# Three Essays on Correlated Binary Outcomes: Detection and Appropriate Models

by

Kyle Irimata

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

Approved March 2018 by the
Graduate Supervisory Committee:

Jeffrey R. Wilson, Chair
Jennifer Broatch
Ioannis Kamarianakis
Ming-Hung Kao
Mark Reiser

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Correlation is common in many types of data, including those collected through longitudinal studies or in a hierarchical structure. In the case of clustering, or repeated measurements, there is inherent correlation between observations within the same group, or between observations obtained on the same subject. Longitudinal studies also introduce association between the covariates and the outcomes across time. When multiple outcomes are of interest, association may exist between the various models. These correlations can lead to issues in model fitting and inference if not properly accounted for. This dissertation presents three papers discussing appropriate methods to properly consider different types of association. The first paper introduces an ANOVA based measure of intraclass correlation for three level hierarchical data with binary outcomes, and corresponding properties. This measure is useful for evaluating when the correlation due to clustering warrants a more complex model. This measure is used to investigate AIDS knowledge in a clustered study conducted in Bangladesh. The second paper develops the Partitioned generalized method of moments (Partitioned GMM) model for longitudinal studies. This model utilizes valid moment conditions to separately estimate the varying effects of each time-dependent covariate on the outcome over time using multiple coefficients. The model is fit to data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to investigate risk factors of childhood obesity. In the third paper, the Partitioned GMM model is extended to jointly estimate regression models for multiple outcomes of interest. Thus, this approach takes into account both the correlation between the multivariate outcomes, as well as the correlation

due to time-dependency in longitudinal studies. The model utilizes an expanded weight matrix and objective function composed of valid moment conditions to simultaneously estimate optimal regression coefficients. This approach is applied to Add Health data to simultaneously study drivers of outcomes including smoking, social alcohol usage, and obesity in children.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

Table Page

# CHAPTER 1

# INTRODUCTION

Binary outcome data are collected in many disciplines and present unique problems as compared to the analysis of continuous outcome data. For instance, researchers may be interested in the study of hospital readmission for patients. In addition to collecting patient information on readmission, researchers may also obtain measurements on covariates such as age, weight, gender or number of diseases. Beyond the difficulties introduced by the type of outcome, such studies can also involve various types of correlation that can further complicate the analysis. One such type of correlation exists between the responses taken on the subjects. These associations, referred to as intraclass correlation, can arise due to repeated measurements on the same subjects, similarities that exist between subjects, or because of hierarchical data sampling structures. In the case of hierarchical data, correlation can exist at multiple levels of clustering. For example, in the study of hospital readmission, repeated measurements may be taken on the same patients, leading to associations between their observations. The patients may in turn have similarities in their outcomes if they have the same primary care physician, which introduces an additional level of association. Correlation may also exist between the outcome and the covariates across time, resulting in time-dependent covariates. When measurements are collected over time, there can sometimes be either a lagged or fall-off effect of the covariate on the outcome as the study progresses. For the hospital readmission example, the number of diseases a patient has can affect his or her probability of readmission. However, the number of diseases can continue to affect the patient in the future as well. The presence of either type of association can have a

significant impact on the fit of logistic regression models, which often rely on the assumption of independence.

### *1.1. Generalized Linear Models*

Generalized linear models are well known for data with independent observations, with an assumed distribution from the exponential family. Consider a random variable $y_i$ such that $i = 1, .., n$; with a distribution from the exponential family, so the log-likelihood function is of the form (Smyth 1989),

$$l(\theta_i, \phi_i^{-1}, \omega_i : y_i) = \sum_i [\omega_i \phi_i^{-1} (y_i \theta_i - b(\theta_i)) - c(y_i, \omega_i \phi_i^{-1})]$$

where

$$c(y_i, \omega_i \phi_i^{-1}) = \omega_i \phi_i^{-1} a(y_i) - \frac{1}{2} s(-\omega_i \phi_i^{-1}) + t(y_i)$$

and $\phi_i$ is unknown and the functions $a(y_i)$ and $b(\theta_i)$ are known.

For the generalized linear model, it is useful to consider the marginal form, which relates the effect of some number of covariates to the mean response. Denote the mean by $\mu_i = E(Y_i) = b'(\theta_i)$ and the $var(y_i) = \sigma_i^2 = [\omega_i \phi_i^{-1} v(\mu_i)]$ where $v(\mu_i) = b''(\theta_i)$ and $b'(\cdot)$ is the first derivative and $b''(\cdot)$ is the second derivative of $b(\cdot)$. The marginal model is given by

$$g(\mu_i) = \boldsymbol{x_i}' \boldsymbol{\beta}$$

where $\boldsymbol{x_i'} = (x_1, \ldots\ldots, x_p)'$ is the vector of covariates and $\boldsymbol{\beta}$ is the vector of regression parameters. The link function $g$ is a monotone and differentiable function, which serves as a transformation on the mean response (Donner 1986).

The logistic regression model, which utilizes a logit link, is a type of generalized linear model and is useful in the analysis of binary data. It relates the log-odds, or logit of the outcome to some number of predictors through regression parameters. We express the logistic regression model as

$$logit(Pr(Y_i = 1|\boldsymbol{\beta})) = \boldsymbol{x_i}'\boldsymbol{\beta}.$$

The generalized linear model, and thus in turn the logistic regression model, requires the assumption that all observations are independent, which is often not the case.

### *1.2. Generalized Linear Mixed Models*

Generalized linear mixed models are an extension of the traditional generalized linear model and incorporate random effects to account for cluster level effects, or correlation in the data. Similar to the generalized linear model, the generalized linear mixed model utilizes a link function to relate a response to some number of covariates. A logit link is most commonly utilized in the case of binary outcome data.

Consider the hierarchical logistic regression model with three levels where there are random effects at two of those levels such that

$$logit\big(Pr\big(Y_{ijk} = 1\big|\boldsymbol{\beta}, \mu_i, \mu_{ij}\big)\big) = \boldsymbol{x_k}'\boldsymbol{\beta} + \mu_i + \mu_{ij}$$

where $\mu_i$ and $\mu_{ij}$ are normally distributed, each with mean 0 and respective variances $\sigma_A^2$ and $\sigma_{B(A)}^2$ (where $\sigma_A^2$ denotes the variance within the primary clusters and $\sigma_{B(A)}^2$ denotes the variance of the secondary units within the primary units) and where $\boldsymbol{x_k}' = (x_{1k}, \dots, x_{pk})'$ is the vector of covariates at the first level of the observational units and $\boldsymbol{\beta}$ is the corresponding vector of regression parameters (Donner 1986). It is customary to

assume that, conditioned on $\mu_i$ and $\mu_{ij}$, the likelihood for the $i^{th}$ subject involves integrating out the random intercept and is given by:

$$L_i = L\left(Y_{ijk} = 1 \middle| \boldsymbol{\beta}, \sigma_A^2, \sigma_{B(A)}^2\right)$$

$$= \iint \prod_k^{n_{ij}} Pr\left(Y_{ijk} = y_{ijk} \middle| \boldsymbol{\beta}, \mu_i, \mu_{ij}\right) \Phi_{\sigma_A^2}(\mu_i)\, \Phi_{\sigma_{B(A)}^2}(\mu_{ij}) d\mu_i d\mu_{ij}$$

$$= \iint \prod_k^{n_{ij}} Pr\left(Y_{ijk} = y_{ijk} \middle| \boldsymbol{\beta}, z_i, z_{ij}\right) \Phi(z_i)\Phi(z_{ij}) dz_i dz_{ij}$$

where $\mu_i = \sigma_A z_i$, $\mu_{ij} = \sigma_{B(A)} z_{ij}$, $z_i \sim N(0,1)$ and $z_{ij} \sim N(0,1)$, which are obtained by transforming from $\mu_i \sim N(0, \sigma_A^2)$ and $\mu_{ij} \sim N\left(0, \sigma_{B(A)}^2\right)$, each represented respectively by $\Phi_{\sigma_A^2}$ and $\Phi_{\sigma_{B(A)}^2}$. As we cannot take the joint likelihood through the product of the $Y_{ijk}$ as they are not independent, we utilize the conditional distribution. Thus, the total likelihood for the marginal is the product of these terms resulting in $L = \prod_1^n L_i$ for all the observations. The maximum likelihood estimates are often obtained through numerical integration, such as through Gauss-Hermite polynomials (Lesaffre and Spiessens 2001).

### *1.3. Accounting for Correlation*

In this work, we introduce methods for measuring and accounting for correlation in binary outcome data. We provide a method for measuring intraclass correlation in the presence of multiple levels of nesting in hierarchical data. This approach is useful for identifying when the association between responses merits the use of a more complex model, in practice. We also provide a new method, called the Partitioned generalized method of moments (Partitioned GMM), for modeling time-dependent covariates in

longitudinal studies. The Partitioned GMM utilizes valid moment conditions to separately estimate the effect of time-dependent covariates on the outcome both within the same time-period as well as at lagged time-periods. Finally, we develop a simultaneous GMM model with partitioned coefficients to account for the correlation between multiple responses of interest, while also estimating the potentially varying effects of time-dependent covariates on each response.

CHAPTER 2

# IDENTIFYING INTRACLASS CORRELATION NECESSITATING HIERARCHICAL MODELING

Kyle M. Irimata, Jeffrey R. Wilson

***Abstract***

Hierarchical binary outcome data with three levels, such as disease remission for patients nested within physicians, nested within clinics are frequently encountered in practice. One important aspect in such data is the correlation that occurs at each level of the data. In parametric modeling, accounting for these correlations increases the complexity. These models may also yield results that lead to the same conclusions as simpler models. We developed a measure of intraclass correlation at each stage of a three-level nested structure and identified guidelines for determining when the dependencies in hierarchical models need to be taken into account. These guidelines are supported by simulations of hierarchical data sets, as well as the analysis of AIDS knowledge in Bangladesh from the 2011 Demographic Health Survey. We also provide a simple rule of thumb to assist researchers faced with the challenge of choosing an appropriately complex model when analyzing hierarchical binary data.

## *2.1. Introduction*

Hierarchical binary data has become increasingly commonplace in many settings and has in turn created the critical challenge of choosing the most appropriate model, without introducing unnecessary complexity. For instance, researchers may be interested

in whether or not a disease is in remission and may collect data on individual patients, each nested within physicians, who are each nested within clinics. This type of correlation within clustered data can lead to incorrect interpretation in data analysis if not properly accounted for (Liang and Zeger 1986). McMahon, Pouget and Tortu (2006) also showed that the strength of the within cluster dependence of the observations is an important consideration when determining an appropriate level of model complexity. Random effects models are often used when the presence of correlation between observations is expected due to the hierarchical structure of the data. Although these models account for associations in the data, they also introduce some complications in both model fitting and interpretation.

The intraclass correlation coefficient (ICC) is a quantitative measure of the similarity among observations within classes or clusters. For hierarchical or multilevel data, the ICC provides a summary of the overall strength of the association amongst responses within a cluster. It is frequently used to quantify the familial aggregation of disease in genetic epidemiological studies (Cohen 1980; Liang, Zeger, and Qaqish 1992). The challenge with intraclass correlation when dealing with binary data lies within the fact that the variance depends on the mean, or in other words the probability. Many approaches do not provide easily accessible methods for addressing clustering at multiple levels of a hierarchy.

Many estimators of the intraclass correlation at each level have been derived for continuous response data (Donner 1986). However, methods for estimating the intraclass correlation coefficient for binary response data are comparatively less investigated. Zou and Donner (2004), amongst others (Ridout, Demétrio, and Firth 1999), provided a

thorough review of techniques for a two-level nested model. Common approaches for estimation are the application of ANOVA type estimators (Elston 1977), the Pearson correlation coefficient (Donner 1986; Mudelsee 2003), moment estimators (Kleinman 1973), or the use of kappa-type measures (Fleiss and Cuzick 1979; Mak 1988). Other approaches include Bayesian hierarchical models (Tan, et al. 1999), pseudo-likelihood or quasi-likelihood approaches (Nelder and Pregibon 1987), as well as less common approaches (Oman and Zucker 2001). O'Connell and McCoach (2008) presented a simple measure of intraclass correlation for binary outcomes, which relies upon the assumption of a logistic distribution on the residuals with variance of 3.29 (Ene, et al. 2015; Snijders and Bosker 1999). The ICC has also been investigated in various sampling designs (Bodian 1994), in applied settings (Gulliford, Ukoumunne, and Chinn 1999), and as a tool for determining the design effect (Cunningham and Johnson 2016).

Many researchers have applied closed-form asymptotic variance formulae for point estimators of the ICC to binary outcome data arising in clusters of variable size (Bloch and Kraemer 1989; Bodian 1994; Cunningham and Johnson 2016; Fleiss and Cuzick 1979). Simulation studies have also shown that confidence intervals based on the estimator provide coverage levels close to nominal over a wide range of parameter combinations (Fleiss and Cuzick 1979). However, existing methods for the analysis of dichotomous outcomes generally have not addressed correlation in models with more than one level of clustering.

In this paper, we derived tests for each level of correlation in a three-stage model with two levels of clustering, which can be readily extended to higher levels. These tests can be used to determine at what level or levels the intraclass correlation needs to be

taken into account prior to fitting a complex model. We adapted an ANOVA type estimator of the ICC to binary outcomes for the case of a hierarchical, unbalanced design with three levels (Kirk 1982; Elston 1977; Ridout, Demétrio, and Firth 1999; Zou and Donner 2004). These estimators provide overall summaries of the association at each of the levels of the data and utilize estimates of each of the variance components, including the residual variance. Such estimators have been shown to perform well as long as the data are not extremely unbalanced (Swallow and Monahan 1984) or the correlation very small (Donner and Koval 1980). We also derived large sample properties and simulated multilevel data to determine thresholds at which each level of clustering the intraclass correlation must be taken into account. Additionally, we provided Cramer's V squared (Cicchetti 1994) as a simple rule of thumb approximation for the ICC.

In Section 2, we review correlated outcomes and present a hierarchical logistic model. In Section 3, we develop ANOVA type intraclass correlation estimators for the first and second levels of a hierarchy. The results of a simulation study to explore the effects of correlation are presented in Section 4. In Section 5, we analyze data from the 2011 Bangladesh Demographic Health Survey and provide comparisons to alternative approaches. Some conclusions and discussions are provided in Section 6.

### *2.2. Hierarchical Models*

#### *2.2.1 Generalized Linear Models*

Generalized linear models are often used to analyze data with independent observations with an assumed distribution from the exponential family. These models are useful in relating predictors to the mean of the response using a link function. The logistic

regression model, which utilizes a logit link, is a member of this family of models and is useful in the analysis of binary data (Dobson 2002). The generalized linear model, and hence the logistic regression model, rely on the assumption of independent observations, which is not the case with hierarchical data; thus it is not appropriate for analyzing multilevel data.

### *2.2.2 Hierarchical Logistic Regression Models*

Hierarchical data often arise as a result of cluster sampling, such as when subjects are recruited from several practices or practitioners (Adams, et al. 2004). In such cases, the responses within clusters may be correlated due to similarities in characteristics or outcomes (Smyth 1989). Correlation will also result if individuals within clusters interact and tend to conform, or if they are all influenced by cluster-level characteristics.

Consider a nested three-level structure with binary outcomes, where the observational units at the first level are nested within secondary units (effect B), which are nested within the primary units (effect A). Each level of nesting results in association that can be measured by an intraclass correlation. Hierarchical logistic regression models extend the ordinary logistic regression model, as they take these correlations into account by introducing a random effect for each cluster level (Dobson 2002). However, these models also introduce additional complexity. As discussed by McMahon, et al (2006), the extent to which the observations within a cluster are correlated is thus useful in determining whether to utilize more complicated models. They discussed a variety of methods for evaluating the amount of intraclass correlation present. However, those

methods did not fully address guidelines for when they are meaningful, nor did they consider hierarchical models with more than two levels.

### *2.3. Multiple Intraclass Correlation*

#### *2.3.1 Intraclass Correlation*

The intraclass correlation coefficient quantifies the similarity of individuals within groups and provides an index of aggregation. Liang and Zeger (1986) noted that ignoring this dependency often leads to incorrect conclusions in data analysis. Further, the degree to which the clustering is present may impact the analysis of the data.

Consider a random sample of binary outcomes, denoted by $Y_{ijk}$ based on a three-level hierarchical structure. Assume that exchangeability is present among the outcomes at level one, with the probability of a success, $Pr(Y_{ijk} = 1) = \pi_{ij}$ for $i = 1, \dots, a; j = 1, \dots, b_i; k = 1, \dots, n_{ij}$. In essence, the observations within a given cluster can be reordered without affecting the joint distribution. Some researchers view exchangeability as a generalization of the assumption of independent, identically distributed distributions. Also assume that observations from different primary clusters are independent, and that observations from different secondary clusters are independent, conditioned on the primary level of clustering. Observations from the same secondary cluster within a given primary cluster are assumed to be correlated with common correlation $\rho_{B(A)} = corr(Y_{ijk}, Y_{ijk'})$ for k ≠ k′, while observations from the same primary cluster are assumed to be correlated with common correlation $\rho_A = corr(Y_{ij.}, Y_{ij'.})$ for $j \neq j'$. We define the overall estimator for the probability of success as $\hat{\pi} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} Y_{ijk}}{N}$, where

$N = \sum_{i=1}^{a} N_i = \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}$ is the total number of observations in the sample. For example, in the study of disease remission, we may consider the patients to be the individual observations, who are in turn nested within doctors at the secondary stage of clustering, each of whom is nested within hospitals at the primary stage of clustering. We explore ANOVA type estimators in addressing intraclass correlation at the primary and secondary levels (Sahai and Ojeda 2007).

*2.3.2 Multiple Intraclass Correlation Estimator*

We derived an ANOVA type estimator of the intraclass correlation at the secondary level of clustering within the primary level of clustering to be:

$$\hat{\rho}_{B(A)} = \frac{r_3}{r_1} * \frac{MSB(A) - MSE}{MSA + \left(\frac{r_3 - r_2}{r_1}\right) MSB(A) + \left(\frac{r_1 r_3 - r_1 + r_2 - r_3}{r_1}\right) MSE}$$

and the estimator for the intraclass correlation at the primary level of clustering as:

$$\hat{\rho}_A = \frac{MSA - \left(\frac{r_2}{r_1}\right) MSB(A) + \left(\frac{r_2 - r_1}{r_1}\right) MSE}{MSA + \left(\frac{r_3 - r_2}{r_1}\right) MSB(A) + \left(\frac{r_1 r_3 - r_1 + r_2 - r_3}{r_1}\right) MSE}$$

for the constants $r_1 = \frac{N - \sum_{i=1}^{a} \frac{\sum_{j=1}^{b_i} n_{ij}^2}{N_i}}{b-a}$, $r_2 = \frac{\sum_{i=1}^{a} \frac{\sum_{j=1}^{b_i} n_{ij}^2}{N_i} - \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}^2}{a-1}$ and $r_3 = \frac{N - \frac{1}{N} \sum_{i=1}^{a} N_i^2}{a-1}$, where $a$ denotes the number of primary clusters and $b$ denotes the total number of secondary clusters. Mimicking the usual notation in ANOVA, we obtained

$$MSA = \frac{1}{a-1}\left[S_2 - \frac{1}{N} S_1^2\right]$$

$$MSB(A) = \frac{1}{b-a}[S_3 - \mathrm{S}_2]$$

$$MSE = \frac{1}{N-b}[S_1 - \mathrm{S}_3]$$

where $S_1 = \sum_{i=1}^{a}\sum_{j=1}^{b_i}\sum_{k=1}^{n_{ij}} Y_{ijk}$, $S_2 = \sum_{i=1}^{a}\frac{\left(\sum_{j=1}^{b_i}\sum_{k=1}^{n_{ij}} Y_{ijk}\right)^2}{N_i}$ and $S_3 =$

$\sum_{i=1}^{a}\sum_{j=1}^{b_i}\frac{\left(\sum_{k=1}^{n_{ij}} Y_{ijk}\right)^2}{n_{ij}}$. The vector $(S_1, S_2, S_3)$ can be shown to be asymptotically distributed as a multivariate normal distribution (Zou and Donner 2004), which after algebraic manipulation has the variance-covariance matrix given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} Var(S_1) & 2\pi Var(S_1) & 2\pi Var(S_1) \\ 2\pi Var(S_1) & 4\pi^2 Var(S_1) & 4\pi^2 Var(S_1) \\ 2\pi Var(S_1) & 4\pi^2 Var(S_1) & 4\pi^2 Var(S_1) \end{pmatrix}$$

where

$$\mathrm{Var}(\mathrm{S}_1) = N\pi(1-\pi) + \rho_{B(A)}\pi(1-\pi)\sum_{i=1}^{a}\sum_{j=1}^{b_i} n_{ij}\left(n_{ij}-1\right) +$$

$$\rho_A\pi(1-\pi)\sum_{i=1}^{a}\sum_{j\neq j'}^{b_i}\sqrt{n_{ij}n_{ij'}\left(1+\rho_{B(A)}\left(n_{ij}-1\right)\right)\left(1+\rho_{B(A)}\left(n_{ij'}-1\right)\right)}.$$

Through the use of the delta method, the asymptotic distributions for $\hat{\rho}_{B(A)}$ and $\hat{\rho}_A$ are

$$\sqrt{N}\left(\hat{\rho}_{B(A)} - \rho_{B(A)}\right) \rightarrow N\left(0, \boldsymbol{\Phi}_{\boldsymbol{B(A)}}{}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi}_{\boldsymbol{B(A)}}\right)$$

and

$$\sqrt{N}(\hat{\rho}_A - \rho_A) \to N\left(0, \boldsymbol{\Phi_A}^T \boldsymbol{\Sigma} \boldsymbol{\Phi_A}\right)$$

(Zou and Donner 2004) respectfully, where

$$\boldsymbol{\Phi_{B(A)}} = \begin{pmatrix} \frac{\partial \hat{\rho}_{B(A)}}{\partial S_1} \\ \frac{\partial \hat{\rho}_{B(A)}}{\partial S_2} \\ \frac{\partial \hat{\rho}_{B(A)}}{\partial S_3} \end{pmatrix} \text{ and } \boldsymbol{\Phi_A} = \begin{pmatrix} \frac{\partial \hat{\rho}_A}{\partial S_1} \\ \frac{\partial \hat{\rho}_A}{\partial S_2} \\ \frac{\partial \hat{\rho}_A}{\partial S_3} \end{pmatrix}$$

are the vectors of partial derivatives evaluated at the expected values:

$$E[S_1] = N\pi,$$

$$E[S_2] = 4\pi^2 a + 4\pi^3(1-\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij} - 1\right) + 4\pi^3 \rho_A (1 -$$

$$\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij} - 1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'} - 1\right)\right)} - N\pi^2,$$

and

$$E[S_3] = \pi(1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2$$

This leads to the variances of our estimators as

$$\text{var}\left(\hat{\rho}_{B(A)}\right) = \boldsymbol{\Phi_{B(A)}}^T \boldsymbol{\Sigma} \boldsymbol{\Phi_{B(A)}} = \frac{var(S_1)}{\lambda_1^4} * \left(\frac{r_3}{r_1}\right)^2 [1 - 4\pi + 4\pi^2] \left(\frac{1}{N-b}\right)^2 (\lambda_1 - d_2\lambda_2)^2$$

and

$$\text{var}(\hat{\rho}_A) = \boldsymbol{\Phi_A}^T \boldsymbol{\Sigma} \boldsymbol{\Phi_A} = \frac{var(S_1)}{\lambda_1^4} (1 + 4\pi^2 - 4\pi) \left(\left(\frac{r_2 - r_1}{r_1(N-b)}\right)\lambda_1 - \left(\frac{d_2}{N-b}\right)\lambda_3\right)^2$$

for the constants $d_1 = \frac{r_3 - r_2}{r_1}$ and $d_2 = \frac{r_1 r_3 - r_3 - r_1 + r_2}{r_1}$ and where

$$\lambda_1 = -\frac{1}{(a-1)} N\pi^2 + \frac{d_2}{N-b} N\pi + \left(\frac{1}{a-1} - \frac{d_1}{b-a}\right)\pi^2 \Bigg( 4a + 4\pi(1 -$$

$$\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij} - 1\right) + 4\pi\rho_A(1 -$$

$$\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij} - 1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'} - 1\right)\right)} - N \Bigg) +$$

$$\left(\frac{d_1}{b-a} - \frac{d_2}{N-b}\right)\left((1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2\right)$$

$$\lambda_2 = -\frac{1}{N-b} N\pi - \frac{1}{b-a}\pi^2 \Bigg( 4a + 4\pi(1-\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij} - 1\right) +$$

$$4\pi\rho_A(1-\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij} - 1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'} - 1\right)\right)} -$$

$$N \Bigg) + \left(\frac{1}{N-b} + \frac{1}{b-a}\right)\left(\pi(1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2\right)$$

and

$$\lambda_3 = -\frac{1}{(a-1)} N\pi^2 + \frac{r_2 - r_1}{r_1(N-b)} N\pi + \pi^2 \left(\frac{1}{a-1} + \frac{r_2}{r_1(b-a)}\right) \Bigg[ 4a + 4\pi(1 -$$

$$\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij} - 1\right) + 4\pi\rho_A(1 -$$

$$\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij} - 1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'} - 1\right)\right)} - N \Bigg] -$$

$$\left(\frac{r_2}{r_1(b-a)} + \frac{r_2 - r_1}{r_1(N-b)}\right)\left[\pi(1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2\right].$$

By a second application of the delta method, the distribution after a logarithmic transformation is

$$\sqrt{N}\left(log(\hat{\rho}_{B(A)}) - log(\rho_{B(A)})\right) \rightarrow N\left(0, \left(\frac{1}{\rho_{B(A)}}\right)^2 \boldsymbol{\Phi_{B(A)}}^T \boldsymbol{\Sigma} \boldsymbol{\Phi_{B(A)}}\right)$$

and

$$\sqrt{N}(log(\hat{\rho}_A) - log(\rho_A)) \rightarrow N\left(0, \left(\frac{1}{\rho_A}\right)^2 \boldsymbol{\Phi_A}^T \boldsymbol{\Sigma} \boldsymbol{\Phi_A}\right)$$

Further discussion of these derivations are provided in the appendix.

*2.3.3 Inference for Intraclass Correlation*

We utilized the results of Section 3.2 to develop methods for comparing the ICC to an established threshold. In particular, we sought to establish methods for evaluating the hypotheses:

$$\mathrm{H}_{01}: \rho_A = \rho_0 \text{ vs. } \mathrm{H}_{\mathrm{a}1}: \rho_A > \rho_0$$

$$\mathrm{H}_{02}: \rho_{B(A)} = \rho_0 \text{ vs. } \mathrm{H}_{\mathrm{a}2}: \rho_{B(A)} > \rho_0$$

These tests allowed us to determine when the strength of the intraclass correlation at each level was large enough to warrant the use of a more complex model. We utilized large sample properties for testing these hypotheses against results established by a simulation study. The asymptotic normal test statistics for these hypotheses are

$$Z_A = \frac{log\left(\frac{\hat{\rho}_A}{\rho_0}\right)}{\sqrt{\frac{1}{N\hat{\rho}_A^2}\widehat{var}(\hat{\rho}_A))}} \text{ and } Z_{B(A)} = \frac{log\left(\frac{\hat{\rho}_{B(A)}}{\rho_0}\right)}{\sqrt{\frac{1}{N\hat{\rho}_{B(A)}^2}\widehat{var}(\hat{\rho}_{B(A)})}}$$

respectively. Donner and Donald (1988) showed that the Pearson correlation is not adequate for testing correlated observations but a simple adjustment can be made. More appropriately the one sided confidence intervals are

$$\left[0, exp\left\{log(\hat{\rho}_A) + Z_\alpha \sqrt{\frac{1}{N\hat{\rho}_A^2}\widehat{var}(\hat{\rho}_A)}\right\}\right]$$

and

$$\left[0, log\left(\hat{\rho}_{B(A)}\right) + Z_\alpha \sqrt{\frac{1}{N\hat{\rho}_{B(A)}^2}\widehat{var}(\hat{\rho}_{B(A)})}\right]$$

based on extension of earlier results (Zou and Donner 2004). These calculations can be completed using our R program available at http://www.public.asu.edu/~jeffreyw.

*2.3.4 Cramer's V and Rule of Thumb*

Cicchetti (1994) presented comparisons between the ICC and the product-moment correlation and claimed that the product-moment correlation places an upper limit on the maximum ICC. We used Cramer's V squared ($V^2$) in a similar manner as an easy to calculate, but conservative approximation for our intraclass correlation measure.

Cramer's V (Cramer 1946; Liebetrau 1983) is a popular coefficient for evaluating relationships between nominal variables, regardless of the dimensions of the contingency table, based on the value of the chi-squared test of independence. Kirk (1982) provided an in depth discussion of the properties of this estimator. For a two-dimensional table based on the number of clusters by the two-category response, we define $V^2 = X^2/\mathrm{N}$,

where N is the sample size. The statistic $V^2$ is the mean square canonical correlation between the clusters and the binary response, and ranges between 0 and 1, where larger values of $V^2$ indicate a stronger association. In our example discussed in Section 5, we found that $V_A^2 = 0.025$, based on the 7 by 2 table, where 7 is the number of clusters (or divisions).

While our R-program can easily compute $\hat{\rho}_A$ and $\hat{\rho}_{B(A)}$ it may not necessarily be convenient for researchers in more applied settings. As such we propose the use of $V^2$ as a quick approximation of the strength of the association. Further support for the use of this measure as an approximation is provided in Section 4.2.

### *2.4. Simulation Study and Hypothesis Testing*

#### *2.4.1 Simulation Study*

We conducted a simulation study to determine thresholds for the ICC at which the associations in the data should not be ignored. Each simulated dataset comprised of 25 primary units, each containing between eight and fifteen secondary units. Each of these secondary units contained between two and forty observations. We created an intercept term for each level of clustering from a normal distribution with a mean of zero and a user defined variance term for each iteration in order to incorporate a clustering effect. One continuous and one binary covariate were adopted into the simulation for each observational unit according to a multivariate random normal distribution where the binary predictor was obtained as a dichotomization of one of the predictors. The probability produced from the combined effect of the fixed and random effects was used to generate a binary outcome according to a Bernoulli distribution.

We simulated 84,000 hierarchical data sets in R and calculated $\hat{\rho}_{B(A)}$ and $\hat{\rho}_A$ at each level of clustering for each dataset. Both these correlations were constrained between 0 and 0.30 for our simulations. A standard logistic regression and a three-level hierarchical model with random intercepts at each level were fit to each dataset. The significance of the predictors was noted for each analysis based on a significance level of $\alpha = 0.05$ and used to obtain the rate of agreement between the two models. The analyses were determined to agree if both analyses indicated that a certain predictor was significant ($p \leq 0.05$), or if both analyses indicated that the predictor was not significant ($p > 0.05$). The percentage agreement, organized according to the ICC at level 2 versus level 3, is provided in Table 2.4.1.

Table 2.4.1. Percentage Agreement for the Continuous/Binary Predictor

| Continuous/Binary Predictor | | **Correlation at secondary cluster (Level 2)** | | | | | |
|---|---|---|---|---|---|---|---|
| | | **[0,0.05)** | **[0.05,0.1)** | **[0.1,0.15)** | **[0.15,0.2)** | **[0.2,0.25)** | **[0.25,0.3)** |
| **Correlation at primary cluster (Level 3)** | **[0,0.05)** | 98.4/98.3 | 93.9/95.7 | 90.9/93.1 | 89.7/92.2 | 86.5/89.2 | 81.4/85.0 |
| | **[0.05,0.1)** | 93.7/95.8 | 90.5/93.4 | 87.5/91.4 | 85.4/87.9 | 82.2/84.3 | 77.9/79.8 |
| | **[0.1,0.15)** | 91.9/93.4 | 88.6/90.5 | 85.6/87.4 | 81.0/83.9 | 77.9/79.9 | 74.5/75.8 |
| | **[0.15,0.2)** | 88.7/90.9 | 85.1/87.3 | 81.2/83.6 | 77.9/79.7 | 74.0/75.0 | 70.0/72.7 |
| | **[0.2,0.25)** | 84.4/87.4 | 81.9/83.7 | 78.0/80.1 | 73.9/75.4 | 70.6/71.3 | 67.5/67.5 |
| | **[0.25,0.3)** | 81.8/83.7 | 77.4/79.9 | 73.8/76.8 | 71.5/73.4 | 67.9/68.4 | 76.2/61.9 |

This simulation showed that the percentage agreement decreased as the correlation increased at either stage of clustering. In the presence of correlation at both levels of clustering, the discrepancy between the logistic regression model and the three-level hierarchical model was more pronounced. Once both measures of correlation increased beyond 0.10, the agreement rate fell below 90%. When the correlation at either the primary or secondary units increased beyond 0.15 the agreement rate for the

continuous predictor fell below 90% regardless of the amount of correlation present at the other level. Conservative researchers should consider more complex models when the amount of association increases beyond 0.10. For most situations, a more complex model should be investigated once the intraclass correlation increases beyond 0.15.

*2.4.2 Upper Bound for Intraclass Correlation – Simulation Study*

Based on a simulated study of 70,000 data sets, we found that $\hat{V}^2_{B(A)} \geq \hat{\rho}_{B(A)}$; thus $\hat{V}^2_{B(A)}$ provides a conservative, easy to use estimate of association for the ICC for the secondary clusters. We also found that $\hat{V}^2_A$ was larger than $\hat{\rho}_A$ for lower levels of association at the primary cluster. In fact, even for larger values of $\hat{\rho}_A$, the value of $\hat{V}^2_A$ was very close; within an allowable error of 0.005. We found that $\hat{V}^2_A + 0.005 \geq \hat{\rho}_A$ for correlation below 0.20 and that within an allowable error of 0.008 that $\hat{V}^2_A + 0.008 \geq \hat{\rho}_A$ for all correlation levels below 0.30. Given these results, we concluded that $\hat{V}^2$ provides a generally conservative approximation of the ICC at each stage of the hierarchy. Thus, $\hat{V}^2$ provides a useful approximation in conjunction with previous results for determining appropriate model complexity.

***2.5. Data Example***

We illustrated the use of these intraclass correlation measures through the analysis of data from the 2011 Bangladesh Demographic Health Survey. These data contained information on 17,457 women from 600 different villages at the secondary stage of clustering, within seven different divisions at the primary level of clustering. The villages at the secondary level of clustering were all of approximately the same size, and

generally corresponded to individual villages in Bangladesh, while the divisions at the primary level of clustering represent administrative regions. The binary outcome was whether the respondent had knowledge of AIDS. We utilized five covariates, which included religion (Islam, Hinduism, Buddhism and Christianity), patient's age at the time of the interview, education level (none, primary, or secondary and higher), number of living children, and whether the individual lived in a rural or urban location. Access to this dataset can be requested from http://www.dhsprogram.com/data/new-user-registration.cfm (NIPORT 2011).

We calculated the ICC for both the village and division level of clustering as well as the respective 90% and 95% confidence intervals. We also calculated the value of $\hat{V}^2$ for each stage of clustering. The results are given in Table 2.5.1.

Table 2.5.1. Intraclass Correlation, Confidence Intervals and $V^2$

| | $\hat{\rho}$ | **Confidence Intervals ($\rho$)** | | $\hat{V}^2$ |
|---|---|---|---|---|
| | | **90%** | **95%** | |
| **District Level (A)** | 0.028 | $(0, 0.106)$ | $(0, 0.155)$ | 0.025 |
| **Village Level (B(A))** | 0.160 | $(0, 0.197)$ | $(0, 0.209)$ | 0.208 |

The 95% confidence interval for ICC at the primary unit for the district level of clustering does not include our cut-off value of 0.15, although the 90% confidence interval does. Thus, at the 5% significance level, we cannot be confident that a model which ignores the district level of clustering is adequate, although at the 10% significance level, we may conclude that a simpler model is sufficient. Both the 90% and 95% confidence interval for the ICC at the secondary units for the village level of clustering include our cut-off value of 0.15. Therefore, there is strong correlation at the village

level, suggesting the need for a more complex model. Hence, at the 10% significance level, we concluded the village level of clustering needed to be accounted for, while the district level of clustering did not.

We also found that both $\hat{V}_A^2$ and $\hat{V}_{B(A)}^2$ provided good approximations for the ICC at each level and also led to similar conclusions. Given the small size of $\hat{V}_A^2$, there is little reason to believe that the district level of clustering needs to be accounted for in the model. Since $\hat{V}_{B(A)}^2$ is relatively large, we should consider models which account for the correlation at the village level.

We fitted a standard logistic regression model to our data to evaluate the covariates under the assumption of independence, and found that religion, age, education, children and location were all significant predictors of AIDS knowledge. We also fitted a two-level hierarchical model with one random intercept term for the effects of villages, and found that education, children and location were significant predictors of AIDS knowledge. In addition, we fitted a three-level hierarchical model with two random intercept terms, one for the effect of districts and one for the effect of villages. For this analysis, education, children and location were significant predictors of AIDS knowledge. The results of these analyses are summarized in Table 2.5.2.

Table 2.5.2. P-Values for the Predictors of AIDS Knowledge

| | **Religion** | **Age** | **Education** | **Children** | **Location** |
|---|---|---|---|---|---|
| **Logistic** | 0.0277 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| **Logistic One Intercept** | 0.81223 | 0.2545 | <0.0001 | <0.0001 | <0.0001 |
| **Logistic Two Intercepts** | 0.9748 | 0.2799 | <0.0001 | <0.0001 | <0.0001 |

We noted that the results of the independence model differed from the results of the two-level hierarchical model. Under the assumption of independence, religion and age, among other predictors, were significant predictors of AIDS knowledge; however, neither of these two predictors were significant once the village level correlation was taken into account. With respect to the significance of predictors, the results of the three-level hierarchical model are identical to those obtained by the two-level model.

The results of these data analyses are in agreement with our simulation results. The correlation at the village level was significant based on the confidence interval for the ICC, as well as $\hat{V}_{B(A)}^2$; thus a random effect should be included to account for the association within the secondary units. However, the correlation at the district level was fairly inconsequential, as illustrated by the confidence interval for the ICC, as well as $\hat{V}_A^2$; therefore, a second random intercept for the district level would not be beneficial. A model which assumes independence is also inappropriate, thus the logistic regression model with one random intercept for the secondary units is best for these data.

As a comparison, we calculated the ICC for each level using an approach discussed by O'Connell and McCoach (2008). Using their method, a three-level hierarchical model with two random intercepts and no predictors was fitted and estimates of the variances for each random term were obtained, where the error variance was assumed to be 3.29. For this approach, $\hat{\rho}_A^* = 0.0348$ and $\hat{\rho}_{B(A)}^* = 0.236$, which are both similar to the values produced using our ANOVA estimator. However, their approach relies on an additional assumption that a logistic regression has an error variance of 3.29. Our approach provides an improved, data driven approximation of the error variance without the blanket requirement of such additional assumptions.

## *2.6. Conclusions*

Multilevel binary outcome data are encountered in a variety of disciplines; however, approaches to evaluate the effect of the clustering when there is more than one level of nesting are not as well documented or researched. The correlation induced at each level of clustering can have a significant effect on data analysis results and often must be taken into account at each level. However, accounting for each level of clustering incorporates a certain degree of challenge with model fitting and interpretation as they involve more complexity. Thus, accounting for the intraclass correlation is a choice between simplicity versus accuracy. We obtained ANOVA type estimators of intraclass correlation for binary outcomes at two levels of nesting, with respective asymptotic properties. Our estimators directly estimate the error variance, without any additional assumptions and further can be applied to the analysis of unbalanced data. These estimators are useful for identifying when more complex models are necessary. In particular, our simulation results showed that more complex models should be considered when the ICC is larger than 0.15 in most cases, and for more conservative researchers when the ICC is greater than 0.10. We also provided simple approximations using a $V^2$ measure, which offers a quick, conservative approximation for the ICC. Both of these approaches are useful for reducing unnecessary model complexity while also ensuring that predictors are not mistakenly identified as significant.

## *2.7. Appendix*

### *2.7.1 Variance Derivations:*

We have that

$$\hat{\rho}_{B(A)} = \frac{r_3}{r_1} * \frac{MSB(A) - MSE}{MSA + \left(\frac{r_3 - r_2}{r_1}\right) MSB(A) + \left(\frac{r_1 r_3 - r_1 + r_2 - r_3}{r_1}\right) MSE}$$

$$\hat{\rho}_{A} = \frac{MSA - \left(\frac{r_2}{r_1}\right) MSB(A) + \left(\frac{r_2 - r_1}{r_1}\right) MSE}{MSA + \left(\frac{r_3 - r_2}{r_1}\right) MSB(A) + \left(\frac{r_1 r_3 - r_1 + r_2 - r_3}{r_1}\right) MSE}$$

for the constants $r_1 = \frac{N - \sum_{i=1}^{a} \frac{\sum_{j=1}^{b_i} n_{ij}^2}{N_i}}{b-a}$, $r_2 = \frac{\sum_{i=1}^{a} \frac{\sum_{j=1}^{b_i} n_{ij}^2}{N_i} - \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{b_i} n_{ij}^2}{a-1}$ and $r_3 = \frac{N - \frac{1}{N} \sum_{i=1}^{a} N_i^2}{a-1}$, where $a$ denotes the number of primary clusters and $b$ denotes the total number of secondary clusters. Mimicking the usual notation in ANOVA, we obtained

$$MSA = \frac{1}{a-1}\left[S_2 - \frac{1}{N} S_1^2\right]$$

$$MSB(A) = \frac{1}{b-a}\left[S_3 - S_2\right]$$

$$MSE = \frac{1}{N-b}\left[S_1 - S_3\right]$$

where $S_1 = \sum_{i=1}^{a} \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} Y_{ijk}$, $S_2 = \sum_{i=1}^{a} \frac{\left(\sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} Y_{ijk}\right)^2}{N_i}$ and $S_3 = \sum_{i=1}^{a} \sum_{j=1}^{b_i} \frac{\left(\sum_{k=1}^{n_{ij}} Y_{ijk}\right)^2}{n_{ij}}$

Through the assumption of independence between observations from different primary clusters, as well as the assumption of independence between observations from different secondary clusters, given the primary level of clustering, we find that the vector

$(S_1, S_2, S_3)$ has the variance-covariance matrix given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(\mathrm{S}_1) & \mathrm{cov}(\mathrm{S}_1, \mathrm{S}_2) & \mathrm{cov}(\mathrm{S}_1, \mathrm{S}_3) \\ \mathrm{cov}(\mathrm{S}_1, \mathrm{S}_2) & \mathrm{Var}(\mathrm{S}_2) & \mathrm{cov}(\mathrm{S}_2, \mathrm{S}_3) \\ \mathrm{cov}(\mathrm{S}_1, \mathrm{S}_3) & \mathrm{cov}(\mathrm{S}_2, \mathrm{S}_3) & \mathrm{Var}(\mathrm{S}_3) \end{pmatrix} =$$

$$\begin{pmatrix} \mathrm{Var}(\mathrm{S}_1) & 2\pi\mathrm{Var}(\mathrm{S}_1) & 2\pi\mathrm{Var}(\mathrm{S}_1) \\ 2\pi\mathrm{Var}(\mathrm{S}_1) & 4\pi^2\mathrm{Var}(\mathrm{S}_1) & 4\pi^2\mathrm{Var}(\mathrm{S}_1) \\ 2\pi\mathrm{Var}(\mathrm{S}_1) & 4\pi^2\mathrm{Var}(\mathrm{S}_1) & 4\pi^2\mathrm{Var}(\mathrm{S}_1) \end{pmatrix}.$$

The variances and covariances can be calculated using

$$cov\left(Y_{ijk}, Y_{ijk'}\right) = corr\left(Y_{ijk}, Y_{ijk'}\right)\sigma_{Y_{ijk}}\sigma_{Y_{ijk'}} = \rho_{B(A)}\pi(1-\pi)$$

and

$$cov\left(Y_{ij.}, Y_{ij'.}\right) = corr\left(Y_{ij.}, Y_{ij'.}\right)\sigma_{Y_{ij.}}\sigma_{Y_{ij'.}}$$

$$= \rho_A\pi(1-\pi)\sqrt{n_{ij}n_{ij'}\left(1+\rho_{B(A)}\left(n_{ij}-1\right)\right)\left(1+\rho_{B(A)}\left(n_{ij'}-1\right)\right)}.$$

For illustration, the variance of $S_1$ is derived as:

$$Var(S_1) = Var\left(\sum_{i=1}^{a}\sum_{j=1}^{b_i}\sum_{k=1}^{n_{ij}} Y_{ijk}\right)$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b_i}\sum_{k=1}^{n_{ij}} Var\left(Y_{ijk}\right) + \sum_{i=1}^{a}\sum_{j=1}^{b_i}\sum_{k\neq k'}^{n_{ij}} cov\left(Y_{ijk}, Y_{ijk'}\right) + \sum_{i=1}^{a}\sum_{j\neq j'}^{b_i} cov\left(Y_{ij}, Y_{ij'}\right)$$

$$= N\pi(1-\pi) + \rho_{B(A)}\pi(1-\pi)\sum_{i=1}^{a}\sum_{j=1}^{b_i} n_{ij}\left(n_{ij}-1\right) + \rho_A\pi(1-$$

$$\pi)\sum_{i=1}^{a}\sum_{j\neq j'}^{b_i}\sqrt{n_{ij}n_{ij'}\left(1+\rho_{B(A)}\left(n_{ij}-1\right)\right)\left(1+\rho_{B(A)}\left(n_{ij'}-1\right)\right)}.$$

*2.7.2 Asymptotic Distributions:*

As discussed by Zou and Donner (2004), the vector $(S_1, S_2, S_3)$ can be shown to be asymptotically distributed as a multivariate normal distribution with variance-covariance matrix as given previously. Through application of the delta method, the asymptotic distributions for $\hat{\rho}_{B(A)}$ and $\hat{\rho}_A$ are

$$\sqrt{N}\left(\hat{\rho}_{B(A)} - \rho_{B(A)}\right) \rightarrow N\left(0, \boldsymbol{\Phi_{B(A)}}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi_{B(A)}}\right)$$

and

$$\sqrt{N}(\hat{\rho}_A - \rho_A) \rightarrow N\left(0, \boldsymbol{\Phi_A}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi_A}\right)$$

respectfully, where

$$\boldsymbol{\Phi_{B(A)}} = \begin{pmatrix} \frac{\partial\hat{\rho}_{B(A)}}{\partial S_1} \\ \frac{\partial\hat{\rho}_{B(A)}}{\partial S_2} \\ \frac{\partial\hat{\rho}_{B(A)}}{\partial S_3} \end{pmatrix} \text{ and } \boldsymbol{\Phi_A} = \begin{pmatrix} \frac{\partial\hat{\rho}_A}{\partial S_1} \\ \frac{\partial\hat{\rho}_A}{\partial S_2} \\ \frac{\partial\hat{\rho}_A}{\partial S_3} \end{pmatrix}$$

are the vectors of partial derivatives evaluated at the expected values:

$$E[S_1] = N\pi,$$

$$E[S_2] = 4\pi^2 a + 4\pi^3(1-\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij}-1\right) + 4\pi^3 \rho_A (1 -$$

$$\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij}-1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'}-1\right)\right)} - N\pi^2,$$

and

$$E[S_3] = \pi(1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2$$

This leads to the variances of our estimators as

$$\text{var}\left(\hat{\rho}_{B(A)}\right) = \boldsymbol{\Phi_{B(A)}}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{\Phi_{B(A)}} = \frac{var(S_1)}{\lambda_1^4} * \left(\frac{r_3}{r_1}\right)^2 [1 - 4\pi + 4\pi^2] \left(\frac{1}{N-b}\right)^2 (\lambda_1 - d_2\lambda_2)^2$$

and

$$\text{var}(\hat{\rho}_A) = \boldsymbol{\Phi_A}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{\Phi_A} = \frac{var(S_1)}{\lambda_1^4} (1 + 4\pi^2 - 4\pi) \left( \left(\frac{r_2 - r_1}{r_1(N-b)}\right) \lambda_1 - \left(\frac{d_2}{N-b}\right) \lambda_3 \right)^2$$

for the constants $d_1 = \frac{r_3 - r_2}{r_1}$ and $d_2 = \frac{r_1 r_3 - r_3 - r_1 + r_2}{r_1}$ and where

$$\lambda_1 = -\frac{1}{(a-1)} N\pi^2 + \frac{d_2}{N-b} N\pi + \left(\frac{1}{a-1} - \frac{d_1}{b-a}\right) \pi^2 \Bigg( 4a + 4\pi(1 -$$

$$\pi)\rho_{B(A)} \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j=1}^{b_i} n_{ij}\left(n_{ij}-1\right) + 4\pi\rho_A(1 -$$

$$\pi) \sum_{i=1}^{a} \frac{1}{N_i} \sum_{j \neq j'}^{b_i} \sqrt{n_{ij} n_{ij'} \left(1 + \rho_{B(A)}\left(n_{ij}-1\right)\right)\left(1 + \rho_{B(A)}\left(n_{ij'}-1\right)\right)} - N \Bigg) +$$

$$\left(\frac{d_1}{b-a} - \frac{d_2}{N-b}\right) \left( (1-\pi)\left[b + N\rho_{B(A)} - b\rho_{B(A)}\right] + N\pi^2 \right)$$

$$\lambda_2 = -\frac{1}{N-b}N\pi - \frac{1}{b-a}\pi^2\Bigg(4a + 4\pi(1-\pi)\rho_{B(A)}\sum_{i=1}^{a}\frac{1}{N_i}\sum_{j=1}^{b_i} n_{ij}\big(n_{ij}-1\big) +$$

$$4\pi\rho_A(1-\pi)\ \sum_{i=1}^{a}\frac{1}{N_i}\sum_{j\neq j'}^{b_i}\sqrt{n_{ij}n_{ij'}\Big(1+\rho_{B(A)}\big(n_{ij}-1\big)\Big)\Big(1+\rho_{B(A)}\big(n_{ij'}-1\big)\Big)} -$$

$$N\Bigg) + \Big(\frac{1}{N-b}+\frac{1}{b-a}\Big)\big(\pi(1-\pi)\big[b+N\rho_{B(A)}-b\rho_{B(A)}\big]+N\pi^2\big)$$

and

$$\lambda_3 = -\frac{1}{(a-1)}N\pi^2 + \frac{r_2-r_1}{r_1(N-b)}N\pi + \pi^2\Big(\frac{1}{a-1}+\frac{r_2}{r_1(b-a)}\Big)\Bigg[4a + 4\pi(1-$$

$$\pi)\rho_{B(A)}\sum_{i=1}^{a}\frac{1}{N_i}\sum_{j=1}^{b_i} n_{ij}\big(n_{ij}-1\big) + 4\pi\rho_A(1-$$

$$\pi)\ \sum_{i=1}^{a}\frac{1}{N_i}\sum_{j\neq j'}^{b_i}\sqrt{n_{ij}n_{ij'}\Big(1+\rho_{B(A)}\big(n_{ij}-1\big)\Big)\Big(1+\rho_{B(A)}\big(n_{ij'}-1\big)\Big)} - N\Bigg] -$$

$$\Big(\frac{r_2}{r_1(b-a)}+\frac{r_2-r_1}{r_1(N-b)}\Big)\big[\pi(1-\pi)\big[b+N\rho_{B(A)}-b\rho_{B(A)}\big]+N\pi^2\big].$$

By a second application of the delta method, the distribution after a logarithmic transformation is

$$\sqrt{N}\big(\log(\hat{\rho}_{B(A)}) - log(\rho_{B(A)})\big) \rightarrow N\left(0, \left(\frac{1}{\rho_{B(A)}}\right)^2 \boldsymbol{\Phi_{B(A)}}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi_{B(A)}}\right)$$

and

$$\sqrt{N}(\log(\hat{\rho}_A) - \log(\rho_A)) \rightarrow N\left(0, \left(\frac{1}{\rho_A}\right)^2 \boldsymbol{\Phi_A}^{\boldsymbol{T}}\boldsymbol{\Sigma}\boldsymbol{\Phi_A}\right)$$

# CHAPTER 3

# PARTITIONED GMM LOGISTIC REGRESSION MODELS FOR LONGITUDINAL DATA

Kyle M. Irimata, Jennifer Broatch, Jeffrey R. Wilson

***Abstract***

Correlation is inherent in longitudinal studies due to the repeated measurements on subjects, as well as due to time-dependent covariates in the study. In the National Longitudinal Study of Adolescent to Adult Health (Add Health), data was repeatedly collected on children in grades 7-12 across four waves. Thus, observations obtained on the same adolescent were correlated, while predictors were correlated with current and future outcomes such as obesity status, amongst other health issues. Previous methods, such as the generalized method of moments (GMM) approach have been proposed to estimate regression coefficients for time-dependent covariates. However, these approaches combined all valid moment conditions to produce an averaged parameter estimate for each covariate and thus assumed that the effect of each covariate on the response was constant across time. This assumption is not necessarily optimal in applications such as Add Health or health-related data. Thus, we depart from this assumption and instead use the Partitioned GMM approach to estimate multiple coefficients for the data based on different time-periods. These extra regression coefficients are obtained using a partitioning of the moment conditions pertaining to each respective relationship. This approach offers a deeper understanding and appreciation into the effect of each covariate on the response. We conduct simulation studies, as well

as analyses of obesity in Add Health, rehospitalization in Medicare data, and depression scores in a clinical study. The Partitioned GMM methods exhibit benefits over previously proposed models with improved insight into the non-constant relationships realized when analyzing longitudinal data.

### *3.1. Introduction*

It is common in many fields, such as health and health related research, to observe subjects or units over time, while also measuring covariates at each visit. For example, the National Longitudinal Study of Adolescent to Adult Health (Add Health) is a study of a nationally representative sample of adolescents in grades 7-12 in the United States, which was collected on a cohort of students over four waves, with the first wave beginning in 1994-1995. This type of study produces longitudinal data, which allow for the testing of more involved hypotheses, with improved efficiency of estimates, as compared to cross-sectional or time-series data (Hsiao 2007). However, the design also complicates the statistical analysis because the observations are no longer independent due to the repeated measurements. The collection over time also introduces interdependence of the covariates and the responses across time. For instance, a child's depression level in the Add Health study may affect his or her obesity status at the time of measurement, as well as obesity status during a future measurement. The correlation between repeated measurements in longitudinal studies has been addressed using marginal models, such as generalized estimating equations (GEE) (Zeger and Liang 1986) or using a subject-specific approach such as mixed modeling(Breslow and Clayton 1993). However, there are comparatively fewer appropriate methods to account for the

time-dependent covariates caused by the correlation between the covariates and response across time. Many approaches for accounting for time-dependent covariates also assume that a constant level of association exists at all time-periods, though in practice this may not hold in longitudinal data. For example, in a study of the effect of high epoetin alpha dosage on mortality amongst elderly hemodialysis patients discussed by Zhang, et al (2009), the dosage is administered based on targeted levels of hematocrit. The levels of hematocrit are thus related to both current and previous doses of epoetin (Heagerty and Comstock 2013).

In this paper, we present a generalized method of moments (GMM) model for time-dependent covariates. This method utilizes a partitioning of the moment conditions to represent the varying impacts of each covariate on the responses across time. The model allows the strength of the impact of the covariate to vary due to time, and utilizes a reconfigured, lower diagonal data matrix. Thus, we provide a model with multiple regression coefficients rather than using a linear combination of the associations, which may impact the overall results. These multiple regression coefficients provide a more complete description of the relationship between the covariates and the response and avoid the potential averaging of positive and negative, or strong and weak relationships.

### *3.1.1 Longitudinal Data and Marginal Models*

Longitudinal data consist of merging inter-individual differences and intra-individual dynamics and have several advantages over cross-sectional or time-series data. One advantage is that it allows researchers to study the dynamic parts of a model. Longitudinal data also provides more accurate predictions of individual outcomes since it

pools the data, “borrowing strength” from other observations, instead of providing predictions of individual outcomes based on the data on the individual (Diggle, et al. 2002).

Marginal or population averaged models, such as GEE, are commonly used in the analysis of such data to address the mean response as a function of covariates. These models utilize a working correlation structure, based on some presumed relationship, but do not distinguish between valid or invalid moment conditions. The correlation is assumed to exist due to the repeated measures taken on the subjects; however, the subjects themselves are assumed independent of one another.

In the case of longitudinal data with time-dependent covariates, Pepe and Anderson (1994) showed that the GEE approach is valid and provides consistent estimates if the independent working correlation matrix is utilized. This is also the case if future applications of the results require the expectation of the response as a function of the current covariates. However, Lai and Small (2007) showed that, although the independent working correlation matrix may be a safe choice, it does not provide efficient estimates. In this paper, we introduce extra regression parameters to analyze longitudinal data, such as the Add Health data, while taking into account the intricate relationships that exist at varying degrees due to the measurements of obesity collected across time.

### *3.1.2 Lagged Models*

Lagged models are often used with longitudinal or time series data. These models incorporate previous values of the dependent or independent variable from earlier time-

periods to account for autocorrelation in the data (Keele and Kelly 2005). However, when there is serial correlation, these models can produce biased estimates. Further, the introduction of a lagged dependent variable sometimes suppresses the effects of the covariates in the model, and often lacks reasonable causal interpretation (Achen 2001). Keele and Kelly (2005) showed that the use of a lagged dependent variable is inappropriate in certain circumstances, though it remains one of the best approaches for addressing time series data. As Diggle, et al (2002) noted, the appropriate use of such predictors depends in part on the goals of the analysis.

Generalized estimating equations have been proposed as an extension to lagged covariate models (Zeger and Liang 1986) with an appropriately selected working correlation matrix. Schildcrout and Heagerty (2005) suggested the use of the independent working correlation matrix for lagged GEE models in order to ensure consistent parameter estimates, although this may also lead to relative inefficiency. We extend the lagged covariate model with a generalized method of moments approach to address the correlation induced by time-dependent covariates.

### *3.1.3 GMM Models*

The generalized method of moments estimator was introduced by Hansen (1982) with applications to econometrics. Lai and Small (2007) and Lalonde, Wilson, and Yin (2014) showed that the generalized method of moments (GMM) model is a great choice when there are time-dependent covariates. Although these methods, amongst others (Guerra, et al. 2012; Zhou, et al. 2014; Chen and Westgate 2017), have been proposed for logistic regression models with time-dependent covariates, these models did not

distinguish between the strength and type of association between the responses and covariates at different time-periods. These approaches instead combined all associations to provide one parameter estimate for each covariate, regardless of the varying strength or direction of the association.

We introduce a partitioned generalized method of moments approach for separating regression coefficients, to distinguish the effect of covariates on the outcome when they are observed in the same time-period from the effect when they are observed in different time-periods. The partitioned model combines the features of lagged models, and the characteristics of GMM models to describe the varying strength of the relationships between the covariates and the responses over time. In Section 2, we review existing methods for longitudinal data, with an emphasis on GMM models. In Section 3, we provide the Partitioned GMM framework, which is used to determine the varying impact of the covariates at different periods on the response. We present the results of a simulation study in Section 4 to demonstrate the performance of the Partitioned GMM model. Applications to Add Health data, Medicare rehospitalizations, and depression scores in a clinical study are discussed in Section 5.

### *3.2. Marginal Regression Modeling with Time-Dependent Covariates*

Marginal models have been introduced to address the challenges created due to time-dependent covariates. Guerra, Shults, Amsterdam and Ten-Have (2012) presented a logistic regression model for longitudinal data, relating the mean of the response with covariates based on the subjects under measure. They adjusted for the correlation due to the repeated measurements on each subject using a maximum likelihood method for time-

independent and time–dependent covariates. Selig, Preacher and Little (2012) accounted for the impact of time-dependent covariates by using several different functional forms, thereby presenting lag-moderated associations. For covariates measured at varying times, they evaluated the difference on one covariate as it relates to the difference on another covariate. Müller and Stadtmuller (2005) introduced a generalized functional linear regression model where the predictor is a random function, which relied on dimension reduction using orthogonal expansion. Our model incorporates a special case of this model, and relies on generalized method of moments to estimate the regression coefficients.

Zhou, Lefante, Rice, and Chen (2014) introduced a method using the modified quadratic inference function. They used alternative forms for the working correlation matrix, with a different form for each type of covariate corresponding to the valid moment conditions. Their approach improves consistency and efficiency, although it relied on a single regression parameter to describe the relationship between each covariate and the response. Chen and Westgate (2017) provided a new GMM approach which utilized a modified weight matrix based on linear shrinkage to help avoid singularity. They also introduced a modified GEE approach with an adjusted working correlation matrix to eliminate biased equations, along with a model selection approach to identify an appropriate model based on the data type. This model was useful in improving properties of the regression parameter estimates; however, their proposed method considered only models where the effect of each predictor on the response is constant across time.

*3.2.1 GMM Models with Covariate Classification*

Lai and Small (2007) used a marginal model for longitudinal continuous data with generalized method of moments (GMM) to account for the time-dependent covariates. They considered repeated observations, with response $y_{it}$ for subject $i$ at time $t$, whose marginal distribution follows a generalized linear model, given the time-dependent vector of covariates $\boldsymbol{x}_{it.} = (x_{it1}, \dots, x_{itJ})$. They assumed that the observations $y_{is}$ and $y_{kt}$ are independent when $i \neq k$ but not necessarily when $i = k$ and $s \neq t$. Thus, observations from different subjects were assumed independent, while observations from the same subject were not. In obtaining estimates of the regression coefficients, Lai and Small made use of the moment conditions such that

$$E\left[\frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \beta_j}\{y_{it} - \mu_{it}(\boldsymbol{\beta})\}\right] = 0 \qquad (3.2.1)$$

for appropriately chosen $s$, $t$, and $j$, where $\mu_{it}(\boldsymbol{\beta}) = E[\{y_{it}|X_{it}\}]$ denotes the expectation of $y_{it}$ based on the vector of covariate values $\boldsymbol{x}_{it.}$ associated with the vector of parameters $\boldsymbol{\beta}$ in the systematic component that describes the marginal distribution of $y_{it}$. Their model made full use of the valid moment conditions in groups based on time-dependent covariates, to obtain estimates.

The benefit for this approach was identifying the appropriate moment conditions associated with the covariates, as methods such as GEE with time-dependent covariates will omit some valid moment conditions. Lai and Small classified covariates as type I, type II or type III, based on which moment conditions were considered valid. Covariates for which Equation (3.2.1) holds for all $s$ and $t$ were designated as type I. A covariate was considered type II when Equation (3.2.1) holds for all $s \geq t$, but fails for some $s < t$.

Thus, type II covariates observed at time $s$ affect the outcome at future time $t$, though there is no feedback from the outcome onto the predictors. Covariates for which Equation (3.2.1) does not hold for any $s > t$ were designated as Type III covariates. These covariates can occur when feedback is present between the outcome at previous time-periods and the covariate at future time-periods. Each type of covariate utilized a different set of moment conditions to estimate the corresponding regression coefficient. These models assumed that the strength and direction of the association between the response and the covariate in any two-different time-periods remains the same. This assumption omits the effect of doses in a patient's care over time, though there is a differential effect as time progresses. Thus, applications of this approach are limited in longitudinal data.

### *3.2.2 GMM Models with Ungrouped Moment Conditions*

As an alternative to the grouping of moments based on covariate type, Lalonde, Wilson, and Yin (2014) introduced a method to ignore the classification and to instead look at the validity of each moment separately. In their individual approach to identifying valid moments, they relied on bivariate correlations to determine validity of the corresponding moment condition. These valid moments were used to obtain estimates of the regression coefficients. They assumed that all moments when the predictor and response are observed in the same time-period, $s = t$ are valid, and tested the remaining $T(T-1)$ moment conditions individually for validity. The moment condition (3.2.1) was considered valid when $\rho_{x_t,e_s} = 0$, that is when the correlation between the residual observed at time s, denoted by $e_s$, and the covariate observed at time *t*, denoted by $x_t$

where $s \neq t$ was zero. Thus, this GMM approach (Lalonde, Wilson, and Yin 2014) accounted for the feedback created between the outcomes at a particular time-period and the predictors at later time-periods. However, similar to Lai and Small (2007), they grouped the valid moments to obtain an estimate of a single regression coefficient to represent the overall effect of a given covariate.

### *3.3. Partitioned Coefficients with Time-Dependent Covariates*

We present the Partitioned GMM as an alternative approach to existing GMM models (Lai and Small 2007; Lalonde, Wilson, and Yin 2014; Zhou, et al. 2014) for time-dependent covariates. We utilize the test for valid moment conditions presented by Lalonde, Wilson and Yin (2014), as well as the type II covariate proposed by Lai and Small (2007), though the approach can be readily extended to incorporate alternative moment condition selection techniques. Instead of grouping all valid moment conditions to obtain an average effect of the covariate on the response, we partition the moment conditions and separate the effects of the covariates on the responses across time. This partitioning produces extra regression parameters for each covariate. The moment conditions are grouped based on the relationship of interest, and on the time lag between the covariate and the response. The Partitioned GMM is best applicable to data without many repeated observations, relative to the number of observations (Stoner, Leroux, and Puumala 2010).

#### *3.3.1 Partitioned GMM Model*

The Partitioned GMM model accounts for the relationships between the outcomes

observed at time $t$, $Y_t$ and the j[th] covariate observed at time $s$, $X_{js}$ for $s \leq t$. In fitting this model, for each time-dependent covariate $X_j$ measured at times $1, 2, \ldots, T$; for subject $i$ and the $j^{th}$covariate, the data matrix is reconfigured as a lower triangular matrix,

$$\mathbf{X}_{ij} = \begin{bmatrix} 1 & X_{ij1} & 0 & \ldots & 0 \\ 1 & X_{ij2} & X_{ij1} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & X_{i1T} & X_{ij(T-1)} & \ldots & X_{ij1} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & X_{ij}^{[0]} & X_{ij}^{[1]} & \ldots & X_{ij}^{[T-1]} \end{bmatrix}$$

where the superscript denotes the difference, $t - s$ in time-periods between the response time $t$ and the covariate time $s$. Thus, the model is given by

$$g(\mu_{it}) = \beta_0 + \beta_j^{tt} X_{ij}^{[0]} + \beta_j^{[1]} X_{ij}^{[1]} + \beta_j^{[2]} X_{ij}^{[2]} \ldots + \beta_j^{[T-1]} X_{ij}^{[T-1]} \qquad (3.3.1)$$

and in matrix notation $g(\boldsymbol{\mu_i}) = \boldsymbol{X_{ij}\beta}_j$, where the $\boldsymbol{X}_{ij}$ matrix denotes the systematic component and the mean response is $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})'$ dependent on the regression coefficients $\beta_j = \left(\beta_0, \beta_j^{tt}, \beta_j^{[1]}, \beta_j^{[2]}, \ldots, \beta_j^{[T-1]}\right)$. The coefficient $\beta_j^{tt}$ denotes the effect of the covariate $X_{jt}$ on the response $Y_t$ during the $t^{th}$ 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 $X_{js}$ 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 identified 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 $X_{js}$ on $Y_t$ across a one time-period lag. In general, each of the $J$ time-dependent covariates yields a maximum of $T$ partitions of $\beta_j$. Let $\boldsymbol{\beta}$ be the concatenation of the parameters associated with each of the $J$ covariates. Thus, for a

model with *J* covariates, the data matrix $\boldsymbol{X}$ will have a maximum dimension of $N$ by $T$, and $\boldsymbol{\beta}$ is a vector of maximum length $J \times (T + 1)$.

The extra regression parameters naturally lead to questions regarding singularity of the data matrix. We argue that this phenomenon is similar to the use of Generalized Estimating Equations (GEE) with correlated data. Stoner, Leroux, and Puumala (2010) found that when the size of each cluster is large relative to the number of clusters, marginal models such as GEE with flexible correlation structures may not converge, while fixed working correlation structures may produce estimates that are not efficient. We analogously found that the use of extra regression parameters produced reliable estimates when the number of clusters are large in comparison to the number of time-periods.

### *3.3.2 Partitioned GMM Estimation*

Consider $y_i$ for $i = 1, \dots, N$; to be an independent and identically distributed random variable with mean $\mu_{it}$ at time *t*, and let $\boldsymbol{\beta_0}$ denote the vector of regression parameters. Let $\boldsymbol{T_j}$ be the $T \times T$ matrix, which specifies which moment conditions are valid for the $j^{\text{th}}$ covariate, as determined by the desired approach. Thus, elements in $\boldsymbol{T_j}$ take on the value of one when there is valid moment condition according to Equation (3.2.1), and takes a value of zero when the moment is not valid for the $j^{\text{th}}$ covariate. The $\frac{1}{2}T(T-1)$ moments pertaining to cases when $s > t$ are set to zero. The elements in $\boldsymbol{T_j}$ are partitioned into up to *T* separate $T \times T$ matrices denoted by $\boldsymbol{T_{jk}}$ for $k = 1, \dots, T-1$. The information for the *T* moment conditions when $s = t$, occurring when the response

and the covariate are observed in the same time-period, are contained in $\boldsymbol{T_{j0}}$, an identity matrix. Information for the moment conditions occurring when the response is observed one time-period after the covariate, $t - s = 1$ are contained in the matrix $\boldsymbol{T_{j1}}$. To accommodate the adjusted data vector $X_{ij}^{[1]}$ discussed in Section 3.1, shift each element in $\boldsymbol{T_{j1}}$ forward by one column such that the valid moment conditions in $\boldsymbol{T_{j1}}$ exist only on the diagonal, with zero otherwise. Each of the remaining matrices $\boldsymbol{T_{jk}}$ are created similarly.

Let $\boldsymbol{T_{vjk}}$ be the reshaped $1\ x\ T^2$ vector of the elements in $\boldsymbol{T_{jk}}$. Concatenate the row vectors for all covariates and lagged effects to form the matrix $\boldsymbol{T_{shape}}$, which is of maximum dimension $(J \times T) \times T^2$. Let $N_v$ be the number of ones in $\boldsymbol{T_{shape}}$, or equivalently the total number of valid moment conditions. Let $\Omega_{tt} \in [x_s, y_t;\ s = t]$ and for $s < t$, consider each valid moment condition where $\Omega_{st} \in [x_s, y_t\ ;\ s \neq t]$. There are $T$ members in $\Omega_{\mathrm{tt}}$ and one member for each of $\Omega_{\mathrm{st}}$. Thus the fitted model to (3.3.1) is

$$\mu_{it}(\beta) = \beta_0 + \beta_j^{tt} X_{ij}^{[0]} + \sum_{\boldsymbol{k=1}}^{\boldsymbol{T-1}} \beta_j^{[k]} X_{ij}^{[k]} \,|_{\boldsymbol{valid\ moments}}$$

Let $\boldsymbol{g}_i$ be an $N_v\ x\ 1$ vector composed of the values of all valid moment conditions for subject $i$, computed at the initial value $\boldsymbol{\beta_0}$. Each element in $\boldsymbol{g}_i$ is calculated as $\frac{\partial \mu_{is}(\boldsymbol{\beta_0})}{\partial \beta_j^{[k]}}[y_{it} - \mu_{it}(\boldsymbol{\beta_0})]$ such that the corresponding element in $\boldsymbol{T_{jk}}$ takes value 1 for $k = 1, \ldots, T-1$. Let $\boldsymbol{G_n}$ be the $N_v\ x\ 1$ vector consisting the sample average of all valid moment conditions, such that

$$\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathbf{g}_{\mathrm{i}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \frac{\partial \mu_{\mathrm{is}}(\boldsymbol{\beta_0})}{\partial \beta_{\mathrm{j}}^{[\mathrm{k}]}}[\mathrm{y_{it}} - \mu_{\mathrm{it}}(\boldsymbol{\beta_0})].$$

The optimal weight matrix $\boldsymbol{W_n}$ is computed as $\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}_i\boldsymbol{g}_i^T\right)^{-1}$, which is of dimension $N_v\ x\ N_v$. Then, the GMM regression estimator is

$$\widehat{\boldsymbol{\beta}}_{GMM} = \underset{\boldsymbol{\beta_0}}{argmin}\ \boldsymbol{G_n}(\boldsymbol{\beta_0})^T\ \boldsymbol{W_n}(\boldsymbol{\beta_0})\ \ \boldsymbol{G_n}(\boldsymbol{\beta_0}),$$

which is the argument minimizing the quadratic objective function. The asymptotic variance of the estimator $\widehat{\boldsymbol{\beta}}_{GMM}$ is

$$\left[\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\frac{\partial \mathbf{g}_{\mathrm{i}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\mathbf{W}_{\mathbf{n}}(\boldsymbol{\beta})\left(\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\frac{\partial \mathbf{g}_{\mathrm{i}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)\right]^{-1},$$

evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_{GMM}$.

*Logistic Regression Model:* In the case of the logistic regression model, the mean is given by

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

so the valid elements in $\boldsymbol{g}_i$ each take the form:

$$\frac{\partial \mu_{\mathrm{is}}(\boldsymbol{\beta_0})}{\partial \beta_{\mathrm{j}}^{[\mathrm{k}]}}[\mathrm{y_{it}} - \mu_{\mathrm{it}}(\boldsymbol{\beta_0})] = \mathrm{x_{isj}}\mu_{\mathrm{is}}(\boldsymbol{\beta_0})[1 - \mu_{\mathrm{is}}(\boldsymbol{\beta_0})][\mathrm{y_{it}} - \mu_{\mathrm{it}}(\boldsymbol{\beta_0})].$$

Thus, for the asymptotic variance, in the case of logistic regression, each $N_v\ x\ 1$ vector $\frac{\partial \mathbf{g}_{\mathrm{i}}(\boldsymbol{\beta})}{\partial \beta_{\mathrm{l}}^{[\mathbf{k}]}}$ in the matrix

$$\frac{\partial \boldsymbol{g}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left[\frac{\partial \boldsymbol{g}_i(\boldsymbol{\beta})}{\partial \beta_j^{[1]}}, \ldots, \frac{\partial \boldsymbol{g}_i(\boldsymbol{\beta})}{\partial \beta_j^{[T-1]}}\right]$$

is obtained as

$$\frac{\partial\left\{\left[\frac{\partial\mu_{is}(\boldsymbol{\beta})}{\partial\beta_j^{[k]}}\right][y_{it}-\mu_{it}(\boldsymbol{\beta})]\right\}}{\partial\beta_l^{[m]}} = x_{isj}\mu_{is}(\boldsymbol{\beta})[1-\mu_{is}(\boldsymbol{\beta})]\left\{x_{isl}[1-2\mu_{is}(\boldsymbol{\beta})][y_{it}-\mu_{it}(\boldsymbol{\beta})]-\right.$$

$$\left. x_{itj}\mu_{it}(\boldsymbol{\beta})[1-\mu_{it}(\boldsymbol{\beta})]\right\},$$

where $j = 1, \ldots, \mathrm{J}$, $l = 1, \ldots J$, $k = 1, \ldots, T-1$ and $m = 1, \ldots, T-1$.

*Normal distribution model:* Similarly, for the normal error model, the moment conditions in $\boldsymbol{g}_i$ take the form

$$\frac{\partial\mu_{is}(\boldsymbol{\beta_0})}{\partial\beta_j^{[k]}}[y_{it}-\mu_{it}(\boldsymbol{\beta_0})] = x_{isj}[y_{it}-\mu_{it}(\boldsymbol{\beta_0})],$$

for the valid moment conditions. The asymptotic variance is computed using the $N_v\ x\ J$ matrix

$$\frac{\partial\boldsymbol{g}_i(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \left[\frac{\partial\boldsymbol{g}_i(\boldsymbol{\beta})}{\partial\beta_j^{[1]}}, \quad \ldots, \quad \frac{\partial\boldsymbol{g}_i(\boldsymbol{\beta})}{\partial\beta_j^{[T-1]}}\right],$$

where each of the $N_v\ x\ 1$ vectors $\frac{\partial\boldsymbol{g}_i(\boldsymbol{\beta})}{\partial\beta_j^{[\boldsymbol{k}]}}$ is computed as

$$\frac{\partial\left\{\left[\frac{\partial\mu_{is}(\boldsymbol{\beta})}{\partial\beta_j^{[k]}}\right][y_{it}-\mu_{it}(\boldsymbol{\beta})]\right\}}{\partial\beta_l^{[m]}} = -x_{isj}x_{isl},$$

for $j = 1, \ldots, \mathrm{J}$, $l = 1, \ldots J$, $k = 1, \ldots, T-1$ and $m = 1, \ldots, T-1$.

### *3.3.3 Types of Partitioned GMM Models*

We present two Partitioned GMM models, one based on the Lalonde, Wilson and Yin (2014) approach to identifying valid moment conditions (Partitioned-LWY) and the other based on the Lai and Small (2007) approach to identifying valid moment conditions

using covariate classification (Partitioned-LS). The partitioning approach is used in conjunction with either of the two methods for identifying and selecting valid moment conditions.

The Lalonde, Wilson and Yin (2014) approach evaluates each moment condition individually, and thus improves estimation. On the other hand, the introduction of lagged parameter estimates depends on comparatively fewer moment conditions, at the different segments. As such, it is possible at times that these lagged parameters will not always be estimable if certain moment conditions are not valid. The Lai and Small (2007) method depends on the grouping of moments based on covariate type. Thus, under this set of moment conditions, all lagged parameters are estimable, although the estimation may rely on moments that are not valid. In Section 4, we conduct a simulation study to evaluate the performance of these two methods.

### *3.4. Simulation Study*

We conducted a simulation study to examine the performance of the Partitioned GMM and to compare the model to non-partitioned models. We simulated a Bernoulli random variable with the mean response dependent on a continuous type II time-dependent covariate (Lai and Small 2007). We assigned a time-dependent covariate weight of 1.5 and with correlation induced by random effects distributed according to a normal distribution, with mean 0 and variance 1. The data were simulated with complete observations at $T = 3$ time-periods for $N = 100,\ 250$ and $500$ subjects. The regression parameters were set to $\beta_0 = 0$, $\beta^{tt} = 0.3$, $\beta^{[1]} = 0.5$ and $\beta^{[2]} = 0.7$, thus there was a strong lagged effect of the covariate on the response across time.

We fit two Partitioned GMM models to each simulated set of data, following the Lalonde, Wilson and Yin (2014) moment approach (Partitioned-LWY) model, and using the Lai and Small (2007) moment approach (Partitioned-LS) model. Following Zeger and Liang (1991), we fit a lagged GEE model (Lagged-GEE) with an independent working correlation matrix. In addition, we fit "one-parameter per covariate" models, including Lalonde, Wilson and Yin (2014) GMM (LWY-GMM), Lai and Small (2007) GMM (LS-GMM) and GEE with an independent working correlation matrix (GEE-IND). In the partitioned models, the coefficient $\beta_j^{tt}$ is comparable with LWY-GMM, LS-GMM and GEE-IND. Average parameter estimates (APE), and coverage probabilities based on 95% confidence intervals are recorded for each model in Table 3.4.1.

Table 3.4.1. Simulated Coverage Probabilities and Average Parameter Estimates **(**APE) for partitioned and cross-sectional approaches

| | | **Intercept = 0** | | $\boldsymbol{\beta^{tt} = 0.3}$ | | $\boldsymbol{\beta^{[1]} = 0.5}$ | | $\boldsymbol{\beta^{[2]} = 0.7}$ | |
|---|---|---|---|---|---|---|---|---|---|
| | **Method** | **Coverage** | **APE** | **Coverage** | **APE** | **Coverage** | **APE** | **Coverage** | **APE** |
| **N=100** | **Partitioned LWY** | 0.94 | 0.00 | 0.91 | 0.29 | 0.93 | 0.52 | 0.91 | 0.56 |
| | **Partitioned LS** | 0.94 | 0.00 | 0.91 | 0.26 | 0.92 | 0.43 | 0.92 | 0.66 |
| | **Lagged-GEE** | 0.96 | 0.00 | 0.93 | 0.29 | 0.93 | 0.46 | 0.94 | 0.76 |
| | **LWY-GMM** | 0.72 | -0.63 | 0.07 | 13.44 | | | | |
| | **LS-GMM** | 0.41 | -3.07 | 0.02 | 25.99 | | | | |
| | **GEE-IND** | 0.96 | 0.00 | 0.24 | 0.51 | | | | |
| **N=250** | **Partitioned LWY** | 0.95 | 0.00 | 0.91 | 0.28 | 0.96 | 0.53 | 0.80 | 0.40 |
| | **Partitioned LS** | 0.95 | 0.00 | 0.91 | 0.26 | 0.92 | 0.44 | 0.91 | 0.64 |
| | **Lagged-GEE** | 0.96 | 0.00 | 0.90 | 0.25 | 0.93 | 0.43 | 0.93 | 0.63 |
| | **LWY-GMM** | 0.88 | -0.17 | 0.00 | 4.04 | | | | |
| | **LS-GMM** | 0.58 | -0.44 | 0.00 | 15.65 | | | | |
| | **GEE-IND** | 0.96 | 0.00 | 0.02 | 0.51 | | | | |
| **N=500** | **Partitioned LWY** | 0.98 | 0.00 | 0.85 | 0.31 | 0.98 | 0.51 | 0.40* | 0.32* |
| | **Partitioned LS** | 0.98 | 0.00 | 0.88 | 0.26 | 0.90 | 0.43 | 0.90 | 0.61 |
| | **Lagged-GEE** | 0.98 | 0.00 | 0.87 | 0.25 | 0.90 | 0.43 | 0.91 | 0.61 |
| | **LWY-GMM** | 0.96 | -0.08 | 0.00 | 1.09 | | | | |
| | **LS-GMM** | 0.72 | -0.18 | 0.00 | 10.05 | | | | |
| | **GEE-IND** | 0.98 | 0.00 | 0.00 | 0.50 | | | | |

*Denotes results obtained from 30 or fewer simulations

The cross-sectional models (LWY-GMM, LS-GMM, and GEE-IND), with one parameter per covariate provide poor coverage and poor parameter estimates. This poor

performance is attributed to the combining of all moment conditions, regardless of strength or direction, which is inherent in these cross-sectional models. As the relationship between the covariate and the response changes over time, a single, a single cross-sectional parameter is incapable of providing accurate and reliable estimates. However, the partitioned models overcome these challenges using estimates for each lagged coefficient.

The Partitioned-LS and Partitioned-LWY models produce comparable results, with some notable differences. The Partitioned-LWY model generally provides better coverage of the true parameter. The Partitioned-LS model produces coverage of at least 88% across all simulation settings, for all regression parameters. The coverage for both approaches was less than nominal, which has previously been discussed as a disadvantage of the simulation approach (Lalonde, Wilson, and Yin 2014). In general, the Partitioned-LWY produced parameter estimates that were less biased than those produced by the Partitioned-LS. However, in cases when the Partitioned-LS outperforms the Partitioned-LWY, the discrepancy was pronounced, with the Partitioned-LWY yielding largely more biased estimates. The Partitioned-LWY model relies on valid moment conditions using an ungrouped approach, and thus some coefficients are not always estimable. The two partitioned approaches both generally perform better than the Lagged-GEE, with the Partitioned-LWY model providing the preferable model overall.

In Section 5, we fit these two partitioned models to three numerical examples using a SAS MACRO (Cai and Wilson 2015; Cai and Wilson 2016) which is available at http://www.public.asu.edu/~jeffreyw. Additional discussions pertaining to the use of this macro are provided in the Supporting Information.

### *3.5. Numerical Examples*

We revisited three numerical examples. One example modeled obesity in children using the National Longitudinal Study of Adolescent to Adult Health (Add Health) (Harris and Udry 2016). A second example was a study of patient rehospitalization (Lalonde, Wilson, and Yin 2014; Jencks, Williams, and Coleman 2009) using Medicare data. A third example focused on a clinical study of depression scores (Reisby, et al. 1977). Each of these examples was analyzed using Partitioned GMM models, with moment conditions selected using the Lalonde, Wilson and Yin approach (Partitioned-LWY) and the Lai and Small approach to obtain moment conditions (Partitioned-LS). We compared the results to the one-parameter per covariate models obtained from Lalonde, Wilson and Yin GMM (LWY-GMM), Lai and Small GMM (LS-GMM), and GEE with an independent (GEE-IND) working correlation matrix. We also provided comparisons to the Lagged-GEE model using an independent working correlation structure.

#### *3.5.1 Add Health Data*

There are efforts in place to reduce and understand childhood obesity in the United States, with 17% childhood obesity ((NCCOR) 2014). We fit the Partitioned GMM models to the National Longitudinal Study of Adolescent to Adult Health (Add Health) to investigate the relationships between risk factors and obesity in adolescents. These data were originally collected from students in grades 7 to 12, beginning in 1994-1995. The students were measured at three time-periods after their initial enrolment, resulting in four measurements, producing information on 2,712 students at each of the four time-periods. The binary outcome measures obesity status based on each student's BMI (Pu, Fang, and Wilson 2017). The time-dependent covariates were depression scale,

number of hours spent watching television, physical activity level and whether the student was a social alcohol drinker. The data included a time-independent predictor for race, denoting white or non-white. The identification of valid moment conditions using the Lalonde, Wilson and Yin (2014) approach is given in Table 3.5.1.1.

Table 3.5.1.1. Moment Conditions for the Add Health Study

| | **Depression** | | | | **TV Hrs** | | | |
|---|---|---|---|---|---|---|---|---|
| | **s=1** | **s=2** | **s=3** | **s=4** | **s=1** | **s=2** | **s=3** | **s=4** |
| **t=1** | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **t=2** | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| **t=3** | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| **t=4** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| | **Activity** | | | | **Alcohol** | | | |
| | **s=1** | **s=2** | **s=3** | **s=4** | **s=1** | **s=2** | **s=3** | **s=4** |
| **t=1** | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **t=2** | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| **t=3** | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| **t=4** | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

For the one-parameter models, LWY-GMM, LS-GMM and GEE-IND, the results vary. The LS-GMM approach identifies all covariates (race, depression, TV Hrs, physical activity level and social alcohol drinking) as significant in predicting obesity. The LWY-GMM model does not find race and alcohol as significant. The GEE-IND model does not race as significant indicator of obesity. These results are included in Table 3.5.1.2.

Table 3.5.1.2. Cross-sectional Parameter Estimates and p-Values for the Add Health Study

| | LS-GMM | | LWY-GMM | | GEE-IND | |
|---|---|---|---|---|---|---|
| **Parameter** | **Est.** | **p-Value** | **Est.** | **p-Value** | **Est.** | **p-Value** |
| **Intercept** | -2.369 | **<.001** | -2.059 | **<.001** | -1.737 | **<.001** |
| **Race** | 0.31 | **0.003** | 0.176 | 0.056 | 0.114 | 0.164 |
| **Depression** | 1.019 | **<.001** | 0.841 | **<.001** | 0.678 | **<.001** |
| **TV Hrs** | 0.017 | **<.001** | 0.016 | **<.001** | 0.012 | **<.001** |
| **Activity** | -0.854 | **<.001** | -0.683 | **<.001** | -0.474 | **<.001** |
| **Alcohol** | 0.244 | **<.001** | 0.124 | 0.068 | 0.147 | **0.032** |

In the Partitioned GMM and the lagged models for the cross-sectional periods, the Partitioned-LS, the Partitioned-LWY and lagged GEE models find depression level and hours spent watching television to be significant. The Partitioned-LWY model finds race, depression level, hours spent watching television and physical activity level as significant in predicting obesity status. Both Partitioned GMM models identify depression level as having significant one time-period lagged effects, though the Partitioned-LWY model also identifies physical activity level as significant at a one time-period lag. Across a two time-period lag, the Partitioned-LS model finds depression level, hours spent watching television, and physical activity level as significant and the Partitioned-LWY model finds physical activity level and social alcohol drinking as significant. Under the Partitioned-LS model, depression level and hours spent watching television are significant predictors across a three time-period lag, and under the Partitioned-LWY model, physical activity level is significant across a three time-period lag. Due to the lack of valid moment conditions based on the Lalonde, Wilson and Yin (2014) method, some lagged relationships are not estimable. The discrepancies between the two analyses are attributed to the different moment conditions employed in obtaining parameter estimates. Although

the Lagged-GEE has similarities with the partitioned models, the results are different. These differences can also be attributed to the use of non-existent moment conditions due to the fixed independent working correlation structure. The estimates and p-values for these two partitioned models as well as the Lagged-GEE, are reported in Table 3.5.1.3.

Table 3.5.1.3. Partitioned Parameter Estimates and p-Values for the Add Health Study

| | | Partitioned-LS | | Partitioned-LWY | | Lagged-GEE | |
|---|---|---|---|---|---|---|---|
| | | **Est.** | **p-Value** | **Est.** | **p-Value** | **Est.** | **p-Value** |
| | **Intercept** | -3.076 | **<.001** | -3.025 | **<.001** | -2.526 | **<.001** |
| | **Race** | 0.074 | 0.433 | 0.222 | **0.020** | 0.067 | 0.456 |
| Cross-sectional | **Depression** | 0.384 | **<.001** | 0.501 | **<.001** | 0.137 | 0.166 |
| | **TV Hrs** | 0.015 | **<.001** | 0.015 | **<.001** | 0.013 | **<.001** |
| | **Activity** | -0.059 | 0.057 | -0.165 | **<.001** | -0.144 | **<.001** |
| | **Alcohol** | -0.060 | 0.414 | 0.010 | 0.895 | -0.124 | 0.064 |
| Lagged one period | **Depression** | 0.315 | **<.001** | 0.582 | **<.001** | 0.290 | **<.001** |
| | **TV Hrs** | 0.002 | 0.216 | 0.004 | 0.089 | 0.004 | **0.046** |
| | **Activity** | -0.028 | 0.197 | -0.095 | **<.001** | -0.021 | 0.350 |
| | **Alcohol** | 0.078 | 0.189 | 0.046 | 0.476 | 0.025 | 0.670 |
| Lagged two periods | **Depression** | 0.661 | **<.001** | - | - | 0.692 | **<.001** |
| | **TV Hrs** | 0.013 | **<.001** | - | - | 0.010 | **<.001** |
| | **Activity** | 0.069 | **0.002** | 0.180 | **<.001** | 0.075 | **0.001** |
| | **Alcohol** | 0.068 | 0.283 | 0.295 | **<.001** | 0.008 | 0.893 |
| Lagged three periods | **Depression** | 0.417 | **<.001** | - | - | 0.432 | **<.001** |
| | **TV Hrs** | 0.012 | **<.001** | - | - | 0.012 | **<.001** |
| | **Activity** | 0.019 | 0.493 | 0.158 | **<.001** | -0.009 | 0.766 |
| | **Alcohol** | 0.017 | 0.822 | - | - | 0.057 | 0.466 |

*3.5.2 Medicare Readmission Data*

Patient rehospitalization within 30 days of discharge for the same diagnosis is a key measure for hospital reimbursements under Medicare. We examined a Medicare dataset to address questions about rehospitalization. The data contained information on 1,625 patients who were admitted to a hospital 4 times. Thus, each subject had three observations indicating the number of days to rehospitalization. The models investigated

the probability of an individual returning to the hospital within 30-days. The covariates were time-dependent, including number of diagnoses (NDX), number of procedures (NPR), length of stay (LOS), and whether the patient had coronary atherosclerosis (DX101) (Jencks, Williams, and Coleman 2009). We utilized the LWY approach (2014) to identify valid moment conditions to fit the Partitioned-LWY. Moment conditions where $s > t$ were not considered, while all moment conditions where $s = t$ are considered valid. Valid moment conditions are denoted by '1' in Table 3.5.2.1. The data were also analyzed using the LS approach assuming Type II covariates.

Table 3.5.2.1. Moment Conditions for the Medicare Study

| | **NDX** | | | **NPR** | | | **LOS** | | | **DX101** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **s=1** | **s=2** | **s=3** | **s=1** | **s=2** | **s=3** | **s=1** | **s=2** | **s=3** | **s=1** | **s=2** | **s=3** |
| **t=1** | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| **t=2** | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| **t=3** | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |

The cross-sectional models (LWY-GMM, LS-GMM and GEE-IND) with one regression parameter per covariate, are used to analyze the Medicare readmission data. These three approaches all identify the number of diagnoses and the length of stay as significant predictors of hospital readmission. The results of these analyses are given in Table 3.5.2.2.

Table 3.5.2.2. Cross-sectional Parameter Estimates and p-Values for the Medicare Study

| | **LS-GMM** | | **LWY-GMM** | | **GEE-IND** | |
|---|---|---|---|---|---|---|
| | **Est.** | **p-value** | **Est.** | **p-value** | **Est.** | **p-value** |
| Intercept | -0.629 | **<.001** | -0.614 | **<.001** | -0.574 | **<.001** |
| NDX | 0.055 | **<.001** | 0.057 | **<.001** | 0.062 | **<.001** |
| NPR | -0.024 | 0.206 | -0.024 | 0.203 | -0.022 | 0.242 |
| LOS | 0.051 | **<.001** | 0.046 | **<.001** | 0.034 | **<.001** |
| DX101 | -0.043 | 0.646 | -0.048 | 0.606 | -0.094 | 0.311 |

The relationships across time in the Medicare data are modelled with the Partitioned-LS and Partitioned-LWY models, and the results of these approaches are similar. These models identify the number of diagnoses and length of stay as significant when the response and the predictor are observed in the same time-period, as well as across a one time-period lag. Length of stay is significant under the Partitioned-LS model at a two time-period lag. Because no valid moment conditions for this particular relationship are identified under the Lalonde, Wilson and Yin approach, the Partitioned-LWY is not able to produce estimates for this parameter. While the Lagged-GEE produces similar results, length of stay is not identified as significant across a two time-period lag. This discrepancy is due to the use of non-existent moment conditions in the GEE model. The results for these three models are given in Table 3.5.2.3.

Table 3.5.2.3. Partitioned Parameter Estimates and p-Values for the Medicare Study

| | | **Partitioned-LS** | | **Partitioned-LWY** | | **Lagged-GEE** | |
|---|---|---|---|---|---|---|---|
| | | **Est.** | **p-Value** | **Est.** | **p-Value** | **Est.** | **p-Value** |
| | **Intercept** | -0.482 | **<.001** | -0.479 | **<.001** | -0.470 | **<.001** |
| Cross-sectional | **NDX** | 0.062 | **<.001** | 0.062 | **<.001** | 0.069 | **<.001** |
| | **NPR** | -0.030 | 0.124 | -0.031 | 0.110 | -0.020 | 0.287 |
| | **LOS** | 0.048 | **<.001** | 0.049 | **<.001** | 0.030 | **<.001** |
| | **DX101** | -0.063 | 0.512 | -0.066 | 0.489 | -0.086 | 0.361 |
| Lagged one period | **NDX** | -0.047 | **<.001** | -0.047 | **<.001** | -0.041 | **<.001** |
| | **NPR** | -0.012 | 0.605 | -0.016 | 0.490 | -0.019 | 0.389 |
| | **LOS** | 0.018 | **0.022** | 0.019 | **0.036** | 0.017 | **0.030** |
| | **DX101** | 0.034 | 0.752 | 0.032 | 0.769 | 0.009 | 0.933 |
| Lagged two periods | **NDX** | 0.007 | 0.657 | 0.023 | 0.098 | 0.018 | 0.259 |
| | **NPR** | -0.048 | 0.112 | -0.030 | 0.291 | -0.043 | 0.154 |
| | **LOS** | 0.029 | **0.044** | - | - | 0.014 | 0.325 |
| | **DX101** | 0.025 | 0.864 | -0.041 | 0.774 | -0.029 | 0.842 |

*3.5.3 Depression Score Data*

Reisby, et al (1977) examined the relationship between Imipramine (IMI) and Desipramine (DMI) plasma levels and clinical response in 52 depressed inpatients. The study spanned four weeks and focused on changes in Hamilton Depression Scores. Each subject received the same dose of IMI at the end of each week, and measurements were taken on DMI and depression levels (Goetgeluk and Vansteelandt 2008). In addition to the time-dependent covariates IMI and DMI, the study included a time-independent covariate for each subject's gender. Moment conditions identified as valid using the Lalonde, Wilson and Yin (2014) approach are included in Table 3.5.3.1.

Table 3.5.3.1. Moment Conditions for the Depression Score Study

| | Gender | | | | IMI | | | | DMI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **s=1** | **s=2** | **s=3** | **s=4** | **s=1** | **s=2** | **s=3** | **s=4** | **s=1** | **s=2** | **s=3** | **s=4** |
| **t=1** | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **t=2** | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| **t=3** | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| **t=4** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

The one parameter models, LS-GMM, LWY-GMM and GEE-IND, produce different results, with each approach identifying one predictor as significant. Gender was not significant in any of the models, though the signs of the coefficient under each model often disagreed. These results are given in Table 3.5.3.2.

Table 3.5.3.2. Cross-sectional Parameter Estimates and p-Values for the Depression Score Study

| | LS-GMM | | LWY-GMM | | GEE-IND | |
|---|---|---|---|---|---|---|
| | **Est.** | **p-Value** | **Est.** | **p-Value** | **Est.** | **p-Value** |
| **Intercept** | 4.881 | **<.0001** | 5.661 | **0.0366** | 5.774 | 0.1485 |
| **SEX** | 0.046 | 0.9167 | -1.323 | 0.2279 | -0.549 | 0.7615 |
| **IMI** | -2.862 | **<.0001** | -0.568 | 0.1528 | -0.945 | 0.2945 |
| **DMI** | 0.094 | 0.0076 | -1.633 | **0.0028** | -2.064 | **0.0008** |

The Lalonde, Wilson and Yin(2014) method identifies all moment conditions for the time-dependent covariates as valid, thus the results of the Partitioned-LS and Partitioned-LWY models are identical. For both partitioned approaches, IMI and DMI are significant in predicting depression scores at a one time-period lag. Thus, the results of the cross-sectional models vary from the partitioned approaches. We also see that the results of the Lagged-GEE vary from the two partitioned methods. Under the Lagged-GEE, only DMI at a two time-period lag is identified as significant. Table 3.5.3.3 presents the results for the partitioned-LS, Partitioned-LWY and Lagged-GEE models.

Table 3.5.3.3. Partitioned Parameter Estimates and p-Values for the Depression Score Study

| | | Partitioned-LS | | Partitioned-LWY | | Lagged-GEE | |
|---|---|---|---|---|---|---|---|
| | | **Est.** | **p-Value** | **Est.** | **p-Value** | **Est.** | **p-Value** |
| | Intercept | 7.454 | 0.174 | 7.454 | 0.174 | 4.897 | 0.231 |
| | Gender | -3.040 | 0.277 | -3.040 | 0.277 | -0.502 | 0.780 |
| Cross-sectional | IMI | -0.660 | 0.105 | -0.660 | 0.105 | -0.935 | 0.359 |
| | DMI | -1.749 | 0.174 | -1.749 | 0.174 | -1.337 | 0.070 |
| Lagged one period | IMI | -0.844 | **<.001** | -0.844 | **<.001** | -0.727 | 0.248 |
| | DMI | 0.626 | **<.001** | 0.626 | **<.001** | 0.389 | 0.490 |
| Lagged two periods | IMI | 0.353 | 0.918 | 0.353 | 0.918 | 0.569 | 0.271 |
| | DMI | -1.052 | 0.705 | -1.052 | 0.705 | -1.103 | **0.020** |
| Lagged three periods | IMI | 0.246 | 0.967 | 0.246 | 0.967 | 0.814 | 0.549 |
| | DMI | -0.450 | 0.930 | -0.450 | 0.930 | -1.050 | 0.350 |

*3.5.4 Consequences*

Most models are easily fit when there is independence among the observations. However, the presence of correlation whether among the observations, or induced through future effects of responses and covariates, or from the correlation among covariates impact the efficiency of the estimates through the variance. Thus, it is important to include valid moments in the computation of the estimates and their efficiency. The results from these three examples and the simulation study reveal that identifying the valid moment conditions is essential, especially when one wants to identify the relationships across time. Moreover, it is evident that while identifying the valid moments is essential, the methods used to determine the significance of the covariate are also important. If these valid moments are combined to obtain estimates for a single regression coefficient, then the true relationships may be distorted. Combining

valid moments from different responses in one time-period with covariates in a different time-period mask the individual impact.

Though non-partitioned models may produce the same results as the Partitioned GMM model in the cross-sectional portion of the data, these results are only circumstantial. In fact, the partitioned and non-partitioned methods are utilizing different sets of information, based on the moment conditions. The non-partitioned models use an averaging of all information between the covariate and the response to produce a cross-sectional estimate, while the partitioned model utilizes only valid moment conditions occurring when the response and covariate are observed in the same time-period. Thus, the non-partitioned model condenses a comparatively larger amount information into a single parameter, and are more likely to present significant results than the partitioned. These non-partitioned models also inherently assume that the relationship between each covariate and the response remains the same over time. Though the lagged-GEE model somewhat overcomes this limitation, the parameter estimation relies on non-existent moment conditions. Thus, the Partitioned GMM models provide more insight into the data by separating the cross-sectional and lagged relationships, while also utilizing valid moment conditions. The grouping of moment conditions based on the time elapsed between observation of the covariate and the response forces the estimation of each regression parameter to rely only on information pertaining to that particular relationship.

Among the Partitioned GMM models, the valid moments are determined based on the method used. The Partitioned-LS approach provides parameter estimates for all lagged relationships based on the covariate classification method. However, that grouping may include some invalid moments. In contrast, the Partitioned–LWY approach utilizes

an individual identification method for each moment condition, and is thus less likely to include invalid moments. Both Partitioned GMM models provide improved understanding regarding the effects of time-dependent covariates on the outcome over time.

### *3.6. Conclusions*

Correlation inherent in repeated measures on subjects present several challenges as compared to the analysis of cross-sectional data. However, the correlation caused by time-dependent covariates introduces an added challenge. Lai and Small (2007), Lalonde, Wilson and Yin (2014), and Zhou, Lefante, Rice and Chen (2014), among others, have presented models addressing the feedback effects due to time-dependent covariates. However, these models do not distinguish between the cross-sectional from the lagged relationships and rather present an overall effect of the covariate on the responses. The Partitioned GMM models separately identifies cross-sectional and lagged effects of the covariates, while also utilizing only valid moment conditions. Thus, the Partitioned GMM provides a more complete description of the complex effects of time-dependent covariates on outcomes.

CHAPTER 4

# SIMULTANEOUS GMM MODELS WITH TIME-DEPENDENT COVARIATES

Kyle M. Irimata, Elsa Vazquez Arreola, Jeffrey R. Wilson

***Abstract***

We propose a simultaneous generalized method of moments (GMM) model for multiple outcomes with partitioned coefficients to account for time-dependent covariates in longitudinal studies. This approach relies on valid moment conditions to ensure efficient parameter estimation, while the partitioned coefficients provide insight into the effect of each covariate on the outcome across time. Using a concatenation of the valid moment conditions, we extend the dimension of the objective function to account for the correlation between the multiple outcomes. This marginal model also has the benefit of avoiding the need for any additional distributional assumptions, as is often used in joint modeling. We apply our approach to simultaneously investigate risk factors of smoking, social alcohol drinking and obesity among adolescents in the United States and provide comparisons to separately fitted models to illustrate the impact of correlation between the outcome variables.

## *4.1. Introduction*

Longitudinal data are common to many disciplines and are useful for investigating how an individual's responses vary as time progresses. As a result of the repeated measurements taken over time, there is inherent correlation between the repeated measurements, as well as between the covariates in one period and the outcome

in a different time. Researchers may also be interested in simultaneous investigation of more than one outcome, which introduces additional correlation between these outcome variables. For example, in the National Longitudinal Study of Adolescent to Adult Health (Add Health) (Harris and Udry 2016), students are followed starting in 1994-1995, and are measured at three later time-periods. Inference may be simultaneously desired on multiple, correlated outcomes, such as smoking and alcohol. The longitudinal nature of the study results in time-dependent covariates such as depression level and physical activity level with both immediate and carry-over effects on either, or both outcomes as time progresses. An appropriate model for this type of data needs to account for all forms of correlation.in order to provide valid inferences. We refer to the separate outcome variables as outcomes, while the repeated measurements on each of those outcomes are referred to as responses. In this paper, we propose a simultaneous generalized method of moments (GMM) model to account for the multiple outcomes as well as the time-dependency inherent to longitudinal data.

### *4.1.1 Time-Dependent Covariates on a Single Response Variable*

The use of marginal models with time-independent covariates on a single response has been well established; however, appropriate models for time-dependent covariates are still emerging and have recently received considerable attention. The generalized estimating equations (GEE) (Zeger and Liang 1986) model, which utilizes a working correlation structure to account for the association between repeated measurements, is a popular type of marginal model for correlated data. However, Hu (1998) and Pepe and Anderson (1994) showed that estimates from GEE models with

arbitrary working correlation structures may lack consistency in the presence of time-dependent covariates, and thus proposed the use of the independent working correlation structure.

Lai and Small (2007) presented generalized method of moments (GMM) as an alternative approach to fitting marginal models with time-dependent covariates, which produced more efficient estimators as compared to GEE with the independent working correlation structure. Their GMM approach utilized a three-type classification scheme for the time-dependent covariates to identify and use estimating equations which are not utilized by the GEE approach with an independent working correlation structure. Zhou, Lefante, Rice, and Chen (2014), based on the classification method introduced by Lai and Small (2007), provided a modified approach for addressing time-dependent covariates using the modified quadratic inference function. Lalonde, Wilson and Yin (2014) extended the GMM model for time-dependent covariates with a method for testing each moment condition separately for validity. This approach allowed the moment conditions to vary, without the need to specify a specific covariate type.

Although the GMM models provide desirable properties, these models inherently assumed that the effect of each covariate on the response was constant over time, which may not be a reasonable assumption in practice. Several models that incorporate lagged coefficients have been proposed to produce separate estimates for the effect of each covariate on the outcome across various time-periods (Muller and Stadtmuller 2005; Selig, Preacher, and Little 2012; Zeger and Liang 1986; Keele and Kelly 2005). Irimata, Broatch and Wilson (2018) introduced a flexible partitioned GMM model to allow the effect of each time-dependent covariate on the outcome to vary over time. This model

was a special case of the functional model discussed by Müller and Stadtmuller (2005). Though this model maintained the desirable properties of GMM models and allowed for additional insight into the effect of each time-dependent covariate, this model was not built to model more than one outcome variable. Thus, we propose an extension of the partitioned GMM, which can simultaneously estimate regression equations for multiple outcomes while accounting the correlation between these outcome variables, without the need to assume a multivariate distribution.

### *4.1.2 Simultaneous Models with Distributional Assumptions*

Multivariate data are common in many disciplines. Separate models for longitudinal data can be fit to address the multiple response variable; however, statistical inferences based on separate analyses ignore the correlation between the multiple outcomes, leading to potentially inefficient parameter estimates (Berridge and Crouchley 2011; Fitzmaurice, Laird, and Ware 2004). Thus, the effect of a covariate is not accurately captured, if we ignore the interdependence of the multiple outcomes.

The multivariate longitudinal modeling (Bandyopadhyay, Ganguli, and Chatterjee 2011; Fieuws and Verbeke 2004; Fieuws, Verbeke, and Molenberghs 2007) is one approach for addressing such limitations. These approaches account for the interdependence in the multivariate outcomes through a multivariate distribution thereby improving efficiency and reduce bias (Gueorguieva 2001; McCulloch 2008). However, because these approaches require a known distribution, they rely on certain assumptions and inference may not be valid if the distribution is misspecified (Wu, et al. 2012). The challenges with separate models can be also be lessened with joint models (Liu, et al.

2010; Maciejewski and Maynard 2004; Xu and Zeger 2001). For instance, the multiple models are often linked through the random effect to account for the association between the responses (Ghebremichael 2015; Liu, et al. 2010). Song, Davidian and Tsiatis (2002) proposed an extension of this approach using a semiparametric approach which required only that the random effects have a smooth density. The use of joint latent class models has also been proposed for longitudinal data (Proust-Lima, et al. 2014). In econometrics, multiple-equation GMM has been used to fit multiple GMM models simultaneously (Hayashi 2000). However, these approaches are not applicable to analyses including time-dependent covariates, and often rely on additional distributional assumptions.

We propose the use of a joint model; however, unlike previous studies, this approach does not require any distributional assumptions, and includes time-varying effects of each covariate using a partitioning of regression parameters. In Section 2, we review correlated responses and generalized method of moments estimators. In Section 3, we introduce a simultaneous GMM approach with partitioned coefficients. In Section 4, the proposed model is applied to the Add Health study. In Section 5, we provide discussions and conclusions.

### *4.2. Generalized Method of Moments Models*

The GMM estimator is used frequently in econometrics (Hansen 1982) and is increasingly useful in statistical modeling. Many researchers have demonstrated the efficacy of the GMM model in addressing time-dependent covariates using a single regression parameter for each covariate (Chen and Westgate 2017; Lai and Small 2007; Lalonde, Wilson, and Yin 2014). Irimata, Broatch and Wilson (2018) utilized a

partitioned GMM model to evaluate time-dependent covariates using multiple regression coefficients per covariate.

Consider a study with repeated observations over $T$ time-periods on $N$ subjects with $J$ covariates. Let $y_{it}$ denote the outcome measured on the $i^{th}$ subject at the $t^{th}$ time-period with marginal density, given the time-varying vector of covariates $\boldsymbol{x}_{it.}$, following a generalized linear model. The observations $y_{is}$ and $y_{kt}$ are assumed to be independent when $i \neq k$, but may be correlated when $i = k$. Denote the conditional expectation of the random variable $y_{it}$ given the vector of covariates $\boldsymbol{x}_{it.}$ with the vector of regression parameters $\boldsymbol{\beta}$ as $\mu_{it}(\boldsymbol{\beta})$. We utilize the moment condition

$$E\left[\frac{\partial \mu_{is}(\boldsymbol{\beta})}{\partial \beta_j}\{y_{it} - \mu_{it}(\boldsymbol{\beta})\}\right] = 0, \qquad (4.2.1)$$

as discussed by Lai and Small (Lai and Small 2007), for appropriately selected $s, t$ and $j$.

### *4.2.1 Single Parameter GMM*

The generalized method of moments models using a single parameter for each covariate are now common (Chen and Westgate 2017; Qu, Lindsay, and Li 2000; Zhou, et al. 2014; Lai and Small 2007). In this paper, we concentrate on the framework established by Lalonde, Wilson and Yin (2014), though the methods discussed in later sections can be extended to accommodate alternative approaches.

Lai and Small (2007) introduced a marginal model using GMM estimation for time-dependent covariates. Their method relied on the classification of each covariate as either Type I, Type II or Type III, which in turn provided information regarding the moment conditions, which could be used to estimate the regression parameter. A

covariate was designated as Type I if Equation (4.2.1) held for all $s$ and all $t$, and thus all moment conditions were considered valid. A Type II covariate satisfied Equation (4.2.1) whenever $s \geq t$, but failed for some $s < t$, with a common example including covariates are not independent of future responses, but without any type of feedback effect from the response back onto the covariate. A covariate was classified as Type III when Equation (4.2.1) does not hold for any $s > t$, and often occur when there is a feedback process from the outcome onto the covariate.

Lalonde, Wilson and Yin (2014) discussed extensions to the GMM model for time-dependent covariates, using a hypothesis test to separately evaluate each moment condition for validity. The test allowed for all valid moment conditions to be utilized in model estimation, without requiring that each covariate be classified by type. They showed that the validity of Equation (4.2.1) is equivalent to the correlation between the residual, $e_s$ observed at time $s$, and the covariate, $x_t$ observed at time t, being equal to zero. They assumed that all moment conditions arising from $s = t$ were are valid. All other $T(T-1)$ moment conditions were tested for validity using the hypothesis $H_0$: $\rho_{x_t,e_s} = 0$; thus the moment conditions were considered valid under the null hypothesis. This hypothesis was tested using a Z-statistic using the pairwise correlations between the residual and covariate at each time-period. The identified valid moment conditions were used to fit the regression model.

### *4.2.2 Partitioned GMM*

Though single parameter per covariate GMM models are useful for modeling time-dependent covariates, these models limit the effect of each covariate to be the same

in each time-period, which may not be an ideal assumption in practice. Irimata, Broatch and Wilson (2018) proposed a partitioned GMM model which utilized multiple parameters for each time-dependent covariate to separately represent the varying relationships between each covariate and the responses over time. They utilized a partitioning method, based on the difference in time-period between the covariate and response. The moment conditions were grouped based on this partitioning, which led to the lower diagonal design matrix

$$\mathbf{X}_{\mathrm{ij}} = \begin{bmatrix} 1 & \mathrm{X_{ij1}} & 0 & \dots & 0 \\ 1 & \mathrm{X_{ij2}} & \mathrm{X_{ij1}} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \mathrm{X_{i1T}} & \mathrm{X_{ij(T-1)}} & \dots & \mathrm{X_{ij1}} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathrm{X_{ij}^{[0]}} & \mathrm{X_{ij}^{[1]}} & \dots & \mathrm{X_{ij}^{[T-1]}} \end{bmatrix}$$

which was used in the marginal model

$$g(\mu_{it}) = \beta_0 + \beta_j^{tt} X_{ij}^{[0]} + \beta_j^{[1]} X_{ij}^{[1]} + \beta_j^{[2]} X_{ij}^{[2]} \dots + \beta_j^{[T-1]} X_{ij}^{[T-1]} \quad (4.2.2)$$

for subject $i$, with appropriate link function $g(\cdot)$. The coefficient $\beta_j^{tt}$ denoted the cross-sectional effect of the $j^{\text{th}}$ covariate on the outcome observed in the same time-period. The parameters $\beta_j^{[1]}, \dots \beta_j^{[T-1]}$ denoted the effect of the $j^{\text{th}}$ covariate on the outcome observed across a 1 to $T-1$ time-period lag, respectively. For example, $\beta_j^{[1]}$ represented the relationship between $X_j$ and the outcome across a one time-period lag. They identified valid moment conditions using either the Type II covariate proposed by Lai and Small (2007), or the test for validity introduced by Lalonde, Wilson and Yin (2014), though their simulation studies suggested that the Lalonde, Wilson and Yin method produces less biased estimates, with better parameter coverage. This model was best suited to data with

few time-periods, relative to the number of subjects, similar to restrictions with GEE models (Stoner, Leroux, and Puumala 2010). We expand on this framework to consider multiple simultaneous responses of interest with a joint model fitting procedure.

### *4.3 Joint Modeling for Correlated Binary Responses*

Though models, such as those based on GMM, are useful for accounting for time-dependent covariates, the separate fit of these models to multiple outcomes ignores the correlation between these multiple outcome variables. The use of separate models leads to inefficient estimates and the results of the analysis (Berridge and Crouchley 2011; Fitzmaurice, Laird, and Ware 2004). The multiple-equation GMM model accounts for the fit of more than one outcome (Hayashi 2000), however the model does not account for time-dependent covariates in longitudinal data. In this section, we present an expanded objective function to estimate regression parameters based on the partitioned GMM model (Irimata, Broatch, and Wilson 2018), for modeling multivariate outcomes using simultaneous model fitting, while also accounting for the time-dependent covariates.

#### *4.3.1 Model and Estimators*

Consider a regression model with time-dependent covariates, and the method of Lalonde, Wilson and Yin (2014) to identify the set of valid moment conditions, though other methods for valid moment identification can be used. Let M denote the number of response variables of interest, where $\boldsymbol{Y}_{(m)i} = \left(y_{(m)i1}, \dots, y_{(m)iT}\right)'$ is the $T\ x\ 1$ vector of outcomes associated with the $i^{\text{th}}$ subject $i = 1, 2, \dots, N$; for the $m^{\text{th}}$ outcome variable and let $\boldsymbol{\mu}_{(m)i} = (\mu_{(m)i1}, \dots, \mu_{(m)iT})'$ denote the corresponding mean vector. Let $\boldsymbol{X}_{(m)i} =$

$\begin{bmatrix} 1 & X_{(m)i11} & \cdots & X_{(m)iJ1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{(m)i1T} & \cdots & X_{(m)iJT} \end{bmatrix}$ denote the matrix of $J$ covariates for model $m$, where at time $t$, the row vector is given by $\mathbf{X}_{(\mathrm{m})\mathrm{i.t}} = \left(\mathrm{X}_{(\mathrm{m})\mathrm{i1t}}, \ldots, \mathrm{X}_{(\mathrm{m})\mathrm{iJt}}\right)$ and the column vector for the $j^{\text{th}}$ covariate is $\boldsymbol{X}_{(m)ij.} = \left(X_{(m)ij1}, \ldots, X_{(m)ijT}\right)'$. To account for the time-dependent covariates and to evaluate lagged effects, we utilize a lower triangular data matrix, similar to the method discussed by Irimata, Broatch and Wilson (2018), such that the data matrix for the $j^{\text{th}}$ covariate in the $m^{\text{th}}$ model is

$$\mathbf{X}'_{(\mathrm{m})\mathrm{ij}} = \begin{bmatrix} 1 & \mathrm{X}_{(\mathrm{m})\mathrm{ij1}} & 0 & \ldots & 0 \\ 1 & \mathrm{X}_{(\mathrm{m})\mathrm{ij2}} & \mathrm{X}_{(\mathrm{m})\mathrm{ij1}} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & \mathrm{X}_{(\mathrm{m})\mathrm{i1T}} & \mathrm{X}_{(\mathrm{m})\mathrm{ij}(\mathrm{T}-1)} & \ldots & \mathrm{X}_{(\mathrm{m})\mathrm{ij1}} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathrm{X}^{[0]}_{(\mathrm{m})\mathrm{ij}} & \mathrm{X}^{[1]}_{(\mathrm{m})\mathrm{ij}} & \ldots & \mathrm{X}^{[\mathrm{T}-1]}_{(\mathrm{m})\mathrm{ij}} \end{bmatrix}$$

where the bracketed superscripts denote the differences in observed time-periods between the $m^{\text{th}}$ outcome and the covariate. Thus, the model is

$$g\left(\mu_{(m)it}\right) = \beta_{(m)0} + \beta^{tt}_{(m)j} X^{[0]}_{(m)ij} + \beta^{[1]}_{(m)j} X^{[1]}_{(m)ij} + \beta^{[2]}_{(m)j} X^{[2]}_{(m)ij} \ldots + \beta^{[T-1]}_{(m)j} X^{[T-1]}_{(m)ij}$$

for appropriate link function $g(\cdot)$, and $g\left(\boldsymbol{\mu}_{(\boldsymbol{m})\boldsymbol{i}}\right) = \boldsymbol{X}_{(\boldsymbol{m})\boldsymbol{ij}}\boldsymbol{\beta}_{(\boldsymbol{m})\boldsymbol{j}}$ in matrix form. The interpretation of each of the regression coefficients is analogous to Model (4.2.2). Thus, $\beta^{tt}_{(m)j}$ denotes the cross-sectional effect of the $j^{\text{th}}$ covariate in the $m^{\text{th}}$ regression model, while $\beta^{[1]}_{(\mathrm{m})\mathrm{j}} \ldots \beta^{[\mathrm{T}-1]}_{(\mathrm{m})\mathrm{j}}$ denote the lagged effects across a $1, \ldots, T-1$ time-period lag.

Let $M = 2$, with two simultaneous outcomes, which for subject $i$ are the vectors $\boldsymbol{Y}_{(\boldsymbol{1})\boldsymbol{i}}$ and $\boldsymbol{Y}_{(\boldsymbol{2})\boldsymbol{i}}$, though the approach is analogous for larger M (Hayashi 2000). Let $\boldsymbol{\beta}_{(\boldsymbol{1})}$ and $\boldsymbol{\beta}_{(\boldsymbol{2})}$ be vectors of unknown regression parameters corresponding to the functions

$\boldsymbol{f}(y_{(1)i}, \boldsymbol{\beta}_{(1)}) = \begin{bmatrix} f_1(y_{(1)i}, \boldsymbol{\beta}_{(1)}) \\ \vdots \\ f_{G_1}(y_{(1)i}, \boldsymbol{\beta}_{(1)}) \end{bmatrix}$ and $\boldsymbol{h}(y_{(2)i}, \boldsymbol{\beta}_{(2)}) = \begin{bmatrix} h_1(y_{(2)i}, \boldsymbol{\beta}_{(2)}) \\ \vdots \\ h_{G_2}(y_{(2)i}, \boldsymbol{\beta}_{(2)}) \end{bmatrix}$ for every

subject i, where the number of moment conditions, $G_1$ and $G_2$ may not be equal. Then there are $G = \sum_{m=1}^{M} G_m$ moment conditions which take the form $\boldsymbol{f}(y_{(1)i}, \boldsymbol{\beta}_{(1)})$ and $\boldsymbol{h}(y_{(2)i}, \boldsymbol{\beta}_{(2)})$, such that

$$E\begin{bmatrix} \boldsymbol{f}(y_{(1)i}, \boldsymbol{\beta}_{(1)}) \\ \boldsymbol{h}(y_{(2)i}, \boldsymbol{\beta}_{(2)}) \end{bmatrix} = E\begin{bmatrix} f_1\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \\ \vdots \\ f_{G_1}\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \\ h_1\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \\ \vdots \\ h_{G_2}\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \end{bmatrix} = \begin{bmatrix} E\left(f_1\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right)\right) \\ \vdots \\ E\left(f_{G_1}\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right)\right) \\ E\left(h_1\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right)\right) \\ \vdots \\ E\left(h_{G_2}\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right)\right) \end{bmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

for all $i$. We define the sample analogue of this moment condition across all subjects as

$$\begin{matrix} \boldsymbol{F_N}(y_{(1)}, \boldsymbol{\beta}_{(1)}) \\ \boldsymbol{H_N}(y_{(2)}, \boldsymbol{\beta}_{(2)}) \end{matrix} = \begin{matrix} \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} f_1\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \\ \vdots \\ f_{G_1}\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \end{bmatrix} \\ \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} h_1\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \\ \vdots \\ h_{G_2}\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \end{bmatrix} \end{matrix} = \frac{1}{N}\begin{bmatrix} \sum f_1\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \\ \vdots \\ \sum f_{G_1}\left(y_{(1)i'}, \boldsymbol{\beta}_{(1)}\right) \\ \sum h_1\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \\ \vdots \\ \sum h_{G_2}\left(y_{(2)i'}, \boldsymbol{\beta}_{(2)}\right) \end{bmatrix}$$

Define the $G \times G$ positive definite weight matrix as $W_N = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix}$, which is computed using the cross product of the moment conditions for each subject. This matrix accounts for the covariance that exists between the two models through the inclusion of the elements in $W_{12}$. Thus,

$$\mathrm{W_N} = \frac{1}{\mathrm{N}}\left[\sum_{\mathrm{i=1}}^{\mathrm{N}}\begin{pmatrix} \mathbf{f}(\mathrm{y_{(1)i}}, \boldsymbol{\beta}_{(1)}) \\ \mathbf{h}(\mathrm{y_{(2)i}}, \boldsymbol{\beta}_{(2)}) \end{pmatrix}\begin{pmatrix} \mathbf{f}(\mathrm{y_{(1)i}}, \boldsymbol{\beta}_{(1)}) \\ \mathbf{h}(\mathrm{y_{(2)i}}, \boldsymbol{\beta}_{(2)}) \end{pmatrix}^{\mathrm{T}}\right].$$

Define the objective function such that:

$$\mathrm{Q_N}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{F_N}\left(\mathrm{y}_{(1)}, \boldsymbol{\beta}_{(1)}\right) \\ \mathbf{H_N}\left(\mathrm{y}_{(2)}, \boldsymbol{\beta}_{(2)}\right) \end{pmatrix}^{\mathrm{T}} \mathrm{W_N^{-1}} \begin{pmatrix} \mathbf{F_N}\left(\mathrm{y}_{(1)}, \boldsymbol{\beta}_{(1)}\right) \\ \mathbf{H_N}\left(\mathrm{y}_{(2)}, \boldsymbol{\beta}_{(2)}\right) \end{pmatrix}.$$

The GMM estimator of $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{(1)} \\ \boldsymbol{\beta}_{(2)} \end{pmatrix}$ is the vector of regression parameters that minimizes $Q_N(\boldsymbol{\beta})$, thus

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GMM}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\mathbf{GMM}(1)} \\ \widehat{\boldsymbol{\beta}}_{\mathbf{GMM}(2)} \end{pmatrix} = \mathrm{argmin}_{\beta \epsilon \mathcal{B}} \mathrm{Q_N}(\boldsymbol{\beta}).$$

This minimum is obtained using nonlinear optimization, such as Newton-Raphson or conjugate gradient method. The estimates are obtained using a continuously updating procedure in which each successive estimate of $\boldsymbol{\beta}$ is obtained with a weight matrix calculated using the estimate for $\boldsymbol{\beta}$ from the previous iteration.

In calculating the asymptotic variance, let

$$\hat{A} = \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \dfrac{\partial f\left(y_{(1)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM}(1)}\right)}{\partial \boldsymbol{\beta}_{(1)}} \\ \dfrac{\partial h\left(y_{(2)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM}(2)}\right)}{\partial \boldsymbol{\beta}_{(2)}} \end{pmatrix}$$

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

By extension, for the general case $m = M$,

$$W_N = \frac{1}{N} \left[ \sum_{i=1}^{N} \begin{pmatrix} \boldsymbol{f}\left(y_{(1)i}, \boldsymbol{\beta}_{(1)}\right) \\ \boldsymbol{h}\left(y_{(2)i}, \boldsymbol{\beta}_{(2)}\right) \\ \vdots \\ \boldsymbol{k}\left(y_{(m)i}, \boldsymbol{\beta}_{(\boldsymbol{M})}\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{f}\left(y_{(1)i}, \boldsymbol{\beta}_{(1)}\right) \\ \boldsymbol{h}\left(y_{(2)i}, \boldsymbol{\beta}_{(2)}\right) \\ \vdots \\ \boldsymbol{k}\left(y_{(m)i}, \boldsymbol{\beta}_{(\boldsymbol{M})}\right) \end{pmatrix}^{T} \right]$$

such that

$$Q_N(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{F_N}(y_{(1)}, \boldsymbol{\beta_{(1)}}) \\ \boldsymbol{H_N}(y_{(2)}, \boldsymbol{\beta_{(2)}}) \\ \vdots \\ \boldsymbol{K_N}(y_{(m)}, \boldsymbol{\beta_{(M)}}) \end{pmatrix}^T W_N^{-1} \begin{pmatrix} \boldsymbol{F_N}(y_{(1)}, \boldsymbol{\beta_{(1)}}) \\ \boldsymbol{H_N}(y_{(2)}, \boldsymbol{\beta_{(2)}}) \\ \vdots \\ \boldsymbol{K_N}(y_{(m)}, \boldsymbol{\beta_{(M)}}) \end{pmatrix}$$

and thus $\widehat{\boldsymbol{\beta}}_{GMM} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM(1)}} \\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM(2)}} \\ \vdots \\ \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM(M)}} \end{pmatrix} = argmin_{\beta \epsilon \mathcal{B}} Q_N(\boldsymbol{\beta}).$

For logistic regression models on the $m^{\text{th}}$ outcome variable, the mean response is

$$\mu_{(m)it}(\boldsymbol{\beta_{(m)}}) = \frac{exp(\boldsymbol{x}_{(m)it.}\boldsymbol{\beta_{(m)}})}{1 + exp(\boldsymbol{x}_{(m)it.}\boldsymbol{\beta_{(m)}})}$$

and thus, each element in $\boldsymbol{f}(y_{(1)i}, \boldsymbol{\beta_{(1)}})$ *or* $\boldsymbol{h}(y_{(2)i}, \boldsymbol{\beta_{(2)}})$ is

$$\frac{\partial \mu_{(m)is}(\boldsymbol{\beta_{(m)}})}{\partial \beta_{(m)j}^{[k]}}[y_{(m)it} - \mu_{(m)it}(\boldsymbol{\beta_{(m)}})]$$

$$= x_{(m)isj}\mu_{(m)is}(\boldsymbol{\beta_{(m)}})[1 - \mu_{(m)is}(\boldsymbol{\beta_{(m)}})][y_{(m)it} - \mu_{(m)it}(\boldsymbol{\beta_{(m)}})]$$

Therefore, each row of $\begin{pmatrix} \frac{\partial f(y_{(1)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM(1)}})}{\partial \boldsymbol{\beta_{(1)}}} \\ \frac{\partial h(y_{(2)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM(2)}})}{\partial \boldsymbol{\beta_{(2)}}} \end{pmatrix}$ in the asymptotic variance of a logistic regression model is computed using

$$\frac{\partial \left\{ \left[ \frac{\partial \mu_{(m)is}(\boldsymbol{\beta_{(m)}})}{\partial \beta_{(m)j}^{[k]}} \right] [y_{(m)it} - \mu_{(m)it}(\boldsymbol{\beta_{(m)}})] \right\}}{\partial \beta_{(m)l}^{[k']}} = x_{(m)isj}\mu_{(m)is}(\boldsymbol{\beta_{(m)}})[1 - \mu_{(m)is}(\boldsymbol{\beta_{(m)}})] \{x_{(m)isl}[1 -$$

$$2\mu_{(m)is}(\boldsymbol{\beta_{(m)}})][y_{(m)it} - \mu_{(m)it}(\boldsymbol{\beta_{(m)}})] - x_{(m)itj}\mu_{(m)it}(\boldsymbol{\beta_{(m)}})[1 - \mu_{(m)it}(\boldsymbol{\beta_{(m)}})]\},$$

where $j = 1, \ldots, \mathrm{J}$, $l = 1, \ldots J$, $k = 1, \ldots, T - 1$ and $k' = 1, \ldots, T - 1$.

Analogously, for normal error models on the $m^{\text{th}}$ outcome variable, the elements in $\boldsymbol{f}\left(y_{(1)i}, \boldsymbol{\beta}_{(\mathbf{1})}\right)$ *or* $\boldsymbol{h}\left(y_{(2)i}, \boldsymbol{\beta}_{(\mathbf{2})}\right)$ are

$$\frac{\partial \mu_{(m)is}\left(\boldsymbol{\beta}_{(\boldsymbol{m})}\right)}{\partial \beta_{(m)j}^{[k]}}\left[y_{(m)it} - \mu_{(m)it}\left(\boldsymbol{\beta}_{(\boldsymbol{m})}\right)\right] = x_{(m)isj}\left[y_{(m)it} - \mu_{(m)it}\left(\boldsymbol{\beta}_{(\boldsymbol{m})}\right)\right]$$

and thus, in calculating the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM}}$, each row of $\begin{pmatrix} \frac{\partial f\left(y_{(1)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM}(\mathbf{1})}\right)}{\partial \boldsymbol{\beta}_{(\mathbf{1})}} \\ \frac{\partial h\left(y_{(2)i}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{GMM}(\mathbf{2})}\right)}{\partial \boldsymbol{\beta}_{(\mathbf{2})}} \end{pmatrix}$ is computed as

$$\frac{\partial\left\{\left[\frac{\partial \mu_{(m)is}\left(\boldsymbol{\beta}_{(\boldsymbol{m})}\right)}{\partial \beta_{(m)j}^{[k]}}\right]\left[y_{(m)it} - \mu_{(m)it}\left(\boldsymbol{\beta}_{(\boldsymbol{m})}\right)\right]\right\}}{\partial \beta_{(m)l}^{[k']}} = -x_{(m)isj}x_{(m)isl}$$

for $j = 1, \ldots, \mathrm{J}$, $l = 1, \ldots J$, $k = 1, \ldots, T-1$ and $k' = 1, \ldots, T-1$. We provide a SAS macro to fit this model at https://github.com/kirimata/Simultaneous-GMM.

### *4.4 Numerical Example*

Alcohol use and smoking have long been identified as strongly related (Guydish, et al. 2011; Kozlowski, Jelinek, and Pope 1986). Recent studies, such as the National Comorbidity Study have reported smoking prevalence of 56.1% for Americans with alcohol disorders (Lasser, et al. 2000). We analyzed data from the National Longitudinal Study of Adolescent Health (Add Health) (Harris and Udry 2016). These data come from a longitudinal study of health-related behaviors in adolescents. The measurements were collected from 80 different high schools and 52 middle schools in the U.S., with information on 2,712 different students gathered across four waves starting in 1994 and ending in 2008. We investigated two binary outcomes representing smoking status and

social alcohol drinking at each measurement. The data included time-dependent covariates for physical activity level, depression level and self-reported health status. We also included a time-independent covariate, representing race as white or non-white. For further comparison, we considered a third binary outcome representing obesity status, as calculated from each child's BMI.

*4.4.1 Simultaneous Modeling of Smoking and Alcohol Use*

We considered two outcome variables for smoking status and social alcohol drinking, thus for this analysis $M = 2$. There was substantial association between these two outcome variables, with correlation $\hat{\phi} = 0.40$, and with $V^2 = \chi^2/N = 0.161$. Thus, since $V^2$ exceeded the 0.15 threshold discussed by Irimata and Wilson (2017), there was significant correlation between these outcomes, warranting consideration in the model. We fitted a simultaneous GMM model to estimate the regression for the two outcome variables simultaneously. In addition we consider two partitioned GMM models separately, using the Lalonde, Wilson and Yin (2014) method to identify valid moment conditions. There were differences in the significance of the covariates between these two models. In particular, for the social alcohol usage model, we saw that depression had a significant one-period lagged coefficient when two separate models were fit. However, this relationship was not significant when the correlation between smoking and social alcohol usage was accounted for through the simultaneous GMM. We found that depression level was significant at a two time-period lag under the simultaneous GMM model, though the two separate models did not identify this relationship as significant. Thus, strong correlation between smoking status and social alcohol drinking affects the

conclusions from a binary logistic regression model. In particular, the fit of two separate models is not the best choice in this scenario as it neglects the association between the outcomes, consistent with findings based on the correlation of the two outcomes (Irimata and Wilson 2017). The parameter estimates and standard errors from the Simultaneous GMM model, as well as the two separate Partitioned GMM models are included in Table 4.4.1. Significant parameter estimates are denoted by bolded entries.

Table 4.4.1. Parameter Estimates and Standard Errors (SE) for the Smoking and Social Alcohol Models in Add Health Data

| Outcome | Time-Period | Parameter | Simultaneous GMM | | Separate Models | |
|---|---|---|---|---|---|---|
| | | | Estimate | SE | Estimate | SE |
| Smoking Status | Cross-Sectional | **Intercept** | **0.849** | **0.152** | **0.824** | **0.155** |
| | | **Race** | **-0.757** | **0.068** | **-0.716** | **0.068** |
| | | **Activity** | -0.023 | 0.022 | -0.019 | 0.022 |
| | | **Depression** | **0.806** | **0.086** | **0.811** | **0.088** |
| | | **Health** | **-0.259** | **0.030** | **-0.260** | **0.030** |
| | Lagged One Period | **Activity** | -0.002 | 0.022 | 0.007 | 0.022 |
| | | **Depression** | 0.002 | 0.089 | 0.018 | 0.091 |
| | | **Health** | **-0.089** | **0.023** | **-0.101** | **0.024** |
| | Lagged Two Periods | **Activity** | **0.063** | **0.023** | **0.063** | **0.024** |
| | | **Depression** | **1.154** | **0.101** | **1.171** | **0.101** |
| | Lagged Three Periods | **Activity** | -0.002 | 0.028 | 0.004 | 0.028 |
| | | **Depression** | **-0.587** | **0.094** | **-0.606** | **0.094** |
| Social Alcohol Use | Cross-Sectional | **Intercept** | 0.208 | 0.156 | 0.207 | 0.159 |
| | | **Race** | **-0.562** | **0.065** | **-0.555** | **0.066** |
| | | **Activity** | **-0.067** | **0.023** | **-0.071** | **0.023** |
| | | **Depression** | **0.864** | **0.101** | **0.877** | **0.103** |
| | | **Health** | **-0.110** | **0.029** | **-0.110** | **0.030** |
| | Lagged One Period | **Activity** | 0.009 | 0.023 | -0.002 | 0.023 |
| | | **Depression** | -0.207 | 0.114 | **-0.257** | **0.113** |
| | | **Health** | 0.004 | 0.024 | 0.016 | 0.024 |
| | Lagged Two Periods | **Activity** | 0.039 | 0.032 | 0.033 | 0.033 |
| | | **Depression** | **0.384** | **0.167** | 0.253 | 0.155 |
| | | **Health** | **0.199** | **0.039** | **0.228** | **0.039** |
| | Lagged Three Periods | **Depression** | -0.204 | 0.186 | -0.313 | 0.177 |
| | | **Health** | **0.134** | **0.040** | **0.164** | **0.041** |

*4.4.2 Simultaneous Modeling of Smoking, Alcohol Use, and Obesity*

We investigated obesity status in the children, while considering smoking and social alcohol usage. We calculated the partial correlations among the three outcome variables. Conditional on obesity, smoking and social alcohol drinking remained strongly correlated at a level of 0.399. Neither smoking nor social alcohol drinking had a strong

correlation with obesity, with $V^2 = 0.0041$ and $V^2 = 0.0046$, respectively (Irimata and Wilson 2017). Thus, we expect that the effect of simultaneously modeling obesity with smoking and social alcohol use is small, given the weak correlation. The partial correlations are provided in Table 4.4.2.

Table 4.4.2. Partial Correlations / $V^2$ Between Smoking, Social Alcohol Usage and Obesity in Add Health Data

| | **Smoking** | **Alcohol** | **Obesity** |
|---|---|---|---|
| **Smoking** | 1 | 0.399 / 0.161 | 0.040 / 0.0041 |
| **Alcohol** | 0.399 / 0.161 | 1 | 0.046 / 0.0046 |
| **Obesity** | 0.040 / 0.0041 | 0.046 / 0.0046 | 1 |

We fitted the Simultaneous GMM for the three outcome variables. We compared the results obtained to the results from three separate Partitioned GMM models. Similar to the model which focused only on smoking and social alcohol usage, the results of these two approaches differed. Under the separate Partitioned GMM models, depression was significant at a one time-period lag in the social alcohol use model, though this was not significant under the simultaneous GMM. The simultaneous GMM identified depression as significant at a two time-period lag in the social alcohol use model, though the fit of separate Partitioned GMM models. Notably, the significance of the predictors in the obesity model did not vary between the two analyses, and the previously discussed discrepancies are the same as those identified Section 4.1. Thus, the correlation between smoking and social alcohol drinking continued to produce differences due to the magnitude of association between these outcomes. However, the inclusion of obesity had very little effect on the results, as implied by the small amount of correlation between

obesity and the other two outcomes. The respective parameter estimates and standard errors from the Simultaneous GMM model and the three separate Partitioned GMM models are given in Table 4.4.3, with significant regression parameters bolded.

Table 4.4.3. Parameter Estimates and Standard Errors (SE) for the Smoking, Social Alcohol and Obesity Models in Add Health Data

| Outcome | Time-Period | Parameter | Simultaneous GMM | | Separate Models | |
|---|---|---|---|---|---|---|
| | | | Estimate | SE | Estimate | SE |
| Smoking Status | Cross-Sectional | **Intercept** | **0.823** | **0.148** | **0.824** | **0.155** |
| | | **Race** | **-0.791** | **0.067** | **-0.716** | **0.068** |
| | | **Activity** | -0.021 | 0.021 | -0.019 | 0.022 |
| | | **Depression** | **0.687** | **0.082** | **0.811** | **0.088** |
| | | **Health** | **-0.240** | **0.029** | **-0.260** | **0.030** |
| | Lagged One Period | **Activity** | 0.001 | 0.022 | 0.007 | 0.022 |
| | | **Depression** | -0.051 | 0.086 | 0.018 | 0.091 |
| | | **Health** | **-0.076** | **0.023** | **-0.101** | **0.024** |
| | Lagged Two Periods | **Activity** | **0.075** | **0.023** | **0.063** | **0.024** |
| | | **Depression** | **1.181** | **0.098** | **1.171** | **0.101** |
| | Lagged Three Periods | **Activity** | 0.032 | 0.028 | 0.004 | 0.028 |
| | | **Depression** | **-0.704** | **0.092** | **-0.606** | **0.094** |
| Social Alcohol Use | Cross-Sectional | **Intercept** | 0.271 | 0.154 | 0.207 | 0.159 |
| | | **Race** | **-0.586** | **0.065** | **-0.555** | **0.066** |
| | | **Activity** | **-0.062** | **0.023** | **-0.071** | **0.023** |
| | | **Depression** | **0.794** | **0.097** | **0.877** | **0.103** |
| | | **Health** | **-0.120** | **0.029** | **-0.110** | **0.030** |
| | Lagged One Period | **Activity** | 0.035 | 0.023 | -0.002 | 0.023 |
| | | **Depression** | -0.209 | 0.111 | **-0.257** | **0.113** |
| | | **Health** | -0.008 | 0.023 | 0.016 | 0.024 |
| | Lagged Two Periods | **Activity** | 0.032 | 0.032 | 0.033 | 0.033 |
| | | **Depression** | **0.331** | **0.156** | 0.253 | 0.155 |
| | | **Health** | **0.217** | **0.038** | **0.228** | **0.039** |
| | Lagged Three Periods | **Depression** | -0.244 | 0.179 | -0.313 | 0.177 |
| | | **Health** | **0.154** | **0.040** | **0.164** | **0.041** |
| Obesity | Cross-Sectional | **Intercept** | -0.252 | 0.230 | -0.104 | 0.228 |
| | | **Race** | **0.415** | **0.103** | **0.361** | **0.103** |
| | | **Activity** | **-0.567** | **0.054** | **-0.566** | **0.054** |
| | | **Depression** | **0.562** | **0.118** | **0.589** | **0.117** |
| | | **Health** | **-0.526** | **0.047** | **-0.559** | **0.047** |
| | Lagged One Period | **Depression** | **1.553** | **0.146** | **1.469** | **0.143** |
| | | **Health** | **-0.201** | **0.042** | **-0.183** | **0.041** |

### *4.5 Conclusions*

In longitudinal data analysis, there are multiple types of correlation that must be accounted for. There is correlation due to the repeated measures, and correlation between the covariates and the response over time. When there are multiple outcome variables of interest, the association between these outcomes must also be taken into account. The Simultaneous GMM model provides an alternative to joint modeling of logistic regression models with time-dependent covariates that avoids the need for additional distribution assumptions. This model accounts for the time-dependent covariates using partitioned coefficients to represent the effect of each predictor in different segments of time. The model also accounts for the correlation between the multiple outcomes of interest using a simultaneous minimization of an extended weight matrix, with concatenated moment conditions. In future works, this model can be used to explicitly model the feedback from the outcome onto the covariates simultaneously.

### *4.6 Acknowledgements*

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data

files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

# CHAPTER 5

# CONCLUSIONS

In this dissertation work, I present three papers addressing correlation in binary outcome data. The first paper addresses a measure of intraclass correlation for three level hierarchical data. This measure provides a practical approach to estimating the association inherent due to clustering, as well as useful thresholds for identifying when the complexity of a mixed model is necessary in practice. This work has been published in *Journal of Applied Statistics*. The second paper of this work provides the Partitioned GMM approach to explicitly model the effects of time-dependent covariates on an outcome of interest over time. Specifically, this method incorporates additional parameters to explain the effect of each covariate on the outcome across time. This paper has been submitted, and is under review with *Statistics in Medicine*. The third paper discusses a simultaneous GMM model to jointly fit models for more than one outcome, while also accounting for the varying effects of time-dependent covariates. This model utilizes an extension of the Partitioned GMM in conjunction with an extended objective function to take into account the association that exists between multiple outcomes in the same data. This manuscript will be submitted to *Statistics*. Taken together, these three works provide useful methods of measuring and accounting for correlation inherent in longitudinal and hierarchical data structures.

REFERENCES

(NCCOR), National Collaborative on Childhood Obesity Research. 2014. *Annual Report 2014*.

Achen, Christopher 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.

Adams, G., et al. 2004. "Patterns of Intra-Cluster Correlation from Primary Care Research to Inform Study Design and Analysis." *Journal of Clinical Epidemiology* 57, no. 8 (Aug): 785-794. http://dx.doi.org/10.1016/j.jclinepi.2003.12.013.

Bandyopadhyay, S., B. Ganguli, and A. Chatterjee. 2011. "A Review of Multivariate Longitudinal Data Analysis." *Stat Methods Med Res* 20, no. 4 (Aug): 299-330. http://dx.doi.org/10.1177/0962280209340191.

Berridge, D.M., and R. Crouchley. 2011. *Multivariate Generalized Linear Mixed Models Using R*: CRC Press.

Bloch, D. A., and H. C. Kraemer. 1989. "2 X 2 Kappa Coefficients: Measures of Agreement or Association." *Biometrics* 45, no. 1 (Mar): 269-87. https://www.ncbi.nlm.nih.gov/pubmed/2655731.

Bodian, C. A. 1994. "Intraclass Correlation for Two-by-Two Tables under Three Sampling Designs." *Biometrics* 50, no. 1 (Mar): 183-93. https://www.ncbi.nlm.nih.gov/pubmed/8086601.

Breslow, N.E., and D.G. Clayton. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88: 9-25.

Cai, Katherine, and Jeffrey R Wilson. 2015. *How to Use Sas® for Gmm Logistic Regression Models for Longitudinal Data with Time-Dependent Covariates*. *SAS Global Forum*. Dallas, TX: SAS Institute.

Cai, Katherine and Jeffrey R Wilson. 2016. *Sas Macro for Generalized Method of Moments Estimation for Longitudinal Data with Time-Dependent Covariates*. *SAS Global Forum*. Las Vegas, NV: SAS Institute.

Chen, I. Chen, and Philip M. Westgate. 2017. "Improved Methods for the Marginal Analysis of Longitudinal Data in the Presence of Time-Dependent Covariates." *Statistics in Medicine* 36, no. 16: 2533-2546. http://dx.doi.org/10.1002/sim.7307.

Cicchetti, Domenic Vincent. 1994. "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology." *Psychological Assessment* 6, no. 4: 284-290. http://dx.doi.org/10.1037/1040-3590.6.4.284.

Cohen, B. H. 1980. "Chronic Obstructive Pulmonary Disease: A Challenge in Genetic Epidemiology." *American Journal of Epidemiology* 112, no. 2: 274-288.

Cramer, H. 1946. *Mathematical Methods of Statistics*, Current Contents/Physical Chemical & Earth Sciences. Princeton, NJ: Princeton University Press.

Cunningham, T. D., and R. E. Johnson. 2016. "Design Effects for Sample Size Computation in Three-Level Designs." *Stat Methods Med Res* 25, no. 2 (Apr): 505-19. http://dx.doi.org/10.1177/0962280212460443.

Diggle, Peter J., et al. 2002. *Analysis of Longitudinal Data*. Oxford, United Kingdom: Oxford University Press.

Dobson, Annette J. 2002. *An Introduction to Generalized Linear Models*. 2nd ed., Chapman & Hall/Crc Texts in Statistical Science Series. Boca Raton: Chapman & Hall/CRC.

Donner, A. 1986. "A Review of Inference Procedures for the Intraclass Correlation Coefficient in the One-Way Random Effects Model." *International Statistical Review* 54, no. 1 (Apr): 67-82. http://dx.doi.org/10.2307/1403259.

Donner, A., and A. Donald. 1988. "The Statistical Analysis of Multiple Binary Measurements." *J Clin Epidemiol* 41, no. 9: 899-905. https://www.ncbi.nlm.nih.gov/pubmed/3183697.

Donner, A., and J. J. Koval. 1980. "The Estimation of Intraclass Correlation in the Analysis of Family Data." *Biometrics* 36, no. 1 (Mar): 19-25. https://www.ncbi.nlm.nih.gov/pubmed/7370372.

Elston, RC. 1977. "Response to Query: Estimating “Heritability” of a Dichotomous Trait." *Biometrics* 33: 232-233.

Ene, Mihaela, et al. 2015. "Multilevel Models for Categorical Data Using Sas Proc Glimmix: The Basics." Paper presented at the SAS Global Forum, Dallas, TX.

Fieuws, S., and G. Verbeke. 2004. "Joint Modelling of Multivariate Longitudinal Profiles: Pitfalls of the Random-Effects Approach." *Stat Med* 23, no. 20 (Oct 30): 3093-104. http://dx.doi.org/10.1002/sim.1885.

Fieuws, S., G. Verbeke, and G. Molenberghs. 2007. "Random-Effects Models for Multivariate Repeated Measures." *Stat Methods Med Res* 16, no. 5 (Oct): 387-97. http://dx.doi.org/10.1177/0962280206075305.

Fitzmaurice, G.M., N.M. Laird, and J.H. Ware. 2004. *Applied Longitudinal Analysis*: Wiley.

Fleiss, JL, and J Cuzick. 1979. "The Reliability of Dichotomous Judgments: Unequal Numbers of Judges Per Subject." *Applied Psychological Measurement* 3: 537-542.

Ghebremichael, Musie. 2015. "Joint Modeling of Correlated Binary Outcomes: Hiv-1 and Hsv-2 Co-Infection." *Journal of Applied Statistics* 42, no. 10 (2015/10/03): 2180-2191. http://dx.doi.org/10.1080/02664763.2015.1022138.

Goetgeluk, Sylvie, and Stijn Vansteelandt. 2008. "Conditional Generalized Estimating Equations for the Analysis of Clustered and Longitudinal Data." *Biometrics* 64, no. 3: 772-780. http://dx.doi.org/10.1111/j.l541-0420.2007.00944.x.

Gueorguieva, R. 2001. "A Multivariate Generalized Linear Mixed Model for Joint Modelling of Clustered Outcomes in the Exponential Family." *Statistical Modelling* 1, no. 3: 177-193. http://dx.doi.org/10.1177/1471082x0100100302.

Guerra, Matthew W., et al. 2012. "The Analysis of Binary Longitudinal Data with Time-Dependent Covariates." *Statistics in Medicine* 31, no. 10: 931-948. http://dx.doi.org/10.1002/sim.4465.

Gulliford, M. C., O. C. Ukoumunne, and S. Chinn. 1999. "Components of Variance and Intraclass Correlations for the Design of Community-Based Surveys and Intervention Studies: Data from the Health Survey for England 1994." *Am J Epidemiol* 149, no. 9 (May): 876-83. https://www.ncbi.nlm.nih.gov/pubmed/10221325.

Guydish, Joseph, et al. 2011. "Smoking Prevalence in Addiction Treatment: A Review." *Nicotine & Tobacco Research* 13, no. 6: 401-411. http://dx.doi.org/10.1093/ntr/ntr048.

Hansen, Lars P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50, no. 4: 1029-1054. http://dx.doi.org/10.2307/1912775.

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

Hayashi, Fumio. 2000. *Econometrics*. Princeton, N.J.: Princeton, N.J. : Princeton University Press.

Heagerty, Patrick J., and Bryan A. Comstock. 2013. "Exploration of Lagged Associations Using Longitudinal Data." *Biometrics* 69, no. 1 (01/22): 197-205. http://dx.doi.org/10.1111/j.1541-0420.2012.01812.x.

Hsiao, Cheng. 2007. "Panel Data Analysis—Advantages and Challenges." *TEST* 16, no. 1: 1-22. http://dx.doi.org/10.1007/s11749-007-0046-x.

Hu, F. B., et al. 1998. "Comparison of Population-Averaged and Subject-Specific Approaches for Analyzing Repeated Binary Outcomes." *Am J Epidemiol* 147, no. 7 (Apr 01): 694-703. https://www.ncbi.nlm.nih.gov/pubmed/9554609.

Irimata, Kyle M, Jennifer Broatch, and Jeffrey R Wilson. 2018. "Partitioned Gmm Logistic Regression Models for Longitudinal Data." *Manuscript Submitted for Publication*.

Irimata, Kyle M, and Jeffrey R Wilson. 2017. "Identifying Intraclass Correlation Necessitating Hierarchical Modeling." *Journal of Applied Statistics*.

Jencks, S. F., M. V. Williams, and E. A. Coleman. 2009. "Rehospitalizations among Patients in the Medicare Fee-for-Service Program." *N Engl J Med* 360, no. 14 (Apr 02): 1418-28. http://dx.doi.org/10.1056/NEJMsa0803563.

Keele, Luke, and Nathan J. Kelly. 2005. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14, no. 2: 186-205. http://dx.doi.org/10.1093/pan/mpj006.

Kirk, Roger E. 1982. *Experimental Design : Procedures for the Behavioral Sciences*. 2nd ed. Monterey, Calif.: Brooks/Cole Pub. Co.

Kleinman, J. C. 1973. "Proportions with Extraneous Variance: Single and Independent Samples." *Journal of the American Statistical Association* 68, no. 341: 46-54. http://dx.doi.org/10.2307/2284137.

Kozlowski, Lynn T., Larry C. Jelinek, and Marilyn A. Pope. 1986. "Cigarette Smoking among Alcohol Abusers: A Continuing and Neglected Problem." *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique* 77, no. 3: 205-207. http://www.jstor.org.ezproxy1.lib.asu.edu/stable/41989220.

Lai, Tze L, and D Small. 2007. "Marginal Regression Analysis of Longitudinal Data." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 69, no. 1: 79-99.

Lalonde, Trent L, Jeffrey R Wilson, and Jianqiong Yin. 2014. "Gmm Logistic Regression Models for Longitudinal Data with Time-Dependent Covariates and Extended Classifications." *Stat Med* 33, no. 27 (Nov 30): 4756-69. http://dx.doi.org/10.1002/sim.6273.

Lasser, K, et al. 2000. "Smoking and Mental Illness: A Population-Based Prevalence Study." *Journal of the American Medical Association* 284, no. 20: 2606-2610. http://dx.doi.org/10.1001/jama.284.20.2606.

Lesaffre, Emmanuel, and Bart Spiessens. 2001. "On the Effect of the Number of Quadrature Points in a Logistic Random Effects Model: An Example." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50, no. 3: 325-335. http://dx.doi.org/10.1111/1467-9876.00237.

Liang, K. Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73, no. 1 (Apr): 13-22. http://dx.doi.org/10.1093/biomet/73.1.13.

Liang, K. Y., S. L. Zeger, and B. Qaqish. 1992. "Multivariate Regression Analyses for Categorical Data." *Journal of the Royal Statistical Society Series B-Methodological* 54, no. 1: 3-40.

Liebetrau, Albert M. 1983. *Measures of Association*, Sage University Papers Series Quantitative Applications in the Social Sciences. Beverly Hills: Sage Publications.

Liu, L., et al. 2010. "A Flexible Two-Part Random Effects Model for Correlated Medical Costs." *J Health Econ* 29, no. 1 (Jan): 110-23. http://dx.doi.org/10.1016/j.jhealeco.2009.11.010.

Maciejewski, M. L., and C. Maynard. 2004. "Diabetes-Related Utilization and Costs for Inpatient and Outpatient Services in the Veterans Administration." *Diabetes Care* 27 Suppl 2 (May): B69-73.

Mak, T. K. 1988. "Analysing Intraclass Correlation for Dichotomous Variables." *Applied Statistics-Journal of the Royal Statistical Society Series C* 37, no. 3: 344-352. http://dx.doi.org/10.2307/2347309.

McCulloch, Charles. 2008. "Joint Modelling of Mixed Outcome Types Using Latent Variables." *Statistical Methods in Medical Research* 17, no. 1: 53-73. http://dx.doi.org/10.1177/0962280207081240.

McMahon, J. M., E. R. Pouget, and S. Tortu. 2006. "A Guide for Multilevel Modeling of Dyadic Data with Binary Outcomes Using Sas Proc Nlmixed." *Computational Statistics & Data Analysis* 50, no. 12 (Aug): 3663-3680. http://dx.doi.org/10.1016/j.csda.2005.08.008.

Mudelsee, M. 2003. "Estimating Pearson's Correlation Coefficient with Bootstrap Confidence Interval from Serially Dependent Time Series." *Mathematical Geology* 35, no. 6 (Aug): 651-665. http://dx.doi.org/10.1023/b:matg.0000002982.52104.02.

Muller, Hans-Georg, and Ulrich Stadtmuller. 2005. "Generalized Functional Linear Models." *Ann. Statist.* 33, no. 2 (2005/04): 774-805. http://dx.doi.org/10.1214/009053604000001156.

Nelder, J. A., and D. Pregibon. 1987. "An Extended Quasi-Likelihood Function." *Biometrika* 74, no. 2 (Jun): 221-232. http://dx.doi.org/10.1093/biomet/74.2.221.

NIPORT. 2011. *Bangladesh Demographic and Health Survey 2011*. Edited by Mitra & Associates NIPORT, ICF International. Calverton, MD.

O'Connell, Ann A., and D. Betsy McCoach. 2008. *Multilevel Modeling of Educational Data*, Quantitative Methods in Education and the Behavioral Sciences. Charlotte, NC: IAP.

Oman, S. D., and D. M. Zucker. 2001. "Modelling and Generating Correlated Binary Variables." *Biometrika* 88, no. 1 (Mar): 287-290. http://dx.doi.org/10.1093/biomet/88.1.287.

Pepe, Margaret S, and Garnet L Anderson. 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, no. 4: 939-951.

Proust-Lima, C., et al. 2014. "Joint Latent Class Models for Longitudinal and Time-to-Event Data: A Review." *Stat Methods Med Res* 23, no. 1 (Feb): 74-90. http://dx.doi.org/10.1177/0962280212445839.

Pu, Jie, Di Fang, and Jeffrey R. Wilson. 2017. "Impact of Communities, Health, and Emotional-Related Factors on Smoking Use: Comparison of Joint Modeling of Mean and Dispersion and Bayes' Hierarchical Models on Add Health Survey." *BMC Medical Research Methodology* 17, no. 1 (February 03): 20. http://dx.doi.org/10.1186/s12874-017-0303-y.

Qu, A., B.G. Lindsay, and B. Li. 2000. "Improving Generalized Estimating Equations Using Quadratic Inference Functions." *Biometrika* 87, no. 4: 823-836.

Reisby, N., et al. 1977. "Imipramine: Clinical Effects and Pharmacokinetic Variability." *Psychopharmacology (Berl)* 54, no. 3 (Nov 15): 263-72. https://www.ncbi.nlm.nih.gov/pubmed/413143.

Ridout, M. S., C. G. Demétrio, and D. Firth. 1999. "Estimating Intraclass Correlation for Binary Data." *Biometrics* 55, no. 1 (Mar): 137-48. https://www.ncbi.nlm.nih.gov/pubmed/11318148.

Sahai, H, and M Ojeda. 2007. *Analysis of Variance for Random Models, Volume 2: Unbalanced Data: Theory, Methods, Applications and Data Analysis*. New York, NY: Springer.

Schildcrout, Jonathan S., and Patrick J. Heagerty. 2005. "Regression Analysis of Longitudinal Binary Data with Time-Dependent Environmental Covariates: Bias and Efficiency." *Biostatistics* 6, no. 4: 633-652. http://dx.doi.org/10.1093/biostatistics/kxi033.

Selig, J.P., K.J. Preacher, and T.D. Little. 2012. "Modeling Time-Dependent Association in Longitudinal Data: A Lag as Moderator Approach." *Multivariate Behavioral Research* 47, no. 5: 697-716.

Smyth, G. K. 1989. "Generalized Linear Models with Varying Dispersion." *Journal of the Royal Statistical Society Series B-Methodological* 51, no. 1: 47-60.

Snijders, T. A. B., and R. J. Bosker. 1999. *Multilevel Analysis : An Introduction to Basic and Advanced Multilevel Modeling*. London ; Thousand Oaks, Calif.: Sage Publications.

Song, Xiao, Marie Davidian, and Anastasios A. Tsiatis. 2002. "A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data." *Biometrics* 58, no. 4: 742-753. http://dx.doi.org/10.1111/j.0006-341X.2002.00742.x.

Stoner, J. A., B. G. Leroux, and M. Puumala. 2010. "Optimal Combination of Estimating Equations in the Analysis of Multilevel Nested Correlated Data." *Statistics in medicine* 29, no. 4: 464-473. http://dx.doi.org/10.1002/sim.3776.

Swallow, W. H., and J. F. Monahan. 1984. "Monte-Carlo Comparison of Anova, Mivque, Reml and Ml Estimators of Variance Components." *Technometrics* 26, no. 1: 47-57. http://dx.doi.org/10.2307/1268415.

Tan, M., et al. 1999. "A Bayesian Hierarchical Model for Multi-Level Repeated Ordinal Data: Analysis of Oral Practice Examinations in a Large Anaesthesiology Training Programme." *Stat Med* 18, no. 15 (Aug): 1983-92. https://www.ncbi.nlm.nih.gov/pubmed/10440881.

Wu, Lang, et al. 2012. "Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues." *Journal of Probability and Statistics* 2012: 17. http://dx.doi.org/10.1155/2012/640153.

Xu, Jane, and Scott L. Zeger. 2001. "The Evaluation of Multiple Surrogate Endpoints." *Biometrics* 57, no. 1: 81-87. http://dx.doi.org/10.1111/j.0006-341X.2001.00081.x.

Zeger, Scott L, and Kung Y Liang. 1986. "Longitudinal Data Analysis for Discrete and Continuous Outcomes." *Biometrics* 42, no. 1: 121-130.

Zeger, Scott L, and Kung Y Liang. 1991. "Feedback Models for Discrete and Continuous Time Series." *Statistica Sinica* 1: 51-64.

Zhang, Yi, et al. 2009. "Estimated Effect of Epoetin Dosage on Survival among Elderly Hemodialysis Patients in the United States." *Clinical Journal of the American Society of Nephrology : CJASN* 4, no. 3 (10/01/received 12/15/accepted): 638-644. http://dx.doi.org/10.2215/CJN.05071008.

Zhou, Y., et al. 2014. "Using Modified Approaches on Marginal Regression Analysis of Longitudinal Data with Time-Dependent Covariates." *Stat Med* 33, no. 19 (Aug 30): 3354-64. http://dx.doi.org/10.1002/sim.6171.

Zou, G., and A. Donner. 2004. "Confidence Interval Estimation of the Intraclass Correlation Coefficient for Binary Outcome Data." *Biometrics* 60, no. 3 (Sep): 807-11. http://dx.doi.org/10.1111/j.0006-341X.2004.00232.x.

# APPENDIX A

## ARTICLE USAGE

All co-authors at time of this writing are listed in each chapter. All co-authors listed in this dissertation have granted permission for the included articles to be used in this dissertation.