\phi}$ of the same dimension as the vector of parameters $\boldsymbol{\theta}$ and the proposal distribution for $\boldsymbol{\theta}$ is based on $\boldsymbol{\phi}$.

In Hamiltonian Monte Carlo the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is combined with an independent distribution of the auxiliary variables $p(\boldsymbol{\phi})$ resulting in a joint distribution $p(\boldsymbol{\theta},\boldsymbol{\phi}) = p(\boldsymbol{\phi})p(\boldsymbol{\theta}|y)$. Random draws are obtained using this joint distribution. Still interest is placed only on $\boldsymbol{\theta}$**,** and $\boldsymbol{\phi}$ values are discarded, since they are only added to move across the parameter space faster. Apart from the posterior distribution, this algorithm also requires the gradient of its log,

$$\frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_1}, \frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_2}, \dots, \frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_k}\right)$$

The Hamiltonian Monte Carlo sampling algorithm is described in Gelman et.al (2013) as follows. For iteration t $(t = 1, \dots, T)$ steps taken are:

1. Sample $\boldsymbol{\phi}$ from $p(\boldsymbol{\phi})$, where $\boldsymbol{\phi} \sim multivariate\ normal(\mathbf{0}, \boldsymbol{M})$ and $\boldsymbol{M}$ is usually a diagonal matrix

2. Update $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ simultaneously repeating L leapfrog steps which consist of the following:

    a. Let $\boldsymbol{\phi} = \boldsymbol{\phi} + \frac{\varepsilon}{2}\frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \boldsymbol{\theta}}$

    b. Let $\boldsymbol{\theta} = \boldsymbol{\theta} + \varepsilon M^{-1}\boldsymbol{\phi}$

    c. Update $\boldsymbol{\phi}$ one more time by letting $\boldsymbol{\phi} = \boldsymbol{\phi} + \frac{\varepsilon}{2}\frac{\partial \log p(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \boldsymbol{\theta}}$

3. Let $\boldsymbol{\theta}^{(t-1)}$ and $\boldsymbol{\phi}^{(t-1)}$ be the values that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ had before starting the leapfrog process and $\boldsymbol{\theta}^{*}$ and $\boldsymbol{\phi}^{*}$ their values after the L leapfrog steps, then calculate the ratio

$$r = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})}{p(\boldsymbol{\theta}^{(t-1)})} \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(t-1)})}$$

4. Draw $u \sim uniform\ (0,1)$

   If $u < r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$

   If $u > r$ then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$

In this algorithm, ε is a scaling factor for the probability distribution of the momentum variables and is known as stepsize. Gelman et.al. (2013) recommend that the product of the number of leapfrog steps and the scaling factor $L\varepsilon = 1$. If convergence is not achieved, the number of leapfrog steps should be increased, and the scaling factor should be reduced. Depending on the choices on distribution for the momentum variables, the scaling factor and the number of leapfrog steps, the Hamiltonian Monte Carlo sampling algorithm might converge to the posterior distribution faster than the Metropolis-based algorithms, making it more efficient in Bayesian computation, but that might not always be the case (Vazquez Arreola & Wilson, 2019).

1.5 Accounting for Time-dependent Covariates and their Associations with One or Multiple Binary Outcomes

In this work, methods are used to study the changing associations between time-dependent covariates and binary outcomes over time. These are marginal models using partitioned coefficients to describe the association.

First, an alternative approach to the Partitioned GMM model with Bayesian estimates that allows the assessment of the current and lagged effects of the time-dependent covariates on binary outcomes is developed. This approach is particularly useful when the Partitioned GMM model provides numerically unstable regression estimates or standard errors.

Second, a method for simultaneously modeling two or more binary longitudinal outcomes while accounting for all forms of correlation among and between outcomes coming from the same subject is developed. This approach helps us to understand the changing associations of time-dependent covariates across time with the multiple outcomes through partitioned coefficients. It is utilized when multiple binary outcomes measure an overall trait of a subject to get a complete picture of such trait.

Finally, a two-part model with feedback that allows one to measure the direct effect and the feedback effect of time-dependent covariates on the outcomes and vice versa. However, it is not always that a feedback effect is interpretable or makes sense.

CHAPTER 2

PARTITIONED MVM MARGINAL MODEL WITH BAYES ESTIMATES FOR CORRELATED DATA AND TIME-DEPENDENT COVARIATES

Abstract

It is customary in the analysis of longitudinal data that one examines the relationship between the outcomes and the risk factors which often include time-dependent covariates. The fit of existing marginal models for such data does not always capture all the types of extra variation among the responses and the covariates. However, when they do, they are often not fully explored. This paper proposes a Partitioned Method of Valid Moments (MVM) marginal regression model with Bayes estimates, using lagged coefficients to fit to longitudinal data, with time-dependent covariates using composite likelihoods. The model is flexible and readily attainable in obtaining estimates of the regression coefficients for time-dependent covariates. A simulation study is conducted to evaluate the properties of the model coefficients. The fit of modeling cognitive impairment diagnosis in NACC Alzheimer survey data and modeling obesity status in the Add Health survey data are explored. Sensitivity analyses in these examples were conducted to evaluate the impact of the prior distribution on the posterior inferences.

2.1 Introduction

Approximately 16 million people lived with cognitive impairment in the United States, in 2011 (*Cognitive Impairment: A Call for Action*, 2011). Such ailments affect people's cognitive functions, interfering with their everyday activities. It results in their inability to live independently, costing patients' family members and government significant in the vicinity of millions of dollars. Time-dependent covariates such as depression, thyroid disorder diagnosis and traumatic brain injury are risk factors of cognitive impairment. In the analysis of longitudinal data, an examination of the changing effects of depression, thyroid disorder diagnosis and traumatic brain injury in cognitive impairment diagnosis across time while controlling for race, gender, and age was conducted. This research proposes a model that addresses the correlation between response and time dependent covariates, with limited number of valid moments by obtaining Bayes estimates.

Health surveys are often collected longitudinally, with observations obtained from each respondent at multiple time-periods. Such methods of collection result in observations that are correlated due to the repeated measurements on each respondent. This induced intraclass correlation makes it impossible to obtain a joint likelihood function for the observations. It is customary to address such correlation in two ways.

One approach is to use the conditional likelihood principles in a subject-specific model based on two or more distributions. It depends on the number of random effects, with one distribution for the outcomes conditional on those random effects. Though the random effects are frequently used in such modeling, the likelihood conditioned on the random effects, as a remedy does not provide insight into the population-averaged mean (Laird & Ware, 1982).

A second approach is to alter the distribution in the random component of the three components designation in a generalized linear model. This is done through a mean-variance relation thereby treating the correlation as a nuisance. Such a method is based on moments as in the generalized estimating equation (GEE) method (Liang & Zeger, 1986). The GEE method is a robust approach which produces efficient estimates of the marginal mean parameters when the working correlation structure is correctly specified. However, they are consistent and distribute as an asymptotic normal estimators even when the working correlation matrix is misspecified (Zeger & Liang, 1986). This approach leads to population-averaged model, which is used to study the marginal mean. This approach is based on the quasi-likelihood principles.

Generalized method of moments (GMM) models are also an attractive alternative for fitting marginal models to longitudinal data. GMM is common in econometrics modeling (Hall, 2005; Hansen, 1982; Hansen et al., 1996), when the likelihood is difficult or impossible to obtain, as it improves the estimation efficiency. Asymptotic theory for GMM estimators has resulted in great interest through the population moment conditions (McFadden, 1989). Qu et al. (2000) and Lai and Small (2007) used a GMM estimator to fit marginal regression models to analyze longitudinal data as an advantage over the use of GEE estimates. Lalonde, Wilson, and Yin (2014) extended this method, and introduced a hypothesis test to identify valid moment conditions.

In some cases, when fitting models to longitudinal data, the repeated measurements at different time-periods induce correlation due to the covariate in one-time-period affecting responses in other time-periods. These relationships across different time-periods lead to additional regression coefficients that give added information at different times. Irimata, Broatch, and Wilson (2019) built upon the GMM framework proposed by Lalonde, Wilson and Yin (2014) and introduced a unique method to fit population-averaged models to correlated data with time-dependent covariates. Their method relies on GMM estimation with a derived partitioned data matrix. This Partitioned GMM model identifies separate coefficients based on the use of valid moments to address the varying impacts of the covariates over time. Thus, multiple regression coefficients are produced to identify changing relationships between time-dependent covariates and outcomes, rather than averaging these relationships across time into a single parameter estimate.

In the fit of the Partitioned GMM marginal model for longitudinal data with its time-dependent covariates, the convergence criterion for the optimization algorithm are not always met, especially in small sample sizes. One may receive a warning indicating that there are numerical instabilities in the parameter estimates and their standard errors. In such cases as an alternative, a Partitioned MVM marginal model with Bayesian estimates is useful.

The remainder of this paper is organized as follows. In Section 2, an introduction of the notation, a review of the Bayesian estimation principles in a generalized linear model framework, along with relevant models for time-dependent covariates are reviewed. In Section 3, the Partitioned MVM marginal model with Bayesian estimates for longitudinal data with time-dependent covariates is developed. In Section 4, a simulation study to determine the performance of the proposed model is conducted. In Section 5, two numerical examples, Alzheimer's data from NACC database (Beekly et al., 2007) and Add Health data (Harris & Udry, 2016) are analyzed. A comparison of the fit of the Partitioned MVM model with Bayesian estimates and the fit of the Partitioned frequentist GMM model is made, and the advantages and disadvantages noted. A sensitivity analysis in these examples is undertaken to evaluate the impact of the prior distribution on the posterior inferences. Some remarks are made in Section 6.

2.2 Background

2.2.1 Marginal Regression Modeling with Time-Dependent Covariates

Lai and Small (2007) fitted marginal models with time-dependent covariates while identifying valid moments through a process as belonging to one of three groups. Their method of identification of valid moments made use of a marginal model for longitudinal continuous data with GMM through moment conditions where,

$$\mathrm{E}\left[\frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \beta_j}\{y_{it} - \mu_{it}(\boldsymbol{\beta})\}\right] = 0 \tag{2.1}$$

for time-points $s = 1, \ldots, T$ and $t = 1, \ldots, T$ and covariate j, where $\mu_{it}(\boldsymbol{\beta}) = \mathrm{E}\{[y_{it}|\boldsymbol{x}_{it}]\}$ denotes the expectation of response y_{it} based on the covariate $\boldsymbol{x}_{it.}$ and the ordinary least squares regression parameters $\boldsymbol{\beta}$ in the systematic component. Then, the valid moments are computed to obtain an estimate of a single regression coefficient to represent the overall effect of a given covariate on the response. Others have developed similar approaches including Zhou, Lefante, Rice, and Chen (2014) who utilized a modified quadratic inference function, and Chen and Westgate (2017) who utilized a modified weight matrix based on linear shrinkage.

Lalonde, Wilson and Yin (2014) introduced an extension of this GMM approach. In addition, they provided a statistical test to identify valid moment conditions when using time-dependent covariates on binary responses and extended the classification of Lai and Small (2007) to type IV. However, their method determines which moment conditions are valid using a correlation test. This approach consists of first fitting a logistic regression model at each time-period t, such that $\text{logit}(p_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\beta}$, where $\boldsymbol{x}'_{it}$ represents the vector of covariates for subject *i* measured at time *t*. Individual correlation tests are conducted to determine whether the correlation between residual e_t at time *t* and covariate x_{js} measured at time *s* with $(s \neq t)$ is significantly different from zero. If the statistical test determines that such correlation is not significantly different from zero, then the moment condition between y_t and covariate x_{js} is declared as valid. This approach recognizes that moment conditions for time-independent covariates and time-dependent covariates measured at the same time as the outcome are valid.

2.2.2 Partitioned Coefficients with Time-Dependent Covariates

Irimata, Broatch, and Wilson (2019) modeled the effect of time-dependent covariates on binary responses through a derived partitioned data matrix with GMM estimation. They partitioned the data matrix to incorporate a set of valid moment conditions based on a time-spacing measure. Valid moment conditions were identified using the statistical tests by Lalonde, Wilson and Yin (2014); however, rather than grouping the valid moments to obtain an average effect of the covariate on the response, they separated out the effects of the covariates on the responses across time-periods. This creates additional regression parameters, thus providing insight into the time-varying effects of that covariate on the response at different time-periods. This model is a special case of the generalized linear regression model proposed by Müller and Stadtmuller (2005). In this paper, we propose the use of Bayesian estimates to obtain these partitioned regression parameter estimates as an alternative to the Partitioned GMM model.

2.2.3 Bayesian Estimates

Bayesian methods of analysis rely heavily on the Bayes' theorem and the likelihood principle. Given the prior distribution, statistical inferences are based on a posterior distribution of the model parameters within the linear estimating function family (Hoff, 2009). The Bayesian estimation procedure is a useful alternative, especially in small datasets or in complex set of modeling with extra parameters, as is the case when one considers the time-dependent covariates with valid moments (Efron, 2015).

Yin (2009) proposed the Bayesian GMM, through the derivation of moments obtained from the working correlation matrix and used it to obtain a quadratic objective function, in the usual GMM framework (Hansen, 1982). This objective function, along with prior distributions was used in the Markov chain Monte Carlo procedure in order to sample from the posterior distribution. In addition, Yin (2009) examined the properties of the Bayesian GMM under the linear regression model for repeated measurements with correlated errors. However, under the Bayesian GMM approach the regression parameters, $\boldsymbol{\beta}$, and the weighting optimal matrix, $\boldsymbol{W_N}(\boldsymbol{\beta})$, are updated concurrently making the surface of the quasi-posterior distribution complicated and causing the Markov Chain Monte Carlo (MCMC) algorithm to become inefficient and unstable (Tanaka, 2020; Yin et al., 2011). Therefore, even in cases when the number of moment conditions is much smaller than the sample size, a Bayesian GMM estimator can be ill-posed (Tanaka, 2020). As such this research deviates from the Bayesian GMM approach. Instead, a Partitioned Method of Valid Moments (MVM) model with Bayesian estimates is proposed. This method provides parameter estimates at least as efficient as estimates from the frequentist partitioned GMM model or from the GEE model with lagged covariates and an independent working correlation matrix.

2.2.4 Composite Likelihoods

Varin, Reid and Firth (2011) stated that composite likelihoods are referred to by different names, for example pseudo likelihood (Molenberghs & Verbeke, 2005), approximate likelihood (Stein et al., 20014), and quasi-likelihood (Glasbey, 2001; Hjort et al., 1994; Hjort & Varin, 2008). Methods based on composite likelihood are called limited information methods, in the psychometric literature (Varin, 2008; Varin et al., 2011). Varin, Reid and Firth (2011) developed a review of recent developments in the theory and application of composite likelihood where they emphasized the current state of knowledge on efficiency and robustness of composite likelihood inference. They mentioned that the set of application areas include longitudinal data analysis.

The simplest composite likelihood for inference in marginal regression models is the pseudo likelihood built under working independence assumptions (Varin et al., 2011),

$$\mathcal{L}_{ind}(\boldsymbol{\beta};\boldsymbol{y}) = \prod_{i=1}^{N} f(\boldsymbol{y}_i;\boldsymbol{\beta}),$$

often referred to in the literature as the independence likelihood (Chandler & Bate, 2007). The independence likelihood allows inference only on marginal regression parameters (Varin et al., 2011). If parameters related to correlation are also of interest one has to model blocks of observations, as in the pairwise likelihood (Cox & Reid, 2004; Varin, 2008; Varin et al., 2011)

$$\mathcal{L}_{pair}(\boldsymbol{\beta};\boldsymbol{y}) = \prod_{i=1}^{N-1}\prod_{k=i+1}^{N} f(\boldsymbol{y}_i\boldsymbol{y}_k;\boldsymbol{\beta})$$

with its extension built from larger sets of observations (Caragea & Smith, 2007). This paper makes use of composite likelihood.

Varin, Reid and Firth (2011) stated one of the motivations for choosing composite likelihoods is usually computational since they help avoid computing or modelling the joint distribution of a possibly high-dimensional response vector. Another reason they provided for using composite likelihoods, is the notion of robustness under possible misspecification of the higher order dimensional distributions similar to the type of robustness achieved by generalized estimating equations, but different to robust point estimation (Varin et al., 2011). Composite likelihoods can also be used to construct joint distributions in settings where there are not clear high dimensional distributions or where the likelihood surface can be much smoother than the full joint likelihood and easier to maximize (G. Liang & Yu, 2003; Varin et al., 2011). When the high dimensional characteristics of the model are not fully specified, one allows for a less complex structure on the parameter space, reduces computational efforts, and takes advantage of the robustness properties of the composite likelihoods (Varin et al., 2011).

2.3 Partitioned MVM Bayesian Marginal model

2.3.1 Inference for Correlated Data using the Independence Likelihood

In this paper, the problem of accounting for clustering in longitudinal studies when conducting inference is undertaken using the independence log likelihood. Chandler and Bate (2007) demonstrated the properties of independence estimating equations that adjust the independence loglikelihood function in the presence of clustering. Their adjustment relies on a robust sandwich estimator of the parameter covariance matrix that fits the model assuming the observations are independent and then adjusts the estimated standard errors to account for the correlation. Their estimators are solutions to the independence estimating equations by Joe (1997) and their approach contends positively with established techniques based on independence estimating equations. For generalized linear models, their method corresponds to that of using an independence working correlation structure when fitting a generalized estimating equations model.

Consider a longitudinal study where N subjects are observed at T different time periods, then $\boldsymbol{y}_i = (y_{i1}, \dots y_{iT})'$ is the vector of T observed outcomes on the i^{th} subject. For each subject i, denote $\Im_i$ as a conditioning set relevant to $\boldsymbol{y}_i$, in our case the time-dependent and time-independent covariates. Assume that the observed outcomes are drawn from a parametric family of joint distributions, indexed by parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and with density functions factorized as $\prod_{i=1}^{N} f_i(\boldsymbol{y}_i|\Im_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$ (Chandler & Bate, 2007).

According to Chandler and Bate (2007), the densities $\{ f_{it} : t = 1, \ldots, T\}$ of the components of $\boldsymbol{y}_i$ are assumed to depend on $\boldsymbol{\beta}$ but not on $\boldsymbol{\alpha}$ when conditioning on the covariates $(\Im_i)$. Therefore, $\boldsymbol{\alpha}$ parameterizes the within-subject correlation, and $\boldsymbol{\beta}$ parameterizes the marginal structure. In our case $\boldsymbol{\beta}$ represents the regression coefficients. Our interest lies on the inference about $\boldsymbol{\beta}$. In the analysis of longitudinal studies subjects (clusters) are assumed to be independent.

Chandler and Bate (2007) suggested that by letting the joint distributions for each subject i be given by $f_i(\boldsymbol{y_i}|\Im_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$ $\{ i = 1, \ldots, N\}$, inference can be based on the full loglikelihood function

$$\ell_F(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \log f_i(\boldsymbol{y}_i|\Im_i; \boldsymbol{\beta}, \boldsymbol{\alpha}) \qquad (3.1)$$

However, when the goal of the analysis is inference about the marginal regression parameters $\boldsymbol{\beta}$ and the correlation parameters $\boldsymbol{\alpha}$ are of no interest, the use of the full likelihood function may not be practical (Chandler & Bate, 2007). Such is the case when fitting marginal models to longitudinal data, where it is accepted to proceed as though, given the covariates $\{\Im_i\}$, the observations are independent. Then the marginal regression coefficients $\boldsymbol{\beta}$ are estimated by maximizing the independence loglikelihood function

$$\ell_I(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \log f_i(\boldsymbol{y_i}|\Im_i; \boldsymbol{\beta}, \boldsymbol{\alpha}) \qquad (3.2).$$

Chandler and Bate (2007) used the score equations $U(\boldsymbol{\beta})$ corresponding to $\ell_I(\boldsymbol{\beta})$ to determine unique root estimators of the independence estimating equations

$$U(\boldsymbol{\beta}) = \frac{\partial \ell_I}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} U_i(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}(\boldsymbol{\beta}) = 0.$$

2.3.2 Partitioned GMM Estimation

Consider a model with a data matrix of time-dependent covariates that originated from an identification of the valid moments (Irimata et al., 2019). Thus, consider a derived partitioned data matrix $\boldsymbol{X}_{ij}^{[\,]}$ whose dimension depends on the number of repeated measures on the response, T. These relationships exist between the outcomes $\boldsymbol{Y}_{*t}$ observed at time t, and the j^th^ covariate $\boldsymbol{X}_{*js}$ observed at time s, for $s \leq t \leq T$ and $j = 1, .., J$. Each time-dependent covariate $\boldsymbol{X}_{*j*} = (\boldsymbol{X}_{*j1}, \dots, \boldsymbol{X}_{*jT})$ is measured at time-periods $t = 1, 2, \dots, T$; for subject i. Thus, the data matrix is restructured into a lower triangular matrix,

$$\boldsymbol{X}_{ij}^{[]} = \begin{bmatrix} 1 & X_{ij1} & 0 & \dots & 0 \\ 1 & X_{ij2} & X_{ij1} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{ijT} & X_{ij(T-1)} & \dots & X_{ij1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X}_{ij}^{[0]} & \boldsymbol{X}_{ij}^{[1]} & \dots & \boldsymbol{X}_{ij}^{[T-1]} \end{bmatrix},$$

where the superscript denotes the difference in time-periods, $|t - s| > 0$, between the response measured at time t and the covariate measured at time s, and $\mathbf{1}$ is the vector of ones. The regression model,

$$g(\mu_{it}) = \beta_0 + \beta_{IN} X_{iIN} + \beta_j^{tt} X_{ij}^{[0]} + \beta_j^{[1]} X_{ij}^{[1]} + \beta_j^{[2]} X_{ij}^{[2]} \dots + \beta_j^{[T-1]} X_{ij}^{[T-1]} \qquad (3.3)$$

where X_{iIN} is the time independent covariate and β_{IN} is the regression coefficient. In matrix notation,

$$g(\boldsymbol{\mu_i}) = \boldsymbol{X}_{IN} \boldsymbol{\beta}_{IN} + \boldsymbol{X}_{\boldsymbol{ij}}^{[\,]} \boldsymbol{\beta}_j,$$

where the $\boldsymbol{X}_{ij}^{[\,]}$ matrix is the lower triangular matrix obtained from the original data matrix, and the mean $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iT})'$ depends on the regression coefficients $\boldsymbol{\beta}_j = \left(\beta_0, \beta_j^{tt}, \beta_j^{[1]}, \beta_j^{[2]}, \dots, \beta_j^{[T-1]}\right)$. The regression coefficient β_j^{tt} denotes the effect of the covariate $\boldsymbol{X}_{*jt}$ on the response $\boldsymbol{Y}_{*t}$, when both are observed in time-period t. When the covariate is observed in a time-period prior to the outcome, or in other words when $s < t$, the lagged effect of the covariate $\boldsymbol{X}_{*js}$ on the response $\boldsymbol{Y}_{*t}$ across a $(t - s)$ period lag is given by the coefficients $\beta_j^{[1]}, \beta_j^{[2]}, \dots, \beta_j^{[T-1]}$. The coefficient $\beta_j^{[1]}$ for instance, denotes the effect of the covariate on the response across a single time-period lag. Thus, the effect of the covariate on the response varies across different time-period lags.

Let the vector $\boldsymbol{\beta}$ denote the concatenation of the parameters associated with each of the *J* covariates and $\boldsymbol{X}_{ij}^{[\,]}$ denote the column-bound data matrix of the lower-triangular matrix. Each of the J time-dependent covariates yield up to T partitions corresponding to $\boldsymbol{\beta}_j$. Thus, $\boldsymbol{X}_{i*}^{[\,]} = \left(\boldsymbol{X}_{i1}^{[\,]}, \dots . \boldsymbol{X}_{iJ}^{[\,]}\right)$ will have a dimension of $(J \times T) + 1$ by N, and $\boldsymbol{\beta}$ will be a vector of length $(J \times T) + 1$. In the presence of time-dependent covariates, it is necessary to consider studying lagged relationships between the covariate and the outcome as the outcome might depend on one or several previous values of the covariate (Schildcrout & Heagerty, 2005).

When modeling longitudinal data, the association between a time-dependent covariate and the outcome cannot reasonably be assumed to be only direct and instantaneous, since it is likely to be cumulative over a certain period of time and to depend on past measurements of the covariate. One approach is to use a lagged model based on splines which are estimated using Bayesian estimates (Obermeier et al., 2015). The primary advantage of using models with lagged covariates is the potential to accurately understand the detailed dependence of the outcome on the full history of time-dependent covariate measurements (Heagerty & Comstock, 2013). One way to properly model longitudinal binary outcomes with time dependent covariates is to include the appropriate lagged values of the covariate (Heagerty, 2002).

2.3.3 Partitioned MVM Inferences with Bayes Estimates

Once the partitioning of the data is complete and the valid moments are identified, Bayes principles are applied to the resulting model. The partitioning of the data matrix and its lagged components with valid moment conditions accounts for some of the correlation in the data. The adjustment unlike the GEE based models takes place in the systematic component of the model. Moreover, when the expectation of an outcome, $\mathrm{y_{it}}$, depends on previous values of the outcome, y_{is} $(s < t)$, then the time-dependent covariate x_{ijs} measured at time s affects the expected value of the outcome y_{it} at time t (Lalonde et al., 2014).

Consider the likelihood function, assuming that the marginal mean of y_{it} is affected by current and previous values of the time-dependent covariates while accounting for time-independent covariates. Then, using (3.2), the independence likelihood, the proposed model

$$g(\mu_{it}) = \beta_0 + \beta_{IN}X_{iIN} + \beta_1^{tt}X_{i1t} + \beta_1^{[1]}X_{i1[t-1]} + \cdots + \beta_1^{[t-1]}X_{i11} + \cdots + \beta_J^{tt}X_{iJt} + \beta_J^{[1]}X_{iJ[t-1]} + \cdots + \beta_J^{[t-1]}X_{iJ1}$$

has likelihood

$$L^c(\boldsymbol{y}|\boldsymbol{X}_{***}^{[]}, \boldsymbol{\beta}) = \prod_{i=1}^{N} f_Y(y_{i1}, \dots y_{iT}|\boldsymbol{1}, \boldsymbol{X}_{i1}^{[0]}, \boldsymbol{X}_{i1}^{[1]}, \dots, \boldsymbol{X}_{i1}^{[T-1]}, \dots, \boldsymbol{X}_{iJ}^{[0]}, \boldsymbol{X}_{iJ}^{[1]}, \dots, \boldsymbol{X}_{iJ}^{[T-1]}, \boldsymbol{\beta})$$

$$= \prod_{i=1}^{N}\prod_{t=1}^{T} P\left(Y_{it} = 1|X_{iIN}, X_{i11}, X_{i12}, \dots, X_{i1t}, \dots, X_{iJ1}, X_{iJ2}, \dots, X_{iJt}, \beta_0, \beta_{IN}, \beta_1^{tt}, \beta_1^{[1]}, \dots, \beta_1^{[t-1]}, \dots, \beta_J^{tt}, \beta_J^{[1]}, \dots, \beta_J^{[t-1]}\right)$$

$$= \prod_{i=1}^{N}\prod_{t=1}^{T}\left(1 + e^{-logit(\mu_{it})}\right)^{-y_{it}}\left(1 + e^{logit(\mu_{it})}\right)^{y_{it}-1}$$

Let $\pi(\boldsymbol{\beta}) = \pi(\beta_0, \beta_{IN}, \beta_1^{tt}, \beta_1^{[1]}, \beta_1^{[2]}, \dots, \beta_1^{[T-1]}, \dots, \beta_J^{tt}, \beta_J^{[1]}, \beta_J^{[2]}, \dots, \beta_J^{[T-1]})$ denotes the prior distribution for the coefficients in the partitioned matrix. The prior distribution for the vector of regression coefficients, $\boldsymbol{\beta}$ is a multivariate normal distribution. The regression coefficients are assumed independent and their prior distributions are assumed normally distributed such that $\beta_j^{tt} \sim N(\mu_{j0}, \sigma_{j0}^2)$ and $\beta_j^{[t-s]} \sim N(\mu_{j0}, \sigma_{j0}^2)$ for $(t-s) = 1, \dots, T-1$. The parameter μ_{j0} is the prior mean and σ_{j0}^2 is the prior variance. A prior distribution (same or different) for each of the coefficients $(\beta_j^{tt}, \beta_j^{[1]}, \beta_j^{[2]}, \dots, \beta_j^{[T-1]})$ for the covariate $X_{ij}^{[t-s]}$ with valid moments. In cases, when there is no known prior information about the effect of a covariate on the outcome of interest, one can use non-informative priors (Efron, 2015). Then, given the likelihood function of the data, the posterior distribution of $\boldsymbol{\beta}$ is

$$\pi\left(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}_{i**}^{[]}\right) \propto L^C\left(\boldsymbol{y}|\boldsymbol{X}_{***}^{[]}, \boldsymbol{\beta}\right) \pi(\boldsymbol{\beta}) .$$

The posterior distribution of $\boldsymbol{\beta}$, the vector of regression coefficients is unknown, but one can use the Monte Carlo Markov Chain to draw samples.

2.4 Simulation Study

A simulation study to examine the performance of the Partitioned MVM marginal model with Bayesian estimates is conducted. A comparison is made of its performance to that of the partitioned GMM model and the lagged GEE model with the independent working correlation matrix. Simulated datasets with sample sizes $N \in \{25, 50, 100, 500, 1000\}$ subjects and time periods $T \in \{3, 5\}$ are generated. The simulated data are generated from a Bernoulli random variable with mean a function of current and lagged covariates of type II (Diggle et al., 2002; Irimata et al., 2019; Lai & Small, 2007). Simulated data under each of the scenarios generated by $(N \cap T)$ for each of the combinations of the 1,000 datasets are generated. Assign time-dependent covariates with weights between $[2,\ 4]$ to estimate delayed effect (lag-1), with time-dependent covariate with weights between $[1\ ,2]$ to estimate further delayed effect (lag-2). The correlation is induced by random effects distributed normal as $N\ (0, 1)$. The regression coefficients are set to $\beta_0 = 0$, $\beta^{tt} = 0.5$, $\beta^{[1]} = 0.3$ and $\beta^{[2]} = 0.1$. Thus, the association between the outcome and the covariate is strongest at the immediate effect (cross-sectional) and weak for delayed (lag 1) and weaker for further delayed effects (lag 2). The prior distributions are $N\ (0, 10000)$.

The percentage coverage, the root mean square error (RMSE), the bias and the percent converged for the parameter estimates are used to compare performances on three models: the Partitioned MVM marginal model with Bayes estimates, the partitioned GMM marginal model, and the lagged GEE model with working independence (Nakata & Tonetti, 2014).

The results for the three models when $T = 3$ using RMSE and percent coverage are given in Table 2.1, bias and percentage of datasets where models converged are presented in Table 2.2. The Partitioned MVM marginal model with Bayes estimates performs better than both the Partitioned GMM model in terms of percentage of coverage for sample sizes of $(N = 25,\ 50,\ 100)$. The percent coverage between the Partitioned MVM marginal model with Bayes estimates and the lagged GEE model with working independence are similar in the settings. For the parameter estimates with the sample sizes used, the RMSE for the Partitioned MVM marginal model with Bayes estimates are similar to those for the partitioned GMM model and the lagged GEE model with independent working correlation matrix. Thus, providing estimates as efficient as those of the other two models. For all sample sizes, the bias of the Partitioned MVM model with Bayes estimates is between the bias of the partitioned GMM model and the lagged GEE model with independent working correlation matrix.

For $(T = 5)$, the time-dependent covariate's weight is held between $[2, 4]$ for delayed effect, the time-dependent covariate's weight is held between $[1,\ 2]$ for a further delayed effect (lag 2), for furthermost delayed effect weight is held between $[0.5, 1]$ (lag 3), and at a level higher than furthermost with a weight $[0,\ 0.5]$ (lag 4). The regression coefficients are set to $\beta_0 = 0, \beta^{tt} = 0.7, \beta^{[1]} = 0.5, \beta^{[2]} = 0.3, \beta^{[3]} = 0.1$ and $\beta^{[4]} = 0.05$. Thus, the association between the outcome and the covariate is weakest at furthest delayed (lag 4), and stronger at smaller lags (less delayed effects). Similar results for $(T = 5)$ are similar to those obtained for cases $(T = 3)$, Tables 2.3 and 2.4.

Table 2.1

Simulation Results for T=3

Sample size	Parms	Partitioned GMM		Partitioned Bayes		GEE independent	
		Coverage	RMSE	Coverage	RMSE	Coverage	RMSE
25	$\beta_0 = 0$	66.77	5.181	93.00	0.366	93.42	0.331
	$\beta^{tt} = 0.5$	69.54	3.956	91.60	0.431	94.57	0.363
	$\beta^{[1]} = 0.3$	68.48	6.258	92.20	2.230	94.46	1.285
	$\beta^{[2]} = 0.1$	57.39	16.499	85.90	26.758	88.09	6.810
50	$\beta_0 = 0$	84.70	0.904	92.90	0.237	93.79	0.228
	$\beta^{tt} = 0.5$	86.70	1.131	92.50	0.270	93.49	0.247
	$\beta^{[1]} = 0.3$	81.00	5.094	92.80	0.934	93.09	0.852
	$\beta^{[2]} = 0.1$	71.34	11.585	91.00	5.276	91.59	4.305
100	$\beta_0 = 0$	91.50	0.180	94.80	0.151	95.20	0.148
	$\beta^{tt} = 0.5$	92.60	0.206	94.90	0.166	95.50	0.161
	$\beta^{[1]} = 0.3$	90.40	3.600	94.20	0.580	95.20	0.557
	$\beta^{[2]} = 0.1$	85.65	6.009	95.30	1.549	95.30	1.331
500	$\beta_0 = 0$	94.50	0.070	94.30	0.067	94.60	0.067
	$\beta^{tt} = 0.5$	94.00	0.076	93.50	0.072	94.30	0.071
	$\beta^{[1]} = 0.3$	94.40	0.273	94.00	0.251	94.10	0.248
	$\beta^{[2]} = 0.1$	94.02	0.711	95.40	0.565	95.50	0.551
1000	$\beta_0 = 0$	94.70	0.049	95.20	0.046	95.10	0.046
	$\beta^{tt} = 0.5$	93.40	0.054	93.10	0.051	93.70	0.051
	$\beta^{[1]} = 0.3$	93.70	0.188	93.60	0.176	94.40	0.175
	$\beta^{[2]} = 0.1$	94.94	0.463	95.00	0.405	95.10	0.401

Table 2.2

Bias and Percentage of Datasets that Converged in T=3

Sample size	Parms	Partitioned GMM		Partitioned Bayes		GEE independent	
		Bias	% converged	Bias	% converged	Bias	% converged
25	$\beta_0 = 0$	-0.142	97.50	0.001	100	0.001	95.70
	$\beta^{tt} = 0.5$	-1.025		-0.154		-0.081	
	$\beta^{[1]} = 0.3$	-1.258		-0.365		-0.100	
	$\beta^{[2]} = 0.1$	-3.905		-7.921		-1.520	
50	$\beta_0 = 0$	-0.001	100	0.009	100	0.007	99.90
	$\beta^{tt} = 0.5$	-0.233		-0.074		-0.040	
	$\beta^{[1]} = 0.3$	-0.960		-0.107		-0.043	
	$\beta^{[2]} = 0.1$	-2.696		-1.269		-0.661	
100	$\beta_0 = 0$	-0.001	100	0.006	100	0.005	100
	$\beta^{tt} = 0.5$	-0.042		-0.018		-0.004	
	$\beta^{[1]} = 0.3$	-0.463		-0.083		-0.054	
	$\beta^{[2]} = 0.1$	-1.079		-0.320		-0.084	
500	$\beta_0 = 0$	0.003	100	0.003	100	0.002	100
	$\beta^{tt} = 0.5$	-0.010		-0.007		-0.003	
	$\beta^{[1]} = 0.3$	-0.013		-0.009		-0.005	
	$\beta^{[2]} = 0.1$	-0.036		-0.036		0.000	
1000	$\beta_0 = 0$	0.000	100	0.000	100	0.000	100
	$\beta^{tt} = 0.5$	-0.002		-0.001		0.002	
	$\beta^{[1]} = 0.3$	-0.017		-0.011		-0.012	
	$\beta^{[2]} = 0.1$	-0.033		-0.017		-0.011	

Table 2.3

Simulation Results for $T = 5$

Sample Size	Parms	Partitioned GMM		Partitioned Bayes		GEE independent	
		Coverage	RMSE	Coverage	RMSE	Coverage	RMSE
25	$\beta_0 = 0$	60.82	1.020	93.00	0.407	93.89	0.366
	$\beta^{tt} = 0.7$	62.03	1.141	88.50	0.543	95.18	0.427
	$\beta^{[1]} = 0.5$	64.30	2.151	90.20	2.046	94.85	1.386
	$\beta^{[2]} = 0.3$	50.27	5.654	84.90	7.786	86.17	4.702
	$\beta^{[3]} = 0.1$	93.93	13.712	84.60	21.645	66.56	11.747
	$\beta^{[4]} = 0.05$	35.10	35.729	74.80	50.797	41.16	46.503
50	$\beta_0 = 0$	67.52	0.515	93.20	0.257	94.07	0.241
	$\beta^{tt} = 0.7$	69.15	0.708	92.10	0.305	94.41	0.268
	$\beta^{[1]} = 0.5$	66.47	1.295	93.40	1.119	93.61	0.965
	$\beta^{[2]} = 0.3$	48.25	3.108	90.70	3.888	88.25	3.755
	$\beta^{[3]} = 0.1$	34.58	7.810	87.50	8.839	73.32	10.432
	$\beta^{[4]} = 0.05$	39.13	20.199	79.50	22.355	47.43	41.755
100	$\beta_0 = 0$	71.40	0.491	94.30	0.166	94.70	0.161
	$\beta^{tt} = 0.7$	70.02	0.290	90.50	0.203	93.17	0.188
	$\beta^{[1]} = 0.5$	65.43	0.915	95.20	0.644	94.60	0.621
	$\beta^{[2]} = 0.3$	53.39	2.022	92.20	2.178	92.97	1.775
	$\beta^{[3]} = 0.1$	35.87	4.536	89.90	4.623	83.08	5.262
	$\beta^{[4]} = 0.05$	28.54	15.214	77.60	11.159	58.41	19.297
500	$\beta_0 = 0$	91.21	0.862	93.80	0.074	94.20	0.073
	$\beta^{tt} = 0.7$	91.25	0.089	93.10	0.079	94.10	0.078
	$\beta^{[1]} = 0.5$	91.14	0.365	95.80	0.251	95.50	0.252
	$\beta^{[2]} = 0.3$	80.52	1.013	94.20	0.619	94.30	0.612
	$\beta^{[3]} = 0.1$	62.69	2.363	95.80	1.375	93.00	1.391
	$\beta^{[4]} = 0.05$	50.64	4.962	95.10	3.179	85.40	3.282
1000	$\beta_0 = 0$	94.64	0.966	95.20	0.048	94.90	0.048
	$\beta^{tt} = 0.7$	94.10	0.061	85.20	0.066	85.80	0.065
	$\beta^{[1]} = 0.5$	94.41	0.218	93.90	0.193	93.00	0.195
	$\beta^{[2]} = 0.3$	88.93	0.676	95.60	0.398	96.20	0.398
	$\beta^{[3]} = 0.1$	79.69	1.434	97.50	0.900	97.30	0.898
	$\beta^{[4]} = 0.05$	62.81	3.522	89.50	2.095	91.20	2.025

Table 2.4

Bias and Percentage of Datasets that Converged in T=5

Sample Size	Parms	Partitioned GMM		Partitioned Bayes		GEE independent	
		Bias	% converge	Bias	% converge	Bias	% converge
25	$\beta_0 = 0$	0.033	74.80	-0.001	100	0.015	62.20
	$\beta^{tt} = 0.7$	-0.370		-0.214		-0.094	
	$\beta^{[1]} = 0.5$	-0.460		-0.591		-0.209	
	$\beta^{[2]} = 0.3$	-1.398		-0.027		-1.030	
	$\beta^{[3]} = 0.1$	-0.839		-4.769		0.177	
	$\beta^{[4]} = 0.05$	-0.102		5.225		-2.631	
50	$\beta_0 = 0$	0.025	85.90	-0.017	100	-0.017	87.70
	$\beta^{tt} = 0.7$	-0.220		-0.124		-0.064	
	$\beta^{[1]} = 0.5$	-0.227		-0.190		-0.091	
	$\beta^{[2]} = 0.3$	-0.553		-1.155		-0.599	
	$\beta^{[3]} = 0.1$	-0.870		-1.663		-1.641	
	$\beta^{[4]} = 0.05$	-0.302		1.745		-3.922	
100	$\beta_0 = 0$	0.024	91.60	-0.007	100	-0.006	98.10
	$\beta^{tt} = 0.7$	-0.104		-0.087		-0.052	
	$\beta^{[1]} = 0.5$	-0.206		-0.015		0.014	
	$\beta^{[2]} = 0.3$	-0.474		-0.501		-0.206	
	$\beta^{[3]} = 0.1$	-0.451		-0.626		-0.721	
	$\beta^{[4]} = 0.05$	-1.248		1.661		-2.499	
500	$\beta_0 = 0$	0.061	96.70	-0.005	100	-0.004	100
	$\beta^{tt} = 0.7$	-0.020		-0.025		-0.019	
	$\beta^{[1]} = 0.5$	-0.079		0.016		0.023	
	$\beta^{[2]} = 0.3$	-0.289		-0.027		0.027	
	$\beta^{[3]} = 0.1$	-0.540		-0.196		-0.171	
	$\beta^{[4]} = 0.05$	-0.287		-0.060		0.071	
1000	$\beta_0 = 0$	0.057	97.10	0.002	100	0.003	100
	$\beta^{tt} = 0.7$	-0.012		-0.026		-0.022	
	$\beta^{[1]} = 0.5$	-0.027		0.037		0.041	
	$\beta^{[2]} = 0.3$	-0.125		0.024		0.043	
	$\beta^{[3]} = 0.1$	-0.204		0.029		0.023	
	$\beta^{[4]} = 0.05$	-0.240		0.290		0.073	

2.5 Numerical Examples

An analysis of two numerical datasets [NACC (Beekly et al., 2007) Alzheimer's data and Add health (Harris & Udry, 2016) survey data] using Partitioned MVM marginal models with Bayes estimates is presented. When possible, a comparison of results with the results obtained using a Partitioned frequentist GMM model is made for convenience. The valid moment conditions are determined using a statistical test (Lalonde et al., 2014). The ***%partitioned*DataMatrix** SAS macro is utilized to obtain the valid moment conditions for time-dependent covariates with the fit of the Partitioned MVM marginal models with Bayes estimates.

Having obtained the partitioned data matrix, the Hamiltonian Monte Carlo sampling algorithm is used through the RStan (Gabry et al., https://cran.r-project.org/web/packages/rstan/index.html; Stan Development Team, https://mc-stan.org/docs/2_19/stan-users-guide/index.html) R-package to fit the Partitioned MVM marginal models with Bayes estimates. Three chains each with 1,000 burn in iterations and 1,000 sampling iterations with thinning=1 are used. The chain convergence is evaluated using visual plots, $\hat{R}$ statistic and effective samples sizes of 1,000 or higher. The Markov Chains for the models converge to the same posterior region. For the parameters, the effective sample size, the minimum number of independent draws from the posterior distribution, are achieved. The statistic, measuring whether the between-chain variance is considerably larger than the within-chain variance, $\hat{R}$ is equal to one. This suggests convergence of the chains (Givens & Hoeting, 2013). The code to fit the Partitioned MVM marginal models with Bayes estimates in R and SAS is given in the

Appendix. While the ***%Partitioned*GMM** SAS macro (Irimata & Wilson, 2018) is utilized to fit a frequentist Partitioned GMM model with time-dependent covariates.

2.5.1 National Alzheimer's Coordinating Center Data

Data from the National Alzheimer's Coordination Center (NACC) (Beekly et al., 2007) are fit using the Partitioned MVM marginal models with Bayes estimates and the Partitioned frequentist GMM model. The data include 1,106 patients, each with four annual visits between 2005 and 2015. Race, gender and age are time-independent covariates. Depression diagnosis, thyroid disease and traumatic brain injury diagnosis are time-dependent covariates.

Prior information pertaining to cognitive impairment and its relationships with the covariates in the model are based on what follows. Rocca et.al. (2011) found that men are significantly less likely to present cognitive impairment than women (OR=0.77 with 95% CI (0.66, 0.89). Blacks (OR=2.38 with 95% CI (1.97, 2.86)) are significantly more likely to have cognitive impairment when compared to whites. Paterniti et.al. (2002) discovered that those with depressive symptoms are significantly more likely to suffer cognitive impairment (OR=1.07 with 95% CI (1.01, 1.13)). Hogervorst et.al. (2008) showed that those with thyroid disorders are more likely to suffer cognitive impairment (OR=1.22 with 95% CI (0.88, 1.69)). LoBue et.al (2016) found that when adjusting for several patients' characteristics, those who had a history of traumatic brain injury are more likely to present mild cognitive impairment (OR=1.14 with 95% CI (0.94, 1.37)). Sheffield and Peek (2011) found that age increased the likelihood of suffering cognitive impairment (OR=1.12 with 95% CI (1.11, 1.13)). This information led to prior distribution for gender (male vs female) as $N(-0.26, 1)$, for race (white vs black) as $N(-0.87, 1)$ and for age as $N(0.12, 1)$. For the coefficients of depression, the prior is $N(0.07, 1)$, for thyroid disorders the prior is $N(0.20, 1)$ and for traumatic brain injury diagnosis the prior is $N(0.13, 1)$. Further, thyroid disorder and traumatic brain injury diagnosis had valid moment conditions at lag-1, lag-2 and lag-3. Depression diagnosis had valid moments at lag-1.

The regression coefficients in the Partitioned MVM marginal models with Bayes estimates for cognitive impairment, in the three Markov Chains converged to the same posterior region, Figure 2.1. A comparison of the prior and posterior distributions for the regression coefficients is made, Figure 2.2. The posterior distribution for the Partitioned MVM marginal models with Bayes estimates are noted in Table 2.5. The model suggests that gender showed differential effects on the probability of diagnosed with cognitive impairment. There is no evidence that diagnoses for thyroid disorder and traumatic brain injury impacted the probability of presenting cognitive disorder. While depression has an immediate and delayed effect on cognitive impairment. Instead, the frequentist GMM model provides unstable estimates and large standard errors for the regression coefficients suggesting non-convergence.

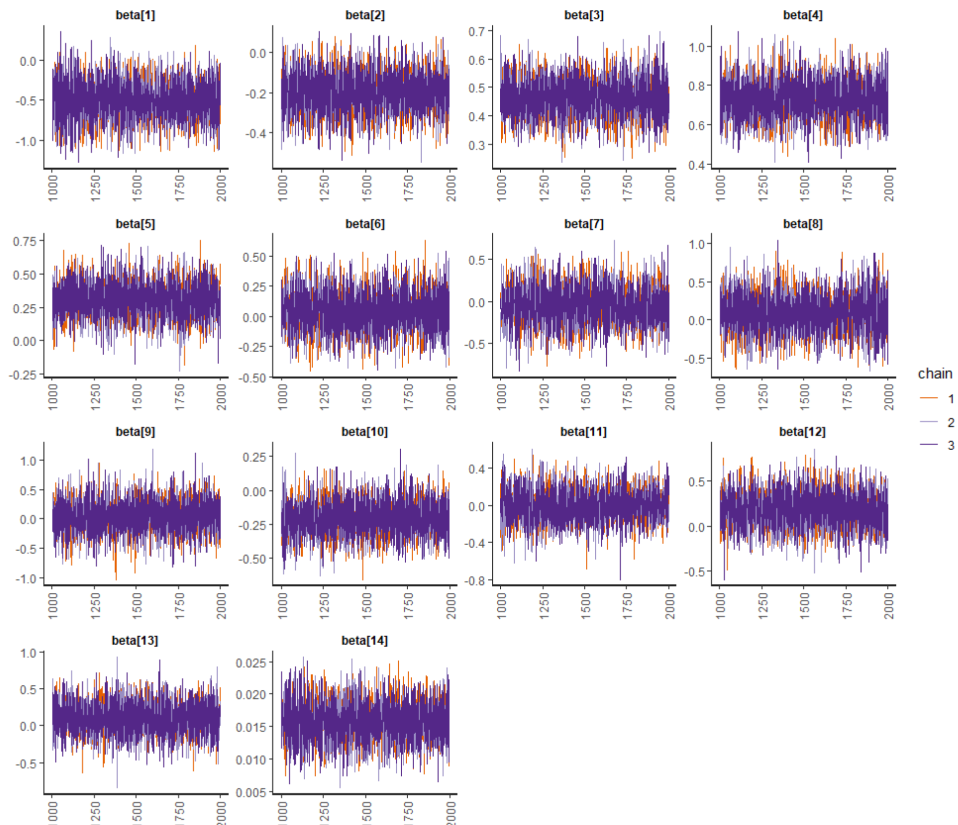
Table 2.5

Partitioned MVM Marginal Models with Bayes Estimates for Alzheimer's Data

Parameter	OR	95% Credible interval		ESS	Significant
White	0.819	0.670	1.000	4653	No
Male	1.584	1.391	1.822	4458	Yes
Age	1.020	1.010	1.020	3684	Yes
Depression	2.075	1.699	2.535	4453	Yes
depression lag-1	1.363	1.041	1.768	3570	Yes
TBI	1.041	0.756	1.448	3214	No
TBI lag-1	0.951	0.607	1.507	2619	No
TBI lag-2	1.094	0.657	1.822	3009	No
TBI lag-3	1.030	0.589	1.804	3615	No
Thyroid	0.811	0.631	1.051	3354	No
Thyroid lag-1	1.000	0.712	1.419	2958	No
Thyroid lag-2	1.197	0.811	1.768	3410	No
Thyroid lag-3	1.105	0.719	1.682	4271	No

Figure 2.1

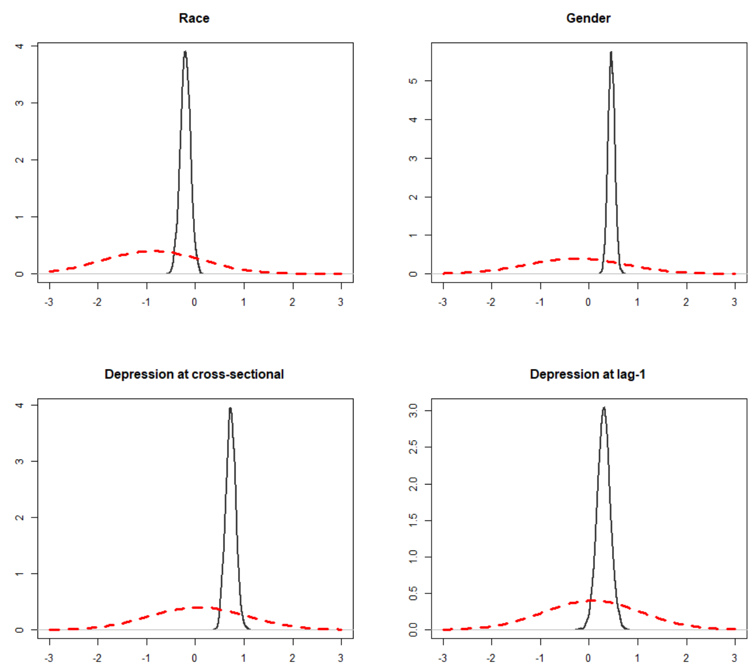
Markov Chains for Coefficients' Posterior Distributions for Cognitive Impairment Diagnosis Model

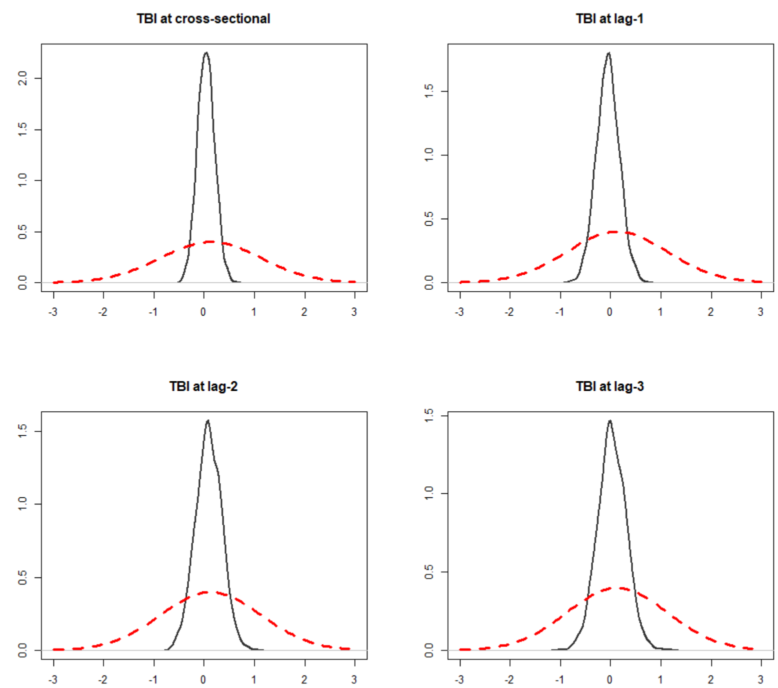


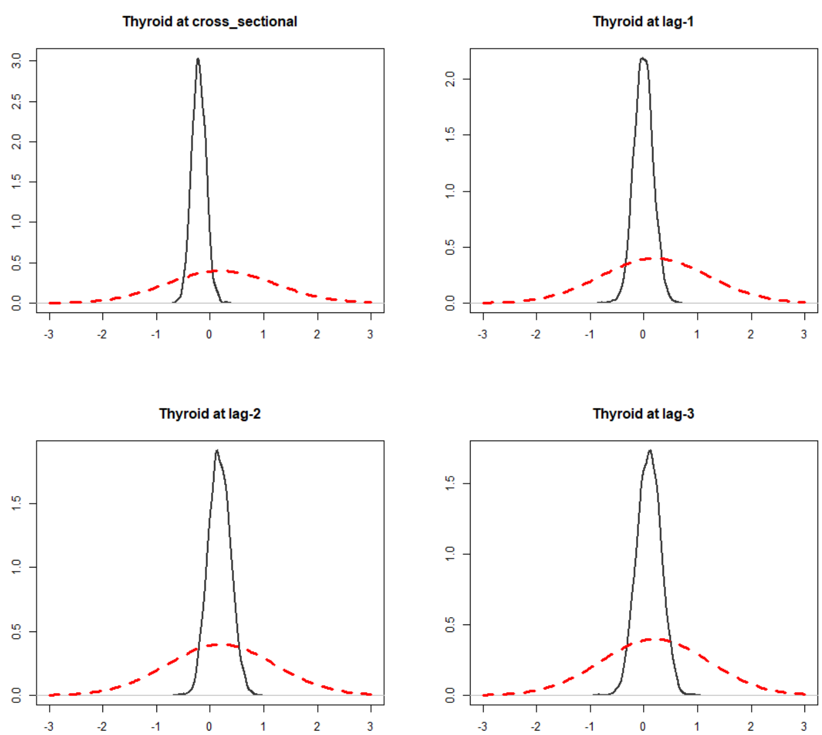
***Note:** beta[1]=intercept, beta[2]=white, beta[3]=male, beta[4]=depression, beta[5]=lag-1 depression, beta[6]=TBI, beta[7]=lag-1 TBI, beta[8]=lag-2 TBI, beta[9]=lag-3 TBI, beta[10]=thyroid, beta[11]=lag-1 thyroid, beta[12]=lag-2 thyroid, beta[13]=lag-3 thyroid, beta[14]=age.

Figure 2.2

Prior and Posterior Distributions for Regression Coefficients in Cognitive Impairment Diagnosis Model



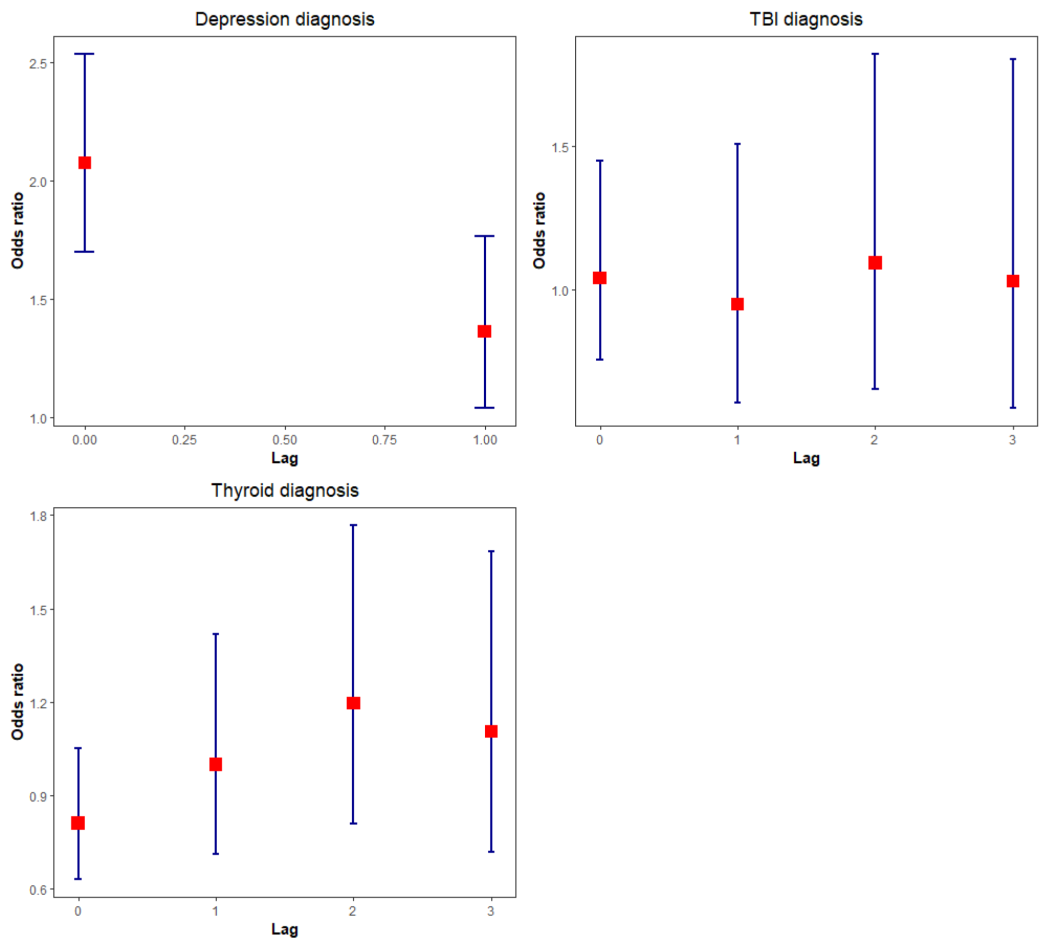

TBI at cross-sectional
TBI at lag-1
TBI at lag-2
TBI at lag-3


Thyroid at cross_sectional
Thyroid at lag-1
Thyroid at lag-2
Thyroid at lag-3

The point estimates and credible intervals for the odds ratios for the impact of depression, traumatic brain injury and thyroid disorder on cognitive impairment over time are displayed, Figure 2.3.

Figure 2.3

Posterior OR and 95% Credible Intervals for Time-dependent Covariates across Time for Cognitive Impairment Diagnosis



A prior sensitivity analysis allows a comparison between the posterior distributions of the regression coefficients based on different prior distributions. The graphs show a superimposition of the posterior distributions(Skene et al., 1986). A measure of prior sensitivity using the Hellinger distance to assess discrepancies between posterior distributions(Roos & Held, 2011) is obtained for modeling cognitive impairment. There are five prior distributions including informative priors with variance of 1 and used as the reference priors. There are three informative priors with the same means as the reference prior but with variances 0.25, 5 and 10. There are also non-informative priors, for regression coefficients with a mean of 0 and the variance of 10000. The Hellinger distances are close to 0, indicating that the posteriors are similar(Roos & Held, 2011). Thus, the choice of the mean for the prior distributions has little effect on the posterior distributions, Figure 2.4. The posterior distributions for informative priors with variance 0.25 are tighter than the others and had higher peaks at the mean or slight shift to the left. The Hellinger distances are smaller than 0.1, except for the posteriors corresponding to the informative priors with variance 0.25 for which Hellinger distances are higher than 0.1, Table 2.6.

Figure 2.4

Posterior Distributions for Prior Sensitivity Analysis for Regression Coefficients when Modeling Cognitive Impairment

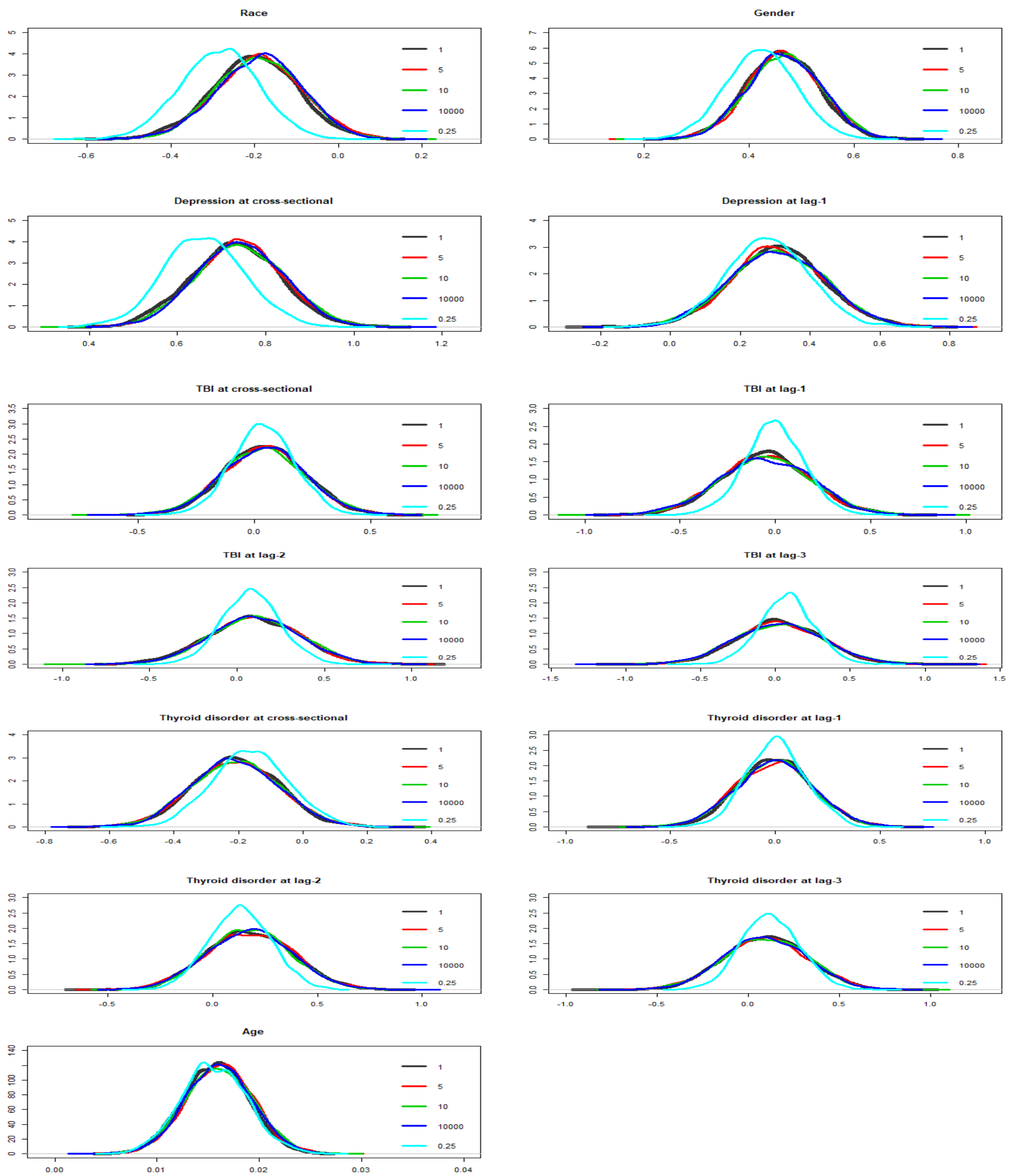


Table 2.6

Hellinger Distances in Posterior Distributions for Cognitive Impairment

Parameter	Noninformative	var=10	var=5	var=0.25
White	0.050	0.048	0.049	0.300
Male	0.033	0.040	0.037	0.211
Age	0.044	0.060	0.060	0.029
Depression	0.053	0.050	0.040	0.258
depression lag-1	0.036	0.034	0.034	0.097
TBI	0.042	0.052	0.035	0.124
TBI lag-1	0.047	0.046	0.033	0.202
TBI lag-2	0.031	0.046	0.032	0.216
TBI lag-3	0.037	0.046	0.033	0.224
Thyroid	0.033	0.039	0.036	0.139
Thyroid lag-1	0.038	0.036	0.042	0.119
Thyroid lag-2	0.032	0.032	0.031	0.188
Thyroid lag-3	0.035	0.036	0.039	0.173

Note: Posterior distributions for non-informative priors and informative priors with variances 5, 10 and 0.25 are compared to posterior distributions with informative priors with variance 1.

2.5.2 Add Health Study

Data obtained from the National Longitudinal Study of Adolescent to Adult Health are analyzed using the Partitioned MVM marginal models with Bayes estimates. There is a need to identify the relationship between several time-dependent covariates on obesity in adolescents until adulthood the United States of America. These data contain information on students in grades 7-12, collected initially in the academic year 1994-1995 (Harris & Udry, 2016). Repeated measures are taken in waves; at baseline and at three follow-up periods. The binary outcome, obesity status, is determined based on BMI. If BMI ≥ 30 then the outcome “obese” takes the value 1, and 0 otherwise. The time-dependent covariates are depression on a continuum, average number of hours watching TV per week, physical activity level, and social drinker status. Race (white vs non-white) is a time-independent covariate.

A statistical test identifies the valid moment conditions for these time-dependent covariates (Lalonde et al., 2014). The partitioning of the time-dependent covariates led to coefficients for measurements at cross-sectional, lag-1, lag-2 and lag-3. In particular, valid moment conditions for time-dependent covariates at lag-1. At lag-2, only the moment conditions for physical activity level and social drinker status are valid. At lag-3, only the moment condition for physical activity level is valid.

The prior relationships between obesity and the covariates are as follows. Caucasian adults are less likely to be obese than non-white (African Americans). Adults among non-depressed (b=-0.696, SE=0.125) and depressed (b=-0.383, SE=0.306) subpopulations (Lincoln et al., 2014). Depression increases the odds of developing obesity (OR=1.58 with 95% CI (1.33, 1.87)). Luppino et.al. (2010). Among children between 3-15 years old, physical activity has a negative relationship with the risk of being obese (OR=0.93 with 95% CI (0.87, 0.98)) (Hong et al., 2016). Children under 18 years of age, who watched TV for one hour daily (OR=1.29 with 95% CI (1.11, 1.50)), two hours daily (OR=1.64 with 95% CI (1.41, 1.89)), and three or more hours daily (OR=2.29 with 95% CI (1.97, 2.67)) are more likely to be obese than those who watched TV less than one hour per day (Singh et al., 2010). This information led to prior distributions for the regression coefficients. In cases where the standard errors for estimated effects in previous studies are too small, assign use standard errors of 1. Such a choice means that the prior distributions would completely cover the tails of the posterior distributions (Givens & Hoeting, 2013). The prior distribution for the regression coefficient for race is $N(-0.113, 1)$, the regression coefficients for depression is $N(0.457, 1)$, the regression coefficients for physical activity level is $N(-0.073, 1)$ and the regression coefficients for television hours is $N(0.255, 1)$. In the absence of consensus, data with conflicting results regarding social alcohol use as it relates to obesity (Traversy & Chaput, 2015), a non-informative prior distribution of $N(0,10000)$ is used.

Figure 2.5 presents trace plots for the three Markov Chains for the regression parameters. It shows that three chains converge to the same posterior region. Figure 2.6 includes a comparison between the prior and posterior distributions for the regression coefficients. It shows that the parameters have prior distributions which completely cover the tails of the posterior distributions. The posterior distribution for the marginal model has parameter as given in Table 2.7.

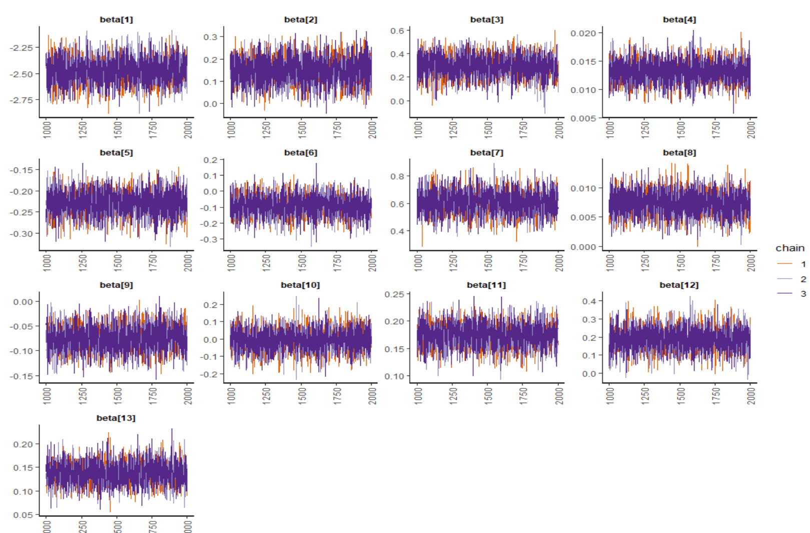
Table 2.7

Results of Partitioned MVM Model with Bayes Estimates for Add Health Data

Parameter	OR	95% Credible Interval		ESS
Race	1.162	1.030	1.297	3834
Alcohol	0.914	0.803	1.041	2950
Lag1 alcohol	0.990	0.878	1.127	3085
Lag 2 alcohol	1.209	1.062	1.377	3320
Depression	1.336	1.116	1.600	2241
Lag1 depression	1.840	1.553	2.160	2673
Tv hours	1.010	1.010	1.020	4308
Lag1 tv hours	1.010	1.000	1.010	4433
Physical activity	0.795	0.748	0.844	2932
Lag1 activity	0.923	0.878	0.970	2761
Lag 2 activity	1.185	1.139	1.234	2579
Lag 3 activity	1.150	1.094	1.197	2567

Figure 2.5

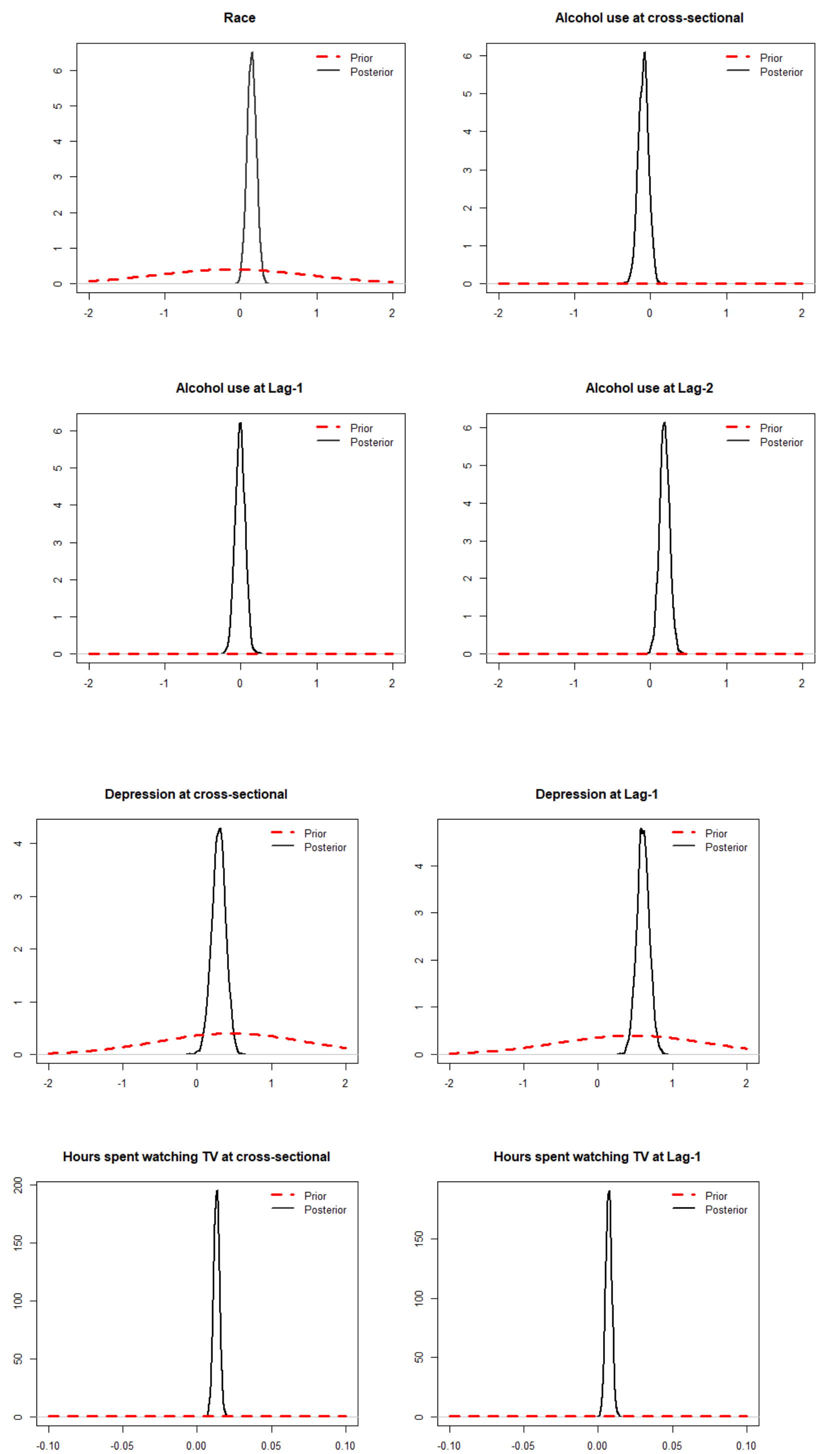
Markov Chains for Coefficients' Posterior Distributions for Obesity

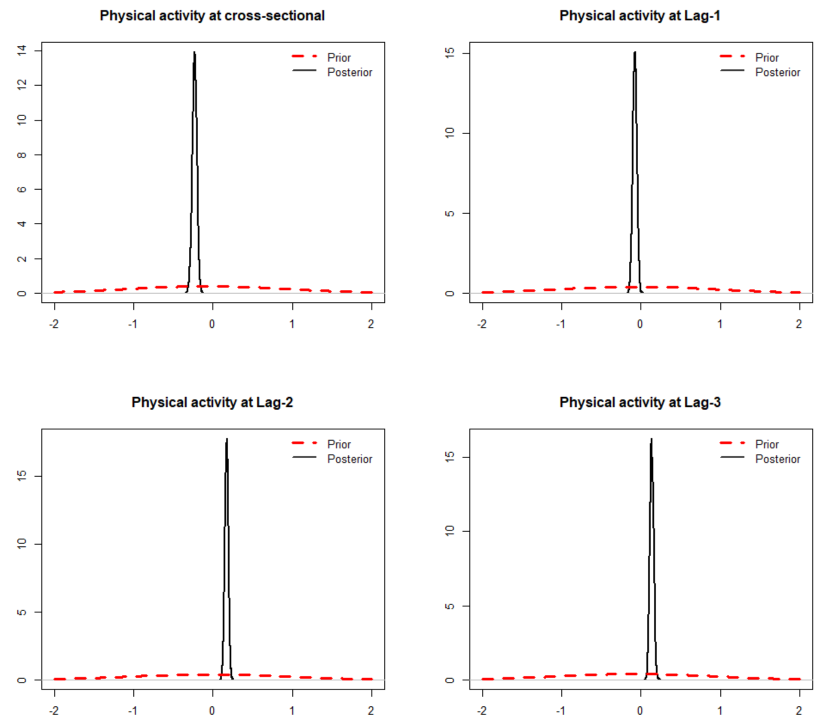


***Note:** beta[1]=intercept, beta[2]=white, beta[3]=depression, beta[4]=TV hours, beta[5]=physical activity, beta[6]=alcohol, beta[7]=lag-1 depression, beta[8]=lag-1 TV hours, beta[9]=lag-1 physical activity, beta[10]=lag-1 alcohol, beta[11]=lag-2 physical activity, beta[12]=lag-2 alcohol, beta[13]=lag-3 physical activity.

Figure 2.6

Prior and Posterior Distributions for Coefficients in Obesity Status



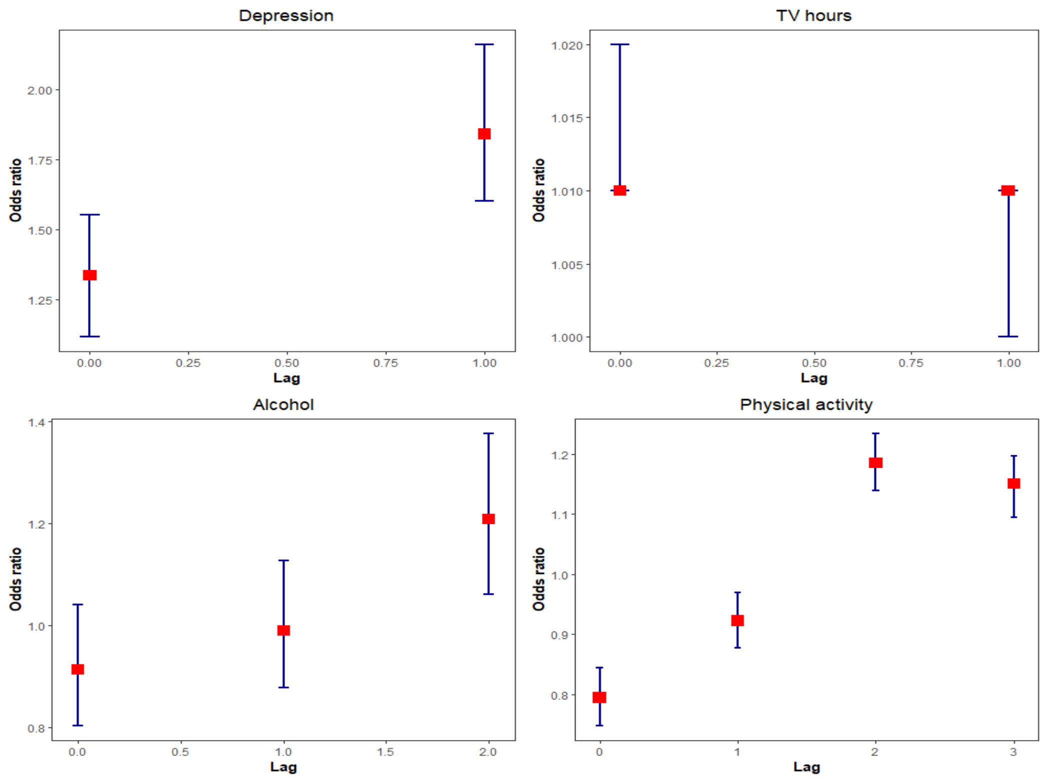


***Note:** Red dotted line represents prior distribution, black solid density represents the posterior distribution

The Partitioned MVM marginal models with Bayes estimates found that social drinking did not have an immediate effect or delayed effect on obesity status. However, there is a significant further delayed effect of social drinking on obesity. Depression had an immediate and delayed effect on obesity. The number of hours spent watching TV had a significant immediate effect on obesity. Physical activity level had an immediate, delayed effects, further delayed and furthermost delayed on obesity. We found that the impact of most time-dependent covariates changed over time, except for the number of hours spent watching TV, Figure 2.7.

Figure 2.7

Posterior OR and 95% Credible Intervals for Time-dependent Covariates



In this analysis, both Partitioned MVM marginal models with Bayes estimates and the Partitioned frequentist GMM produce similar results, Table 2.8. There are no issues with convergence in the Partitioned frequentist GMM as in the first example.

Table 2.8

Comparison of Partitioned MVM Marginal Models with Bayes Estimates and Partitioned GMM (Add Health)

Parameter	Bayesian		Frequentist	
	OR estimate	Significant	OR estimate	Significant
Race	1.162	Yes	1.249	Yes
Alcohol	0.914	No	1.010	No
lag1 alcohol	0.990	No	1.047	No
lag 2 alcohol	1.209	Yes	1.343	Yes
Depression	1.336	Yes	1.650	Yes
lag1 depression	1.840	Yes	1.790	Yes
tv hours	1.010	Yes	1.015	Yes
lag1 tv hours	1.010	No	1.004	No
physical activity	0.795	Yes	0.848	Yes
lag1 activity	0.923	Yes	0.909	Yes
lag 2 activity	1.185	Yes	1.197	Yes
lag 3 activity	1.150	Yes	1.171	Yes

A prior sensitivity analysis is conducted for modeling obesity in the Add health dataset similar to the methods in the previous example. Figure 2.8 shows the superimposed posterior distributions based on several prior distribution choices for the obesity status model. It reveals that the choice of mean or variance for prior distributions have little effect on the posterior distributions. This is validated in Table 2.9. The Hellinger distances between posterior distributions for reference priors and the other priors are smaller than 0.1.

Figure 2.8

Posterior Distributions for Prior Sensitivity for Regression Coefficients of Obesity

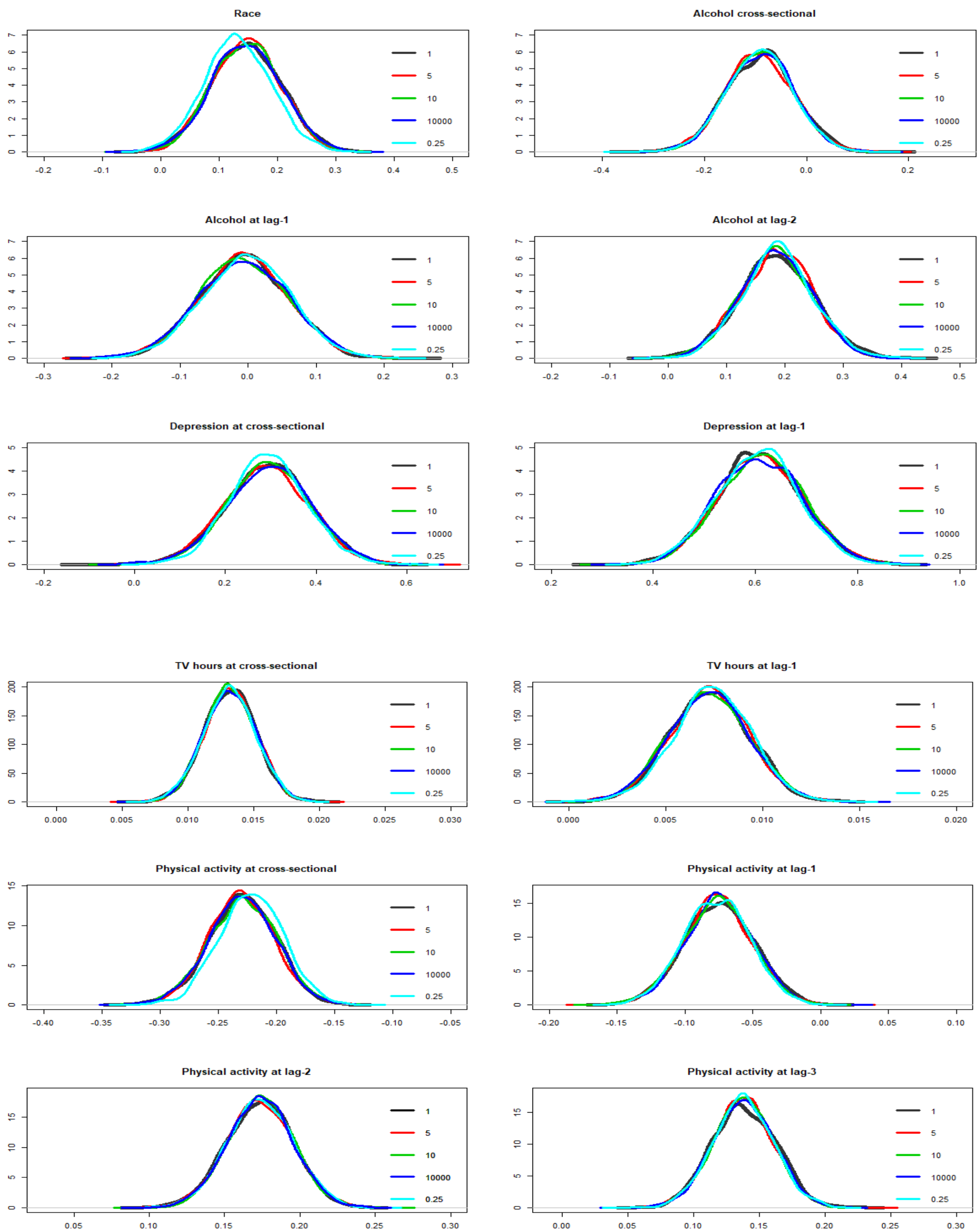


Table 2.9

Hellinger Distances between Posterior Distributions for Obesity

Parameter	Noninformative	var=10	var=5	var=0.25
Race	0.025	0.027	0.034	0.089
Alcohol	0.029	0.029	0.040	0.026
lag1 alcohol	0.031	0.034	0.027	0.036
lag 2 alcohol	0.044	0.043	0.054	0.044
Depression	0.031	0.026	0.041	0.058
lag1 depression	0.037	0.035	0.028	0.034
tv hours	0.025	0.029	0.033	0.027
lag1 tv hours	0.032	0.029	0.037	0.042
physical activity	0.019	0.029	0.035	0.098
lag1 activity	0.032	0.032	0.038	0.038
lag 2 activity	0.024	0.040	0.024	0.019
lag 3 activity	0.036	0.049	0.047	0.044

Note: Posterior distributions for non-informative priors and informative priors with variances 5, 10 and 0.25 are compared to posterior distributions with informative priors with variance 1.

2.6 Discussion

The Partitioned MVM marginal model with Bayes estimates for time-dependent covariates addresses the correlation due to time-dependent covariates through an identification of valid moments from a derived Partitioned matrix. It consists of additional regression covariates and as such, it also addresses the situation of few valid moments at times and avoids non-convergence.

The coefficients provide a model with estimates as efficient to those from the GEE with lagged covariates and working correlation matrix with independence, or as good as the frequentist GMM marginal model. It takes advantage of the Bayesian principles. It is flexible and attainable in obtaining estimates of the regression coefficients for time-dependent covariates. It achieves fast computation and avoiding the issue of non-convergence which remains a challenging problem for the lagged modeling approaches.

Acknowledgements

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

CHAPTER 3

MODELING SIMULTANEOUS RESPONSES WITH NESTED WORKING CORRELATION AND BAYES ESTIMATES FOR DATA WITH TIME-DEPENDENT COVARIATES

Abstract

In the analysis of longitudinal data, it is common to characterize the relationship between the (repeated) response measures and the covariates. However, when the covariates do vary over time (time-dependent covariates) there is extra relation due to the delayed effects that need to be accounted for. Moreover, these studies often consist of simultaneous responses from a subject. However, a joint likelihood function of the simultaneous responses is impossible to determine and so maximum likelihood estimates are unattainable as the observations are correlated. In this paper, a simultaneous modeling of multiple response variables, using a hierarchical working correlation matrix, with Bayesian regression estimates on a partitioned data matrix is developed. A simulation study demonstrating the benefits of this model is conducted. The method using data from Chinese Quality of Health survey data was analyzed. We provide code in R and a SAS macro.

3.1 Introduction

Hierarchically structured data are common in survey data. In particular, clustering of subjects (e.g., patients clustered within hospitals) and longitudinal repeated measurements taken from a subject lead to hierarchical data structure. When analyzing such interdependent data as longitudinal repeated measures, the sampling units are independent, but the repeated measurements on each sampling unit over time are correlated. For clustered or longitudinal binary data, the joint likelihood of the repeated measurements on a subject is typically difficult to formulate. In such cases, we often rely on a quasi-likelihood thereby concentrating on the variance-mean relation through marginal models that help us understand what impacts the mean outcome of the population of interest. The fit using the so-called generalized estimating equation (GEE) (Liang & Zeger, 1986), leads to a robust method that produces consistent and asymptotic normal estimators even with a miss specified working correlation matrix is widely known to statisticians. If the correlation structure is correctly specified, the GEE estimator is efficient, notwithstanding, it gives us a population-averaged or marginal model.

Marginal models are often not the choice with repeated measures data as researchers often resort to subject specific models, in other words, modeling a function of the mean outcome conditional on individual subjects (Hu et al., 1998). Subject-specific models are based on two or more parametric distributions, one or more associated with the random effects and one for the outcomes conditional on the random effects. However, while random effects models are frequently used in modeling longitudinal data, they answer the question about the mean response of an individual with conditional results, rather than the one initially posed by the marginal model. In fact, these random coefficient models are conditioned on the random effects, which largely are random so the notion of conditional on it is sometimes tough to wrap one's mind around (Laird & Ware, 1982; Ware, 1985).

The method of regression estimates using the generalized method of moments (GMM) is a popular technique in econometrics modeling (Hall, 2005; Hansen, 1982; Hansen et al., 1996). The use of GMM estimators and the related asymptotic theory through population moment conditions have gotten lots of attention in statistical research (Hansen, 1982; McFadden, 1989). The GMM method achieves estimation efficiency when the likelihood is difficult or impossible to work with, as is the case with correlated binary observations. The GMM is also used to make inferences to semiparametric models where there are more moment conditions than unknown parameters.

Qu et al. (2000) and Lai and Small (2007) presented a GMM marginal regression model to analyze longitudinal data and demonstrated its advantages over the GEE model. Lalonde, Wilson, and Yin (2014) extended Lai and Small's work in presenting a method to identify valid moments. Interestingly, Irimata, Broatch, and Wilson (2019) presented a reconfiguration of the data matrix to address the correlation due to the varying impact of the covariates on the response at present and future times. The reconfiguration addresses the correlation due to the time-dependent covariates.

These methods have been used when modeling a single response of interest. However, many longitudinal studies have multiple responses of interest measured at each time point. Studies that observe multiple binary responses may focus on the simultaneous occurrence of two or more of these responses (McCulloch, 2008). Sometimes the multiple responses of interest are considered a joint measure of a subject's trait over time. Lipsitz et.al. (2009) studied a joint measure of heart function over time based on the binary responses abnormal heart rate, abnormal blood pressure and abnormal heart wall thickness.

When longitudinal studies focus on the simultaneous occurrence of multiple responses but the analysis focuses on a single response at a time, it provides an incomplete picture (Lipsitz et al., 2009). Fieuws et.al. (2007) and Lipsitz et.al.(2009) argued that to understand longitudinal change in simultaneous responses, one must model them jointly over time. A joint model for multiple longitudinal responses must account for the correlation between repeated measurements on the same response within subjects, and should allow for correlation between measurements on the different responses (Fieuws et al., 2007; Lipsitz et al., 2009). When time-dependent covariates are present, one should also account for the changing effects of such covariates on one or more of the simultaneous responses of interest (Irimata et al., 2019). There are times when one or more time-dependent covariates are thought to affect all of the simultaneous responses. The simultaneous modeling allows the assessment of the overall impact as well as the separate and joint effect of such covariates on the responses over time (McCulloch, 2008).

In this paper, a model is presented that addresses time-dependent covariates effects on simultaneous binary responses. It is a marginal regression model for multiple binary outcomes that relates the marginal probability of each outcome to a set of covariates over time. It provides partitioned regression coefficients for time-dependent covariates. The regression parameters are estimated for each outcome, and the model does not assume all outcomes have the same regression parameters. Also this model depends on the use of Bayesian principles to obtain estimates of regression coefficients as applied to the reconfiguration of the data matrix (Irimata et al., 2019).

The remainder of this paper is organized as follows. In Section 2, a review of the GMM partitioning of the data matrix to address the time-dependent covariates. In Section 3, a simultaneous modelling of responses with Bayes estimates of the coefficients with a hierarchical working correlation on the responses is derived. In Section 4, a simulation study to demonstrate the properties of the simultaneous model coefficients is conducted. The fit of the proposed model is demonstrated with a numerical example using the Chinese Longitudinal Healthy Longevity Survey data in Section 5.

3.2 Background

Consider a longitudinal data structure comprised of one response variable y_{it} and a vector of J covariates $\boldsymbol{x}_{i*t} = (x_{i1t}, \ldots, x_{iJt})$, observed at times $t = 1, \ldots, T$; for subjects $i = 1, \ldots, N$. Such longitudinal data in which measurements are collected repeatedly, with certain regularity, are common in health and health related fields and in social science research to name a few. Assume there is missing data, the missing is completely at random assumption (Lai & Small, 2007; Little & Rubin, 2002). Then, whether a subject's data are missing at a given time t is conditionally independent, given the subject's covariates at time t, $\mathbf{x}_{i*t}$, of the subject's missing outcomes, past outcomes, future outcomes and the covariates at past or future time points (Little & Rubin, 2002).

Without loss of generality and for convenience, assume each subject is observed at each time point. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$ be the $T \times 1$ vector of outcome values associated with the $J \times 1$ covariate vectors $\boldsymbol{x}_{i*1}, \ldots, \boldsymbol{x}_{i*T}$ for the i^{th} subject. For $i \neq i'$, assume $\boldsymbol{y}_i$ and $\mathbf{y}_{i'}$ are independent, but generally the components of $\boldsymbol{y}_i$ are correlated.

At time t, let the i$^{\text{th}}$ subject $(i = 1, 2, \ldots, N)$ be observed with response y_{it} and let $\boldsymbol{x}_{i*t}$ be the corresponding vector of J covariates $(j = 1, 2, .., J)$. Characterize the relationship between y_{it} and $\boldsymbol{x}_{i*t}$ as if the observed values y_{it} come from a distribution belonging to the exponential family. Thus, the density function of y_{it} given $\boldsymbol{x}_{i*t} = x_{i1t}, \ldots, x_{iJt}$ takes the form of a random variable with mean μ and variance σ^2 as a member of the exponential family:

$$\mathrm{f}(y_{it}|\theta_{it}, \phi) = \exp\left\{\frac{(y_{it}\theta_{it} - b(\theta_{it}))}{a(\phi)} + c(y_{it}, \phi)\right\} \tag{2.1}$$

where θ_{it} is the canonical mean parameter, ϕ is the dispersion parameter, and the functions a, b, and c are known. Further,

$$\mu_{it} = E(y_{it}|\boldsymbol{x}_{i*t}) = b'(\theta_{it})$$

and

$$\sigma^2 = \mathrm{a}(\phi)\mathrm{b}''(\theta_{\mathrm{it}})$$

where b' denotes the first derivative and b'' denotes the second derivative. Thus, the mean and variance are related

$$\mathrm{Var}(y_{it}) = a(\phi)\mathrm{V}(\mu_{it})$$

where

$$a(\phi) = \frac{\phi}{w},$$

where w is the weight and $V(\mu_{it})$ is the variance function (McCullagh & Nelder, 1989). Barndorff-Nielsen (1978) and Blaesild (1985) studied generalized linear models and showed that a, b, and θ_{it} are related. A generalized linear model (GLM) consists of a unified framework for various discrete and continuous outcomes (McCullagh & Nelder, 1989).

The generalized estimating equation (GEE) model relies on working correlation structure which can depend on an unknown $s \times 1$ parameter vector $\boldsymbol{\alpha}$. The dimension of the matrix depends on the number of repeats and correlation strength can differ from subject to subject, but the type of the correlation matrix Σ_i for the i^{th} subject is fully specified by $\boldsymbol{\alpha}$. The working covariance matrix of $\boldsymbol{y}_i$

$$\boldsymbol{V}_i = A_i^{\frac{1}{2}} R(\boldsymbol{\alpha}) A_i^{\frac{1}{2}} / \lambda$$

Then, we define the generalized estimating equations as

$$\sum_{i=1}^{N} U_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{S}_i$$

where $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{S}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})$. The generalized estimating equations are similar to the function presented from the quasi-likelihood approach except that in this case $\boldsymbol{V}_i$ is a function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Liang and Zeger (1986) showed conditions under which $\widehat{\boldsymbol{\beta}}$ satisfies $\sum_{i=1}^{N} U_i(\widehat{\boldsymbol{\beta}}, \boldsymbol{\alpha}) = 0$. They include the assumption that the estimating equation is asymptotically unbiased and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically multivariate Gaussian under suitable regularity conditions. They showed that $\widehat{\boldsymbol{\beta}}$ is consistent regardless of whether the actual correlation matrix of is $R_i(\boldsymbol{\alpha})$. Although correct specification of the working correlation structure does not affect consistency, correct specification enhances efficiency.

3.2.1 Marginal Regression Modeling with Time-Dependent Covariates

Lai and Small (2007) fitted marginal models to continuous data with time-dependent covariates through grouping of the valid moments. They presented a marginal model for longitudinal continuous data with GMM estimates through the moment conditions

$$\mathrm{E}\left[\frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \beta_j}\{y_{it} - \mu_{it}(\boldsymbol{\beta})\}\right] = 0 \qquad (2.2)$$

for appropriately chosen times s, t, and predictor j. Let

$$\mu_{it}(\boldsymbol{\beta}) = \mathrm{E}[\{y_{it}|\boldsymbol{x}_{i*t}\}]$$

denote the expectation of y_{it} based on the vector of covariate values $\boldsymbol{x}_{i*t}$ associated with the vector of parameters $\boldsymbol{\beta}$ in the systematic component that describes the marginal distribution of y_{it}. Lai and small (2007) classified time-dependent covariates into three types and determined valid moment conditions based on such classifications. They combined the valid moments to obtain an estimate of a single regression coefficient to represent the overall effect of a given covariate. Lalonde, Wilson, and Yin (2014) introduced a method that identifies valid moments. Other approaches to address the problem include Zhou, Lefante, Rice, and Chen (2014) who utilized a modified quadratic inference approach, while Chen and Westgate (2017) chose a modified weight matrix based on linear shrinkage, and Müller and Stadtmuller (2005) used a generalized functional linear regression model.

3.2.2 Partitioned Coefficients with Time-Dependent Covariates

Irimata, Broatch and Wilson (2019) provided a method based on time-dependent covariates with the use of GMM estimates based on a reconfiguration of the data matrix. Their method identifies valid moments for time-dependent covariates one at a time. However, instead of combining the valid moments into one regression coefficient, they are held separately to provide measures for the cross sectional (immediate effect) effect, lag-1 (delayed) effect, and so on. This method is unique as it differs from other models, which do not separate out the regression coefficients.

While these additional regression coefficients have the advantage of interpretability. At times, the number of valid moments may not be sufficient to provide consistent estimates (Lai & Small, 2007). This problem of having too few equations and not having enough data to obtain the generalized method of moment's estimates for the regression coefficients are addressed with Bayesian estimates.

This partitioning of the data matrix is convenient and readily attainable, as it provides an alternative but interpretable approach to modeling correlated data with time-dependent covariates (Irimata et al., 2019). It produces additional regression parameters for each time-dependent covariate, and provides valuable insight into time-varying relationships.

Irimata, Broatch and Wilson (2019) address a marginal model for cases when there are time-dependent covariates. The valid moments are consistent with the reconfiguration of the data matrix (or partitioned data matrix). The data matrix has dimension depending on the number of repeated measures on the response. The reconfigured matrix is partitioned into a lower triangular matrix. These valid moments are identified as a result of the uncorrelated relationship between the residuals $(Y_t - \mu_t)$ at time t, and the j^{th} covariate value X_{ijs} observed at time s, for $s \leq t$. Thus, for subject i with the j$^{\text{th}}$covariate, each time-dependent covariate $\boldsymbol{X}_{ij}$ is measured at times $1, 2, \ldots, T$, and the partitioned data matrix is reconfigured as a lower triangular matrix,

$$\boldsymbol{X}_{ij} = \begin{bmatrix} 1 & X_{ij1} & 0 & \ldots & 0 \\ 1 & X_{ij2} & X_{ij1} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & X_{iJT} & X_{ij(T-1)} & \ldots & X_{ij1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X}_{ij}^{[0]} & \boldsymbol{X}_{ij}^{[1]} & \ldots & \boldsymbol{X}_{ij}^{[T-1]} \end{bmatrix}$$

where the superscript denotes the difference, $t - s$ in time-periods between the response time t and the covariate time s. Thus, a model for subject i at time t with one time-dependent covariate, $\boldsymbol{X}_{ij}$ and no time-independent covariates is

$$g(\mu_{it}) = \beta_0 + \beta_j^{tt} X_{ijt} + \beta_j^{[1]} X_{ij,t-1} + \beta_j^{[2]} X_{ij,t-2} + \cdots + \beta_j^{[T-1]} X_{ij1} \qquad (2.3)$$

In matrix notation $\text{g}(\boldsymbol{\mu_i}) = \boldsymbol{X}_i' \boldsymbol{\beta}$, where the $\boldsymbol{X}_i$ is the subject's matrix of covariates denoting the systematic component of the model with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})'$ that depends on the regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)$, the concatenation of the parameters associated with each of the J covariates. Where $\boldsymbol{\beta_j} = \left(\beta_j^{tt}, \beta_j^{[1]}, \beta_j^{[2]}, \ldots, \beta_j^{[T-1]}\right)$ and $j = 1, \ldots, J$. The coefficient β_j^{tt} denotes the effect of the covariate $\boldsymbol{X}_{*jt}$ on the response Y_t during the t$^{\text{th}}$ period, or in other words, when the covariate and the outcome are observed in the same time-period. When $s < t$, we denote the lagged effect of the covariate $\boldsymbol{X}_{*js}$ on the response Y_t by the coefficients $\beta_j^{[1]}, \beta_j^{[2]}, \ldots, \beta_j^{[T-1]}$. These additional coefficients allow the effect of the covariate on the response to change across time and to be addressed separately, rather than assuming that the association maintains the same strength and direction over time. For example, the coefficient $\beta_j^{[1]}$ denotes the effect of $\boldsymbol{X}_{*js}$ on Y_t across a one time-period lag. In general, each of the J time-dependent covariates produce a maximum of T partitions of $\boldsymbol{\beta}_j$. Thus, for a model with J covariates, the data matrix $\boldsymbol{X}$ will have a maximum dimension of $N \times \big((J \times T) + 1\big)$ and $\boldsymbol{\beta}$ is a vector of maximum length $\big((J \times T) + 1\big)$.

The moment conditions when $s = t$ have been proven to be always valid (Lai & Small, 2007; Pepe & Anderson, 1994), thus the cross-sectional regression coefficient, β_j^{tt} is always guaranteed to be estimable. However, as the lagged regression coefficients $\beta_j^{[1]}, \dots, \beta_j^{[T-1]}$ rely on moment conditions evaluated using hypothesis testing, they may not be estimable if no valid moment conditions are identified.

3.2.3 Bayes Inferences

This scenario with few cases for valid moment conditions led to the consideration of Bayes principles. Let $D = \{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{N}$ be a dataset, where $\boldsymbol{X}_i$ is a vector of covariates and Y_i is a response. Let $\pi_0(\boldsymbol{\beta})$ be a prior density on a parameter vector $\boldsymbol{\beta}$, and let $\mathrm{P}(Y_i|\boldsymbol{X}_i, \boldsymbol{\beta})$ be the likelihood of observation i given the parameter $\boldsymbol{\beta}$. The Bayesian posterior is calculated as the density

$$P_N(\boldsymbol{\beta}) = \frac{\exp\big(\Psi_N(\boldsymbol{\beta})\big)\pi_0(\boldsymbol{\beta})}{\varepsilon_N}$$

where $\Psi_N(\boldsymbol{\beta}) = \sum_{i=1}^{N} \ln P(Y_i|\boldsymbol{X}_i, \boldsymbol{\beta})$) is the model log-likelihood and

$$\varepsilon_N = \int \exp\big(\Psi_N(\boldsymbol{\beta})\big)\pi_0(\boldsymbol{\beta})d\boldsymbol{\beta}$$

is the marginal likelihood. Thus, we rely on a dataset

$$\widetilde{D} = \left\{\left(\zeta_i, \tilde{X}_i, \tilde{Y}_i\right)\right\}_{i=1}^{M}$$

with $M \leq N$ such that the weighted log-likelihood

$$\widetilde{\Psi}_N(\boldsymbol{\beta}) \sum\nolimits_i^M \zeta_i \ln\ P(\tilde{Y}_i | \widetilde{X}_i, \boldsymbol{\beta})$$

satisfies

$$\left|\Psi_N(\boldsymbol{\beta}) - \widetilde{\Psi}_N(\boldsymbol{\beta})\right| \leq \varepsilon|\Psi_N(\boldsymbol{\beta})|\ \forall \boldsymbol{\beta} \epsilon \boldsymbol{\Theta} \qquad (2.4)$$

If $\widetilde{D}$ satisfies (2.4) then it is an ε-coreset of D, and the approximate posterior

$$\tilde{P}_N(\boldsymbol{\beta}) = \frac{\exp\left(\widetilde{\Psi}_N(\boldsymbol{\beta})\right)\pi_0(\boldsymbol{\beta})}{\tilde{\varepsilon}_N}$$

with

$$\tilde{\varepsilon}_N = \int \exp\left(\widetilde{\Psi_N}(\boldsymbol{\beta})\right)\pi_0(\boldsymbol{\beta})d\boldsymbol{\beta}$$

has a marginal likelihood $\tilde{\varepsilon}_N$ which approximates the true marginal likelihood ε_N. This follows (Huggins et al., 2016), that if

$$\left|\Psi_N(\boldsymbol{\beta}) - \widetilde{\Psi}_N(\boldsymbol{\beta})\right| \leq \varepsilon|\Psi_N(\boldsymbol{\beta})|\ \forall \boldsymbol{\beta} \epsilon \boldsymbol{\Theta}$$

then for any prior $\pi_0(\boldsymbol{\beta})$ such that the marginal likelihoods

$$\tilde{\varepsilon}_N = \int exp\left(\widetilde{\Psi_N}(\boldsymbol{\beta})\right)\pi_0(\boldsymbol{\beta})d\boldsymbol{\beta}$$

are finite, the marginal likelihoods satisfy $|\ln\varepsilon - \ln\tilde{\varepsilon}_N| \leq \varepsilon|\ln\varepsilon|$.

3.3 MVM Marginal Model with Bayes Estimates

The longitudinal studies are often designed to study changes which are measured repeatedly over time (Fieuws et al., 2007). They are often multivariate in responses (several response variables) having repeated measurements on the different responses for each subject. When modeling multiple longitudinal outcomes one should account for two types of correlation: correlations between measurements on different outcomes and correlations among measurements on the same outcomes (Gueorguieva, 2001).

One approach to modeling simultaneous longitudinal outcomes, is to use a set of latent, unobserved, random effects to address correlation for two or more responses (Ghebremichael, 2015; Fang, Sun & Wilson, 2018). Such models assume that the responses shared a common unobservable feature, and as such account for the correlation between the measures taken from the same subject. However, the proposed model takes a different approach, one that would allow interpretation to remain on the marginal mean and not to divert to the conditional mean.

The proposed model accounts for two sets of correlation, one between the simultaneous responses and another for current and future effects of time dependent covariates. The correlation among the simultaneous responses is modeled through a working correlation matrix for within-outcome correlations (e.g., caused by repeated measures of the outcome across time) and an outer working correlation matrix for between-outcome correlations (e.g., caused by outcome responses taken from the same subject). Thus, we adjust the random component as we would in the usual generalized estimating equations.

3.3.1 Simultaneous Responses with Nested Working Correlation Matrix

Consider a model for simultaneous responses that accounts for the two types of correlation among these responses, while addressing the time-dependent covariates. In this model, (on the left side of the model) time is nested within responses and responses are nested within subjects. This gives rise to a two-part GEE partitioned model with normal priors in pursuit of a posterior distribution while making use of a nested working correlation matrix to address the responses.

Let Y_{irt} denote the r^{th} response $(r = 1, \dots R)$ from the i^{th} subject $(i = 1, \dots, N)$ at the t$^{\text{th}}$ period $(\text{t} = 1, \dots, \text{T})$. Then Y_{irt} is a binary random variable and takes on a value of 1 (event) or a value of 0 (non-event). For the i$^{\text{th}}$ subject, measured *T* times, on the r$^{\text{th}}$ response denote the vector of length *T* as $\boldsymbol{Y_{ir}} = (Y_{ir1} \quad Y_{ir1} \dots. \quad Y_{irT})'$. Then, for the i$^{\text{th}}$ subject measured *T* times on the *R* responses there is a vector $\boldsymbol{Y_i} = (\boldsymbol{Y_{i1}} \quad \boldsymbol{Y_{i2}} \dots. \quad \boldsymbol{Y_{iR}})'$ of length $(R \times T)$. Assume that for the vector of length T, $\boldsymbol{Y_{ir}}$, there is a set of covariates (time-independent and time-dependent) that generate its own partitioned data matrix $\boldsymbol{X_{ir}^{[\,]}}$. Let $\boldsymbol{X_i^{[\,]}}$ be a block diagonal design matrix for subject i, composed of the partitioned data matrices of covariates $\boldsymbol{X_{ir}^{[\,]}}$ for $r = 1, \dots, R$; with associated regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta_1} \quad \dots.. \quad \boldsymbol{\beta_R})'$, where $\boldsymbol{\beta_r}$ $(r = 1, \dots R)$ is the vector of regression coefficient associated with $\boldsymbol{X_{ir}^{[\,]}}$ for the r^{th} response.

The *R* simultaneous responses from the same subject i have two levels of correlation, between responses and within responses. The *R* responses are measured *T* times and this gives rise to a square matrix of correlations Σ of dimension $R \times T$ such that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1T} \\ \vdots & \ddots & \vdots \\ \Sigma_{1T} & \cdots & \Sigma_{TT} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{rr'} = \begin{bmatrix} 1 & \cdots & \rho_{1r}^{r} \\ \vdots & \ddots & \vdots \\ \rho_{1r}^{r} & \cdots & 1 \end{bmatrix}.$$

The $\boldsymbol{\Sigma}_{rr'}$ is seen as the innermost correlation or within correlation, while Σ contain the outmost correlation or between responses. The within correlations are the association within each of *R* responses. Then, there is the correlation between the *R* responses. The relations between and within, may each take on different correlation strength but the structure remains the same. This approach to simultaneous modelling of responses allows us to address a set of correlated responses as a marginal or population-averaged model with time-dependent covariates.

To illustrate how the working correlation matrix works in this model, consider an example where data are collected on three binary outcomes, (Y_1, Y_2, Y_3), at three different time periods for each subject. The vector of outcomes for each subject is $\boldsymbol{Y}_i' = (Y_{i11}, Y_{i12}, Y_{i13}, Y_{i21}, Y_{i22}, Y_{i23}, Y_{i31}, Y_{i32}, Y_{i33})$. The correlation matrix contains components measuring the two levels of correlation among the outcomes and is of the form

$$
\begin{array}{c|ccc|ccc|ccc|}
 & Y_{11} & Y_{12} & Y_{13} & Y_{21} & Y_{22} & Y_{23} & Y_{31} & Y_{32} & Y_{33} \\
\hline
Y_{11} & 1 & \upsilon_{11,12} & \upsilon_{11,13} & \alpha_{11,21} & \tau_{11,22} & \tau_{11,23} & \alpha_{11,31} & \tau_{11,32} & \tau_{11,33} \\
Y_{12} & \upsilon_{12,11} & 1 & \upsilon_{12,13} & \tau_{12,21} & \alpha_{12,22} & \tau_{12,23} & \tau_{12,31} & \alpha_{12,32} & \tau_{12,33} \\
Y_{13} & \upsilon_{13,11} & \upsilon_{13,12} & 1 & \tau_{13,21} & \tau_{13,22} & \alpha_{13,23} & \tau_{13,31} & \tau_{13,32} & \alpha_{13,33} \\
\hline
Y_{21} & - & - & - & 1 & \upsilon_{21,22} & \upsilon_{21,23} & \alpha_{21,31} & \tau_{21,32} & \tau_{21,33} \\
Y_{22} & - & - & - & \upsilon_{21,22} & 1 & \upsilon_{22,23} & \tau_{22,31} & \alpha_{22,32} & \tau_{22,33} \\
Y_{23} & - & - & - & \upsilon_{21,23} & \upsilon_{22,23} & 1 & \tau_{23,31} & \tau_{23,32} & \alpha_{23,33} \\
\hline
Y_{31} & - & - & - & - & - & - & 1 & \upsilon_{31,32} & \upsilon_{31,33} \\
Y_{32} & - & - & - & - & - & - & \upsilon_{31,32} & 1 & \upsilon_{32,33} \\
Y_{33} & - & - & - & - & - & - & \upsilon_{31,33} & \upsilon_{32,33} & 1 \\
\end{array}
$$

This correlation matrix is made up for three types of correlation parameters (Alzahrani, 2016):

1. The intra-outcome correlation parameter, $\nu_{rt,rs}$, relates outcome r measured at time t with the same outcome measured at time s
2. The inter-outcome correlation parameter, $\alpha_{rt,r't}$, which relates outcome r measured at time t with outcome r' also measured at time t
3. The cross-correlation parameters, $\tau_{rt,r's}$, which relate outcome r measured at time t with outcome r' measured at time s.

The intra-outcome correlation parameters, $\nu_{rt,rs}$, make up the within outcome correlation structure. The inter-outcome, $\alpha_{rt,r't}$, and cross-correlation parameters $\tau_{rt,r's}$, compose the between outcome correlation structure.

At the within level, the correlation structure may be independent, compound symmetry (exchangeable), auto regressive, unstructured, or self-determined, among others. The within outcome correlation matrices for the three outcomes are equal. If the within outcome correlation structure is

- independent, then all $\nu_{rt,rs} = 0$.
- exchangeable, then all $\nu_{rt,rs} = \rho$.
- autoregressive(1), then $\nu_{r1,r2} = \nu_{r2,r3} = \rho$ and $\nu_{r1,r3} = \rho^2$.
- unstructured, then all $\nu_{rt,rs}$ are different from each other.

In this example, there are three between outcome correlation matrices. At the between level, the structure can be independent, exchangeable or unstructured. If the between outcome correlation structure is

- independent, then all $\alpha_{rt,r't} = 0$ and all $\tau_{rt,r's} = 0$, for all three between outcome correlation matrices.
- exchangeable, then all three between outcome correlation matrices are equal, with $\alpha_{rt,r't} = \rho$ for all t and all $r \neq r'$. The values of $\tau_{rt,r's}$ vary depending on the within outcome correlation structure.
- unstructured, then then all three between outcome correlation matrices are different, with $\alpha_{1t,2t} = \alpha_{12}$, $\alpha_{1t,3t} = \alpha_{13}$ and $\alpha_{2t,3t} = \alpha_{23}$. The values of $\tau_{rt,r's}$ vary depending on the within outcome correlation structure.

The fit of this model consists of two-steps following the identification of the valid moments associated with each response. Define our simultaneous model as

$$\mathbf{E}(\boldsymbol{Y}_i) = \boldsymbol{E}\begin{pmatrix} \boldsymbol{Y}_{i1} \\ \boldsymbol{Y}_{i2} \\ \vdots \\ \boldsymbol{Y}_{iR} \end{pmatrix} = \mathbf{g}\left(\begin{pmatrix} \boldsymbol{X}_{i1}^{[\,]} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_{i2}^{[\,]} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{X}_{iR}^{[\,]} \end{pmatrix}\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_R \end{pmatrix}\right)$$

The estimators of the regression parameters $\boldsymbol{\beta}$ are consistent. Each $\boldsymbol{\beta}_{\boldsymbol{r}}$ represents a vector of regression parameters associated with the partitioned matrix corresponding to the r^{th}

vector of responses, $\boldsymbol{Y_{ir}}$, for $r = 1, \dots R$. Let $\boldsymbol{\Delta}_i$ be the diagonal matrix of the marginal variance of $\boldsymbol{y}_i$. Let $\boldsymbol{E}_i$ be the true correlation matrix, and let $\boldsymbol{\Omega}_i$ be a working correlation matrix which may not be identical to $\boldsymbol{E}_i$.

The partitioned data matrices for the *R* outcomes with $J_1, J_2, \dots, J_R$ time-dependent covariates are $\boldsymbol{X_{i1}^{[\,]}}, \boldsymbol{X_{i2}^{[\,]}}, \dots, \boldsymbol{X_{iR}^{[\,]}}$, respectively; such that

$$\boldsymbol{X_{ir}^{[\,]}} = \begin{bmatrix} 1 & X_{ir11} & 0 & \dots & 0 & X_{ir21} & 0 & \dots & 0 & \cdots & X_{irJ_r1} & 0 & \dots & 0 \\ 1 & X_{ir12} & X_{ir11} & \cdots & 0 & X_{ir22} & X_{ir21} & \cdots & 0 & \cdots & X_{irJ_r2} & X_{irJ_r1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{ir1T} & X_{ir1(T-1)} & \cdots & X_{ir11} & X_{ir2T} & X_{ir2(T-1)} & \cdots & X_{ir21} & \cdots & X_{irJ_rT} & X_{irJ_r(T-1)} & \cdots & X_{irJ_r1} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{1} & \boldsymbol{X_{ir1}^{[0]}} & \boldsymbol{X_{ir1}^{[1]}} & \dots & \boldsymbol{X_{ir1}^{[T-1]}} & \boldsymbol{X_{ir2}^{[0]}} & \boldsymbol{X_{ir2}^{[1]}} & \dots & \boldsymbol{X_{ir2}^{[T-1]}} & \cdots & \boldsymbol{X_{irJ_r}^{[0]}} & \boldsymbol{X_{irJ_r}^{[1]}} & \dots & \boldsymbol{X_{irJ_r}^{[T-1]}} \end{bmatrix}$$

Then for each of the *R* outcomes the vector of logits has components:

$$\begin{aligned} \text{logit}(p_{irt}) &= \boldsymbol{X_{ir}^{[\,]}}[t,]\boldsymbol{\beta}_r \\ &= \beta_{r0} + \beta_{r1}^{tt}X_{ir1t} + \sum_{k=1}^{t-1} \beta_{r1}^{[k]} X_{ir1(t-k)}|_{v.m.} + \beta_{r2}^{tt}X_{ir2t} + \sum_{k=1}^{t-1} \beta_{r2}^{[k]} X_{ir2(t-k)}|_{v.m.} + \cdots \\ &\quad + \beta_{rJ_r}^{tt}X_{irJ_rt} + \sum_{k=1}^{t-1} \beta_{rJ_r}^{[k]} X_{irJ_r(t-k)}\,|_{v.m.} \end{aligned}$$

where $\boldsymbol{X_{ir}^{[\,]}}[t,]$ is the row vector of covariates for outcome r coming from subject i at time t.

Assume that each binary outcome Y_{irt} in $\boldsymbol{Y_{ir}}$ follows a marginal Bernoulli distribution. Then, the vector of success probabilities $\boldsymbol{p_{ir}}$ for outcome r and subject i has components

$$p_{irt} = \frac{\exp(\boldsymbol{X_{ir}^{[\,]}}[t,]\boldsymbol{\beta}_r)}{1 + \exp(\boldsymbol{X_{ir}^{[\,]}}[t,]\boldsymbol{\beta}_r)}$$

The overall vector of marginal probabilities for subject i, $\boldsymbol{p_i'} = (\boldsymbol{p_{i1}'}, \boldsymbol{p_{i2}'}, \dots, \boldsymbol{p_{iR}'})$ also has length $R \times T$ (Lipsitz et al., 2009).

Following Lipsitz et.al. (2009) approach, the generalized estimating equations (GEE) are given by

$$\sum_{i=1}^{N}\left(\left(\frac{\partial \boldsymbol{Y_i}}{\partial \boldsymbol{\beta}}\right)' \boldsymbol{V}_i^{-1}(\boldsymbol{Y_i} - \boldsymbol{p_i})\right) = 0$$

where, $\boldsymbol{p_i}$ is a mean vector of dimension $(\mathrm{R} \times \mathrm{T})$ such that $\boldsymbol{p_i}=\mathrm{E}\,(\boldsymbol{Y_i})$ and the variance covariance matrix is defined as a

$$\boldsymbol{V}_i = \theta \boldsymbol{\Delta_i^{1/2}} \boldsymbol{\Omega}_i \boldsymbol{\Delta_i^{1/2}}$$

where $\boldsymbol{\Delta}_i$ is the diagonal matrix of the marginal variance of $\boldsymbol{Y_i}$, and $\boldsymbol{\Omega}_i$ is a working correlation matrix which may not be identical to $\boldsymbol{E}_i$, the true correlation matrix.

The procedure to estimate the regression coefficients for the model consists of two stages. In the first stage, we estimate the working correlation matrix using frequentist joint GEE, with no covariates in the model and assuming that the intercepts for each of the outcomes are not necessarily the same. In the second stage, use in the estimated working correlation matrix into the Gaussian log-likelihood utilized by Crowder (1985) and Zhang and Paul (2013) when fitting GEE models to binary outcomes:

$$\tilde{\mathrm{L}}(\boldsymbol{y}|\boldsymbol{\beta}) = \left\{-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{Y_i} - \boldsymbol{p_i}(\boldsymbol{\beta}))'\boldsymbol{V_i^{-1}}(\boldsymbol{Y_i} - \boldsymbol{p_i}(\boldsymbol{\beta})) + \log(|2\pi \boldsymbol{V_i}|)\right\}$$

Combine this log-likelihood, $\tilde{\mathrm{L}}(\boldsymbol{y}|\boldsymbol{\beta})$, with $\pi(\boldsymbol{\beta})$, the prior distribution for the vector of coefficients $\boldsymbol{\beta}$ (Chernozhukov & Hong, 2003; Yin, 2009), to form the posterior distribution of $\boldsymbol{\beta}$ such that

$$\tilde{\pi}(\boldsymbol{\beta}|\,\boldsymbol{y}) \propto \tilde{\mathrm{L}}(\boldsymbol{y}|\boldsymbol{\beta})\,\pi(\boldsymbol{\beta})$$

Assume that the prior distribution of the vector of regression coefficients $\pi(\boldsymbol{\beta})$ is multivariate normal (Yin, 2009). Assign normal priors to the elements of $\boldsymbol{\beta}$, and then sample from the posterior distribution accordingly. The working correlation matrix estimated in stage one for all iterations of the Markov Chain Monte Carlo (MCMC) algorithm is used. The process for estimating this model combines the simplicity of estimating working correlation matrix using the frequentist joint GEE, with the desirable properties of MCMC for estimating the regression coefficients (Canapu et al., 2013).

3.3.2 Special Case: Single Response MVM Models with Bayesian estimates

The simultaneous MVM model can be modified to fit a single outcome, if there is only one outcome of interest. The response vector for subject i, $\boldsymbol{Y_{ir}}$, has length T, as does the vector of marginal probabilities $\boldsymbol{p_{ir}}$. The partitioned data matrix contains the covariates and the lags of time-dependent covariates with valid moments for that particular outcome. The dimensions of the working correlation matrix reduce to $(T \times T)$ as the modification of the model is done through this working correlation matrix. When estimating the single response model, we follow the same steps as if we were estimating the simultaneous response model. However, in the first stage, we only estimate the within-outcome working correlation matrix, $\boldsymbol{\Sigma}_{rr'}$, for the outcome of interest. At stage two of the estimation process, replace the estimated within outcome working correlation matrix, $\widehat{\boldsymbol{\Sigma}_{rr'}}$, in the Gaussian log-likelihood at each iteration of the Markov Chain Monte Carlo algorithm to obtain Bayesian estimates for the regression coefficients.

3.4. Simulation Study

A simulation study was conducted to assess the performance of the simultaneous MVM Bayesian model. Simulate two binary outcomes (Y_1, Y_2), each measured at three time points, resulting in 6 correlated binary outcomes per subject. These outcomes are simulated assuming that both are affected by the same continuous time-dependent covariate, but that the covariate is associated with them in different ways. For outcome 1, the regression coefficients are set to $\beta_0 = 0$, $\beta^{tt} = 0.04$, $\beta^{[1]} = 0.12$ and $\beta^{[2]} = 0.2$. For outcome 2, the regression coefficients are set to $\alpha_0 = 0$, $\alpha_1^{tt} = 0.25$, $\alpha_1^{[1]} = 0.15$ and $\alpha_1^{[2]} = 0.05$. The working correlation matrix had an exchangeable within-outcome correlation and an exchangeable between outcome correlation structures. For the within-outcome correlation, the off-diagonal elements are $\rho_w = 0.45$; for the between-outcome correlation the diagonal elements are $\rho_{db} = 0.35$, while the off-diagonal elements were $\rho_{ob} = 0.30$, thus the working correlation matrix used was:

$$\begin{bmatrix} 1 & 0.45 & 0.45 & 0.35 & 0.30 & 0.30 \\ 0.45 & 1 & 0.45 & 0.30 & 0.35 & 0.30 \\ 0.45 & 0.45 & 1 & 0.30 & 0.30 & 0.35 \\ 0.35 & 0.30 & 0.30 & 1 & 0.45 & 0.45 \\ 0.30 & 0.35 & 0.30 & 0.45 & 1 & 0.45 \\ 0.30 & 0.30 & 0.35 & 0.45 & 0.45 & 1 \end{bmatrix}$$

The longitudinal binary outcomes with the exchangeable working correlation within outcomes and the exchangeable working correlation matrix between outcomes are generated following the algorithm by Emrich and Piedmonte (1991) through the SAS function *RandMVBinary*. The sample sizes used are $N \in \{25, 100, 500, 1000\}$ with 200 datasets simulated for each sample size. For each dataset, we fit the simultaneous MVM model with Bayes estimates with the correct working correlation matrix (exchangeable within-outcome and exchangeable between-outcome). The simultaneous MVM model with incorrect working correlation matrix structures was also fit: unstructured within-outcome and independent between-outcome, unstructured within-outcome and unstructured between-outcome, and AR(1) within-outcome and exchangeable between-outcome. These choices allow one to assess the performance of the simultaneous MVM Bayesian model when the working correlation structure among the outcomes is miss specified. The percentage of coverage, the root mean square error (RMSE) and the mean bias for each model is computed (Table 3.1, Table 3.2 and Table 3.3, respectively). Define percentage coverage as the percentage of datasets for which the credible interval of the simultaneous model covers the true value of the parameter.

The percentage of coverage suggest the miss specification of the working correlation matrix did not affect coverage. All models, including the wrong working correlation matrix structures had similar percentages of coverage for all regression parameters. The percentages of coverage were smallest for sample sizes of 25 subjects. However, the percentage of coverage for all parameters for sample sizes of 25 were all greater than 83.00.

The RMSE's for models with miss specified working correlation matrix structures are very similar to the RMSE's for the model with the correct working correlation structures. The RMSE's decreased as the sample size increased. When measuring efficiency of estimators using RMSE, larger sample sizes result in more efficient estimators. In general, the RMSE's for the regression parameters corresponding to the cross-sectional effects were smaller than the RMSE's for parameters corresponding to lag-1 effects, which in turn had smaller RMSE's than the regression parameters corresponding to the lag-2 effects.

For all regression parameters, the mean bias for the models with miss specified working correlation matrix structures are very similar to the mean bias of the model with the correctly specified working correlation matrix. The highest absolute value of the mean bias was 0.059, resulting in the highest percentage bias being 5.9%. This indicates that the simultaneous MVM model with Bayesian estimates performs well in terms of bias, even with incorrect working correlation matrix structures. The sample size did not influence means bias, since for some parameters the mean bias did not decline with higher sample sizes.

We also studied whether the miss specification of the working correlation matrix structure affected the precision of the simultaneous MVM Bayesian model estimators, Table 3.4. A check for the percentage of datasets for which the model with the true working correlation matrix resulted in more precise estimates than the models with miss specified working correlation structures. For each dataset, the model with the true working correlation matrix is more precise than the other model, if the credible interval for the model with the true working correlation matrix contained the true value of the parameter and was narrower than the credible interval of the model with the miss specified working correlation matrix. In general, as the sample size increased, the model with the true working correlation matrix became more precise than the models with incorrect working correlation matrices. Also, the models with incorrect working correlation matrices that accounted for the between outcome correlation (unstructured within and unstructured between, AR(1) within and exchangeable between) were at least as precise as the model with the true working correlation matrix. However, the model with incorrect working correlation matrix that assumed that there was no between outcome correlation (unstructured within, independent within) is less precise than the model with the true working correlation matrix. The precision is not affected by incorrectly specified working correlation matrices as long as the between outcome correlation is accounted for. It is important to note that the precision of this model is affected when the between outcome correlation of the simultaneous outcomes is assumed to not exist.

The results of the simulation study show that the simultaneous MVM model with Bayes estimates preserves the properties of GEE models. Bias, RMSE and percentage of coverage are not affected by miss-specification of the working correlation matrix. However, precision of the model is affected if it is assumed that the simultaneous outcomes are independent of one another. This relates to the efficiency of GEE, where models with independent working correlation matrices are less efficient than models with other working correlation matrix structures.

Table 3.1

Comparing Percentage Coverage between Models using Different Structures of Working Correlation Matrix

Sample size	Outcome	Parameter	Exchangeable, Exchangeable	Unstructured, Independent	Unstructured, Unstructured	AR(1), Exchangeable
25	Y1	$\beta_1^{tt} = 0.04$	83.00	85.00	85.50	86.00
		$\beta_1^{[1]} = 0.12$	91.50	89.50	90.00	86.50
		$\beta_1^{[2]} = 0.20$	89.50	89.50	88.50	88.50
	Y2	$\alpha_1^{tt} = 0.25$	84.00	84.00	85.00	84.00
		$\alpha_1^{[1]} = 0.15$	87.00	88.00	87.50	85.50
		$\alpha_1^{[2]} = 0.05$	87.50	86.00	87.00	86.00
100	Y1	$\beta_0 = 0$	96.50	97.00	97.50	96.00
		$\beta_1^{tt} = 0.04$	92.50	93.00	93.50	95.00
		$\beta_1^{[1]} = 0.12$	91.50	93.50	93.00	91.00
		$\beta_1^{[2]} = 0.20$	97.50	96.50	97.00	96.50
	Y2	$\alpha_0 = 0$	92.00	91.00	93.00	92.00
		$\alpha_1^{tt} = 0.25$	89.50	91.00	89.00	86.50
		$\alpha_1^{[1]} = 0.15$	97.00	96.00	97.00	95.50
		$\alpha_1^{[2]} = 0.05$	94.00	92.50	94.00	96.00
500	Y1	$\beta_0 = 0$	93.00	92.00	93.50	91.50
		$\beta_1^{tt} = 0.04$	99.50	96.50	99.00	98.50
		$\beta_1^{[1]} = 0.12$	94.50	96.00	94.50	95.50
		$\beta_1^{[2]} = 0.20$	98.00	98.00	98.00	96.50
	Y2	$\alpha_0 = 0$	97.00	97.50	97.50	97.00
		$\alpha_1^{tt} = 0.25$	89.00	91.00	90.50	90.50
		$\alpha_1^{[1]} = 0.15$	93.00	93.00	91.50	91.50
		$\alpha_1^{[2]} = 0.05$	94.00	94.00	95.50	97.00
1000	Y1	$\beta_0 = 0$	100.00	100.00	100.00	100.00
		$\beta_1^{tt} = 0.04$	100.00	99.50	100.00	99.50
		$\beta_1^{[1]} = 0.12$	97.00	98.00	85.00	85.00
		$\beta_1^{[2]} = 0.20$	90.50	91.00	90.50	96.50
	Y2	$\alpha_0 = 0$	91.00	95.50	87.50	86.00
		$\alpha_1^{tt} = 0.25$	96.00	99.00	96.00	96.50
		$\alpha_1^{[1]} = 0.15$	98.00	99.50	98.00	93.50
		$\alpha_1^{[2]} = 0.05$	78.00	77.50	78.00	85.50

Table 3.2

Comparing RMSE between Models using Different Structures of Working Correlation Matrix

Sample size	Outcome	Parameter	Exchangeable, Exchangeable	Unstructured, Independent	Unstructured, Unstructured	AR(1), Exchangeable
25	Y1	$\beta_0 = 0$	0.337	0.360	0.338	0.344
		$\beta_1^{tt} = 0.04$	0.274	0.276	0.271	0.278
		$\beta_1^{[1]} = 0.12$	0.328	0.324	0.319	0.340
		$\beta_1^{[2]} = 0.20$	0.471	0.465	0.452	0.495
	Y2	$\alpha_0 = 0$	0.327	0.354	0.330	0.331
		$\alpha_1^{tt} = 0.25$	0.306	0.307	0.300	0.313
		$\alpha_1^{[1]} = 0.15$	0.402	0.421	0.402	0.419
		$\alpha_1^{[2]} = 0.05$	0.540	0.534	0.530	0.567
100	Y1	$\beta_0 = 0$	0.128	0.139	0.128	0.131
		$\beta_1^{tt} = 0.04$	0.114	0.117	0.113	0.114
		$\beta_1^{[1]} = 0.12$	0.132	0.131	0.131	0.134
		$\beta_1^{[2]} = 0.20$	0.176	0.179	0.176	0.183
	Y2	$\alpha_0 = 0$	0.150	0.166	0.151	0.153
		$\alpha_1^{tt} = 0.25$	0.118	0.123	0.117	0.121
		$\alpha_1^{[1]} = 0.15$	0.125	0.130	0.126	0.131
		$\alpha_1^{[2]} = 0.05$	0.192	0.198	0.193	0.198
500	Y1	$\beta_0 = 0$	0.069	0.079	0.069	0.071
		$\beta_1^{tt} = 0.04$	0.039	0.040	0.038	0.038
		$\beta_1^{[1]} = 0.12$	0.054	0.051	0.053	0.055
		$\beta_1^{[2]} = 0.20$	0.069	0.072	0.070	0.072
	Y2	$\alpha_0 = 0$	0.066	0.075	0.066	0.065
		$\alpha_1^{tt} = 0.25$	0.050	0.053	0.050	0.051
		$\alpha_1^{[1]} = 0.15$	0.067	0.063	0.068	0.071
		$\alpha_1^{[2]} = 0.05$	0.081	0.083	0.080	0.083
1000	Y1	$\beta_0 = 0$	0.044	0.046	0.043	0.043
		$\beta_1^{tt} = 0.04$	0.027	0.032	0.027	0.028
		$\beta_1^{[1]} = 0.12$	0.058	0.049	0.059	0.064
		$\beta_1^{[2]} = 0.20$	0.057	0.059	0.057	0.063
	Y2	$\alpha_0 = 0$	0.054	0.055	0.054	0.054
		$\alpha_1^{tt} = 0.25$	0.023	0.024	0.023	0.022
		$\alpha_1^{[1]} = 0.15$	0.049	0.039	0.049	0.051
		$\alpha_1^{[2]} = 0.05$	0.083	0.084	0.081	0.081

Table 3.3

Comparing Bias between Models using Different Structures of Working Correlation Matrix

Sample size	Outcome	Parameter	Exchangeable, Exchangeable	Unstructured, Independent	Unstructured, Unstructured	AR(1), Exchangeable
25	Y1	$\beta_0 = 0$	-0.010	-0.012	-0.007	-0.004
		$\beta_1^{tt} = 0.04$	0.011	0.005	0.006	0.008
		$\beta_1^{[1]} = 0.12$	-0.034	-0.041	-0.024	-0.032
		$\beta_1^{[2]} = 0.20$	0.006	0.023	0.022	-0.003
	Y2	$\alpha_0 = 0$	-0.021	-0.021	-0.019	-0.022
		$\alpha_1^{tt} = 0.25$	-0.052	-0.059	-0.049	-0.050
		$\alpha_1^{[1]} = 0.15$	0.011	-0.002	0.016	0.009
		$\alpha_1^{[2]} = 0.05$	-0.064	-0.057	-0.059	-0.070
100	Y1	$\beta_0 = 0$	-0.017	-0.019	-0.017	-0.019
		$\beta_1^{tt} = 0.04$	-0.002	-0.008	0.000	-0.001
		$\beta_1^{[1]} = 0.12$	0.012	0.001	0.013	0.012
		$\beta_1^{[2]} = 0.20$	0.021	0.016	0.020	0.019
	Y2	$\alpha_0 = 0$	0.007	0.010	0.008	0.008
		$\alpha_1^{tt} = 0.25$	0.012	0.002	0.014	0.013
		$\alpha_1^{[1]} = 0.15$	0.001	-0.008	0.004	0.007
		$\alpha_1^{[2]} = 0.05$	-0.005	-0.006	-0.006	-0.008
500	Y1	$\beta_0 = 0$	-0.012	-0.013	-0.012	-0.014
		$\beta_1^{tt} = 0.04$	0.009	0.002	0.010	0.010
		$\beta_1^{[1]} = 0.12$	0.021	0.006	0.021	0.019
		$\beta_1^{[2]} = 0.20$	0.025	0.021	0.025	0.029
	Y2	$\alpha_0 = 0$	0.008	0.009	0.008	0.008
		$\alpha_1^{tt} = 0.25$	0.009	-0.005	0.010	0.008
		$\alpha_1^{[1]} = 0.15$	0.027	0.012	0.027	0.030
		$\alpha_1^{[2]} = 0.05$	-0.010	-0.016	-0.010	-0.009
1000	Y1	$\beta_0 = 0$	0.002	0.001	0.002	0.001
		$\beta_1^{tt} = 0.04$	-0.013	-0.021	-0.013	-0.012
		$\beta_1^{[1]} = 0.12$	0.046	0.034	0.047	0.050
		$\beta_1^{[2]} = 0.20$	0.037	0.035	0.039	0.044
	Y2	$\alpha_0 = 0$	-0.006	-0.005	-0.006	-0.007
		$\alpha_1^{tt} = 0.25$	0.005	-0.008	0.005	0.004
		$\alpha_1^{[1]} = 0.15$	0.039	0.027	0.039	0.042
		$\alpha_1^{[2]} = 0.05$	-0.052	-0.054	-0.050	-0.049

Table 3.4

Comparing Precision between Simultaneous Model with True Working Correlation Matrix and Three Simultaneous Models with Miss Specified Working Correlation Matrix

Sample size	Outcome	Parameter	Unstructured, Independent	Unstructured, Unstructured	AR(1), Exchangeable
25	Y1	$\beta_0 = 0$	77.00	43.00	13.00
		$\beta_1^{tt} = 0.04$	51.00	31.50	51.00
		$\beta_1^{[1]} = 0.12$	51.50	35.00	42.00
		$\beta_1^{[2]} = 0.20$	48.00	36.00	70.50
	Y2	$\alpha_0 = 0$	68.50	36.50	22.50
		$\alpha_1^{tt} = 0.25$	47.50	28.50	51.00
		$\alpha_1^{[1]} = 0.15$	42.00	25.50	34.50
		$\alpha_1^{[2]} = 0.05$	43.50	33.50	68.00
100	Y1	$\beta_0 = 0$	87.50	44.50	18.00
		$\beta_1^{tt} = 0.04$	72.50	40.50	65.00
		$\beta_1^{[1]} = 0.12$	64.00	40.00	43.50
		$\beta_1^{[2]} = 0.20$	56.50	48.00	89.00
	Y2	$\alpha_0 = 0$	79.50	50.00	17.50
		$\alpha_1^{tt} = 0.25$	66.00	46.00	64.00
		$\alpha_1^{[1]} = 0.15$	63.00	43.50	44.50
		$\alpha_1^{[2]} = 0.05$	55.50	40.00	82.50
500	Y1	$\beta_0 = 0$	91.50	43.00	10.00
		$\beta_1^{tt} = 0.04$	82.50	42.50	78.00
		$\beta_1^{[1]} = 0.12$	73.00	46.50	34.50
		$\beta_1^{[2]} = 0.20$	69.50	51.50	95.50
	Y2	$\alpha_0 = 0$	90.00	48.50	11.00
		$\alpha_1^{tt} = 0.25$	59.00	42.00	74.50
		$\alpha_1^{[1]} = 0.15$	74.00	41.50	53.00
		$\alpha_1^{[2]} = 0.05$	62.00	51.00	87.50
1000	Y1	$\beta_0 = 0$	99.50	40.50	2.00
		$\beta_1^{tt} = 0.04$	80.00	36.50	83.50
		$\beta_1^{[1]} = 0.12$	72.00	47.50	39.00
		$\beta_1^{[2]} = 0.20$	59.00	51.00	90.50
	Y2	$\alpha_0 = 0$	91.00	50.50	23.50
		$\alpha_1^{tt} = 0.25$	80.00	49.50	57.50
		$\alpha_1^{[1]} = 0.15$	77.50	41.00	62.00
		$\alpha_1^{[2]} = 0.05$	43.50	33.00	78.00

3.5. Numerical Examples

A numerical example demonstrating the fit of the simultaneous responses MVM models with Bayesian estimates for time dependent covariates is presented. In that example, quality of life measures as it pertains to Chinese Longitudinal Healthy Longevity Study are analyzed (Zeng et al. 2009).

3.5.1 Simultaneous Responses- Chinese Longitudinal Healthy Longevity Study

A four-response MVM model with Bayesian estimates for time-dependent covariates to the Chinese Longitudinal Healthy Longevity Study is fit. This fit identifies relationships between several covariates (time-independent and time-dependent) and responses that measured one's quality of life. The data are collected in four waves starting in 1998 and continuing in 2000, 2002 and 2005 for Chinese people aged 77 years and older.

The four binary responses are (1) healthy or not, (2) complete physical check as measured by interviewer or not, (3) self-rated quality of life, and (4) self-rated health. The time-dependent covariates are: make their own decisions, consumed vegetables frequently, dress without assistance and having visual difficulties. Gender is also included as a time-independent covariate. Table 3.5 shows how the outcomes are coded as binary.

The working correlation matrix for the four outcomes at the four time periods, using an AR (1) structure for the within-subject correlation and an exchangeable structure for the between-outcome correlation is obtained. The within-outcome correlation matrix is $\begin{bmatrix} 1 & 0.167 & 0.028 & 0.005 \\ 0.167 & 1 & 0.167 & 0.028 \\ 0.028 & 0.167 & 1 & 0.167 \\ 0.005 & 0.028 & 0.167 & 1 \end{bmatrix}$. While the between-outcome correlation matrix is

$$\begin{bmatrix} 0.196 & 0.033 & 0.005 & 0.001 \\ 0.033 & 0.196 & 0.033 & 0.005 \\ 0.005 & 0.033 & 0.196 & 0.033 \\ 0.001 & 0.005 & 0.033 & 0.196 \end{bmatrix}.$$

Table 3.5

Coding Outcomes as Binary

Outcome	Original value	Binary
(1) Interviewer-rated health	Surprisingly healthy Relatively healthy	Good health
	Moderately ill Very ill	Bad health
(2) Complete physical check	Yes	Yes
	Partially able to No	No
(3) Self-rated quality of life	Very good Good	Good
	So so Bad Very bad	Not good
(4) Self-rated health	Very good Good	Good
	So so Bad Very bad	Not good

The valid moment conditions are obtained. Non-informative priors $N(0, 10000)$ for all regression coefficients are instituted. Table 3.6 contains results of the fitted model. Figures 3.1, 3.2, 3.3 and 3.4 show the credible intervals for the odds ratios of the time-dependent covariates on the four outcomes over time. All covariates show significant immediate effects on interviewer-rated health. There are delayed effects for eating vegetables, ability to dress without assistance, having visual difficulties. Ability to make own decisions, frequently consuming vegetables, and having visual difficulties had further delayed significant effects on interviewer-rated health.

Males were significantly more likely to complete a physical check than females. All covariates had a significant impact on the probability of completing a physical check at immediate effects. Delayed effects for consuming vegetables and the ability to dress without assistance had a significant impact on ability to complete a physical check. Further delayed effects ability to make own decision and ability to dress without assistance significantly impacted respondents' probabilities of completing a physical check. Furthermost delayed effects for frequently consuming vegetables and having visual difficulties had a significant impact on ability to complete physical check.

For self-rated quality of life, frequently eating vegetables and having visual difficulties had a significant immediate impact. There were no significant delayed or further delayed effects for any of the predictors on self-rated quality of life. Visual difficulties had significant furthermost delayed effects on self-rated quality of life.

For self-rated health, all covariates are significant immediate impacts for having false teeth. Ability to dress without assistance had significant effects on self-rated health as a further delayed impact. Having visual difficulties significantly impacted self-rated health across a further delayed effect.

Being able to make own decisions and to dress without assistance increased the immediate likelihood of being in good health according to interviewer and to self and to complete a physical check. The frequent consumption of vegetables had positive immediate effects on the four outcomes, while the presence of visual difficulties had negative effects on all outcomes. Vegetable consumption has a significant negative delayed effect on interviewer-rated health and completion of physical check only. The ability to dress without assistance had negative delayed effects on interviewer-rated health, complete physical check and self-rated health. Visual difficulties decreased the likelihood of being in good health according to the interviewer and of good quality of life according to self across a one time period lag. Ability to make own decisions had negative further delayed effects on interviewer-rated health and completion of a physical check. Interviewer-rated health and self-rated health were both negatively impacted by presence of visual difficulties across a two time period lag. In general, when a covariate affected more than one of the simultaneous outcomes measuring quality of life at the same time period, it affected the simultaneous outcomes in a similar fashion.

Table 3.6

Results of Simultaneous MVM Model with Bayes Estimates for Chinese Longevity Study

Outcome	Time-period	Parameter	OR	OR 95% CI		ESS
Interviewer rated health	Immediate Effects	Male	1.041	0.914	1.174	6837
		Own decision	1.916	1.649	2.226	4739
		Vegetables	2.387	2.034	2.829	4638
		able to dress	8.085	6.360	10.074	4298
		visual difficulty	0.343	0.298	0.395	4673
	Delayed Effect	Own decision	0.923	0.811	1.062	3729
		Vegetables	0.741	0.644	0.869	3903
		able to dress	0.423	0.361	0.497	3633
		visual difficulty	0.625	0.533	0.741	4602
	Further Delayed effect	Own decision	0.698	0.560	0.878	6780
		Vegetables	0.651	0.543	0.779	5168
		able to dress	1.051	0.835	1.323	4238
		visual difficulty	0.638	0.507	0.795	4818
	Furthermost delayed effect	Own decision	0.852	0.664	1.105	4214
		Vegetables	1.105	0.819	1.492	4639
		able to dress	0.705	0.472	1.030	3403
		visual difficulty	1.259	0.861	1.804	5006
Ability to complete physical check	Immediate Effects	Male	1.433	1.271	1.616	5091
		Own decision	1.584	1.363	1.840	4325
		Vegetables	1.768	1.492	2.117	4105
		able to dress	5.529	4.393	6.959	4251
		visual difficulty	0.237	0.206	0.275	5339
	Delayed Effect	Own decision	1.083	0.914	1.297	3549
		Vegetables	0.741	0.625	0.887	3983
		able to dress	0.651	0.533	0.803	3633
		visual difficulty	0.896	0.748	1.062	5647
	Further Delayed effect	Own decision	0.589	0.436	0.811	3604
		Vegetables	1.127	0.844	1.537	6170
		able to dress	1.462	1.221	1.768	2823
		visual difficulty	1.000	0.733	1.350	5554
	Furthermost delayed effect	Own decision	0.763	0.566	1.020	4623
		Vegetables	2.014	1.477	2.746	4607
		able to dress	0.803	0.543	1.221	4373
		visual difficulty	0.566	0.415	0.771	4959
		Male	0.980	0.852	1.139	5809

Self-rated quality of life	Immediate Effects	Own decision	1.150	0.980	1.363	4801
		Vegetables	1.462	1.209	1.768	5356
		able to dress	1.297	0.980	1.768	3727
		visual difficulty	0.787	0.657	0.923	5104
	Delayed Effect	Own decision	1.105	0.852	1.433	5554
		Vegetables	0.923	0.803	1.062	4113
		able to dress	0.861	0.549	1.336	3639
		visual difficulty	0.803	0.651	0.990	4667
	Further Delayed effect	Own decision	1.010	0.795	1.284	5113
		able to dress	0.763	0.472	1.209	3907
		visual difficulty	1.020	0.763	1.336	5578
	Furthermost delayed effect	Own decision	0.914	0.664	1.271	4523
		Vegetables	1.174	0.819	1.699	4303
		able to dress	1.051	0.698	1.616	3434
		visual difficulty	1.377	0.905	2.014	4606
Self-rated health	Immediate Effects	Male	1.116	0.961	1.310	4949
		Own decision	1.197	1.010	1.433	4744
		Vegetables	1.632	1.350	1.974	4969
		able to dress	2.858	2.181	3.781	4014
		visual difficulty	0.613	0.522	0.726	4889
	Delayed Effect	Own decision	1.174	0.961	1.462	4953
		Vegetables	1.030	0.835	1.271	4198
		able to dress	0.763	0.583	0.980	4092
		visual difficulty	0.869	0.705	1.073	4244
	Further Delayed effect	Own decision	1.041	0.733	1.522	5066
		Vegetables	0.970	0.748	1.246	3463
		able to dress	0.638	0.482	0.835	2922
		visual difficulty	0.726	0.560	0.961	4961
	Furthermost delayed effect	Own decision	1.094	0.795	1.522	5771
		Vegetables	1.139	0.779	1.665	5347
		able to dress	0.844	0.517	1.336	3935

Figure 3.1

Point Estimates and 95% Credible Intervals for Interviewer-rated Health

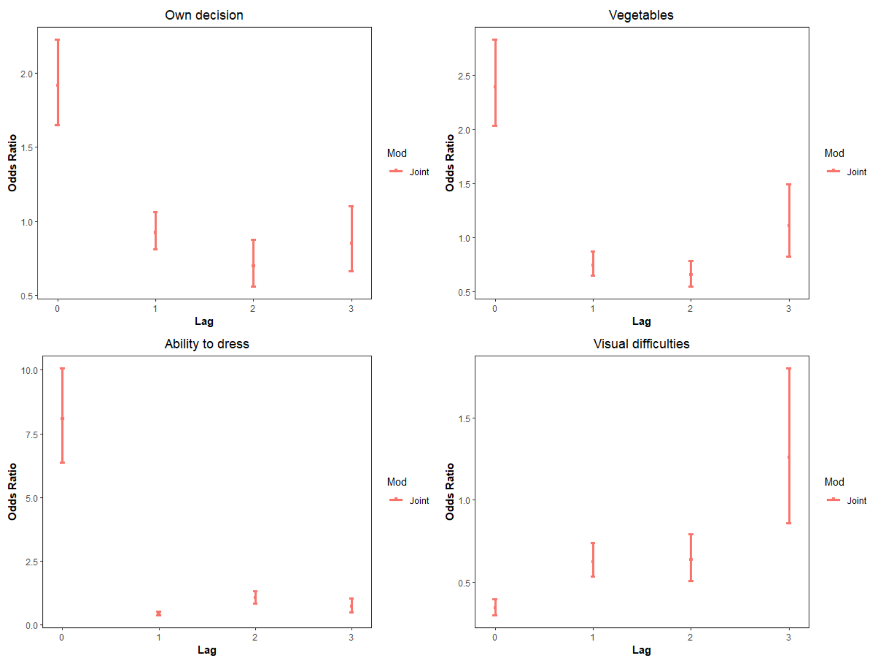


Figure 3.2

Point Estimates and 95% Credible Intervals for Ability to Complete Physical Check

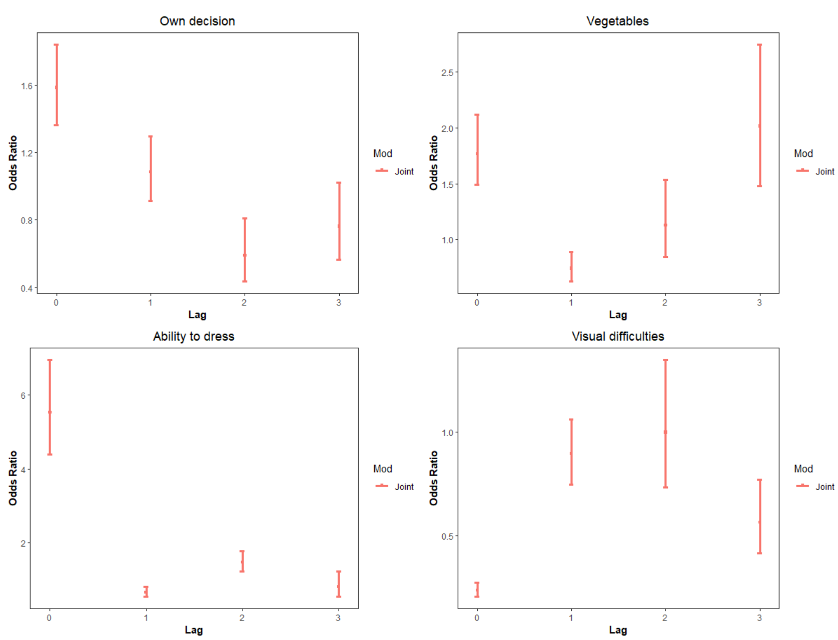


Figure 3.3

Point Estimates and 95% Credible Intervals for Self-rated Quality of Life

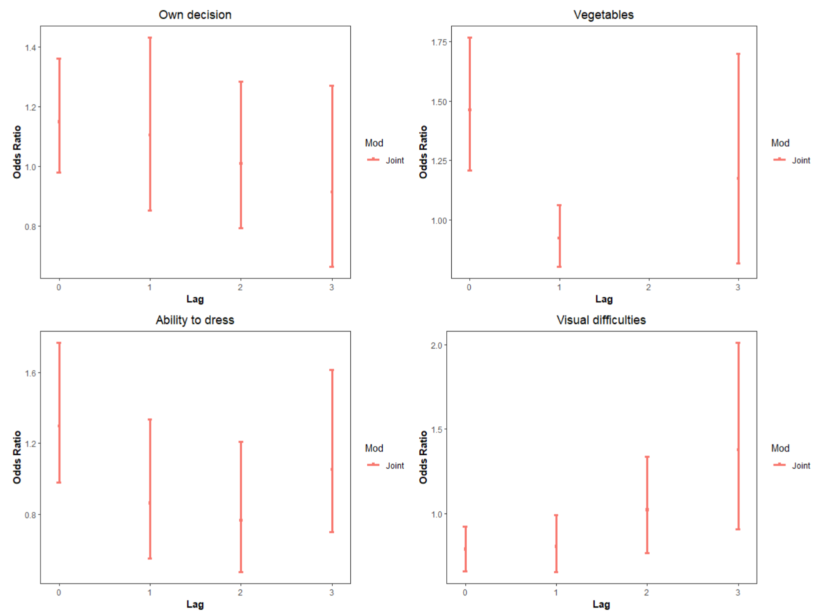
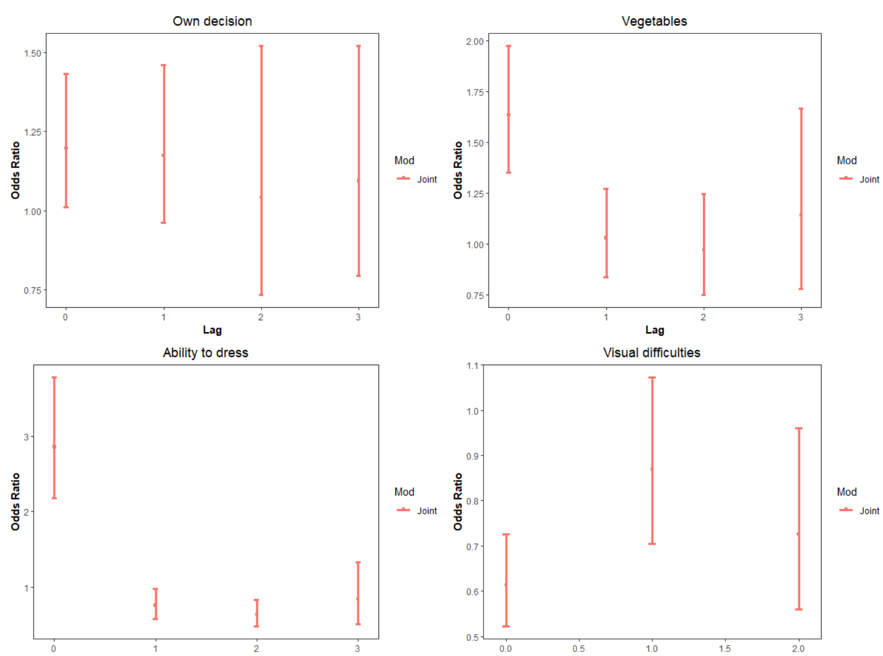


Figure 3.4

Point Estimates and 95% Credible Intervals for Self-rated Health



3.6 Discussion

The proposed simultaneous MVM model with Bayesian estimates is used for marginal inference in longitudinal studies with multiple response variables. It has the unique behavior to partition the time-dependent covariates which addresses correlation caused to the time varying effects of the covariates. In addition, it accounts for the two types of correlation (inner-outer) when looking at multiple responses. However, the model relies on the ability to identify valid moments to obtain consistent estimates of the regression parameters. The expansion of the use of the GEE model affords the same properties of consistency and efficiency afforded to the GEE estimators. The use of Bayes principles affords the model the advantages appreciated with such.

The model regression parameters perform well in terms of percentage coverage, bias, and root mean square error in small samples and better in larger samples. The miss specification of the inner-outer working correlation matrix related does not affect percentage of coverage, bias, or root mean square error. The precision of the regression coefficients is not affected by miss specified working correlation matrices, unless the between outcome correlation of the simultaneous outcomes is ignored by assuming that the responses are independent.

CHAPTER 4

A TWO-PART GMM REGRESSION MODEL FOR FEEDBACK FOR TIME-DEPENDENT COVARIATES

Abstract

Correlated observations in longitudinal studies are often due to repeated measures on the subjects or in the case of clustered data due to hierarchical structure of the design. In addition, correlation may be realized due to the association between responses at a particular time and the predictors at earlier times. Also, there is feedback between response at the present and the covariates at a later time, though this is not always relevant and so is often ignored. In any event, each case must be accounted for as they can have different effects on the regression coefficients. Several authors have provided models that reflect the direct impact and the delayed impact of covariates on the response, utilizing valid moment conditions to estimate relevant regression coefficients. However, there are applications where one cannot ignore the impact of the responses on future covariates. We propose the use of a two-part model to account for the additional feedback thus modeling the direct impact, as well as the delayed impact of the covariates on future responses and vice versa. We demonstrate the use of the two-part model by revisiting child morbidity and its impact on future values of BMI in the Philippines health data. We also present an example where we model obesity status and its feedback effects of physical activity and depression levels using the Add Health dataset.

4.1 Introduction

When analyzing longitudinal data, there is feedback that may go unchecked thus masking the real impact of the covariate. Diggle, Heagerty, Yee and Zeger (2002) explained that in the presence of longitudinal data with time-dependent covariates, there are usually three questions of interest:

1. What is the relationship between the outcome Y_{it} and the covariate X_{ijt} when both are measured at the same time (cross-sectional relationship/association)?
2. Is the outcome at time t, Y_{it}, impacted by the time-dependent covariate measured at time $[t-s]$, $X_{ij[t-s]}$; $(s = 1, 2, .., t-1)$ (lagged covariates related/associated with future values of the outcome)?
3. What factors affect time-dependent covariate at time t, X_{ijt}, does outcome at wave $[t-s]$ associates with time-dependent covariate at time t?

The correlation realized due to the relation between the responses at a one period and the predictors at present or earlier time-periods has been addressed (Irimata et al., 2019). Such models used valid moment conditions from partial regression coefficients based on data relations between response (time t) and covariates (time s) $[X_s, Y_t]$ where $[s <= t]$. This is necessary as response and predictor associated with different time periods do not necessarily provide valid moments. One can use a partitioned data matrix in the systematic component with additional regression coefficients to analyze such data (Irimata et al., 2019). Their model addressed the first two questions of interest when analyzing longitudinal data.

However, in health and health related data there are cases when the feedback (response on covariates in the future) is real. Such is the case the Philippine's study. In that case, the response had an impact on the covariate at later times. Such situation corresponds to the third question. We expand on the partitioned GMM model to allow the responses as measured presently to impact future covariates. We propose to fit a two-part model to answer the three questions together. The first part consists of the partitioned GMM model, while the second part makes use of simultaneous modeling of the responses.

There are two basic approaches used to analyze longitudinal data with time-independent covariates. The subject-specific models, in which heterogeneity in regression parameters is explicitly modelled; and population-averaged models in which the mean response for the population is where the interest lies. The proposed model is of the latter.

4.1.1 General Framework

Consider the longitudinal data for unit/subject i that has been measured T times. Let $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})'$ represent the vector of outcomes for subject i that are associated with the data matrix $\boldsymbol{X}_i = \begin{bmatrix} x_{i11} & \cdots & x_{iJ1} \\ \vdots & \ddots & \vdots \\ x_{i1T} & \cdots & x_{iJT} \end{bmatrix}$. Where the row vector, ${\boldsymbol{x}_{i*t}}' = (x_{i1t}, \dots, x_{iJt})$ represents the covariate values at time t and the column vector $\boldsymbol{x}_{ij*} = (x_{ij1}, \dots, x_{ijT})'$ contains the values of the j^{th} covariate, such that $t = 1, \dots, T$; and $j = 1, \dots, J$. For each subject i, consider a generalized linear model (McCullagh & Nelder, 1989) with link function g and covariate matrix $\boldsymbol{X}_i$ of dimension T by J; where J represents the total number of covariates, with mean $\boldsymbol{\mu_i}$ then $g[\boldsymbol{\mu_i}] = \boldsymbol{X}_i\boldsymbol{\beta}$**.** Let the variance of $\boldsymbol{y_i}$ $\text{var}[\boldsymbol{y_i}] = \boldsymbol{v}(\boldsymbol{\mu_i})\varphi$ where $\boldsymbol{\mu_i}$ is the mean and $\mathbf{v}(.)$ is a function of $\boldsymbol{\mu_i}$. Consider a diagonal matrix $\boldsymbol{A_i}$ based on the elements from the vector $\boldsymbol{v}(\boldsymbol{\mu_i})$. Choose a correlation matrix $\boldsymbol{R}_i$ of dimension T, which represents the relationship among the T responses from subject i. Thus, the variance-covariance matrix for $\boldsymbol{y_i}$ is

$$\boldsymbol{V}_{\boldsymbol{y}_i} = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i\boldsymbol{A}_{i}^{1/2},$$

and sum over all N subjects,

$$\sum_{i=1}^{N} \boldsymbol{D}_i' \, [\boldsymbol{V}_{\widehat{\boldsymbol{y}}_i})]^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

where $\boldsymbol{D_i} = \frac{\partial \mu_i}{\partial \beta}$ so

$$\widehat{\boldsymbol{\beta}} = [\sum_{i=1}^{N} \boldsymbol{D}_i' \, [\boldsymbol{V}_{\boldsymbol{y}_i}]^{-1}\boldsymbol{D}_i]^{-1}[\sum_{i=1}^{N} \boldsymbol{D}_i' \, [\boldsymbol{V}_{\boldsymbol{y}_i}]^{-1}\boldsymbol{y}_i]$$

are quasi-likelihood regression parameter estimates that depend on the mean $\boldsymbol{\mu_i}$ and variance of $\boldsymbol{y}_i$. Moreover, the solution reflects the weighted linear combination of the covariates as included in w_{ij} in computing the coefficients through $\sum_{i=1}^{N}\sum_{j=1}^{T} w_{ij}\, y_{ij}$. Thus, the responses at each time are impacted by the present as well as past covariate values. We want to address the times when the responses in turn provide feedback on covariates in the future. The problem with the $\boldsymbol{R}_i$ matrix is that it assumes the way x_{is} impact y_{it} is the same way y_{is} impacts x_{it}. Also, it includes all the correlations whether or not they are based on valid moment conditions.

4.1.2 Covariate on Response Model

It is not uncommon in health research to observe individuals over time, while taking note of a set of covariates at each visit. In modeling the interdependence in particular the feedback from the responses at time t on the covariate in time $t + s$. Zeger and Liang (1992) showed how generalized estimating equations could be used to study feedback. They used a flexible class of feedback models by presenting the form of the conditional mean and covariance for each response given the covariates and the previous responses.

4.1.3 Marginal Model

There are two so-called basic approaches: subject-specific (SS) models in which heterogeneity in regression parameters is explicitly modelled; and population-averaged (PA) models in which the aggregate response for the population is the focus. The population-averaged response is modelled as a function of covariates without any concerns for subject-to-subject heterogeneity. In fact, the subject specific mixed models use both the method of population-averaged response as well as a distributional assumption concerning the variability among subjects to obtain coefficients estimates. Hence, the interpretation of coefficient between the two models differ (Zeger et al., 1988).

A population-averaged model for longitudinal data with time-dependent covariates has been with the GEE with non-diagonal working correlation structures would be valid with respect to a key condition (Pepe & Anderson, 1994). The key condition is that the regression coefficients $\boldsymbol{\beta}$ based on a population-averaged model $\mathrm{E}\left(\left(Y_{it}\middle|X_{ijt}\right)\right) = g(X_{i*t}\boldsymbol{\beta})$ are estimated using a diagonal working covariance matrix with the understanding that the marginal expected value is $\mathrm{E}\left(\left(Y_{it}\middle|X_{itj}\right)\right) = g\left(\left(Y_{it}\middle|X_{itj}, j = 1, \dots, n_i\right)\right)$, where Y_{it} is the outcome for the ith unit at the tth time and X_{ijt}

is the corresponding covariate. In other words, the condition says that the expectation of the response is a function of the current covariates only.

One approach to treat models with time-dependent covariates may be to include one or more additional terms, $X_{i,t-1}, X_{i,t-2}, \ldots \ldots$ as predictors. However, these do not involve feedback. As other researchers have pointed out, the approach will depend in part on the goals of the analysis (Diggle et al., 2002, Chapter 12). The concerns with GEE that Pepe and Anderson (1994) addressed may arise when the analyst or researcher is unable or unwilling to include additional terms $X_{i,t-1}, X_{i,t-2}, \ldots$ .... in the model. However, the consistency is assured regardless of the validity of the key assumption if a subject's repeated measurements are treated as independent (the independent working correlation is employed). Pepe and Anderson (1994) suggested the use of the independent working correlation when using GEE with time-dependent covariates as a "safe" choice of analysis.

4.1.4 Lagged Models

Lagged dependent models are often used in longitudinal or time series data. These models incorporate the dependent variable from previous time periods to help take into account autocorrelation in the data (Keele & Kelly, 2006). The models include lagged dependent variables or so-called endogenous variable as well as lagged predictor variables. However, when there is serial correlation, these models can produce biased estimates. Moreover, the introduction of a lagged dependent variable sometimes suppresses the effects of the covariates in the model, and often lacks reasonable causal interpretation (Anchen, 2001).

4.1.5 GMM Models

Lai and Small (2007) and Lalonde, Wilson, and Yin (2014) have shown that the generalized method of moments (GMM) model is a good choice when there are time-dependent covariates. However, they distinguish between coefficients that measure on responses from time t and covariates from time s, $\beta_j^{|s-t|}$ and covariates based on data with responses and covariates in the same time-period β_j^{tt} . While they have identified valid cases for estimation, they have rather combined valid cases ($s < t, s = t$) in the model when estimating β_j.

A generalized method of moments approach that partitions the regression coefficients to investigate individually the effect of covariates on the outcome when they are observed in the same time-period, as well as when they are observed in previous time-periods has been developed by Irimata Broatch and Wilson (2019). In Section 2, we review existing GMM regression models. In Section 3, we present a two-part model as an extension of these key models. It allows us to determine the impact of the covariates at different times (when possible) on the response and response on covariate at another time.

An example is as the case in the data were collected by the International Food Policy Research Institute in the Bukidnon Province in the Philippines and focus on quantifying the association between body mass index BMI, and morbidity 4 months into the future. The data were collected in 1984–1985 by surveying 448 households living within a 20-mile radius. Data were collected at four time points, on 4-month intervals (Bhargava, 1994). Lai and Small (2007) focused on the youngest child (1–14years) in each household and only consider those children who have complete data at all-time points. This realized 370 children with three observations. The response variable measure repeatedly was

$$y_{it} = \log\left[\frac{\text{days over last 2 weeks before time t child was sick}}{14.5 - \text{days over last 2 weeks before time t child was sick}}\right]$$

The time dependent predictors were BMI, age and survey time. Gender is a time independent predictor. Lai and Small (2007) gave two reasons suggesting that there may be feedback.

(a) If a child is sick, the child may not eat much and this could affect the child's weight in the future and

(b) Infections have generalized effects on nutrient metabolism and utilization (Martorell & Ho, 1984).

Lai and Small (2007) said both reasons (a) and (b) indicate potential feed-back effect and are most relevant for diarrheal infections (Martorell & Ho, 1984) and the proportion of children in the study who were sick (over a 2-week period) and who had diarrheal infection was only 9%.

4.2 Regression Models with Time-dependent Covariates

4.2.1 Lai and Small Model

Lai and Small (2007) used a marginal model for longitudinal continuous data with generalized method of moments (GMM) model to account for the time-dependent covariates. The model made full use of the valid moment conditions provided by time-dependent covariates, to obtain estimates. This estimation was done as they classified the time-dependent covariates into one of three types: I, II, and III. Through a selective grouping of certain moments, they obtained a single estimate for the regression coefficient associated with a certain covariate.

Consider repeated observations taken over T times on N subjects with J covariates such that $(y_{it}, \boldsymbol{x}_{it.})$ contain measurements at time t for subjects $i = 1, \dots, N$; for covariates $j = 1, \dots, J$; and times $t = 1, \dots, T$; where y_{it} denotes the outcome for subject i at time t, whose marginal distribution given the time-dependent vector $\boldsymbol{x}_{it} = (x_{i1t}, \dots \dots x_{iJt})$ of covariates follows a generalized linear model. We assumed that observations y_{is} and y_{kt} are independent whenever $i \neq k$ but not necessarily when $i = k$ and $s \neq t$. To obtain estimates of the coefficients, Lai and Small (2007) made use of the moment conditions

$$\mathrm{E}\left[\frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \beta_j}\{y_{it} - \mu_{it}(\boldsymbol{\beta})\}\right] = 0$$

for appropriately chosen s,t, and j, where $\mu_{it}(\boldsymbol{\beta})$ denotes expectation of y_{it} based on the vector of covariate values, $\boldsymbol{x}_{it}$ and $\boldsymbol{\beta}$ is the vector of parameters in the systematic component that describes the marginal distribution of y_{it}.

The challenge is to identify the appropriate method of moments associated with the covariates. The fact is that certain valid moment conditions may be omitted in the estimation of the regression coefficient if not properly accounted for. Such is the case when using GEE with working independence estimators with time-dependent covariates.

In order to identify all the valid moment conditions, Lai and Small (2007) relied on their definition of the type of covariate (types I, II, or III) to determine which moment conditions to use when estimating the model. Time-independent covariates were treated as type I. Each type of covariate requires a different set of moment conditions to be used in finding the corresponding coefficient's estimate. However, in their approach with such classification, it is assumed that the association between responses and covariates in any two different time-period remains the same.

4.2.2 Lalonde, Wilson, and Yin Marginal Model

Instead of using the grouping of moments, Lalonde, Wilson, and Yin (2014) adopted a method to identify the validity of each moment separately. However, similar to Lai and Small (2007) they grouped the valid moments to obtain an estimate of a single regression coefficient. In fact, they considered each moment condition separately for validity rather than grouping. They looked at bivariate correlations in determining if the corresponding moment condition is valid to use in obtaining estimates of regression coefficients. In particular, Lalonde, Wilson and Yin (2014) extended the classification of the types of covariates by introducing a type IV.

4.2.3 Irimata, Broatch, and Wilson Marginal Model

Irimata, Broatch and Wilson (2019) instead of grouping all the valid conditions to estimate a particular $\boldsymbol{\beta}_j$ over all time periods postulated that one needs to consider partitioning $\boldsymbol{\beta_j}$ so as to have $\beta_j^{tt}, \beta_j^{1t}, \dots, \beta_j^{st}$. Where β_j^{st} denotes the association of the covariate in period s with the response in period *t* $(s < t)$. They let $|s - t|$ dictate their grouping. Once this partitioning with valid moments were accomplished they uses GMM to obtain the regression coefficients. They argued that the T^2 moment conditions present at most T coefficients based on the possible valid moments.

4.3 Two-Part Model for Impact and Feedback

4.3.1 Model Prediction on Responses with Time-dependent Covariates

Correlated models are often addressed either in the random component, as in the case of GEE model, or in the systematic component, as in the case of generalized linear mixed models, with random effects. We proposed a model that in part addresses the correlation and the feedback through the systematic component but with fixed effects rather than with random effects.

We consider a partitioned generalized method of moments (GMM) model, for $s < t$. Let each covariate X_{ijt} to be measured at times $t = 1, 2, \ldots, T$; resulting for subject i and covariate $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots X_{ijT})'$. Thus, we present the model

$$\mathrm{g}(\mu_{it}) = \beta_0 + \beta_j^{tt} X_{ijt} + \beta_j^{|s-t|=1} X_{ij,s=t-1} \ldots + \beta_j^{|s-t|=t-1} X_{ij,s=1} \qquad [3.1]$$

with $s < t$, so

$$\mathrm{g}(\boldsymbol{\mu_i}) = \boldsymbol{X}_{ij}^{P} \boldsymbol{\beta}_j$$

where the $\boldsymbol{X}_{ij}^{P}$ is a matrix of a column of ones concatenated with a lower diagonal matrix as the systematic component, where $\boldsymbol{\mu}_i = (\mu_{i1} \ldots \mu_{iT})'$ dependent on the regression coefficient $\boldsymbol{\beta}_j = \left(\beta_0, \beta_j^{tt}, \beta_j^{|s-t|=1}, \beta_j^{|s-t|=T-1}\right)$ where s and t goes from 1 to T.

The coefficient β_j^{tt} denotes the effect of the covariate X_{ijt} on the response Y_{it} during the tth period. However, when $s \neq t$ it does not necessarily follow that we should interpret the past, using two different time periods in the same way as when $\boldsymbol{X}_{i,t}$ and $Y_{i,t}$ are in the same time period, $\mathrm{s} = \mathrm{t}$. The impact of a covariate on the response from another period is not intuitively the same as when they are in the same period. This is especially true in health research when time of dose will have impact on the reaction of the patient. Thus, their effects should not all be combined, but rather analyzed through the use of X_{ijs} and Y_{it}. This is best explained by $\beta_j^{|s-t|=1}$ for representing the effect of $X_{s=t-1}$ on Y_t , and by $\beta_j^{|t-s|=2}$ for representing the effect of $X_{s=t-2}$ on Y_t and so on. In general, we can consider the systematic component consisting of J covariates and let $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta_1}, \dots, \boldsymbol{\beta_J})'$ be the parameters associated with those covariates. Thus $\mathbf{X}$ is of maximum dimension NT by $(JT + 1)$ and $\boldsymbol{\beta}$ is a vector of maximum dimension $JT + 1$. The optimal GMM estimator of $\boldsymbol{\beta}$, is $\widehat{\boldsymbol{\beta}}^{GMM}$ obtained using the objective function $\mathbf{h_n'M_nh_n}$ where $\mathbf{h_n}$ is a (N_v) x 1 vector consists of all valid moment conditions, and $\mathbf{M_n}$ is a $(N_v \times N_v)$ weight matrix, where N_v denotes the total number of valid moment conditions.

Let $\boldsymbol{T_{vj}}$ be a square matrix of dimension T that specifies valid moment conditions for the jth covariate. Thus, elements in $\boldsymbol{T_{vj}}$ take on the value of one when there is valid moment, such that the condition:

$$\mathrm{E}_{\boldsymbol{\beta_0}}\left[\frac{\partial \mu_{is}(\boldsymbol{\beta_0})}{\partial \beta_j^{st}}\{y_{it} - \mu_{it}(\boldsymbol{\beta_0})\}\right] = 0,$$

holds for the jth covariate. Elements in $\boldsymbol{T_{vj}}$ are zero when the moment is not valid. Convert $\boldsymbol{T_{vj}}$ into a 1 x $0.5T(T + 1)$ row vector by reshaping it for $j = 1, \dots, J$ and concatenate the rows for all covariates to form $\boldsymbol{T_{shape}}$, a $J \times T^2$matrix. The number of 1's in the $\boldsymbol{T_{shape}}$ matrix, is the total number of valid moment conditions, denoted by N_v. But all $0.5T(T - 1)$ moments are set to zero, pertaining to cases where $s < t$. Our model is based only on the valid moment conditions.

Thus, the fitted model is

$$\mu_{it}(\beta) = \beta_0 + \beta_j^{tt} X_{ijt} + \sum_{\boldsymbol{s=1}}^{\boldsymbol{t-1}} \beta_j^{[s]} X_{ij(t-s)} |_{\textbf{valid moments}}$$

when the valid moments conditions exist and β_j^{tt} denotes the regression parameter for the cases when the valid moment conditions always exist, which represent the effect of the covariate in the same period as the response. The coefficient $\beta_j^{[s]}$ represents of the effect of the covariate on the response when the response is not in the same time period but the moment is valid and in particular when $s < t$. The GMM estimator $\widehat{\boldsymbol{\beta}}^{GMM}$ is the argument to minimize the quadratic objective function $\mathbf{h_n}(\boldsymbol{\beta_0})' \, \mathbf{M_n}(\boldsymbol{\beta_0}) \; \mathbf{h_n}(\boldsymbol{\beta_0})$, such that

$$\widehat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta_0}}{\operatorname{argmin}} \, \mathbf{h_n}(\boldsymbol{\beta_0})' \, \mathbf{M_n}(\boldsymbol{\beta_0}) \; \mathbf{h_n}(\boldsymbol{\beta_0}).$$

and $N_v \times N_v$ weight matrix $\mathbf{M_n}$ is computed as $\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{h}_i \boldsymbol{h}_i'\right)^{-1}$.The asymptotic variance of $\widehat{\boldsymbol{\beta}}_{GMM}$ is computed as

$$\left[\left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \boldsymbol{h}_i(\boldsymbol{\beta})}{\partial \beta_j^{st}}\right)' \boldsymbol{M_n}(\boldsymbol{\beta}) \left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \boldsymbol{h}_i(\boldsymbol{\beta})}{\partial \beta_j^{st}}\right)\right]^{-1},$$

evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_{GMM}$. In the case of logistic regression, the elements take the form:

$$\frac{\partial \mu_{is}(\boldsymbol{\beta_0})}{\partial \beta_j^{st}} [y_{it} - \mu_{it}(\boldsymbol{\beta_0})] = x_{isj} \mu_{is}(\boldsymbol{\beta_0})[1 - \mu_{is}(\boldsymbol{\beta_0})][y_{it} - \mu_{it}(\boldsymbol{\beta_0})] \quad [3.2]$$

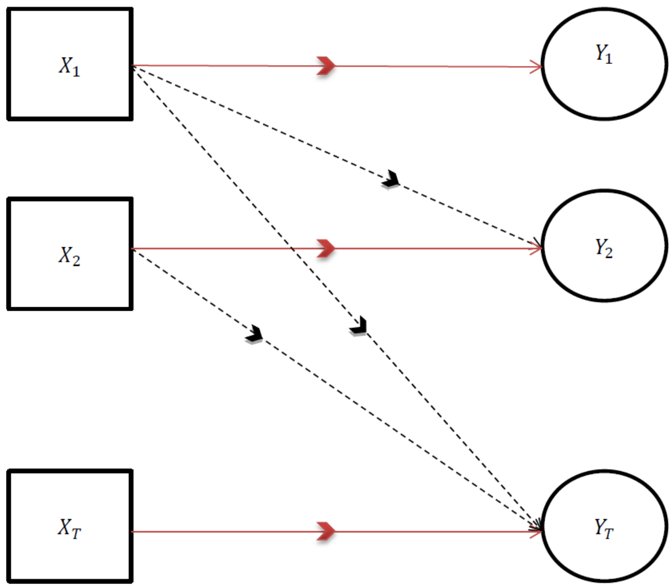
where

$$\mu_{it}(\boldsymbol{\beta_0}) = \frac{\exp(\boldsymbol{X_i \beta})}{1+\exp(\boldsymbol{X_i \beta})},$$

such that the element in row s, column t of $\boldsymbol{T_{vj}}$ takes value 1. Stage 1 is explained by Figure 4.1.

Figure 4.1

Stage 1: Impact of X on Y



4.3.2 Feedback of Responses on Time-dependent Predictors Model

In part 2, we address feedback from the responses to the time-dependent predictors in the model. This is not always of significance. Recall, in the International Food Policy Research Institute in Philippines, Lai and Small gave two reasons that one would consider such feedback. We consider function of the mean of the covariate X on the response Y such that,

$$g(\gamma_{irs}) = \alpha_0 + \alpha_r^{ss} Y_{is} + \alpha_r^{|s-t|=1} Y_{i,t=s-1} \ldots . + \alpha_r^{|s-t|=s-1} Y_{i,t=1} \qquad [3.3]$$

for $t < s$. We address the feedback as in the case of the direct impact through the systematic component but with fixed effects rather than random effects. Thus, in stage one, we presented the model with $s < t$ in [3.3]. In stage 2, we present a similar model with $t < s$ for time-dependent covariates and a function so

$$\mathrm{g}(\boldsymbol{\gamma}_{ir}) = \boldsymbol{Y}_r^{\boldsymbol{P}} \boldsymbol{\alpha}_r$$

where the $\boldsymbol{Y}_r^{\boldsymbol{P}}$ matrix consists of a column of ones concatenated with a diagonal and lower diagonal matrix as the systematic component, with regression coefficient $\boldsymbol{\alpha}_r =$ $\left(\alpha_{r0}, \alpha_r^{|s-t|=0}, \alpha_r^{|s-t|=1}, \alpha_r^{|s-t|=T-1}\right)$ where t and s goes from 1 to T. The lower diagonal matrix in this case is the upper diagonal matrix in stage 1 that was ignored. The impact of a response on the covariate in another period is not intuitively the same as when they are in the same period. There may be delayed effect. This is explained by $\alpha_r^{|s-t|=1}$ for representing the effect of Y_{t-1} on X_t , and by $\alpha_r^{|s-t|=2}$ for representing the effect of Y_{t-2} on X_t and so on.

Define the optimal GMM estimator of $\boldsymbol{\alpha}$, as $\widehat{\boldsymbol{\alpha}}^{GMM}$ which is obtained by solving the objective function $\mathbf{k}_\mathbf{n}'\mathbf{S}_\mathbf{n}\mathbf{k}_\mathbf{n}$ where $\mathbf{k}_\mathbf{n}$ is a (N_{v_α}) x 1 vector consists of all valid moment conditions, and $\mathbf{S}_\mathbf{n}$ is a $\left(N_{v_\alpha}\right)$ x (N_{v_α}) weight matrix, where N_{v_α} denotes the total number of valid moment conditions. Similarly, we can identify the valid moments during stage 1 as the method of valid moments depends on correlations and not a distinction of covariate from response.

Thus, for $s > t$, let $\Omega_{tt} \in [x_s, y_t \in s = t]$ and consider each valid moment condition where $\Omega_{st} \in [x_s, y_t \in s \neq t]$. There are T members in Ω_{tt} and one member for each of Ω_{st}. Thus the fitted model is

$$\gamma_{irs}(\boldsymbol{\alpha}) = \alpha_{r0} + \alpha_r^{ss} Y_{is} + \sum_{\boldsymbol{t}=\mathbf{1}}^{\boldsymbol{s}-\mathbf{1}} \alpha_r^{|t|} Y_{i,s-t} \,|_{\textbf{valid moments}}$$

when the valid moments conditions exist and α_r^{ss} denotes the regression parameter for the cases when the valid moment conditions always exist, which represent the effect of the response in the same period as the covariate. The coefficient $\alpha_r^{|t|}$ represents the effect of the response on the covariate when the response is measured before the covariate is but the moment is valid and in particular $s > t$.

When there is only one j^th^ covariate providing feedback, we obtain GMM estimators for $\hat{\alpha}_j^{ts}$ following a similar process to that in part 1. However, if we have more than one covariate showing feedback, say R, we use a simultaneous GMM model to estimate the regression coefficients $\boldsymbol{\alpha}$ (Irimata et al., 2018). We let X_{irt} denote the r^{th} time-dependent covariate ($r = 1, \dots R$) from the i^{th} subject ($i = 1, \dots, N$) measured at the t^{th} period ($t = 1, \dots, T$). For the i^{th} subject, measured T times, the vector $\boldsymbol{X_{ir}} = (X_{ir1} \quad X_{ir2} \dots \quad X_{irT})'$ contains the measurements on the r^{th} time-dependent covariate. Then, for the i^{th} subject measured T times on the R time-dependent covariates with feedback there is a vector $\boldsymbol{X_i} = (\boldsymbol{X_{i1}} \quad \boldsymbol{X_{i2}} \dots \quad \boldsymbol{X_{iR}})'$ of length $(R \times T)$. Assume that each vector $\boldsymbol{X_{ir}}$ of length T of time-dependent covariates has its own partitioned data matrix $\boldsymbol{Y_{ir}^p}$, a block diagonal partitioned data matrix of lagged outcomes such that

$$\boldsymbol{Y_{ir}^p} = \begin{bmatrix} Y_{i1} & 0 & 0 & 0 \\ Y_{i2} & Y_{i1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ Y_{iT} & Y_{i,T-1} & \dots & Y_{i1} \end{bmatrix}$$

with associated regression coefficients $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1} \quad \dots \quad \boldsymbol{\alpha_R})'$. Where $\boldsymbol{\alpha_r}$ $(r = 1, \dots R;)$ is the vector of regression coefficients associated with $\boldsymbol{Y_{ir}^p}$ for the r^{th} time-dependent covariate.

We model the feedback from the outcome to each of the R time-dependent covariates using the simultaneous GMM. We have vectors $\boldsymbol{G_1}, \boldsymbol{G_2}, \dots, \boldsymbol{G_R}$ containing the sample average of the valid moment conditions $N_{v1}, N_{v2}, \dots, N_{vR}$ associated with the vectors of regression coefficients $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \dots, \boldsymbol{\alpha_R}$, respectively. So that

$$\boldsymbol{G_1}(\boldsymbol{\alpha_1}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g_{i1}}$$

$$\boldsymbol{G_2}(\boldsymbol{\alpha_2}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g_{i2}}$$

$$\vdots$$

$$\boldsymbol{G_R}(\boldsymbol{\alpha_R}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g_{iR}}$$

The $\boldsymbol{g_{ir}}$'s are vectors of length N_{vr} and contain the values of all the valid moment conditions for subject i used to estimate the feedback of the outcome on the r^{th} time-dependent covariate.

We estimate the regression coefficients $\boldsymbol{\alpha}$, using the GMM estimation method by minimizing the quadratic form

$$Q_N(\boldsymbol{\alpha}) = \boldsymbol{G}_N(\boldsymbol{\alpha})'\boldsymbol{W}_N(\boldsymbol{\alpha})\boldsymbol{G}_N(\boldsymbol{\alpha})$$

Such that $\boldsymbol{G}_N(\boldsymbol{\alpha}) = \left(\boldsymbol{G_1}(\boldsymbol{\alpha_1}), \boldsymbol{G_2}(\boldsymbol{\alpha_2}), \dots, \boldsymbol{G_R}(\boldsymbol{\alpha_R})\right)'$ is a vector of length $M = N_{v1} + N_{v2} + \cdots + N_{vR}$ and $\boldsymbol{W}_N(\boldsymbol{\alpha})$ is the optimal weight matrix of dimension $M \times M$ with

$$\boldsymbol{W}_N(\boldsymbol{\alpha}) = \left[\frac{1}{N}\left(\sum_{i=1}^{N} \begin{bmatrix} \boldsymbol{g_{i1}} \\ \boldsymbol{g_{i2}} \\ \vdots \\ \boldsymbol{g_{iR}} \end{bmatrix} \begin{bmatrix} \boldsymbol{g_{i1}} \\ \boldsymbol{g_{i2}} \\ \vdots \\ \boldsymbol{g_{iR}} \end{bmatrix}'\right)\right]^{-1}$$

Thus, $\widehat{\boldsymbol{\alpha}}_{GMM} = \text{argmin}_{\alpha \in \mathrm{A}}\, Q_N(\boldsymbol{\alpha})$.

When the r^{th} time-dependent covariate is binary the elements of $\boldsymbol{g_{ir}}$ take a similar form to those in [3.2]. If the r^{th} time-dependent covariate is continuous the elements of $\boldsymbol{g_{ir}}$ have the form

$$\boldsymbol{g_{ir}} = y_{it}[x_{irs} - \gamma_{irs}(\boldsymbol{\alpha_r})]$$

We calculate the asymptotic variance of $\widehat{\boldsymbol{\alpha}}$ by first obtaining

$$\hat{A} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} \frac{\partial \boldsymbol{g_{i1}}(\widehat{\boldsymbol{\alpha_1}})}{\partial \boldsymbol{\alpha_1}} \\ \frac{\partial \boldsymbol{g_{i2}}(\widehat{\boldsymbol{\alpha_2}})}{\partial \boldsymbol{\alpha_2}} \\ \vdots \\ \frac{\partial \boldsymbol{g_{iR}}(\widehat{\boldsymbol{\alpha_R}})}{\partial \boldsymbol{\alpha_R}} \end{bmatrix}$$

Where $\hat{A}$ is the vector of partial derivatives evaluated at $\widehat{\boldsymbol{\alpha}}_{GMM}$ and the asymptotic variance of $\widehat{\boldsymbol{\alpha}}_{GMM}$ is given by $Var(\widehat{\boldsymbol{\alpha}}_{GMM}) = \left(\hat{A}W_N^{-1}\hat{A}\right)^{-1}$ with W_N^{-1} evaluated at $\widehat{\boldsymbol{\alpha}}_{GMM}$.

The partial derivatives in $\hat{A}$ can be calculated separately for each $\widehat{\boldsymbol{\alpha_r}}$. If the r^{th} time-dependent covariate is binary the rows of $\frac{\partial \boldsymbol{g_{ir}}(\widehat{\boldsymbol{\alpha_r}})}{\partial \boldsymbol{\alpha_r}}$ are calculated using the following expression:

$$\frac{\frac{\partial \gamma_{irt}(\boldsymbol{\alpha_r})}{\partial \alpha_r^{[k]}}\left(x_{irs} - \gamma_{irs}(\boldsymbol{\alpha_r})\right)}{\partial \alpha_r^{[k']}}$$

$$= y_{it}\gamma_{irt}(\boldsymbol{\alpha_r})[1 - \gamma_{irt}(\boldsymbol{\alpha_r})]\{y_{it}[1 - 2\gamma_{irt}(\boldsymbol{\alpha_r})][x_{irs} - \gamma_{irs}(\boldsymbol{\alpha_r})]$$
$$- y_{is}\gamma_{irs}(\boldsymbol{\alpha_r})[1 - \gamma_{irs}(\boldsymbol{\alpha_r})]\}$$

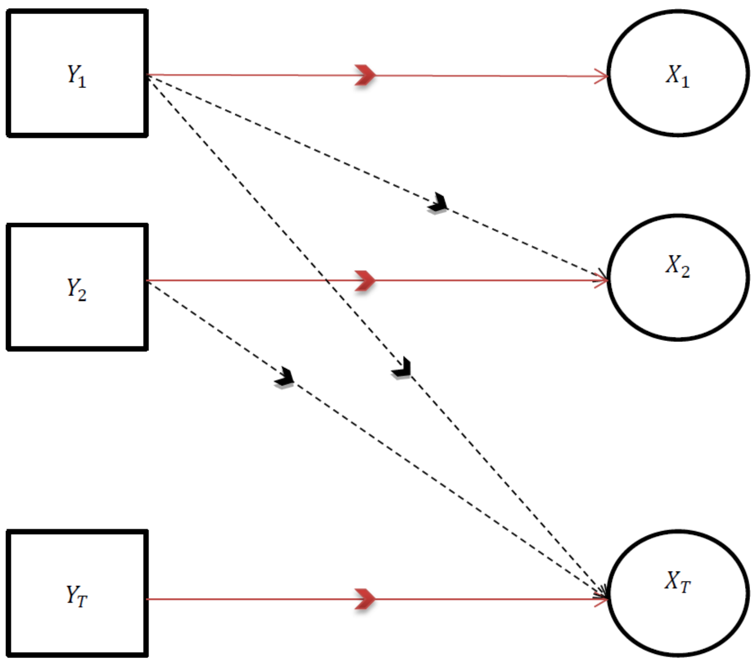
When the r^{th} time-dependent covariate is continuous the rows of $\frac{\partial \boldsymbol{g_{ir}}(\widehat{\boldsymbol{\alpha_r}})}{\partial \boldsymbol{\alpha_r}}$ can be computed as:

$$\frac{\frac{\partial \gamma_{irt}(\boldsymbol{\alpha}_r)}{\partial \alpha_r^{[k]}}\left(x_{irs} - \gamma_{irs}(\boldsymbol{\alpha}_r)\right)}{\partial \alpha_r^{[k']}} = -y_{it}y_{is}$$

For $k = 0, 1, \dots, T-1$ and $k = 0, 1, \dots, T-1$.

Figure 4.2

Stage 2: Impact of Y on X



The two-part model consists of two submodels: a marginal generalized linear model for the responses and a similar marginal generalized linear model for the time-dependent covariate. In stage one, as shown in Figure 1, for the time-dependent covariates X on the response Y for T times depicting the impact of current effects, and delayed effects on Y. Thus, we write two-part model as

$$\mu_{it}(\beta) = \beta_0 + \boldsymbol{X_0} + \beta_j^{tt} X_{jt} + \sum_{\boldsymbol{s<t}}^{\boldsymbol{T}} \beta_j^{|s-t|=T-1} X_{jt} \,|_{\textbf{valid moments}}$$

With time-independent variable $\boldsymbol{Z_0}$ with valid moments, and time-dependent predictors with possible feedback

$$\gamma_{irs}(\alpha) = \alpha_0 + \sum_{\boldsymbol{s>t}}^{\boldsymbol{T}} \alpha_r^{|s-t|=T-1} Y_{irs} \,|_{\textbf{valid moments}}$$

4.4 Numerical Examples

4.4.1 Child Morbidity in the Philippines

We revisited the data collected by the International Food Policy Research Institute in the Bukidnon Province in the Philippines to check what impacted children's morbidity. Our outcome variable was whether the child was sick. Our time-dependent covariate was body mass index (BMI). We also include gender and age as covariates in our model. We first investigated how our covariates impacted children's likelihood of being sick using Part-1 of our model. We then checked whether sickness status influenced future values of children's BMI using Part-2 of the model.

Table 4.1 presents results for Part-1. We found that age had a significant effect on morbidity ($p<0.0001$). Body mass index had immediate ($p=0.038$), delayed ($p=0.044$) and further delayed effects ($p=0.0001$) on children's likelihood of getting sick. There was an increasing relationship between BMI and morbidity status over time, the more time passed the less BMI increases reduced the likelihood of being sick.

Table 4.1

Modeling Morbidity

Covariate	OR	95% CI		P-value
Gender	1.140	0.847	1.534	0.388
Age	0.979	0.969	0.990	<0.0001
BMI cross-sectional	0.885	0.788	0.993	0.038
BMI lag-1	0.980	0.961	0.999	0.044
BMI lag-2	1.035	1.014	1.057	0.001

We also investigated whether there was a feedback effect from child morbidity to BMI, Table 4.2. We found that whether a child was sick or not did not significantly impact the child's BMI in the future. However, our overall model indicated that if children were sick in time 1 or time 2 than their BMI decreased in the next time-period. For those children who were sick in time 1, their BMI decreased in time 2 which increased their likelihood of being sick in time 2 and time 3.

Table 4. 2

Feedback from Morbidity to BMI

	Effect	95% CI		p-value
Intercept	15.433	13.353	17.514	<0.0001
Sick cross-sectional	-0.095	-2.846	2.656	0.946
Sick lag-1	-0.249	-3.257	2.758	0.871
Sick lag-2	0.097	-3.105	3.299	0.953

4.4.2 Add Health Study

We used data from the first four waves of the Add health study to determine what risk factors associated with obesity status over time, this is Part 1 of the model. The risk factors studied included time-independent and time-dependent risk factors. Our time-independent risk factors included race (white vs non-white) and gender (male vs female). The time-dependent risk factors included social alcohol use, physical activity level, depression level and number of hours spent watching television. We also studied whether obesity status associated with future values of the risk factors physical activity level and depression level, the Part-2 of the model.

We found that when fitting Part-1 of the model, all time-dependent risk factors had valid moment conditions across a one time period lag. However, only social alcohol use and physical activity level had valid moment conditions across a two-time period lag and across a three time period lag only physical activity level had valid moment conditions. Table 4.3 presents the results for Part-1 of the model. We can observe that physical activity level ($p<0.0001$), depression level ($p<0.0001$) and number of hours spent watching television ($p<0.0001$) had significant cross-sectional associations with obesity status. Physical activity level ($p<0.0001$) and depression ($p<0.0001$) associated significantly with obesity status across a one time period lag. Social alcohol use ($p<0.0001$) and physical activity level ($p<0.0001$) associated with obesity status across a two-time period lag. Physical activity level ($p<0.0001$) associated with obesity status across a three time period lag.

We studied whether obesity status significantly associated with future physical activity level, Table 4.4. We found that obesity status significantly associated with physical activity levels in the next three time-periods. There was an increasing association between obesity status and physical activity levels, the more time distance (lags) between obesity status and physical activity levels the less obesity status reduced physical activity levels.

We found that physical activity reduced the likelihood of being obese at the cross-sectional and lag-1 time periods. Those who were obese reported less physical activity levels in the next time period, which resulted in them being more likely to be obese which again reduced their physical activity levels. In general, being obese had a negative impact in the physical activity levels for the next three time periods.

Table 4.3

Modeling Obesity Status

Time period	Risk factor	OR	95% CI		p-value
Cross-sectional effects	Race	1.239	1.027	1.496	0.025
	Gender	0.922	0.773	1.101	0.369
	Alcohol	1.019	0.874	1.188	0.812
	Activity	0.849	0.796	0.906	<0.0001
	Depression	1.727	1.410	2.115	<0.0001
	TV hours	1.016	1.012	1.020	<0.0001
Delayed effects	Alcohol	1.041	0.917	1.182	0.531
	Activity	0.909	0.866	0.955	<0.0001
	Depression	1.788	1.454	2.197	<0.0001
	TV hours	1.004	1.000	1.008	0.059
Further delayed effects	Alcohol	1.333	1.166	1.523	<0.0001
	Activity	1.193	1.144	1.243	<0.0001
Furthermost delayed effects	Activity	1.174	1.125	1.224	<0.0001

Table 4.4

Feedback from Obesity Status to Physical Activity Level

	Effect	95% CI		p-value
Obesity cross-sectional	-0.800	-0.872	-0.729	<0.0001
Obesity lag-1	-0.415	-0.469	-0.361	<0.0001
Obesity lag-2	-0.207	-0.247	-0.168	<0.0001
Obesity lag-3	-0.070	-0.111	-0.028	0.001

We also studied the feedback effect of obesity status on future values of depression levels, Table 4.5. Obesity status significantly associated with depression levels in the next two time periods. There was a changing association between obesity status and depression levels over time, being obese decreased depression levels in the next time-period, but increased depression levels at the two-time period lag. We also observed that the more depressed a person was the more likely that person was of being obese in the present in the next time-period, however, obesity status not always resulted in higher depression levels. Our analysis shows a complicated relationship between obesity status and depression levels, such results have been observed in the past (Newman & Robertson, 2018).

Table 4.5

Feedback from Obesity Status to Depression Level

	Effect	95% CI		p-value
Obesity cross-sectional	0.275	0.252	0.299	<0.0001
Obesity lag-1	-0.138	-0.172	-0.104	<0.0001
Obesity lag-2	0.088	0.047	0.129	<0.0001
Obesity lag-3	0.031	-0.026	0.087	0.285

4.5 Conclusions

The correlation inherent in measures recorded over time as affected by time-dependent covariates presents a set of extra challenges as compared to the analysis of cross-sectional data. In particular, the changes and feedback presented when the covariates are time-dependent cannot be ignored. Often, the feedback effects go unchecked. However, any modeling of longitudinal data must address the impact from the feedback as well as the immediate and the delayed effects of covariates on the responses. We found that modeling time-dependent covariates allows us to identify the valid moments, which in turn helps determine the significant predictors to consider. While there is merit in the models that address time-dependent covariates (Irimata et al., 2019; Lai & Small, 2007; Lalonde et al., 2014; Zhou et al., 2014), they do not always account for the feedback. The two-part partitioned GMM model presented allows one to account for the feedback across different time-periods. It partitions the regression coefficients and allow us to identify directional and delayed effects.

CHAPTER 5

CONCLUSIONS

In this dissertation, I present three papers that address the modeling of longitudinal binary outcomes while accounting for covariates that are time-dependent.

The first paper presents an alternative model to estimate the current and future effects of time-dependent covariates on binary outcomes when the partitioned GMM model provides numerically unstable estimates and standard errors of the regression coefficients. This model uses the partitioned data matrix to account for the correlation among the outcomes. It is estimated using Markov Chain Monte Carlo algorithms that provide Bayesian estimates of the regression coefficients. It provides more efficient estimates than the partitioned GMM model and as efficient estimates as the GEE model with lagged covariates and an independent working correlation matrix. This was shown by the simulation study.

The second paper introduces an approach for jointly modeling of two or more longitudinal binary outcomes. This model accounts for both the correlation among and between multiple outcomes as well as the changing effects of time-dependent covariates on these outcomes. The model estimation consists of two stages. In the first stage, the working correlation matrix that contains the within and between outcome correlation is estimated using joint GEE. In the second stage, the regression coefficients are estimated using Markov Chain Monte Carlo sampling algorithms, thus providing Bayesian estimates. The regression coefficient estimates are efficient even when the working correlation matrix is miss specified.

In the third paper, I propose a two-part model that investigates the changing effects of time-dependent covariates on a binary outcome as well as feedback from the outcome to the time-dependent covariates. Stage 1 of the model allows one to determine whether there are significant cross-sectional and lagged associations between several time-dependent covariates and a binary outcome over time, while controlling for time-independent covariates. Stage 2 of the model allows one to determine whether the outcome has significant impact on future values of the time-dependent covariates. The partitioned GMM model is used to estimate stage 1 of the model and stage 2 depending if the feedback is practical enough to warrant modeling. If one is interested on the feedback involving two or more time-dependent covariates, then stage 2 of the model is estimated using the simultaneous partitioned GMM model.

Overall, these three papers discuss different approaches to help better understand the relationship between time-dependent covariates and binary outcomes when modeling longitudinal data. These models are marginal mean models. They can be extended to subject-specific models.

REFERENCES

Alzahrani, H. (2016). *Multivariate Binary Longitudinal Data Analysis* [Florida State University]. http://fsu.digital.flvc.org/islandora/object/fsu%3A507619

Anchen, C. H. (2001). *Why lagged dependent variables can suppress the explanatory power of other independent variables*. Annual Meeting of the Political Methodology Science Association, Los Angeles, CA.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Wiley.

Beekly, D. L., Ramos, E. M., Erin, M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., Kukull, W. A., Reiman, E. M., Kowall, N., Landreth, G., Shelanski, M., Welsh-Bohmer, K., Levey, A. I., Potter, H., Ghetti, B., … Raskind, M. (2007). *The National Alzheimer's Coordinating Center (NACC) database: The uniform dataset*. https://www.alz.washington.edu/

Bhargava, A. (1994). Modelling the health of Filipino children. *Journal of the Royal Statistical Society, A*, *157*(3), 417–432.

Blaesild, P., & Jensen, J. L. (1985). Saddlepoint formulas for reproductive exponential families. *Scandinavian Journal of Statistics*, *12*(3), 193–202.

Canapu, M., Gonen, M., & Begg, C. B. (2013). An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in Medicine*, *32*(26), 4550–4566.

Caragea, P. C., & Smith, R. L. (2007). Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, *98*(7), 1417–1440.

Chandler, R. E., & Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, *94*(1), 167–183.

Chen, I. C., & Westgate, P. M. (2017). Improved methods for the marginal analysis of longitudinal data in the presence of time-dependent covariates. *Statistics in Medicine*, *36*(16), 2533–2546.

Chernozhukov, V., & Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, *115*(2), 293–346.

Cognitive impairment: A call for action. (2011). Center for Disease Control and Prevention. https://www.cdc.gov/aging/pdf/cognitive_impairment/cogimp_poilicy_final.pdf

Cox, D. R., & Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, *91*(3), 729–737.

Crowder, M. (1985). Guassian estimation for correlated binary data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *47*(2), 229–237.

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

Efron, B. (2015). Frequentist accuracy of Bayesian estimates. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *77*(3), 617–649.

Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, *45*(4), 302–304.

Fang, D., Sun, R., & Wilson, J. R. (2018). Joint modeling of correlated binary outcomes: The case of contraceptive use and HIV knowledge in Bangladesh. *PLOS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0190917

Fieuws, S., Verbeke, G., & Molenberghs, G. (2007). Random effects models for multivariate repeated measures. *Statistical Methods in Medical Research*, *16*(5), 387–397.

Gabry, J., Goodrich, B., & Guo, J. (n.d.). *Rstan: The R interface to Stan. R-package version 2.18.2*. https://cran.r-project.org/web/packages/rstan/index.html

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (Third Edition). Chapman and Hall.

Ghebremichael, M. (2015). Joint modeling of correlated binary outcomes: HIV-1 and HSV-2 co-infection. *Journal of Applied Statistics*, *42*(10), 2180–2191.

Givens, G. H., & Hoeting, J. A. (2013). *Computational Statistics* (Second Edition). John Wiley & Sons, Inc.

Glasbey, C. A. (2001). Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *50*(2), 143–154.

Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, *1*(3), 177–193.

Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, *14*(3), 262–280.

Harris, K. M., & Udry, J. R. (2016). *National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use]*. In: Inter-university Consortium for Political and Social Research (ICPSR) [distributor]

Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal cateogorical data. *Biometrics*, *58*(2), 342–351.

Heagerty, P. J., & Comstock, B. A. (2013). Exploration of Lagged Associations using Longitudinal Data. *Biometrics*, *69*(1), 197–205.

Hjort, N. L., Omre, H., Frisen, M., Godtliebsen, F., Helgeland, J., Moller, J., Vedel Jansen, E. B., Rudemo, M., & Stryhn, H. (1994). Topics in spatial statistics (with discussion, comments and rejoinder). *Scandinavian Journal of Statistics*, *21*(4), 289–357.

Hjort, N. L., & Varin, C. (2008). ML, PL, QL in Markov chain models. *Scandinavian Journal of Statistics*, *35*(1), 64–82.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer.

Hogervorst, E., Hupper, F., Matthews, F. E., & Brayne, C. (2008). Thyroid function and cognitive decline in the MRC Cognitive Function and Ageing Study. *Psychoneuroendocrinology*, *33*(7), 1013–1022.

Hong, I., Coker-Bolt, P., Anderson, K. R., Lee, D., & Velozo, C. A. (2016). Relationship between physical activity and overweight and obesity in children: Findings from the 2012 National Health and Nutrition Examination Survey National Youth Fitness Survey. *The American Journal of Occupational Theraphy*, *70*(5), 7005180060p1-7005180060p8.

Hu, F., Goldberg, J., Hedeker, D., Flay, B., & Pentz, A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology*, *147*(7), 694–703.

Huggins, J., Campbell, T., & Broderick, T. (2016). Coresets for scalable Bayesian logistic regression. *NIPS Proceedings*. Neural Information Processing Systems, Barcelona, Spain.

Irimata, K. M., Broatch, J., & Wilson, J. R. (2019). Partitioned GMM logistic regression models for longitudinal data. *Statistics in Medicine*, *38*(12), 2171–2183.

Irimata, K. M., Vazquez Arreola, E., & Wilson, J. R. (2018). Simultaneous GMM models with time-dependent covariates. In *Three Essays on Correlated Binary Outcomes: Detection and Appropriate Models*. https://repository.asu.edu/items/49005

Irimata, K. M., & Wilson, J. R. (2018). Using SAS to estimate lagged coefficients with the %partitionedGMM macro. *SAS Global Forum 2018 Conference Proceedings*. SAS Gobal Forum, Denver, CO.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall.

Keele, L., & Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis*, *14*(2), 186–205.

Lai, T. L., & Small, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: A generalised method of moments approach. *Journal of the Royal Statistical Society, Series B*, *69*(1), 79–99.

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

Lalonde, T. L., Wilson, J. R., & Yin, J. (2014). GMM logistic regression models for longitudinal data with time-dependent covariates. *Statistics in Medicine*, *33*(27), 4756–4769.

Liang, G., & Yu, B. (2003). Maximum pseudo likelihood estimation in network tomography. *IEEE Transactions on Signal Processing*, *51*(8), 2043–2053.

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

Lincoln, K. D., Abdou, C. M., & Lloyd, D. (2014). Race and socioeconomic differences in obesity and depression among Black and non-Hispanic White Americans. *Journal of Health Care for the Poor and Underserved*, *25*(1), 257–275.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Sinha, D., Parzen, M., & Lipshultz, S. (2009). Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: An application to AIDS. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *172*(1), 3–20.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (Second). John Wiley & Sons, Inc.

LoBue, C., Denney, D., Hynan, L. S., Rossetti, H. C., Lacritz, L. H., Hart, J., Womack, K. B., Woon, F. L., & Munro-Cullum, D. C. (2016). Self-reported traumatic brain injury and mild cognitive impairment: Increased risk and earlier age of diagnosis. *Journal of Alzheimer's Disease*, *51*(3), 727–736.

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Penninx, B. W., & Zitman, F. G. (2010). Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. *Archives of General Psychiatry*, *67*(3), 220–229.

Martorell, R., & Ho, T. J. (1984). Malnutrition, Morbidity, and Mortality. *Population and Development Review*, *10*, 49–68.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (Second). Chapman and Hall.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, *17*(1), 53–73.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*(5), 995–1026.

Molenberghs, G., & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.

Muller, H.-G., & Stadtmuller, U. (2005). Generalized functional linear models. *Annals of Statistics*, *33*(2), 774–805.

Nakata, T., & Tonetti, C. (2014). *Small sample properties of Bayesian estimators of labor income processes* (Finance and Economics Discussion Series). Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.

Newman, T., & Robertson, C. (2018, November 13). Obesity and depression revisited. *Medical News Today*. https://www.medicalnewstoday.com/articles/323668

Obermeier, V., Scheipl, F., Heumann, C., Wassermann, J., & Küchenhoff, H. (2015). Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *64*(2), 395–412.

Paterniti, S., Verdier-Taillefer, M. H., Dufouil, C., & Alperovich, A. (2002). Depressive symptoms and cognitive decline in elderly people. *The British Journal of Psychiatry*, *181*, 406–410.

Pepe, M., & Anderson, J. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics: Simulation and Computation*, *23*(4), 939–951.

Qu, A., Lindsay, B. G., & Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, *87*(4), 823–836.

Rocca, W. A., Petersen, R. C., Knopman, D. S., Hebert, L. E., Evans, D. A., Hall, K. S., Gao, S., Unverzagt, F. W., Langa, K. M., Larson, E. B., & White, L. R. (2011). Trends in the incidence and prevalence of Alzheimer's disease, dementia and cognitive impairment in the United States. *Alzheimer's & Dementia*, *7*(1), 80–93.

Roos, M., & Held, L. (2011). Sensitivity analysis in Bayesian Generalized Linear Mixed Models for binary data. *Bayesian Analysis*, *6*(2), 259–278.

Schildcrout, J. S., & Heagerty, P. J. (2005). Regression analysis of longitudinal binary data with time-dependent environmental covariates: Bias and efficiency. *Biostatistics*, *6*(4), 633–652.

Sheffield, K. M., & Peek, M. K. (2011). Changes in the prevalence of cognitive impairment among older Americans 1993-2004: Overall trends and differences by Race/Ethnicity. *American Journal of Epidemiology*, *174*(3), 274–283.

Singh, G. K., Siahpush, M., & Kogan, M. D. (2010). Rising social inequalities in US childhood obesity, 2003-2007. *Annals of Epidemiology*, *20*(1), 40–52.

Skene, A. M., Shaw, J. E. H., & Lee, T. D. (1986). Bayesian modelling and sensitivity analysis. *Journal of the Royal Statistical Society, D (The Statistician)*, *35*(2), 281–288.

Stan Development Team. (n.d.). *Stan modeling language users guide and reference manual, Version 2.18.0*. https://mc-stan.org/docs/2_19/stan-users-guide/index.html

Stein, M. L., Chi, Z., & Welty, L. J. (20014). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *66*(2), 275–296.

Tanaka, M. (2020). Adaptive MCMC for Generalized Method of Moments with many moment conditions. *ArXiv*. https://doi.org/arXiv:1811.00722v4

Traversy, G., & Chaput, J. P. (2015). Alcohol consumption and obesity: An update. *Current Obesity Reports*, *4*(1), 122–130.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, *92*, 1–28.

Varin, C., Nancy, R., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*(1), 5–42.

Vazquez Arreola, E., & Wilson, J. R. (2019). Implementing the Hamiltonian Monte Carlo Sampling algorithm in PROC MCMC. *Western Users of SAS Softwared Conference and Educational Forum*.
https://www.lexjansen.com/wuss/2019/195_Final_Paper_PDF.pdf

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, *39*(2), 95–101.

Yin, G. (2009). Bayesian generalized method of moments. *Bayesian Analysis*, *4*(2), 191–207.

Yin, G., Ma, Y., Liang, F., & Yuan, Y. (2011). Stochastic Generalized Method of Moments. *Journal of Computational and Graphical Statistics*, *20*(3), 714–727.

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*(1), 121–130.

Zeger, S. L., & Liang, K.-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, *11*(14–15), 1825–1839.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, *44*(4), 1049–1060.

Zeng, Y., Vaupel, J., Xiao, Z., Liu, Y., & Zhang, Z. (n.d.). *Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998-2012*.

Zhang, X., & Paul, S. (2013). Modified Gaussian estimation for correlated binary data. *Biometrical Journal*, *55*(6), 885–898.

Zhou, Y., Lefante, J., Rice, J., & Chen, S. (2014). Using modified approaches on marginal regression analysis of longitudinal data with time dependent covariates. *Statistics in Medicine*, *33*(19), 3354–3364.