4-2018

Statistical Models for Correlated Data

Xiaomeng Niu

Western Michigan University, niuxiaomeng2006@hotmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Statistics and Probability Commons

Recommended Citation

https://scholarworks.wmich.edu/dissertations/3220

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
STATISTICAL MODELS FOR CORRELATED DATA

by

Xiaomeng Niu

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Statistics
Western Michigan University
April 2018

Doctoral Committee:

Dr. Hyun Keun Cho, Ph.D., Chair
Dr. Magdalena Niewiadowska-Bugaj, Ph.D.
Dr. Joshua Naranjo, Ph.D.
Dr. Jun-Seok Oh, Ph.D.
Correlated data arise frequently in many studies where multiple response variables or repeatedly measured responses within subjects are correlated. My dissertation topic lies broadly in developing various statistical methodologies for correlated types of data such as longitudinal data, clustered data, and multivariate data.

Multiple response variables might be relevant within subjects. A univariate procedure fitting each response separately does not take into account the correlation among responses. To improve estimation efficiency for the regression parameter, this study proposes two estimation procedures by accommodating correlations among the response variables. The proposed procedures do not require knowledge of the true correlation structure nor does it estimate the parameters associated with the correlation. We further propose simple and powerful inference procedures for a goodness-of-fit test that possess the chi-squared asymptotic properties.

For longitudinal count data with overdispersion, the overdispersion parameter plays a significant role in efficient estimation of the regression parameter. In this study, we develop a correlation structure for longitudinal count data in the negative binomial regression model, which is incorporated into a joint estimating equation to estimate both the regression parameter and the overdispersion parameter simultaneously. On the other
hand, inclusion of the overdispersion parameter can hinder efficient estimation and inference for the regression parameter when overdispersion is not present. This study provides new modeling for longitudinal count data and proposes a test detecting the presence of overdispersion.

In clinical trials, the outcome of a disease is measured at baseline and again as a response variable recorded over time during follow-up treatments. Baseline measurement is one of the most important determinants of the proper treatment for patients. This study proposes an nonparametric polynomial regression model to assess the treatment effects over time at various baseline levels. We further propose a model selection procedure based on the empirical log-likelihood method, which identifies the optimality of polynomial at each baseline. In addition, we provide a hypothesis test to assess the therapeutic effects by comparing the treatment groups.
Acknowledgments

I would like to first express my gratitude to my committee chair Dr. Hyunkeun Cho. His guidance, encouragement, and help gave me the courage to overcome the difficulties in completing this dissertation. I am really grateful to him for taking the time to review my work and to guide me through the whole process of writing the dissertation. I am so honored to have him as my advisor.

Secondly, I would like to thank my committee members, Dr. Magdalena Niewiadomska-Bugaj, Dr. Joshua Naranjo, and Dr. Jun-Seok Oh for their valuable suggestion and feedback. In particular, I am thankful to Dr. Magdalena Niewiadomska-Bugaj, the chair of our Statistics department, for giving me the opportunity to be a graduate assistant. While completing my doctorate, the experience I gained from my work is well worth it. Thank you for believing me in these responsibilities.

Lastly and most importantly, I would like to thank my family for supporting me and trusting me even when I doubt myself. Without their support, I had no chance of studying in the United States and finishing my Ph.D. program. I am deeply grateful to my husband, Yuyang Xiao, for his support and encouragement over the past six years.

Xiaomeng Niu
Contents

Acknowledgements ii

List of Tables v

List of Figures vii

1 Introduction 1

2 Simultaneous Estimation and Inference for Multiple Response Variables 5

   2.1 Introduction ................................................. 5
   2.2 Methodologies ............................................. 7
      2.2.1 Estimating Functions in Generalized Multivariate Linear Model . . 7
      2.2.2 Estimation and Inference via Generalized Method of Moments . . 9
      2.2.3 Estimation and Inference via Empirical Likelihood .................. 11
   2.3 Simulation Studies ......................................... 13
   2.4 Application to MSCM Data ................................. 15
   2.5 Discussion ................................................ 16
   2.6 Proofs of Theorems ........................................ 18

3 Efficient Regression Modeling for Correlated and Overdispersed Count Data 28

   3.1 Introduction ................................................ 28
   3.2 Methodologies ............................................. 31
Table of Contents - Continued

3.2.1 Negative Binomial Model for Longitudinal Data .......................... 31
3.2.2 Choice of Basis Matrices .................................................. 34
3.2.3 Test for Detecting Overdispersion ....................................... 36
3.3 Simulation Studies ................................................................. 37
3.4 Application to Transportation Crash Data .................................... 39
3.5 Proofs of Theorems ............................................................... 40

4 Adjusting for Baseline Outcomes in Comparing the Effects of Randomized Treatments 48
4.1 Introduction ........................................................................... 48
4.2 Methodologies ....................................................................... 52
  4.2.1 Estimation of Varying Coefficients ....................................... 52
  4.2.2 Choice of the Most Parsimonious Model ................................. 54
  4.2.3 Hypothesis Test for Comparing Treatment Effects .................. 56
4.3 Simulation Studies ................................................................. 57
4.4 Application to Depression Study .............................................. 59
4.5 Proofs of Theorems ............................................................... 62

References .................................................................................. 73

Appendix

A HSIRB Approval Letter ............................................................. 81
## List of Tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>2.1 Relative efficiency of proposed estimators obtained from generalized method moment (GMM) and empirical likelihood (EL).</td>
</tr>
<tr>
<td>2.2</td>
<td>Proportions of times that $H_0 : \beta_j = 0, j = 1, 2, 3$, is rejected in 1000 simulation runs through Wald test of univariate approach and chi-squared test of generalized method moment (GMM) and empirical likelihood (EL).</td>
</tr>
<tr>
<td>2.3</td>
<td>Description of variables in MSCM data set.</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimated coefficients along with confidence intervals at a 95% level in brackets with * when zero is not included using generalized method moment (GMM), empirical likelihood (EL) and univariate approach.</td>
</tr>
<tr>
<td>3.1</td>
<td>Mean squared errors ($\times 1000$), average length of confidence intervals (LCI), and coverage proportion (CP) at 95% confidence level using negative binomial regression under AR(1), compound symmetry (CS), and independent (IN) structures.</td>
</tr>
<tr>
<td>3.2</td>
<td>Relative efficiency of proposed estimators obtained from negative binomial regression under AR(1), compound symmetry (CS), and independent (IN) structures.</td>
</tr>
<tr>
<td>3.3</td>
<td>Description of variables in crash data.</td>
</tr>
</tbody>
</table>
3.4 Estimated coefficients along with standard errors in brackets marked with * if covariate is statistically significant through negative binomial regression and Poisson regression. .............................................................. 46

4.1 Relative efficiency of proposed estimators obtained from quadratic regression under compound symmetry and AR(1) structures. .................. 68
4.2 Estimated coefficients of depression data adopting linear regression under AR(1) structure. ................................................................. 69
4.3 Estimated covariate coefficients of depression data. ....................... 69
4.4 Test statistic (T), p-value and degrees of freedom (df) of equality test of three treatments: medication (M), psychotherapy (P) and community (C). 70
# List of Figures

3.1 Quantile-Quantile plot under AR(1) structure without overdispersion. . . 47

3.2 Power of the hypothesis test with overdispersion parameter from 0 to 0.01 under AR(1), compound symmetry, and independent structures. . . . . . 47

4.1 Quantile-Quantile plots for chi-square distribution with three degrees of freedom versus proposed test statistic testing equality of polynomial curves at $y_0 = 1$ under compound symmetry and AR(1) structures. . . . . . . 71

4.2 Fitted quadratic polynomial curves of medication (solid curve), psychotherapy (dashed curve) and community (dotted curve). . . . . . . . . . . 72
Chapter 1

Introduction

Correlated data arise frequently in various studies, where measurements on the same subject are repeatedly tracked over time. The generalized regression model does not take into account the correlation within the subjects, which may reduce the efficiency of estimation. To improve the estimation efficiency, we should incorporate the within-subject correlation into the estimating equations. In this thesis, we focus on the development of statistical methodologies for various correlated data to improve the estimation efficiency and conduct the hypothesis test.

Our work is motivated by these three types of correlated data. On is the mother’s stress and children’s morbidity data, mothers and their children were enrolled in the study for one month. The objective of this study is to identify which features influence both a mother’s stress and children’s illness. This is a bivariate responses case, where the mother’s stress and children’s morbidity are highly correlated. The second one is the transportation crash data, which consists of the number of accidents and various features collected annually from the same road segment. It is of particular interest to build a regression model that allows us to investigate causes of car crashes and to anticipate the number of crashes. According to the preliminary exploration, this study belongs to the longitudinal count data analysis, which has the overdispersion issue. The third one is the randomized clinical trials
data, women suffering from depression were randomly assigned to one of three therapies after their baseline severities were measured. The depression score was recorded monthly to evaluate effects of the different treatments on depression longitudinally. The purpose of this study is to explore the three treatment effects over time at each baseline level. To investigate the above-mentioned three types of correlated data, we develop various statistical methodologies as shown in following Chapters.

Multivariate data arise frequently in biomedical and health studies where multiple response variables are collected across subjects. Unlike a univariate procedure fitting each response separately, a multivariate regression model provides a unique opportunity in studying the joint evolution of various response variables. In Chapter 2, we propose two estimation procedures that improve estimation efficiency for the regression parameter by accommodating correlations among the response variables. The proposed procedures do not require knowledge of the true correlation structure nor does it estimate the parameters associated with the correlation. Theoretical and simulation results confirm that the proposed estimators are more efficient than the one obtained from the univariate approach. We further propose simple and powerful inference procedures for a goodness-of-fit test that possess the chi-squared asymptotic properties. Extensive simulation studies suggest that the proposed tests are more powerful than the Wald test based on the univariate procedure. The proposed methods are also illustrated through the mother’s stress and children’s morbidity study.

The objective of Chapter 3 is to propose an efficient estimation procedure in a marginal mean regression model for longitudinal count data and to develop a hypothesis test for detecting the presence of overdispersion. We extend the matrix expansion idea of quadratic inference functions to the negative binomial regression framework that entails accommodating both the within-subject correlation and overdispersion issue. Theoretical and numerical results show that the proposed procedure yields a more efficient estimator asymp-
totically than the one ignoring either the correlation or overdispersion. When the overdispersion is absent in data, the proposed method might hinder the estimation efficiency in practice, yet the Poisson regression based regression model is fitted to the data sufficiently well. Therefore, we construct the hypothesis test that recommends an appropriate model for the analysis of the correlated count data. Extensive simulation studies indicate that the proposed test can identify the effective model consistently. The proposed procedure is also applied to a transportation safety study and recommends the proposed negative binomial regression model.

The main objective of clinical trials is to evaluate a treatment effect of interest on disease. In many clinical trials, a baseline outcome of disease can play a critical role in determining a proper treatment. In Chapter 4, we propose a nonparametric polynomial regression model for longitudinal data and provide an efficient estimation procedure that accommodates the within-subject correlation. The model is fitted by utilizing information from subjects whose baseline outcomes are close to the baseline of interest. The proposed model enjoys the property of a nonparametric model, since the coefficients in the model varies across baseline levels. In addition, the proposed model accesses the treatment effect in accordance with a level of baseline by allowing the degree of a polynomial to vary across values of the baseline outcome. In general, the model with a higher degree of a polynomial provides a better fit to the observed data, yet could be overfitted and complex. Therefore, we propose a model selection procedure based on an empirical log-likelihood ratio, which adaptively identifies the proper degree of a polynomial at different levels of the baseline outcome. At a given baseline level, a hypothesis test is further developed to evaluate the therapeutic effect by comparing the patterns of polynomial curves among treatment groups. The proposed log-likelihood ratio test statistic asymptotically follows a chi-squared distribution under the null hypothesis. In extensive simulation studies, we confirm that the proposed procedure not only selects the true model, but also examines the
difference between the outcomes over time successfully. The proposed entire procedure is illustrated through the analysis of depression data and explores dynamic longitudinal trajectories of depression scores over time across baseline depression scores.
Chapter 2

Simultaneous Estimation and Inference for Multiple Response Variables

2.1 Introduction

Multivariate regression analysis has become increasingly important due to the rise of multivariate data where different types of multiple outcomes are measured from each subject. A univariate procedure decomposes the multivariate data and models each response separately. Although this procedure is simple and popular, it might be inefficient if correlations among different response variables are present. A multivariate regression model is a viable alternative; it is a flexible generalization of the univariate regression that allows us to accommodate the joint information between the response variables.

For the multivariate regression model, Breiman and Friedman (1997) developed the curd and whey approach to improve the prediction performance by utilizing the dependency among responses. Yuan et al. (2007) and Chen and Huang (2012) proposed the high dimensional reduced-rank regression model. Recently, Rothman et al. (2010), Lee and Liu (2012), and Wang (2015) formulated the multivariate regression problem in a penalized log-likelihood framework. The aforementioned methods are mainly applicable when all
response variables are continuous. Mixed discrete and continuous response variables have been analyzed through various methods based on the quasi-likelihood (Prentice and Zhao, 1991; Zhao et al., 1992; Catalano and Ryan, 1992), the likelihood-based (Fitzmaurice and Laird, 1997), and the latent variable model (Sammel et al., 1997; Teixeira-Pinto and Normand, 2009).

In this paper, we develop a set of estimating functions in a generalized multivariate linear regression model and propose two efficient estimation and inference procedures based on the developed estimating functions. The proposed approaches enable us to analyze multivariate data all together regardless of response families, since different link functions are used according to the type of response variables. To accommodate correlations among different response variables, we employ a data driven approach that does not require the exact knowledge of the correlation structure. In addition, we employ the matrix expansion idea of quadratic inference functions (Qu et al., 2000) and develop estimating functions for estimation of the regression parameter of interest. Since the set of estimating functions is over-identified, we adopt the generalized method of moments (Hansen, 1982) and the empirical likelihood (Owen, 1988; Qin and Lawless, 1994), respectively.

Theoretical results indicate that the two proposed estimators are asymptotically consistent and follow the same normal distribution. Moreover, asymptotic estimation efficiency is improved when response variables are correlated and associated with a different set of covariates. Extensive simulation studies also confirm that the proposed approaches outperform the univariate approach in terms of smaller mean squared errors in cases where a set of covariate differs between response variables. When the results of the two proposed approaches are compared in a finite sample, the empirical likelihood approach may yield a more efficient estimator than the generalized method of moments approach. Although both approaches use the same set of estimating functions, the empirical likelihood removes some of the ambiguity stemming from a consistent estimate of the weight matrix in the
generalized method of moments.

For statistical inference on the regression parameter, we provide a plug-in estimator of the asymptotic covariance matrix of the regression parameter. In addition, we propose inference procedures for model diagnostic and goodness-of-fit tests. The proposed test statistics possess the same chi-squared asymptotic properties as in the likelihood ratio test, yet they do not require specifying the likelihood functions. This is especially challenging when the various types of response variables is modeled simultaneously. The results of the simulation studies indicate that the proposed tests are more powerful than the Wald test based on the univariate procedure. The proposed approaches are also applied to the real data set consisting of correlated bivariate response variables: mother’s stress and children’s morbidity.

The paper is organized as follows. In Section 2.2, we develop estimating functions, provide efficient estimation and inference on the regression parameter in the generalized multivariate regression model, and investigate the asymptotic properties of the proposed estimators. Sections 2.3 and 2.4 illustrate various simulation studies and data analysis for the mother’s stress and children’s morbidity study, respectively. We conclude with a discussion in Section 2.5. The theoretical proofs and necessary conditions are placed in Section 2.6.

2.2 Methodologies

2.2.1 Estimating Functions in Generalized Multivariate Linear Model

Suppose that $y_{ij}$ is the $j$th response variable, $j = 1, \ldots, J$, from subject $i$ and $y_{ij}$’s are independent and identically distributed for $i = 1, \ldots, n$, where $J$ is the number of responses and $n$ is the sample size. For a generalized univariate linear model, a conditional mean of $y_{ij}$ is formulated as $\mu_{ij} = E(y_{ij}|x_{ij}) = g(x_{ij}^T \beta_j)$, where $g(\cdot)$ is an inverse link function and
\( x_{ij} \) is a \( p_j \) dimensional predictor vector corresponding to a parameter vector \( \beta_j \) associated with the \( j \)th response. Here the predictor vector \( x_{ij} \) and its dimension \( p_j \) can be different across responses. For ease of presentation, we set \( p_j = p \) for all \( j \).

To accommodate associations among \( J \) responses within the same subject, we stack up \( J \) response variables from the \( i \)th subject as the vector \( Y_i = (y_{i1}, \ldots, y_{ij})^T \) and the corresponding covariates

\[
X_i = \begin{pmatrix}
x_{i1} & 0 & \cdots & 0 \\
0 & x_{i2} & \cdots & 0 \\
& & \ddots & \vdots \\
0 & 0 & \cdots & x_{ij}
\end{pmatrix}.
\]

(2.1)

Therefore, the conditional mean vector of \( Y_i \) can be specified as \( \mu_i = E(Y_i|X_i) = g(X_i^T\beta) \), where \( \beta = (\beta_1^T, \ldots, \beta_J^T)^T \) is a \( pJ \) dimensional parameter vector. Note that the inverse link function can be chosen differently corresponding to the type of response variables. We extend generalized estimating equations (Liang and Zeger, 1986) and provide the equations

\[
\frac{1}{n} \sum_{i=1}^{n} \mu_i^T A_i^{-1/2} R(\alpha)^{-1} A_i^{-1/2}(Y_i - \mu_i) = 0,
\]

(2.2)

where \( \mu_i = \partial \mu_i / \partial \beta \), \( A_i \) is a \( J \times J \) diagonal marginal variance matrix of \( Y_i \) and \( R(\alpha) \) is a working correlation matrix with a nuisance parameter \( \alpha \). This working correlation structure enables us to accommodate correlations among different responses. Estimation for a nuisance parameter \( \alpha \) is required after assuming the type of working correlation structure. However, any fundamental assumption of the working correlation structure is not ideal due to the correlated nature of multiple responses. This motivates us to utilize a data driven approach (Qu and Lindsay, 2003) that approximates the true correlation structure through the sample correlation matrix of \( Y_i \) and decomposes \( R(\alpha)^{-1} \) using the eigenvector decomposition, i.e., \( R(\alpha)^{-1} = b_0 I_J + b_1 B_1 + \ldots + b_J B_J \), where \( B_j = e_j e_j^T \).
\[ j = 1, \ldots, J, \] with an eigenvector \( e_j \) corresponding to the \( j \)th largest eigenvalue of the sample correlation matrix, and \( b_0, b_1, \ldots, b_J \) are unknown constants. Following Qu et al. (2000), we replace \( R(\alpha)^{-1} \) in (2.2) with \( B_j \)'s and extend (2.2) to

\[
h_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} h_i(\beta) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 
\mu_i^T A_i^{-1}(Y_i - \mu_i) \\
\mu_i^T A_i^{-1/2} B_j A_i^{-1/2}(Y_i - \mu_i) \\
\vdots \\
\mu_i^T A_i^{-1/2} B_J A_i^{-1/2}(Y_i - \mu_i)
\end{pmatrix}.
\] (2.3)

Note that every equation in (2.2) can be approximated by a linear combination of elements in (2.3) and the vector \( h_n(\beta) \) does not involve any nuisance parameters associated with the working correlation structure. Moreover, it follows from the moment assumption \( \mu_i = E(Y_i|X_i) \) that \( E\{h_i(\beta)\} = 0 \) under the true parameter. We also note that the conventional estimator ignoring correlations among different responses can be readily obtained by solving the first term in (2.3) only, denoted by \( h_{n1}(\beta) = \frac{1}{n} \sum_{i=1}^{n} h_{i1}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \mu_i^T A_i^{-1}(Y_i - \mu_i) \). The rest of the terms in (2.3), denoted by \( h_{n2}(\beta) \), enable us to incorporate the correlation information and achieve estimation efficiency for the regression parameter. We call \( h_n(\beta) \) estimating functions and provide two approaches based on the estimating functions for estimation and inference on the parameter \( \beta \).

### 2.2.2 Estimation and Inference via Generalized Method of Moments

In general, we cannot set each element in (2.3) to zero simultaneously in estimating the regression parameter, since \( h_n(\beta) \) contains more estimating functions than the number of parameters. Alternatively, we employ quadratic inference functions (Qu et al., 2000) on the basis of the generalized method of moments (Hansen, 1982),

\[
H_n(\beta) = h_n(\beta)^T V_n(\beta)^{-1} h_n(\beta),
\] (2.4)
where a weighting matrix $V_n(\beta) = \frac{1}{n} \sum_{i=1}^n h_i(\beta)h_i(\beta)^T$ is a consistent estimate of the covariance matrix of $h_i(\beta)$. The functions optimally combine the estimating functions in (2.3) by imposing a lower weight to the estimating function with a larger variance. The estimator of $\beta$ can be obtained as $\hat{\beta} = \arg\min_{\beta} H_n(\beta)$. We let $\beta_0$ be the true parameter and $h_i(\beta)$, $i = 1, \ldots, n$, be independent realizations of the process $h(\beta)$.

**Theorem 2.1.** Under the regularity conditions in Section 2.6, the estimator $\hat{\beta}$ satisfies 

$$
\sqrt{n}(\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N(0, \Omega), \text{ as } n \rightarrow \infty, \text{ where } \Omega = \left\{ E(\frac{\partial h}{\partial \beta})^T E(hh^T)^{-1} E(\frac{\partial h}{\partial \beta}) \right\}^{-1}.
$$

Furthermore, $\Omega_i - \Omega$ is positive semi-definite, where $\Omega_i$ is the asymptotic covariance matrix of the estimator under the independent working correlation structure.

Theorem 2.1 shows that the proposed estimator $\hat{\beta}$ is consistent and follows an asymptotic normal distribution with the true mean $\beta_0$ and the covariance matrix $\Omega$. This estimator is the most efficient among estimators obtained from the same linear class of the estimating functions in (2.3), since the limiting covariance matrix $\Omega$ is minimum in the sense of Loewner ordering. This theorem also implies that the proposed approach yields a more efficient estimator than the univariate procedure which does not incorporate the correlation information. More importantly, the efficiency gain does not require the use of the true correlation structure. If the conditional mean of all responses depends on the same set of covariates, i.e. $x_{i1} = \cdots = x_{ij}$ in (2.1) for $i = 1, \ldots, n$, then all estimating functions in $h_{i2}(\beta)$ are linearly dependent to those in $h_{i1}(\beta)$ and therefore $\Omega = \Omega_i$. The asymptotic covariance matrix $\Omega$ can be consistently estimated for statistical inference on the regression parameter as

$$
\left( \sum_{i=1}^n \frac{\partial h_i(\hat{\beta})}{\partial \beta} \right)^T \left( \sum_{i=1}^n h_i(\hat{\beta})h_i(\hat{\beta})^T \right)^{-1} \left( \sum_{i=1}^n \frac{\partial h_i(\hat{\beta})}{\partial \beta} \right)
$$

After fitting the multivariate regression model to data, one often asks if a certain set of covariates are statistically significant or any arbitrary parametric models fit multivariate...
data adequately. To tackle these issues we construct the hypothesis test for the regression parameter based on $H_n(\beta)$, since the functions play an inferential role similar to minus twice the loglikelihood functions. Suppose that $\beta$ is partitioned into two sets $(\alpha, \gamma)$, where $\alpha$ is a $q$ dimensional vector of particular interest and $\gamma$ is a $pJ - q$ dimensional parameter vector. Then the hypotheses are formulated as

$$H_0 : \alpha = \alpha_0 \text{ versus } H_\beta : \alpha \neq \alpha_0, \quad (2.5)$$

where $\alpha_0$ is a constant vector. We propose the test statistic

$$T_n = n\{H_n(\hat{\gamma}|\alpha_0) - H_n(\hat{\beta})\},$$

where $\hat{\gamma} = \arg \min_{\gamma} H_n(\gamma|\alpha_0)$ and $\hat{\beta} = \arg \min_{\beta} H_n(\beta)$. The test statistic is positive systematically and close to 0 under the null hypothesis.

**Theorem 2.2.** Under the regularity conditions in Section 2.6, $T_n \xrightarrow{d} \chi^2_q$ under $H_0$ as $n \to \infty$.

Theorem 2.2 confirms that the test statistic $T_n$ follows the asymptotic chi-squared distribution with $q$ degrees of freedom under the null hypothesis in (2.5).

### 2.2.3 Estimation and Inference via Empirical Likelihood

We propose a viable alternative for the parameter estimation and its inference. Under $E\{h_i(\beta_0)\} = 0$, we employ the empirical likelihood ratio (Qin and Lawless, 1994)

$$L_n(\beta) = \sup \left\{ \prod_{i=1}^{n} nw_i \bigg| \sum_{i=1}^{n} w_i h_i(\beta) = 0, \sum_{i=1}^{n} w_i = 1, 0 \leq w_i \leq 1 \right\}, \quad (2.6)$$

where $w_i$ denotes a point mass assigned to the observations from the $i$th subject. The maximum empirical likelihood estimator can be obtained as $\tilde{\beta} = \arg \max_{\beta} L_n(\beta)$. The
corresponding point masses satisfy \( w_i = n^{-1}\{1 + \lambda^T h_i(\tilde{\beta})\}^{-1} \), where \( \lambda \) is the solution to \( n^{-1} \sum_{i=1}^{n} h_i(\tilde{\beta})\{1 + \lambda^T h_i(\tilde{\beta})\}^{-1} = 0 \). The empirical likelihood based approach does not involve any variance estimation, while the approach using the weighted quadratic function requires a consistent estimate for the optimal weight matrix \( V_n(\beta) \) in (2.4).

**Theorem 2.3.** Under the regularity conditions in Section 2.6, the estimator \( \tilde{\beta} \) holds \( \sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, \Omega) \) as \( n \to \infty \), and \( \Omega - \Omega \) is positive semi-definite.

Theorem 2.3 confirms that \( \tilde{\beta} \) is asymptotically consistent and normal. In addition, the maximum empirical likelihood estimator \( \tilde{\beta} \) is as efficient as the generalized method of moments estimator \( \hat{\beta} \) because their asymptotic covariance matrices \( \Omega \) are identical. Therefore, the empirical likelihood based estimation procedure also yields a more efficient estimator than the univariate procedure ignoring the correlations among the response variables.

We also propose an inference function for a goodness-of-fit test based on the empirical likelihood which parallels the standard parametric with respect to the likelihood ratio statistics. We consider the empirical likelihood ratio (2.6) to test the null hypothesis defined in (2.5) in Section 2.2.2 and formulate an empirical likelihood ratio test statistic as

\[
W_n = 2[\log\{L_n(\tilde{\beta})\} - \log\{L_n(\tilde{\gamma}|\alpha_0)\}],
\]

where \( \tilde{\gamma} \) and \( \tilde{\beta} \) maximize \( L_n(\gamma|\alpha_0) \) and \( L_n(\beta) \), respectively. This is twice the difference in the log-empirical likelihoods under \( H_0 \) and \( H_\alpha \) in (2.5).

**Theorem 2.4.** Under the regularity conditions in Section 2.6, \( W_n \xrightarrow{d} \chi^2_q \) under \( H_0 \) as \( n \to \infty \).

Theorem 2.4 shows that the test statistic \( W_n \) follows the asymptotic chi-squared distribution with \( q \) degrees of freedom under the null hypothesis in (2.5). The empirical
likelihood ratio test rejects the null hypothesis at a significant level $\alpha$ if the value of $W_n$ is larger than the critical value $\chi^2_{1-\alpha,q}$, where $\chi^2_{1-\alpha,q}$ is $1-\alpha$ quantile of $\chi^2_q$.

2.3 Simulation Studies

In this section, we generate three correlated outcomes through the following generalized linear regression model, $\mu_{ij} = g(x^T_{ij}\beta_j)$ for $i = 1, \ldots, 100$ and $j = 1, \ldots, 3$, where each covariate $x_{ij}$ is generated independently from a standard normal distribution and the true parameter is $\beta = (\beta_1, \beta_2, \beta_3)^T = (0.2, 0.4, 0.6)^T$. To evaluate performance of estimation and inference, we first explore three different types of the equally correlated outcomes: all normal responses, all binary responses, and a mixture of one binary and two normal responses; at two strengths: $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$ and $0.8$, respectively, where $\rho_{jj'}$ is the correlation between the outcome $j$ and the outcome $j'$. Moreover, in order to evaluate the performance of the proposed approaches when correlations across three outcomes are different, we also generate the three outcomes based on $\rho_{12} = 0.2$, $\rho_{23} = 0.5$, and $\rho_{13} = 0.8$. Moreover, to assess performance of estimation in case where all normal responses are associated with the same covariate, an intercept and common covariate of $z_i$ are added as $\mu_{ij} = g(1 + x^T_{ij}\beta_j + 0.5z_i)$, where $z_i$ is generated from a standard normal distribution. Note that the R package MASS, bindata, and BinNor are applied to generate three different types of correlated responses, respectively.

We estimate all parameters simultaneously through two proposed approaches and univariate approach that estimates each parameter individually from 1000 simulation runs. Table 2.1 reports the relative efficiency of the proposed estimators to the ones obtained from the univariate method, defined by the ratio of the mean squared errors, $\text{MSE}(\hat{\beta}_j) = \frac{1}{1000} \sum_{i=1}^{1000} \left( \hat{\beta}^{(i)}_j - \beta_j \right)^2$, where $\hat{\beta}^{(i)}_j$ is the estimator from the $i$th simulation. The larger the value of the relative efficiency, the more efficient the proposed estimator.
According to the results in Table 2.1, the proposed estimation approaches outperform the univariate approach in terms of smaller mean squared errors. When three responses are highly correlated, the mean squared errors using the proposed approaches are smaller than those of the univariate procedure regardless of estimating functions under consideration. Moreover, the proposed approaches still yield more efficient estimators than the univariate approach even though the correlations among three responses are all different. These estimation efficiency can be improved in cases where the covariates differ in value for different responses. Otherwise, no efficiency gain is made, i.e., relative efficiencies of an intercept and common covariate are all between 0.9 and 1. When results of the generalized method of moments and the empirical likelihood are compared, they are comparable in all cases under consideration.

For statistical inference, we investigate whether each covariate is significant or not by testing the null hypothesis \( H_0 : \beta_j = 0 \) against the alternative hypothesis \( H_A : \beta_j \neq 0 \) for \( j = 1, 2, \) and 3. We compute the test statistics \( T_n \) and \( W_n \) from 1000 simulation runs and report the true negative rates and the proportion of times that the null hypothesis is rejected, in Table 2.2. The results indicate that the true negative rates based on the two approaches are small under the weak signal, although the empirical likelihood rejects the null hypothesis slightly more than the generalized method of moment. As the size of the coefficient increases, true negative rates also increase and reach one at \( \beta_3 = 0.6 \) in all cases but binary responses. We further compare results with the Wald test (i.e., the null hypothesis is rejected if \( \left( \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \right)^2 > \chi^2_{0.95,1} \), where \( \hat{\beta}_j \) and \( \text{se}(\hat{\beta}_j) \) are the estimated coefficient and its standard error obtained from the univariate approach). Table 2.2 shows that the tests based on the two proposed procedures reject the null hypothesis with higher frequencies compared to the Wald test in all finite sample cases under consideration.
2.4 Application to MSCM Data

We illustrate the proposed approaches through the mother’s stress and children’s morbidity (MSCM) data studied by Alexander and Markowitz in 1986. A total of 167 mothers with their children between the ages of 18 months and 5 years were enrolled in the MSCM study for 28 days. The objective of this study is to identify which features influence both a mother’s stress and children’s illness. Therefore, we treat the proportions of days that a mother gets stressed (stress) and her children are ill (illness) as bivariate response variables. Among the eight covariates provided in Table 2.3, characteristics of mother and children are considered as the control variables for mother’s stress and children’s illness, respectively, while the household size (size) and race (race) are treated as common covariates.

Accordingly, two linear regression models are specified as

\[
\text{illness} = \beta_{10} + \beta_{11} \text{size} + \beta_{12} \text{race} + \beta_{13} \text{chlth} + \beta_{14} \text{gender} + \epsilon,
\]

where \(\text{chlth}\) is children’s health level and

\[
\text{stress} = \beta_{20} + \beta_{21} \text{size} + \beta_{22} \text{race} + \beta_{23} \text{mhlth} + \beta_{24} \text{edu} + \beta_{25} \text{employ} + \beta_{26} \text{marry} + \epsilon,
\]

where \(\text{mhlth}, \text{edu}, \text{employ}, \text{and marry}\) are mother’s characteristics of health level, education level, employment status, and marriage status, respectively. Note that the bivariate response variables are more likely to be correlated, since a mother may suffer from stress when her children are sick. To accommodate this correlation, we estimate all coefficients simultaneously using the generalized method of moment and the empirical likelihood. We also apply the univariate approach that fits the above models separately.

Table 2.4 reports the estimated coefficients and the corresponding 95% confidence
interval along with * if zero is not included. The results confirm that all approaches perform similarly in the children's illness model; the household size and children's race and health level are statistically significant. Specifically, healthy white children living in a large family are less likely to be sick. For the mother's stress model, effects of all covariates except race are statistically significant based on the two proposed approaches, while only three covariates (education status, health level, and household size) are associated with mother's stress when the univariate approach is applied. Thus, we further investigate whether employment level and marriage status influence the mother's stress or not by computing the test statistics $T_n$ and $W_n$ for each covariate individually; $T_n = 4.44$ and $W_n = 22.48$ for an employment level and $T_n = 6.93$ and $W_n = 20.83$ for a marriage status. The inference procedures confirm that both covariates are statistically significant in the mother's stress model, since the test statistics are greater than a critical value of $\chi^2_{0.95,1} = 3.84$. The results correspond to the ones based on the confidence interval discussed above.

### 2.5 Discussion

In this paper, we have proposed two estimation procedures that gain estimation efficiency by incorporating the informative correlation among different responses within the same subject. Our theoretical derivations confirm that both proposed approaches achieve the same amount of estimation efficiency in the sense that the limiting covariance matrices of regression parameters are identical as the sample size goes to infinity. In addition, our simulation studies also indicate that both proposed approaches perform well when the sample size is large. However, the results of estimation might be inefficient in some applications. As an example, if the number of estimating functions is relatively large compared to the number of unknown parameters, so-called a highly over-identified case, the
estimators obtained by the generalized method of moments approach may have large bias and the confidence intervals may have poor coverage proportion (Altonji and Segal, 1996; Imbens, 2002; Westgate, 2012, 2013). In such case, the estimation of the empirical likelihood approach is superior to the generalized method of moments, since it eliminates some of the ambiguity stemming from the weight matrix estimation; its asymptotic bias does not increase with the number of estimating functions, whereas the bias of the generalized method of moments increases. Therefore, with many estimating functions, the bias of empirical likelihood estimators is less than the bias of the generalized method of moments estimators (Imbens, 2002; Newey and Smith, 2004). Alternatively, the bias of the generalized method of moments estimators can be reduced by implementing the bias-correction of the asymptotic covariance matrix based on Taylor series expansions of the estimated regression parameters (Westgate, 2012, 2013).

The problem caused by the highly over-identified case may arise in the proposed generalized method of moments based estimation procedure since the dimension of estimating functions increases as the number of outcomes increases, while the number of regression parameters remains the same. This can be resolved by extending various methodologies either selecting a subset of estimating functions (Andrews, 1999; Okui, 2009; Liao, 2013; DiTraglia, 2016) or combining all estimating functions (Donald et al., 2009; Cho and Qu, 2015). For the empirical likelihood based procedure, although the small bias property of estimators is nice, the precision of estimation can be improved through various approaches such as bootstrap, Bartlett correction and the adjusted empirical likelihood (Horowitz, 1998; DiCiccio et al., 1991; Chen and Huang, 2013).
2.6 Proofs of Theorems

The following regularity conditions are required to establish the asymptotic properties of the proposed estimators:

1. The parameter space \( \Theta \) is compact and \( \beta \in \Theta \).
2. There exists a \( \beta_0 \) such that \( E\{h(\beta)\} = 0 \) for every \( i \) if and only if \( \beta = \beta_0 \).
3. \( \partial h_i(\beta) / \partial \beta \partial^2 h_i(\beta) / \partial \beta^2 \) are continuous in a neighborhood of \( \beta_0 \).
4. The matrices \( V_n \) and \( C_n \) converge almost surely to constant positive definite \( V \) and \( C \), respectively, where \( V^{-1} = C^T C \) and \( V = E\{h(\beta_0)h(\beta_0)^T\} \).

Proof of Theorem 2.1. From condition 4, the quadratic function \( H_n(\beta) \) can be rewritten as

\[
H_n(\beta) = h_n(\beta)^T V_n(\beta)^{-1} h_n(\beta) = h_n(\beta)^T C_n(\beta)^T C_n(\beta) h_n(\beta) = \left| C_n(\beta) \frac{1}{n} \sum_{i=1}^{n} h_i(\beta) \right|^2.
\]

In addition, \( |H_n(\hat{\beta})| < |H_n(\beta_0)| \) holds because \( \hat{\beta} = \text{arg min}_\beta H_n(\beta) \). Hence,

\[
\left| C_n(\hat{\beta}) \frac{1}{n} \sum_{i=1}^{n} h_i(\hat{\beta}) \right|^2 < \left| C_n(\beta_0) \frac{1}{n} \sum_{i=1}^{n} h_i(\beta_0) \right|^2. \tag{A.1}
\]

By the law of large numbers and condition 2, the right hand side of (A.1) converges to 0. Moreover, it follows from the uniform law of large numbers and the continuity mapping theorem that

\[
\left| C_n(\hat{\beta}) \frac{1}{n} \sum_{i=1}^{n} h_i(\hat{\beta}) - C(\hat{\beta}) E\{h(\hat{\beta})\} \right| \rightarrow 0 \text{ almost surely.}
\]

Consequently, \( |C(\hat{\beta}) E\{h(\hat{\beta})\}|^2 \rightarrow 0 \), almost surely. Hence, \( \hat{\beta} \) converges to \( \beta_0 \), almost surely.
By Taylor expansion, we have

\[ h_n(\hat{\beta}) = h_n(\beta_0) + h_n(\beta_0)(\hat{\beta} - \beta_0), \tag{A.2} \]

where \( h_n = 1/n \sum_{i=1}^{n} \partial h_i(\beta)/\partial \beta \) and \( \hat{\beta} \) lies between \( \hat{\beta} \) and \( \beta_0 \). By multiplying (A.2) by \( h_n(\hat{\beta})^T V_n^{-1} \), we obtain

\[ h_n(\hat{\beta})^T V_n^{-1} h_n(\hat{\beta}) = h_n(\hat{\beta})^T V_n^{-1} h_n(\beta_0) + h_n(\hat{\beta})^T V_n^{-1} h_n(\beta_0)(\hat{\beta} - \beta_0). \tag{A.3} \]

Since the left hand side of (A.3) is 0 due to \( \hat{\beta} = \arg \min_\beta H_n(\beta) \), (A.3) can be rewritten as

\[ \sqrt{n}(\hat{\beta} - \beta_0) = -\{h_n(\hat{\beta})^T V_n^{-1} h_n(\beta)\}^{-1} h_n(\hat{\beta})^T V_n^{-1/2} \sqrt{n} V_n^{-1/2} h_n(\beta_0). \]

By the central limit theorem, \( \sqrt{n} V_n^{-1/2} h_n(\beta_0) \overset{d}{\rightarrow} N(0, I) \), where \( I \) is an identity matrix. Consequently, it follows that \( \sqrt{n}(\hat{\beta} - \beta_0) \overset{d}{\rightarrow} N(0, \Omega) \), where \( \Omega = \left\{ \begin{bmatrix} \frac{\partial h}{\partial \beta} \end{bmatrix}^T V^{-1} \begin{bmatrix} \frac{\partial h}{\partial \beta} \end{bmatrix} \right\}^{-1} \).

Next, we prove \( \Omega_l - \Omega \) is positive semi-definite. Note that the estimator based on the univariate procedure is obtained by minimizing \( h_n^T V_n^{-1} h_n \), where \( h_n = 1/n \sum_{i=1}^{n} (h_{n1}^T, h_{n2}^T)^T = (h_{n1}, h_{n2})^T \) and \( V_n = 1/n \sum_{i=1}^{n} h_i(\beta) h_i(\beta)^T \). We orthogonalize \( h_{n2}(\beta) \) against \( h_{n1}(\beta) \) as \( h_{n2}^o(\beta) = h_{n2}(\beta) - V_{11} V_{11}^{-1} h_{n1}(\beta) \) with \( V_{21} = E \{ h_{11}(\beta) h_{12}(\beta) \} \) and \( V_{11} = E \{ h_{11}(\beta) h_{11}(\beta)^T \} \), which satisfies \( E \{ h_{11}(\beta) h_{12}^o(\beta) \} = 0 \). Then the estimator \( \hat{\beta} \) can be obtained by minimizing \( h_n^o(\beta)^T V_n^{o-1} h_n^o(\beta) \), where \( h_n^o(\beta) = \{ h_{n1}(\beta)^T, h_{n2}^o(\beta)^T \}^T \) and \( V^n = E \{ h_n^o(\beta) h_n^o(\beta)^T \} \).

The information matrix of \( \hat{\beta} \) is proportional to

\[ h_n^o)^T V_n^{o-1} h_n^o = (h_{n1}^T, h_{n2}^o)^T \left( \begin{array}{c} V_{11}^{-1} \ 0 \\ 0 \ V_{22}^{o-1} \end{array} \right) \left( \begin{array}{c} h_{n1} \\ h_{n2}^o \end{array} \right) \tag{A.4} \]

\[ = h_{n1}^T V_{11}^{-1} h_{n1} + h_{n2}^{oT} V_{22}^{o-1} h_{n2}^o, \]

where \( V_{22}^{o} = E \{ h_{n2}^{o}(\beta) h_{n2}^{o}(\beta)^T \} \). Similarly, the information matrix of the estimator using the
univariate approach is \( h_{\alpha_1}^T V_{\alpha_1}^{-1} h_{\alpha_1} \), which is the first term of (A.4). Asymptotic estimation gain is achieved, since \( V_{\alpha_2} \) in (A.4) is positive semi-definite.

Proof of Theorem 2.2. We simplify the notations as

\[
\frac{\partial H_n}{\partial \beta} = \dot{H}, \; \frac{\partial^2 H_n}{\partial \beta^2} = \ddot{H}, \; \frac{\partial H_n}{\partial \alpha} = \dot{H}_\alpha, \; \frac{\partial^2 H_n}{\partial \alpha^2} = \ddot{H}_\alpha, \; \frac{\partial H_n}{\partial \gamma} = \dot{H}_\gamma, \; \frac{\partial^2 H_n}{\partial \alpha \partial \gamma} = \ddot{H}_\alpha\gamma, \; \frac{\partial^2 H_n}{\partial \gamma^2} = \dddot{H}_\gamma.
\]

Let \( \beta_0 = (\alpha_0, \gamma_0) \) be the true parameter, \( \bar{\gamma} = \arg\min H_n(\gamma|\alpha_0) \) and \( \hat{\beta} = \arg\min_\beta H_n(\beta) \).

By Taylor expansion, \( H_n(\gamma_0|\alpha_0) - H_n(\bar{\gamma}|\alpha_0) \) can be extended as

\[
H_n(\alpha_0, \gamma_0) - H_n(\alpha_0, \bar{\gamma}) = (\gamma_0 - \bar{\gamma})^T \dot{H}_\gamma(\alpha_0, \bar{\gamma}) + \frac{1}{2} (\gamma_0 - \bar{\gamma})^T \ddot{H}_\gamma(\alpha_0, \bar{\gamma})(\gamma_0 - \bar{\gamma}),
\]

where \( \bar{\gamma} \) lies between \( \gamma_0 \) and \( \bar{\gamma} \). Similarly, we expand \( H_n(\beta_0) - H_n(\hat{\beta}) \) to

\[
\begin{pmatrix}
\alpha_0 - \hat{\alpha} \\
\gamma_0 - \hat{\gamma}
\end{pmatrix}^T
\begin{pmatrix}
\dot{H}_\alpha(\hat{\alpha}, \hat{\gamma}) \\
\dot{H}_\gamma(\hat{\alpha}, \hat{\gamma})
\end{pmatrix} + \frac{1}{2}
\begin{pmatrix}
\alpha_0 - \hat{\alpha} \\
\gamma_0 - \hat{\gamma}
\end{pmatrix}^T
\begin{pmatrix}
\ddot{H}_\alpha(\hat{\alpha}, \hat{\gamma}) & \ddot{H}_\alpha(\hat{\alpha}, \hat{\gamma}) \\
\ddot{H}_\gamma(\hat{\alpha}, \hat{\gamma}) & \ddot{H}_\gamma(\hat{\alpha}, \hat{\gamma})
\end{pmatrix}
\begin{pmatrix}
\alpha_0 - \hat{\alpha} \\
\gamma_0 - \hat{\gamma}
\end{pmatrix},
\]

where \( (\hat{\alpha}, \hat{\gamma}) \) lies between \( (\alpha_0, \gamma_0) \) and \( (\hat{\alpha}, \hat{\gamma}) \). It follows from \( \dot{H}_\gamma(\alpha_0, \bar{\gamma}) = \dot{H}_\beta(\hat{\alpha}, \hat{\gamma}) = 0 \) that \( H_n(\gamma_0|\alpha_0) - H_n(\hat{\beta}) = \)

\[
\frac{1}{2}
\begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix}^T
\dot{H}(\hat{\alpha}, \hat{\gamma})
\begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix} - \frac{1}{2}
\begin{pmatrix}
0 \\
0
\end{pmatrix}^T
\ddot{H}(\alpha_0, \bar{\gamma})
\begin{pmatrix}
0 \\
0
\end{pmatrix}.
\]

By Taylor expansion, we obtain

\[
\dot{H}_\gamma(\alpha_0, \bar{\gamma}) = \dot{H}_\gamma(\alpha_0, \gamma_0) + \ddot{H}_\gamma(\bar{\gamma} - \gamma_0) + O_p(n^{-1}) = 0 \quad \text{and} \quad (A.5)
\]

\[
\dot{H}_\gamma(\hat{\alpha}, \hat{\gamma}) = \dot{H}_\gamma(\alpha_0, \gamma_0) + \ddot{H}_\gamma(\hat{\gamma} - \gamma_0) + \dddot{H}_\gamma(\hat{\alpha} - \alpha_0) + O_p(n^{-1}) = 0. \quad (A.6)
\]
By solving for \((\tilde{\gamma} - \gamma_0)\) from (A.5) and (A.6), we have

\[
\begin{pmatrix}
0 \\
\tilde{\gamma} - \gamma_0
\end{pmatrix} = \begin{pmatrix}
0 & 0 \\
H_{\gamma\gamma}^{-1}H_{\gamma\alpha} & I
\end{pmatrix} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix} + O_p(n^{-1}).
\]

Hence, \(H_n(\alpha_0, \tilde{\gamma}) - H_n(\hat{\alpha}, \hat{\gamma})\) is specified as

\[
\frac{1}{2} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix}^T \left\{ \hat{H}(\alpha, \tilde{\gamma}) - \begin{pmatrix}
0 & \hat{H}_\alpha \hat{H}_{\gamma\gamma}^{-1} \\
0 & I
\end{pmatrix} \hat{H}(\alpha_0, \tilde{\gamma}) \begin{pmatrix}
0 & 0 \\
\hat{H}_{\gamma\gamma}^{-1}H_{\gamma\alpha} & I
\end{pmatrix} \right\} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix}
\]

\[
= \frac{1}{2} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix}^T \begin{pmatrix}
\hat{H}_{\alpha\alpha} - \hat{H}_{\alpha\gamma} \hat{H}_{\gamma\gamma}^{-1} \hat{H}_{\gamma\alpha} & 0 \\
0 & 0
\end{pmatrix} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix}
\]

\[
= \frac{1}{2} \left( \hat{\alpha} - \alpha_0 \right)^T \left( \hat{H}_{\alpha\alpha} - \hat{H}_{\alpha\gamma} \hat{H}_{\gamma\gamma}^{-1} \hat{H}_{\gamma\alpha} \right) (\hat{\alpha} - \alpha_0). \tag{A.7}
\]

It follows from Theorem 2.1, \(\hat{H} \rightarrow 2\Psi\) and \(\Psi = \Omega^{-1}\) that

\[
\sqrt{n} \begin{pmatrix}
\hat{\alpha} - \alpha_0 \\
\hat{\gamma} - \gamma_0
\end{pmatrix} \overset{d}{\rightarrow} N \left( 0, \Psi^{-1} \right) = N \left( \begin{pmatrix}
0 \\
0
\end{pmatrix}, \begin{pmatrix}
\psi_{\alpha\alpha} & \psi_{\alpha\gamma} \\
\psi_{\gamma\alpha} & \psi_{\gamma\gamma}
\end{pmatrix}^{-1} \right)
\]

\[
\sqrt{n}(\hat{\alpha} - \alpha_0) \overset{d}{\rightarrow} N \left( 0, (\psi_{\alpha\alpha} - \psi_{\alpha\gamma} \psi_{\gamma\gamma}^{-1} \psi_{\gamma\alpha})^{-1} \right). \tag{A.8}
\]

The proof is completed by (A.7), (A.8), Slutsky’s theorem, and Theorem 10.2d in Arnold (1981).

\(\square\)

**Proof of Theorem 2.3.** We recall \(w_i = n^{-1} \{ 1 + \lambda^T h_i(\tilde{\beta}) \}^{-1}\) and formulate

\[
S_{n1}(\beta, \lambda) = \sum_{i=1}^{n} w_i h_i(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \lambda^T h_i(\tilde{\beta})} h_i(\beta)
\]

and

\[
21
\]
\[ S_{n2}(\beta, \lambda) = \sum_{i=1}^{n} w_i \left\{ \frac{\partial h_i(\beta)}{\partial \beta} \right\}^T \lambda = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \lambda^T h_i(\beta)} \left\{ \frac{\partial h_i(\beta)}{\partial \beta} \right\}^T \lambda. \]

By Taylor expansion, we have

\[ 0 = S_{n1}(\tilde{\beta}, \tilde{\lambda}) = S_{n1}(\beta_0, 0) + \frac{\partial S_{n1}(\beta_0, 0)}{\partial \beta} (\tilde{\beta} - \beta_0) + \frac{\partial S_{n1}(\beta_0, 0)}{\partial \lambda} (\tilde{\lambda} - 0) + O_p(n^{-1/2}) \quad (A.9) \]

\[ 0 = S_{n2}(\tilde{\beta}, \tilde{\lambda}) = S_{n2}(\beta_0, 0) + \frac{\partial S_{n2}(\beta_0, 0)}{\partial \beta} (\tilde{\beta} - \beta_0) + \frac{\partial S_{n2}(\beta_0, 0)}{\partial \lambda} (\tilde{\lambda} - 0) + O_p(n^{-1/2}), \]

where \( \tilde{\beta} = \arg \max_{\beta} L_n(\beta) \) and \( \tilde{\lambda} \) is a solution to \( n^{-1} \sum_{i=1}^{n} h_i(\tilde{\beta}) \{ 1 + \lambda^T h_i(\tilde{\beta}) \}^{-1} = 0 \). By rearranging (A.9) and (A.10), the estimators \((\tilde{\beta}, \tilde{\lambda})\) can be expressed as

\[ \begin{pmatrix} \tilde{\lambda} \\ \tilde{\beta} - \beta_0 \end{pmatrix} = Q_n^{-1} \begin{pmatrix} -S_{n1}(\beta_0, 0) + O_p(n^{1/2}) \\ O_p(n^{1/2}) \end{pmatrix}, \]

where

\[ Q_n = \begin{pmatrix} \frac{\partial S_{n1}}{\partial \lambda}(\beta_0, 0) & \frac{\partial S_{n1}}{\partial \beta}(\beta_0, 0) \\ \frac{\partial S_{n2}}{\partial \lambda}(\beta_0, 0) & 0 \end{pmatrix} = \begin{pmatrix} -V_n(\beta_0) & \hat{h}_n(\beta_0) \\ \hat{h}_n(\beta_0)^T & 0 \end{pmatrix}. \]

By formulating for an inverse block matrix and the central limit theorem, we then obtain

\[ \sqrt{n}(\tilde{\beta} - \beta_0) = \{ \hat{h}_n(\beta_0)^T V_n(\beta_0)^{-1} \hat{h}_n(\beta_0) \}^{-1} \hat{h}_n(\beta_0)^T V_n(\beta_0)^{-1/2} \sqrt{n} V_n(\beta_0)^{-1/2} h_n(\beta_0) + o_p(1) \xrightarrow{d} N(0, \Omega), \]

where \( \Omega = \left\{ E \left( \frac{\partial h}{\partial \beta} \right)^T V^{-1} E \left( \frac{\partial h}{\partial \beta} \right) \right\}^{-1} \). The proof of estimation efficiency is the same as that of Theorem 2.1. Thus, it is omitted.

\[ \square \]
Proof of Theorem 2.4. By Taylor expansion and

\[ \lambda = \left\{ \frac{1}{n} \sum_{i=1}^{n} h_i(\beta) h_i(\beta)^{\top} \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} h_i(\beta) \right\} + o_p(1), \]

we have

\[
2 \log \{ L_n(\beta) \} = -2 \sum_{i=1}^{n} \log \{ 1 + \lambda^T h_i(\beta) \} \\
= 2 \sum_{i=1}^{n} \lambda^T h_i(\beta) - \sum_{i=1}^{n} \{ \lambda^T h_i(\beta) \}^2 + o_p(1) \\
= -n h_n(\beta)^T V_n^{-1} h_n(\beta) + o_p(1) \\
= -n H_n + o_p(1).
\]

Thus the log-empirical likelihood ratio test statistic is rewritten as

\[
W_n = 2[\log \{ L_n(\tilde{\beta}) \} - \log \{ L_n(\tilde{\gamma}|\alpha_0) \}] = n\{H_n(\tilde{\gamma}|\alpha_0) - H_n(\tilde{\beta})\} + o_p(1) = T_n + o_p(1).
\]

Hence, \( W_n \) follows the asymptotic distribution of \( T_n \) and the rest of the proof is same as that of Theorem 2.2.

\[ \Box \]
Table 2.1: Relative efficiency of proposed estimators obtained from generalized method moment (GMM) and empirical likelihood (EL).

| Type      | Method | Equal, 0.5 | | Equal, 0.8 | | Unequal |
|-----------|--------|------------|------------|------------|------------|
|           |        | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| Binary    | GMM    | 1.08       | 1.11       | 1.11       | 1.55       | 1.31       | 1.25       | 1.28       | 1.20       | 1.24       |
|           | EL     | 1.10       | 1.15       | 1.16       | 1.48       | 1.33       | 1.32       | 1.25       | 1.24       | 1.32       |
| Mixture   | GMM    | 1.08       | 1.44       | 1.44       | 1.26       | 2.60       | 2.52       | 1.15       | 2.11       | 2.27       |
|           | EL     | 1.14       | 1.44       | 1.43       | 1.33       | 2.62       | 2.48       | 1.19       | 2.07       | 2.24       |
| Normal    | GMM    | 1.34       | 1.32       | 1.37       | 3.00       | 2.89       | 2.88       | 1.35       | 1.40       | 1.44       |
|           | EL     | 1.37       | 1.31       | 1.38       | 2.96       | 2.86       | 2.92       | 1.35       | 1.41       | 1.44       |
| Normal    | GMM    | 1.29       | 1.35       | 1.38       | 2.95       | 3.24       | 3.26       | 1.36       | 1.44       | 1.52       |
| with z_i  | EL     | 1.30       | 1.34       | 1.38       | 2.95       | 3.22       | 3.19       | 1.35       | 1.45       | 1.54       |
Table 2.2: Proportions of times that $H_0 : \beta_j = 0, j = 1, 2, 3,$ is rejected in 1000 simulation runs through Wald test of univariate approach and chi-squared test of generalized method moment (GMM) and empirical likelihood (EL).

<table>
<thead>
<tr>
<th>Type</th>
<th>Method</th>
<th>Equal, 0.5</th>
<th></th>
<th>Equal, 0.8</th>
<th></th>
<th>Unequal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>$\beta_2$</td>
<td>$\beta_3$</td>
<td>$\beta_1$</td>
<td>$\beta_2$</td>
</tr>
<tr>
<td>Binary</td>
<td>GMM</td>
<td>0.15</td>
<td>0.40</td>
<td>0.74</td>
<td>0.19</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>EL</td>
<td>0.22</td>
<td>0.47</td>
<td>0.77</td>
<td>0.24</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Univariate</td>
<td>0.12</td>
<td>0.34</td>
<td>0.64</td>
<td>0.11</td>
<td>0.32</td>
</tr>
<tr>
<td>Mixture</td>
<td>GMM</td>
<td>0.15</td>
<td>0.99</td>
<td>1.00</td>
<td>0.15</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>EL</td>
<td>0.21</td>
<td>0.98</td>
<td>1.00</td>
<td>0.21</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Univariate</td>
<td>0.13</td>
<td>0.97</td>
<td>1.00</td>
<td>0.13</td>
<td>0.97</td>
</tr>
<tr>
<td>Normal</td>
<td>GMM</td>
<td>0.60</td>
<td>0.98</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>EL</td>
<td>0.65</td>
<td>0.98</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Univariate</td>
<td>0.50</td>
<td>0.97</td>
<td>1.00</td>
<td>0.50</td>
<td>0.98</td>
</tr>
<tr>
<td>Normal</td>
<td>GMM</td>
<td>0.68</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>with $z_i$</td>
<td>EL</td>
<td>0.62</td>
<td>0.99</td>
<td>1.00</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Univariate</td>
<td>0.49</td>
<td>0.98</td>
<td>1.00</td>
<td>0.48</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Table 2.3: Description of variables in MSCM data set.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>household size</td>
<td>1=more than 3 people, 0=2 or 3 people</td>
</tr>
<tr>
<td>race</td>
<td>1=non-white, 0=white</td>
</tr>
<tr>
<td>children's health level</td>
<td>0=very poor/poor, 1=fair, 2=good, 3=very good</td>
</tr>
<tr>
<td>gender</td>
<td>1=female, 0=male</td>
</tr>
<tr>
<td>mother's health level</td>
<td>0=very poor/poor, 1=fair, 2=good, 3=very good</td>
</tr>
<tr>
<td>education level</td>
<td>1=high school graduate, 0=less than high school</td>
</tr>
<tr>
<td>employment status</td>
<td>1=employed, 0=unemployed</td>
</tr>
<tr>
<td>marriage status</td>
<td>1=married, 0=other</td>
</tr>
</tbody>
</table>
Table 2.4: Estimated coefficients along with confidence intervals at a 95% level in brackets with * when zero is not included using generalized method moment (GMM), empirical likelihood (EL) and univariate approach.

<table>
<thead>
<tr>
<th>Response Covariate</th>
<th>GMM</th>
<th>EL</th>
<th>Univariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illness intercept</td>
<td>0.225</td>
<td>0.228</td>
<td>0.215</td>
</tr>
<tr>
<td>size</td>
<td>-0.075</td>
<td>-0.074</td>
<td>-0.065</td>
</tr>
<tr>
<td>race</td>
<td>0.055</td>
<td>0.052</td>
<td>0.057</td>
</tr>
<tr>
<td>health</td>
<td>(-0.063, -0.024)*</td>
<td>(-0.063, -0.024)*</td>
<td>(-0.062, -0.019)*</td>
</tr>
<tr>
<td>gender</td>
<td>0.014</td>
<td>0.020</td>
<td>0.018</td>
</tr>
<tr>
<td>Stress intercept</td>
<td>0.175</td>
<td>0.173</td>
<td>0.171</td>
</tr>
<tr>
<td>size</td>
<td>-0.052</td>
<td>-0.048</td>
<td>-0.043</td>
</tr>
<tr>
<td>race</td>
<td>0.006</td>
<td>0.009</td>
<td>0.021</td>
</tr>
<tr>
<td>health</td>
<td>(-0.053, -0.017)*</td>
<td>(-0.054, -0.018)*</td>
<td>(-0.061, -0.021)*</td>
</tr>
<tr>
<td>education</td>
<td>0.039</td>
<td>0.043</td>
<td>0.049</td>
</tr>
<tr>
<td>employ</td>
<td>-0.043</td>
<td>-0.044</td>
<td>-0.028</td>
</tr>
<tr>
<td>marry</td>
<td>0.036</td>
<td>0.037</td>
<td>0.031</td>
</tr>
</tbody>
</table>

27
Chapter 3

Efficient Regression Modeling for Correlated and Overdispersed Count Data

3.1 Introduction

Longitudinal count data arise frequently in transportation safety studies, where measurements on each segment are repeatedly tracked over time. For example, Michigan transportation crash data consist of the number of crashes and various features collected annually from the same segment of the road. It is of particular interest to build a regression model that allows us to investigate causes of car crashes and to anticipate the number of crashes. If overdispersion is negligible, a Poisson regression model is often applied for analysis of count data by assuming that the marginal mean and variance of crash frequencies are equal. When overdispersion is present, which commonly exists in transportation crash studies, the Poisson regression might be too restrictive due to the above-mentioned assumption. A negative binomial regression can be a viable alternative; it tackles the overdispersion problem by accommodating an overdispersion parameter in
regression modeling. In many recent applications to crash frequency analysis, negative binomial regression has been used; see Miaou (1994), Shankar et al., (1995), Poch and Mannering (1996), and Abdel-Aty and Radwan (2000). However, most of the aforementioned works are based on the typical generalized linear modeling which does not take into account correlations among the measurements within the same segment and thereby may result in loss of estimation efficiency.

Recently, several authors have published likelihood-based models for longitudinal count data with overdispersion, and the overdispersion parameter is modeled using a random mean approach (Toscas and Faddy, 2003; Fahrmeir and Echacarria, 2006; Czado and Kolbe, 2007; Molenberghs et al., 2007; Fotouhi, 2008; Grunwald et al., 2011). However, their approaches require to provide a full likelihood model. On the other hand, non-likelihood approaches have also been employed for longitudinal count data. Generalized estimating equations (Liang and Zeger, 1986) and quadratic inference functions (Qu et al., 2000) are commonly used under a marginal mean regression framework. The generalized estimating equations require the first two moments and estimation of a few nuisance parameters associated with a working correlation structure. This yields an asymptotically consistent estimator regardless of the working correlation structure, yet it can be inefficient under the misspecified working correlation structure. To tackle this problem, Qu et al., (2000) extended the inverse of the working correlation structure to a set of predetermined matrices, which avoids additional estimation of the nuisance parameters. The quadratic inference functions can obtain a more efficient estimator than the one ignoring the within-subject correlation, even if the assumed working correlation structure is incorrect.

Thall and Vail (1990) utilized a mixed effects approach to build covariance models and estimated the covariance components through the method of moments. However, the application of their approach is restrictive due to the limited formulation of the covariance structure. Jowaheer and Sutradhar (2002) proposed a joint generalized estimating equa-
tions approach based on a general autocorrelation structure, which is obtained by applying
the Gaussian working matrices. However, their estimation procedure is challenging in prac-
tical applications, because the working correlation structure is complicated especially for
time-dependent covariates. In addition, both of the aforementioned works modeled re-
gression and overdispersion parameters jointly, which makes it necessary to estimate more
nuisance parameters associated with the covariance structure.

In this article, we extend the matrix expansion idea of the quadratic inference functions
to the negative binomial regression framework. This enables us to facilitate various types
of correlation structures without estimating additional nuisance parameters associated with
the covariance structure. In addition, the proposed approach is readily applicable for both
time-independent and time-dependent covariates. Our theoretical investigations indicate
that the proposed estimator is asymptotically more efficient than the one obtained un-
der the Poisson regression framework. Simulation studies also confirm that the proposed
approach outperforms the quadratic inference functions based on the Poisson regression
in all cases with overdispersion. On the other hand, inclusion of the overdispersion pa-
rameter can hinder efficient estimation and inference for the regression parameter when
overdispersion has not occurred as shown in our simulation studies. To recommend a more
proper model for longitudinal count data, we further propose a hypothesis test detecting
the presence of overdispersion. The theoretical study demonstrates that the proposed test
statistic asymptotically follow a chi-square distribution when the overdispersion is absent.
Extensive simulation studies confirm that the proposed test is powerful and identifies a
more suitable model successfully in all cases under consideration. The proposed approach
is also utilized on a transportation crash study and recommends the negative binomial
regression model by confirming the presence of overdispersion.

The paper is organized as follows. In Section 3.2, we propose a new estimation ap-
proach under the negative binomial regression framework and provide a hypothesis test de-
Section 3.3 reports the simulation studies for correlated count data with or without overdispersion. Section 3.4 provides the analysis of the above-mentioned transportation crash data. The theoretical proofs and regularity conditions are placed in Section 3.5.

3.2 Methodologies

3.2.1 Negative Binomial Model for Longitudinal Data

Suppose that $Y_{ij}$ is the count response and $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$ is a $p$-dimensional vector of covariates for subject $i$ measured at time $j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$, where $t_1, \ldots, t_n$ are the number of measurements from the $n$ subjects. Since each subject is repeatedly measured $t_i$ times, these measurements are likely to be correlated and thereby assumed to be dependent, while subjects are independent.

To fit the regression model to count data, the Poisson distribution of $Y_{ij}$ can be considered as

$$f_p(Y_{ij}) = \frac{\theta_i^{Y_{ij}}}{Y_{ij}!} \exp(-\theta_i),$$

where $f_p(\cdot)$ is the probability mass function, $\theta_i = \exp(x_i'\beta)$, and $\beta$ is a $p$-dimensional regression parameter vector of interest. The key feature of this model is that the variance of a response variable equals the mean of the response, i.e., $E(Y_{ij}) = \text{var}(Y_{ij}) = \theta_i$, which is so-called non-overdispersion.

When the response variable is subject to overdispersion, where the variance of the response is greater than the mean, the Poisson regression model might be too restrictive due to the fundamental assumption. A negative binomial regression model is a viable alternative; it accommodates an overdispersion parameter in the model by denoting the
probability mass function of the response variable as

\[ f_{nb}(Y_{ij}) = \frac{\Gamma(\gamma^{-1} + Y_{ij})}{\Gamma(\gamma^{-1}) Y_{ij}!} \left( \frac{1}{1 + \gamma \theta_{ij}} \right)^{\gamma^{-1}} \left( \frac{\gamma \theta_{ij}}{1 + \gamma \theta_{ij}} \right)^{Y_{ij}} \] (3.2)

with \( \theta_{ij} = \exp(x_{ij}'\beta) \) and an overdispersion parameter of \( \gamma \). Note that the overdispersion parameter is a non-negative value and influences the variance as \( \text{var}(Y_{ij}) = \theta_{ij} + \gamma \theta_{ij}^2 \), while the mean still remains \( E(Y_{ij}) = \theta_{ij} \). As \( \gamma \to 0 \), the negative binomial distribution converges to the Poisson distribution. Basically, there are two well-known forms of the negative binomial models namely, the linear (NB1) and the quadratic (NB2) negative binomial models (Cameron and Trivedi, 1986). NB1 has the variance function appearing as \( \theta_{ij} + \gamma \theta_{ij} \) and NB2 owns form \( \theta_{ij} + \gamma \theta_{ij}^2 \). The form of parameterization of the model (3.2) is NB2, which is typically chosen by researchers (Greene, 2007). Since modeling the mean and variance are interdependent, we propose to develop an estimation approach for longitudinal data with overdispersion that simultaneously estimates the regression parameters of our interest as well as the overdispersion parameter.

For joint estimation of \( \beta \) and \( \gamma \), we develop the joint estimating equations by employing the first two moments of the response variable as

\[ \sum_{i=1}^{n} \hat{\mu}_i A_i^{-1/2} \Phi_i^{-1} A_i^{-1/2} (f_i - \mu_i) = 0 \] (3.3)

where \( f_i = (Y_{i1}, \ldots, Y_{it_i}, Y_{i1}, \ldots, Y_{it_i})' \) and \( \mu_i = \{ E(Y_{i1}), \ldots, E(Y_{it_i}), E(Y_{i1}^2), \ldots, E(Y_{it_i}^2) \} \) are \( 2t_i \)-dimensional vectors, \( \hat{\mu}_i = \partial \mu_i / \partial \zeta \) is a \( 2t_i \times (p + 1) \) matrix with \( \zeta = (\beta', \gamma)' \). \( A_i \) is a \( 2t_i \times 2t_i \) diagonal variance matrix of \( f_i \) with \( \text{var}(Y_{ij}) = (\theta_{ij} + \gamma \theta_{ij}^2) \) and \( \text{var}(Y_{ij}^2) = \theta_{ij} + (6 + 7\gamma) \theta_{ij}^2 + (4 + 16\gamma + 12\gamma^2) \theta_{ij}^3 + (4\gamma + 10\gamma^2 + 6\gamma^3) \theta_{ij}^4 \) for \( j = 1, \ldots, t_i \), and \( \Phi_i \) is a true correlation matrix of \( f_i \) which is unknown in practice.

Following Liang and Zeger (1986), we replace \( \Phi_i \) in (3.3) with a working correlation structure, denoted by \( R_i(\alpha) \) with a nuisance parameter vector of \( \alpha \), and provide the
following joint estimating equations

\[ \sum_{i=1}^{n} \mu_i' A_i^{-1/2} R_i(\alpha)^{-1} A_i^{-1/2} (\beta_i - \mu_i) = 0. \]  

(3.4)

In order to obtain the estimators of \( \beta \) and \( \gamma \), specification of the working correlation structure and estimation of \( \alpha \) in \( R_i(\alpha) \) are required prior to solving (3.4). This can be challenging because four pieces of correlation information in \( R_i(\alpha) \) are taken into account; \( \text{cor}(Y_{ij}, Y_{ij}^2) \), \( \text{cor}(Y_{ij}, Y_{iw}) \), \( \text{cor}(Y_{ij}^2, Y_{iw}^2) \), and \( \text{cor}(Y_{ij}, Y_{iw}^2) \) for \( j \neq w \). In this case, an assumption of a simple working correlation structure such as compound symmetry or the first-order autoregressive, denoted by AR(1), might cause a loss of estimation efficiency.

If the unspecified working correlation structure is assumed, then there are \( 2t_i^2 - t_i \) nuisance parameters of \( \alpha \) in \( R_i(\alpha) \) and estimation of \( \alpha \) could be problematic with a large size of \( t_i \).

To tackle this problem, Qu et al., (2000) employed a class of basis matrices to represent the inverse of \( R_i(\alpha) \) in (3.4) as

\[ R_i(\alpha)^{-1} = \sum_{m=1}^{k} b_{im} B_{im}, \]  

(3.5)

where \( B_{im} \) is the so-called basis matrix and \( b_{im} \) is an unknown constant. The choice of basis matrices will be discussed in Section 3.2.2. With the basis matrices, we develop the following inference function

\[ H_n(\beta, \gamma) = \tilde{h}_n(\beta, \gamma)' G_n(\beta, \gamma)^{-1} \tilde{h}_n(\beta, \gamma), \]  

(3.6)

where \( G_n(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} h_i(\beta, \gamma) h_i(\beta, \gamma)' \) and \( \tilde{h}_n(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} h_i(\beta, \gamma) \) with

\[ h_i(\beta, \gamma) = \begin{pmatrix} \mu_i' A_i^{-1/2} B_{i1} A_i^{-1/2} (f_i - \mu_i) \\ \vdots \\ \mu_i' A_i^{-1/2} B_{ik} A_i^{-1/2} (f_i - \mu_i) \end{pmatrix}. \]  

(3.7)
The estimator of $\beta$ and $\gamma$ are obtained simultaneously by minimizing $H_n(\beta, \gamma)$ as $(\hat{\beta}, \hat{\gamma}) = \arg\min_{\beta, \gamma} H_n(\beta, \gamma)$. Remark that the inference function (3.6) does not include the nuisance parameters associated with the working correlation structure.

Under the regularity conditions in Section 3.5, the asymptotic distribution of the proposed estimators is multivariate normal with the true mean vector of $(\beta, \gamma)$ and covariance matrix $\Sigma_{nb} = (\Psi'G^{-1}\Psi)^{-1}$ with $\Psi = E\{\partial h_i(\beta, \gamma)/\partial (\beta, \gamma)\}$ and $G = E\{h_i(\beta, \gamma)h_i(\beta, \gamma)'\}$. In addition, the estimator is the most efficient among the estimators obtained from the same linear class of (3.7) because $\Sigma_{nb}$ reaches the minimum in the sense of Loewner ordering. Moreover, the asymptotic variance of the proposed estimator is no greater than that of the estimator obtained by quadratic inference functions under the Poisson regression framework (Qu et al., 2000), regardless of whether overdispersion is present or not. This ensures that the proposed joint estimation procedure gains estimation efficiency of the regression parameters by not only incorporating the within subject correlation, but also accounting for overdispersion. The proof of the aforementioned asymptotic results is placed in Section 3.5. For statistical inference on $(\beta, \gamma)$, the asymptotic covariance matrix can be evaluated through a plug-in estimation procedure as

$$
\hat{\Sigma}_{nb} = \left[\left\{\frac{1}{n} \sum_{i=1}^{n} \partial h_i(\hat{\beta}, \hat{\gamma}) \right\}' \left\{\frac{1}{n} \sum_{i=1}^{n} h_i(\hat{\beta}, \hat{\gamma})h_i(\hat{\beta}, \hat{\gamma}') \right\}^{-1} \left\{\frac{1}{n} \sum_{i=1}^{n} \partial h_i(\hat{\beta}, \hat{\gamma}) \right\} \right]^{-1}.
$$

### 3.2.2 Choice of Basis Matrices

The choice of valid basis matrices plays an important role in estimation efficiency of the regression parameters in (3.6), since the inverse of the working correlation matrix $R_i(\alpha)$ is presented by a linear combination of the basis matrices in (3.5). In this section, we provide a set of basis matrices that approximates the true correlation structure $\Phi_i$ in (3.3). We denote the true correlation matrix of $Y_i$ as $C_i$ and separate $\Phi_i$ into two pattern matrices, $\Phi_i = Q \otimes C_i + D$, where $Q$ is a $2 \times 2$ matrix with 1 on diagonal and an unknown positive
constant on off-diagonal, $D$ is the remainder of the correlation, and $\otimes$ is the Kronecker product. Note that all elements in $D$ are either close to 0 or exactly 0, which implies that the contribution of the second term is negligible. Therefore, we represent the working correlation structure as $R_i(\alpha) = Q \otimes C_i$ and decompose the inverse of $R_i(\alpha)$ in (3.4) as

$$R_i(\alpha)^{-1} = Q^{-1} \otimes C_i^{-1} = (a_1 M_1 + a_2 M_2) \otimes C_i^{-1},$$

where $M_1$ is a $2 \times 2$ identity matrix, $M_2$ is a $2 \times 2$ matrix with 0 on diagonal and 1 elsewhere, and $a_1$ and $a_2$ are unknown constants. The set of basis matrices is finally determined by the type of the correlation structure, $C_i$, which is often figured out with ease.

For instance, if the measurement times are equispaced and the closer measurements within a subject are more likely to be dependent, as in our real data, $C_i$ is generally assumed to be AR(1). Consequently, the inverse of $C_i$ is decomposed by $C_i^{-1} = \delta_{13} W_{11} + \delta_{12} W_{12} + \delta_{13} W_{13}$, where $W_{11}$ is an identity matrix, $W_{12}$ is a matrix with 1 on the sub-diagonals and 0 elsewhere, and $W_{13}$ is a matrix with 1 on the corners $(1,1)$ and $(t_i,t_i)$ and 0 elsewhere. Since the basis matrix $W_{13}$ involves many 0’s, a loss of estimation efficiency is negligible after omitting $W_{13}$ (Qu et al., 2000). Therefore, the inverse of $R_i(\alpha)$ is represented with four basis matrices,

$$R_i(\alpha)^{-1} = (a_1 M_1 + a_2 M_2) \otimes (\delta_{13} W_{11} + \delta_{12} W_{12})$$

$$= a_1 \delta_{13} (M_1 \otimes W_{11}) + a_2 \delta_{12} (M_2 \otimes W_{12}) + a_1 \delta_{12} (M_1 \otimes W_{12}) + a_2 \delta_{12} (M_2 \otimes W_{12})$$

$$= b_{i1} B_{i1} + b_{i2} B_{i2} + b_{i3} B_{i3} + b_{i4} B_{i4}.$$
modeled as

\[ R_i(\alpha)^{-1} = (a_1 M_1 + a_2 M_2) \otimes (\delta_{i4} W_{i1} + \delta_{i5} W_{i4}) \]

\[ = a_1 \delta_{i4} (M_1 \otimes W_{i1}) + a_2 \delta_{i4} (M_2 \otimes W_{i1}) + a_1 \delta_{i5} (M_1 \otimes W_{i4}) + a_2 \delta_{i5} (M_2 \otimes W_{i4}) \]

\[ = b_{i5} B_{i1} + b_{i6} B_{i2} + b_{i7} B_{i5} + b_{i8} B_{i6}. \]

### 3.2.3 Test for Detecting Overdispersion

In the analysis of count data with overdispersion, Poisson regression modeling results in loss of estimation efficiency of the regression parameter as shown in Section 3.2.1. On the contrary, it is noteworthy that inclusion of the overdispersion parameter under the negative binomial model framework may hinder efficient estimation when overdispersion is absent in practice. Therefore, it is essential to choose a parsimonious model that provides a good fit for longitudinal count data. In this section, we develop a hypothesis test for detecting whether overdispersion is significant. If the overdispersion is present, the proposed estimation approach under the negative binomial regression in (3.2) is the optimal choice for modeling. Otherwise the Poisson regression in (3.1) is a proper model with the equal mean and variance assumption. The proposed hypothesis test statement is

\[ H_0: \text{Poisson regression model vs. } H_A: \text{Negative binomial regression model}. \]

Since the inference function in (3.6) plays a similar role as the log-likelihood ratio test function, we propose a test statistic for these hypotheses based on \( H_n(\beta, \gamma) \). We remark that the Poisson regression is recommended in cases where the overdispersion parameter is zero. In this manner, the test statistic is formulated as

\[ T_n = n \left\{ H_n(\bar{\beta}, \gamma = 0) - H_n(\bar{\beta}, \hat{\gamma}) \right\}, \quad (3.8) \]
where $\tilde{\beta} = \arg\min_{\beta} H_n(\beta | \gamma = 0)$ and $(\tilde{\beta}, \tilde{\gamma}) = \arg\min_{\beta, \gamma} H_n(\beta, \gamma)$. The following theorem shows that $T_n$ possesses the same chi-square asymptotic properties as the likelihood ratio test, yet it does not require specifying the likelihood function, which is challenging especially for correlated count data.

**Theorem 3.1.** Under the regularity conditions in Section 3.5, $T_n \overset{d}{\rightarrow} \chi_1^2$ under $H_0$ as $n \to \infty$.

This theorem illustrates that the proposed test statistic asymptotically follows the chi-square distribution with one degrees of freedom under the null hypothesis. The proposed test statistic can be readily extended to test whether a set of covariates are statistically significant or not. The proof of Theorem 3.1 is provided in Section 3.5.

### 3.3 Simulation Studies

In this section, we evaluate the proposed methodology through 1000 simulated datasets with or without overdispersion. The correlated count response variable is modeled as $\theta_{ij} = \exp(2 + x_{ij}\beta)$ with $\beta = (\beta_1, \beta_2)' = (1, 0.5)'$ and $x_{ij} = (x_{ij1}, x_{ij2})$ for $i = 1, \ldots, 300$ and $j = 1, \ldots, 5$. Each covariate $x_{ijk}$, $k = 1, 2$, is independently and uniformly distributed on $(0, 2)$. In order to generate the correlated and overdispersed count data, the R package `corcounts` is applied by treating the true covariance matrix of the responses as an AR(1) with a correlation coefficient of 0.8 and $\gamma = 0$ and 0.1 as the overdispersion parameter.

We estimate the regression parameters and overdispersion parameter through the proposed negative binomial regression approach under the three different working correlation structures of the response variables: AR(1), compound symmetry, and independent. Table 3.1 provides mean squared errors of the estimators, i.e., $\text{MSE}(\hat{\beta}_q) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}_q^{(i)} - \beta_q)^2$, where $\hat{\beta}_q^{(i)}$, $q = 1, 2$, is the estimate for the $i$th simulation. The results confirm that the mean squared errors are smallest when the AR(1) working correla-
tion structure is employed regardless of the presence of overdispersion in the data. Even if the working correlation structure is misspecified, the proposed approach still outperforms the one ignoring the correlation in terms of smaller mean square errors. We also compute 95% confidence intervals and report the average length of the confidence intervals and the proportion of times the interval includes the true parameter in Table 3.1. All coverage proportions are similar and close to 0.95. This might lead us to conclude that the proposed approach performs similarly regardless of the working correlation structure, however the confidence intervals obtained under the AR(1) and compound symmetry are narrower than those under the independent working structure. This indicates that the proposed approach yields a more efficient and precise estimate by accommodating the within-subject correlation.

The proposed approach is also compared with the Poisson regression approach under the three different working correlation structures: AR(1), compound symmetry, and independent. Table 3.2 provides the relative efficiency of the proposed estimator using the negative binomial approach to one using the Poisson regression, defined by the ratio of the mean squared errors. The larger the value of the relative efficiency, the more efficient the proposed estimator. When overdispersion is present, i.e., $\gamma = 0.1$, the proposed approach outperforms the Poisson approach since the relative efficiencies are greater than one in all cases under consideration. This suggests that estimation efficiency is gained when the response variable is subject to overdispersion.

We further examine the performance of two regression approaches when overdispersion is absent. According to the results in Table 3.2, the Poisson regression approach outperforms the proposed approach regardless of the working correlation structure. This motivates us to conduct the hypothesis test for $H_0 : \text{Poisson regression}$ against $H_A : \text{negative binomial regression}$. We compute the test statistic and reject the null hypothesis if the $p$-value is less than a significance level of 0.05 for the 1000 simulations. When $\gamma = 0$,
the rejection rates are all close to the nominal level of 0.05: 0.044, 0.045, and 0.057 under AR(1), compound symmetry, and independent working correlation structure, respectively.

Figure 3.1 also confirms that the test statistic under the AR(1) follows the chi-square distribution with one degree of freedom sufficiently well when the null hypothesis is true. Moreover, we investigate the power of the proposed test as shown in Figure 3.2. The results show that the rejection rate increases rapidly with the value of the overdispersion parameter and reaches one when $\gamma = 0.006$ under the AR(1) and compound symmetry structures, while the power assuming the working independence is always lower. The results confirm that the proposed test provides the more appropriate regression model for correlated count data successfully.

### 3.4 Application to Transportation Crash Data

In this section, we analyze transportation crash data consisting of 7233 segments from Michigan Traffic Management Center areas to investigate the causes of crashes. The number of crashes and various characteristics at each segment were collected from 2008 to 2013 annually. In the analysis of the data, we treat the number of crashes as the response variable and consider nine features as independent variables. The length of the segment, number of lanes, number of CCTV cameras (CCTV), number of dynamic message signs (DMS), number of microwave vehicle detection systems (MVDS), and annual average daily traffic (AADT) are all numerical variables. The presence of a passing lane, presence of a median, and presence of a shoulder are all categorical variables. The description of these variables are provided in Table 3.3. Here, the AR(1) is considered as the working correlation structure since the equispaced measurement times and the nature of the crash counts measure suggest auto-correlative dependence.

We first conduct the hypothesis test to see if overdispersion is present in the data.
A large value of the test statistic, $T_n = 469$, along with a $p$-value = 0 confirms the presence of overdispersion and therefore suggests the negative binomial regression model. For the purpose of comparison, we also assess the regression parameters in the Poisson regression model. Table 3.4 reports the estimated coefficients and the standard errors of the parameter. We note that we use a * to denote that the 95% confidence interval does not contain zero, indicating a significant result. The results show that the length of the segment, number of lanes, presence of a median, presence of a shoulder, and AADT are statistically significant in both models, while the presence of a median and DMS are significant based on the proposed approach and CCTV is an significant factor based on the Poisson approach.

According to this real life situation, the results of the proposed approach is more persuasive; the presence of a median generally reduces the risk of a crossover collision occurring since it provides more space between the opposite lanes. The dynamic message sign (DMS) is a visual sign that displays information about traffic conditions, travel times, construction, and road incidents. The DMS allows drivers to be aware of road conditions and therefore helps to decrease crash frequencies. On the other hand, CCTV does not provide any information to drivers even though it helps the government to collect road and vehicle information. Our results also correspond to Oh et al. (2015). In summary, the proposed hypothesis test has proposed the negative binomial model that provides a more precise assessment about causes of car crashes.

### 3.5 Proofs of Theorems

The following regularity conditions are required to establish the asymptotic properties of the proposed estimators:

1. The parameter space $\Theta$ is compact and $(\beta, \gamma) \in \Theta$. 

2 There exists a \((\beta_0, \gamma_0)\) such that \(E\{h_i(\beta, \gamma)\} = 0\) for every \(i\) if and only if \((\beta, \gamma) = (\beta_0, \gamma_0)\).

3 \(h_i(\beta, \gamma)\) is continuously differentiable in \((\beta, \gamma)\) and \(1/n \sum_{i=1}^n \partial h_i(\beta, \gamma)/\partial(\beta, \gamma)\) converges in probability to \(\Psi = E\{\partial h_i(\beta, \gamma)/\partial(\beta, \gamma)\}\).

4 \(G_n(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^n h_i(\beta, \gamma)h_i(\beta, \gamma)'\) converges to \(G = E\{h_i(\beta, \gamma)h_i(\beta, \gamma)\}'\) in probability, where \(G\) is positive definite.

Proof of Theorem 3.1. Under \(H_0\), we denote the true parameter as \((\beta_0, \gamma_0) = (\beta_0, 0)\) and also define

\[
\hat{H} = \begin{pmatrix} \frac{\partial H_n}{\partial \beta} \\ \frac{\partial H_n}{\partial \gamma} \end{pmatrix} \quad \text{and} \quad \hat{H}_\beta = \begin{pmatrix} \frac{\partial^2 H_n}{\partial \beta^2} & \frac{\partial^2 H_n}{\partial \beta \partial \gamma} \\ \frac{\partial^2 H_n}{\partial \gamma \partial \beta} & \frac{\partial^2 H_n}{\partial \gamma^2} \end{pmatrix}.
\]

Recall that \(\tilde{\beta} = \text{arg min}_\beta H_n(\beta|\gamma = 0)\) and \((\hat{\beta}, \hat{\gamma}) = \text{arg min}_{\beta, \gamma} H_n(\beta, \gamma)\). By Taylor expansion, we extend \(H_n(\beta_0|\gamma = 0) - H_n(\tilde{\beta}|\gamma = 0)\) to

\[
H_n(\beta_0, 0) - H_n(\tilde{\beta}, 0) = (\beta_0 - \tilde{\beta})' \hat{H}_\beta(\tilde{\beta}, 0) + \frac{1}{2} (\beta_0 - \tilde{\beta})' \hat{H}_\beta(\beta^*, 0) (\beta_0 - \tilde{\beta}),
\]

for some \(\beta^*\) between \(\beta_0\) and \(\tilde{\beta}\). Similarly, we have

\[
H_n(\beta_0, 0) - H_n(\hat{\beta}, \hat{\gamma}) = \begin{pmatrix} \beta_0 & \beta_0 - \hat{\beta} \\ 0 & \hat{\gamma} - \hat{\gamma} \end{pmatrix}' \hat{H}(\hat{\beta}, \hat{\gamma}) + \frac{1}{2} \begin{pmatrix} \beta_0 & \beta_0 - \hat{\beta} \\ 0 & \hat{\gamma} - \hat{\gamma} \end{pmatrix}' \hat{H}(\beta^*, \gamma^*) \begin{pmatrix} \beta_0 & \beta_0 - \hat{\beta} \\ 0 & \hat{\gamma} - \hat{\gamma} \end{pmatrix},
\]

where \((\beta^*, \gamma^*)\) is between \((\beta_0, 0)\) and \((\hat{\beta}, \hat{\gamma})\). Note that \(\hat{H}_\beta(\tilde{\beta}, 0) = \hat{H}(\tilde{\beta}, \hat{\gamma}) = 0\), hence

\[
T_n/n = H_n(\tilde{\beta}|\gamma = 0) - H_n(\hat{\beta}, \hat{\gamma}) = \frac{1}{2} \begin{pmatrix} \beta_0 - \hat{\beta} \\ \hat{\gamma} - \hat{\gamma} \end{pmatrix}' \hat{H}(\beta^*, \gamma^*) \begin{pmatrix} \beta_0 - \hat{\beta} \\ \hat{\gamma} - \hat{\gamma} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \beta_0 - \beta_0 \\ \hat{\gamma} - \hat{\gamma} \end{pmatrix}' \hat{H}(\beta^*, 0) \begin{pmatrix} \beta_0 - \beta_0 \\ \hat{\gamma} - \hat{\gamma} \end{pmatrix}.
\]
By Taylor expansion, we have

$$\dot{H}_\beta(\tilde{\beta}, 0) = \dot{H}_\beta(\beta, 0) + \dot{H}_{\beta\beta}(\tilde{\beta} - \beta) + O_p(n^{-1}) = 0 \quad \text{and} \quad (A.2)$$

$$\dot{H}_\beta(\tilde{\beta}, \tilde{\gamma}) = \dot{H}_\beta(\beta, 0) + \dot{H}_{\beta\beta}(\tilde{\beta} - \beta) + \dot{H}_{\beta\gamma}(\tilde{\gamma} - 0) + O_p(n^{-1}) = 0. \quad (A.3)$$

Solving for $\tilde{\beta} - \beta$ from equations (A.2) and (A.3) gives

$$\begin{pmatrix} \tilde{\beta} - \beta_0 \\ \tilde{\gamma} - 0 \end{pmatrix} = \begin{pmatrix} I & \dot{H}_{\beta\beta}^{-1} \dot{H}_{\beta\gamma} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - 0 \end{pmatrix} + O_p(n^{-1}). \quad (A.4)$$

By substituting (A.4) into (A.1), we obtain

$$\frac{1}{2} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - 0 \end{pmatrix} \left( \begin{pmatrix} \hat{H}(\beta^*, \gamma^*) - \dot{H}(\beta^*, 0) & \dot{H}(\beta^*, 0) \\ \dot{H}_{\beta\beta}^{-1} \dot{H}_{\beta\gamma} & 0 \end{pmatrix} \right) \left( \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - 0 \end{pmatrix} \right) = \frac{1}{2} \hat{\gamma} \left( \dot{H}_{\beta\gamma} - \dot{H}_{\gamma\beta} \dot{H}_{\beta\beta}^{-1} \dot{H}_{\beta\gamma} \right) \hat{\gamma}. \quad (A.5)$$

Since $\dot{H}_n(\beta, \gamma) = 2\dot{h}_n(\beta, \gamma)^T G_n(\beta, \gamma)^{-1} \dot{h}_n(\beta, \gamma) + O_p(1)$ and $\dot{h}_n(\beta, \gamma)^T G_n(\beta, \gamma)^{-1} \dot{h}_n(\beta, \gamma)$ converges to $\Sigma_{nb}^{-1}$ in distribution, (A.5) is asymptotically equal to $\hat{\gamma} \left( \Omega_{\gamma\gamma} - \Omega_{\gamma\beta} \Omega_{\beta\beta}^{-1} \Omega_{\beta\gamma} \right) \hat{\gamma}$, where $\Omega = \Sigma_{nb}^{-1}$. The asymptotic normality can be defined as

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - 0 \end{pmatrix} \xrightarrow{d} N(0, \Omega^{-1}) = N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{\beta\beta} & \Omega_{\beta\gamma} \\ \Omega_{\gamma\beta} & \Omega_{\gamma\gamma} \end{pmatrix}^{-1} \right),$$

$$\sqrt{n} \hat{\gamma} - 0 \xrightarrow{d} N \left( 0, \left( \Omega_{\gamma\gamma} - \Omega_{\gamma\beta} \Omega_{\beta\beta}^{-1} \Omega_{\beta\gamma} \right)^{-1} \right).$$
Therefore, the following distribution holds

\[ n\{H_n(\hat{\beta} | \gamma = 0) - H_n(\hat{\beta}, \hat{\gamma})\} = n\hat{\gamma}(\Omega_{\gamma \gamma} - \Omega_{\gamma \beta} \Omega_{\beta \beta}^{-1} \Omega_{\beta \gamma}) \hat{\gamma} \overset{d}{\to} \chi^2_1. \]

**Proof of asymptotic results in Section 3.2.1.** The proof of the asymptotic normality of the proposed estimator is similar to that of Theorem 3.2 in Hansen (1982). Here we prove that the asymptotic variance of the proposed estimator is no greater than that of the estimator using the quadratic inference function under the Poisson regression framework (Qu, Lindsay, and Li, 2000). Recall that \( R/(\alpha)^{-1} = (a_1 M_1 + a_2 M_2) \otimes C_i^{-1} \), where \( M_1 \) is a 2 \times 2 identity matrix. We then decompose \( M_1 \) as \( M_1 = M_{11} + M_{12} \), where \( M_{11} \) is a matrix with 1 on \((1, 1)\) and 0 elsewhere and \( M_{12} \) is a matrix with 1 on \((2, 2)\) and 0 elsewhere. Then, \( R/(\alpha)^{-1} = (a_1 M_{11} + a_1 M_{12} + a_2 M_2) \otimes C_i^{-1} \), and the corresponding score vector is

\[
\bar{h}_n(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} h_i(\beta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 
\hat{\mu}' A_i^{-1/2}(M_{11} \otimes C_i^{-1}) A_i^{-1/2}(f_i - \mu_i) \\
\hat{\mu}' A_i^{-1/2}(M_{12} \otimes C_i^{-1}) A_i^{-1/2}(f_i - \mu_i) \\
\hat{\mu}' A_i^{-1/2}(M_2 \otimes C_i^{-1}) A_i^{-1/2}(f_i - \mu_i)
\end{pmatrix}. \tag{A.6}
\]

Suppose \( h_{i_1}(\beta, \gamma) \) is specified as the first set of \( h_i(\beta, \gamma) \), and \( h_{i_r}(\beta, \gamma) \) is the remaining two sets. Note that the estimator based on the Poisson regression (Qu, Lindsay and Li, 2000) is obtained by \( h_{i_1}(\beta, \gamma) \) only, and the corresponding asymptotic covariance matrix is defined as \( \Sigma_p = (\Psi_1 G_{11}^{-1} \Psi_1)^{-1} \), where \( \Psi_1 = E\{\partial h_{i_1}(\beta, \gamma)/\partial(\beta, \gamma)\} \) and \( G_{11} = E\{h_{i_1}(\beta, \gamma) h_{i_1}(\beta, \gamma)\}' \). To differentiate the contributions of \( h_{i_1}(\beta, \gamma) \) and \( h_{i_r}(\beta, \gamma) \), we then orthogonalize \( h_{i_r}(\beta, \gamma) \) against \( h_{i_1}(\beta, \gamma) \) as \( h_{i_r}^2(\beta, \gamma) = h_{i_r}(\beta, \gamma) - G_{21} G_{11}^{-1} h_{i_1}(\beta, \gamma) \), where \( G_{21} = E\{h_{i_r}(\beta, \gamma) h_{i_1}(\beta, \gamma)\}' \), which satisfies \( E\{h_{i_r}^2(\beta, \gamma) h_{i_1}(\beta, \gamma)\}' = 0 \). By replac-
ing \( h_{ir}(\beta, \gamma) \) in (A.6) with \( h_{ir}^o(\beta, \gamma) \), we obtain

\[
\Sigma_{nb}^{-1} = \psi' G^{-1} \psi = (\psi_1', \psi_2') \begin{pmatrix}
G_{11}^{-1} & 0 \\
0 & G_{22}^o^{-1}
\end{pmatrix}
\begin{pmatrix}
\psi_1 \\
\psi_2
\end{pmatrix} = \Sigma_p^{-1} + \psi_2' G_{22}^o G_{22}^o \psi_2, \quad (A.7)
\]

where

\[
\psi_2^o = \psi_2 - G_{21} G_{11}^{-1} \psi_1,
\]

\[
G_{22}^o = E \{ h_{ir}^o(\beta, \gamma) h_{ir}^o(\beta, \gamma)' \}
\]

with \( \psi_2 = E \{ \partial h_{ir}(\beta, \gamma) / \partial (\beta, \gamma) \} \). Since \( G_{22}^o \) in (A.7) is positive semidefinite, it consequently follows that \( \sqrt{\Sigma_{nb} \nu} \leq \sqrt{\Sigma_p \nu} \) for any nonzero constant vector of \( \nu \). In a similar manner as above, it can be readily proved that the asymptotic variance of the proposed estimator is no greater than that of the estimator assuming the independent correlation structure.
Table 3.1: Mean squared errors (×1000), average length of confidence intervals (LCI), and coverage proportion (CP) at 95% confidence level using negative binomial regression under AR(1), compound symmetry (CS), and independent (IN) structures.

<table>
<thead>
<tr>
<th></th>
<th>With overdispersion</th>
<th>Without overdispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>LCI</td>
</tr>
<tr>
<td>AR(1)</td>
<td>β₁</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>β₂</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>0.064</td>
</tr>
<tr>
<td>CS</td>
<td>β₁</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>β₂</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>0.061</td>
</tr>
<tr>
<td>IN</td>
<td>β₁</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>β₂</td>
<td>0.298</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Table 3.2: Relative efficiency of proposed estimators obtained from negative binomial regression under AR(1), compound symmetry (CS), and independent (IN) structures.

<table>
<thead>
<tr>
<th></th>
<th>With overdispersion</th>
<th>Without overdispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR(1)</td>
<td>CS</td>
</tr>
<tr>
<td>β₁</td>
<td>1.356</td>
<td>1.165</td>
</tr>
<tr>
<td>β₂</td>
<td>1.390</td>
<td>1.176</td>
</tr>
</tbody>
</table>
### Table 3.3: Description of variables in crash data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash</td>
<td>number of crashes</td>
<td>numerical</td>
</tr>
<tr>
<td>Length</td>
<td>length of segment (mile)</td>
<td>numerical</td>
</tr>
<tr>
<td>Lanes</td>
<td>number of lanes</td>
<td>numerical</td>
</tr>
<tr>
<td>Passlane</td>
<td>presence of pass lane</td>
<td>categorical, 1=present, 0=absent</td>
</tr>
<tr>
<td>Median</td>
<td>presence of median</td>
<td>categorical, 1=present, 0=absent</td>
</tr>
<tr>
<td>Shoulder</td>
<td>presence of shoulder</td>
<td>categorical, 1=present, 0=absent</td>
</tr>
<tr>
<td>CCTV</td>
<td>number of CCTV camera</td>
<td>numerical</td>
</tr>
<tr>
<td>DMS</td>
<td>number of dynamic message signs</td>
<td>numerical</td>
</tr>
<tr>
<td>MVDS</td>
<td>number of microwave vehicle detection systems</td>
<td>numerical</td>
</tr>
<tr>
<td>AADT</td>
<td>annual average daily traffic (in thousands)</td>
<td>numerical</td>
</tr>
</tbody>
</table>

### Table 3.4: Estimated coefficients along with standard errors in brackets marked with * if covariate is statistically significant through negative binomial regression and Poisson regression.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Negative binomial</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.248(0.052)*</td>
<td>1.533(0.063)*</td>
</tr>
<tr>
<td>Length</td>
<td>0.304(0.004)*</td>
<td>0.188(0.010)*</td>
</tr>
<tr>
<td>Lanes</td>
<td>0.200(0.015)*</td>
<td>0.234(0.018)*</td>
</tr>
<tr>
<td>Shoulder</td>
<td>-0.545(0.031)*</td>
<td>-0.483(0.041)*</td>
</tr>
<tr>
<td>AADT</td>
<td>0.033(0.001)*</td>
<td>0.022(0.001)*</td>
</tr>
<tr>
<td>PassLane</td>
<td>-0.004(0.040)</td>
<td>0.107(0.056)</td>
</tr>
<tr>
<td>MVDS</td>
<td>-0.025(0.035)</td>
<td>-0.001(0.021)</td>
</tr>
<tr>
<td>Median</td>
<td>-0.063(0.020)*</td>
<td>-0.050(0.030)</td>
</tr>
<tr>
<td>DMS</td>
<td>-0.378(0.062)*</td>
<td>-0.076(0.054)</td>
</tr>
<tr>
<td>CCTV</td>
<td>0.047(0.047)</td>
<td>0.064(0.029)*</td>
</tr>
<tr>
<td>γ</td>
<td>0.486(0.014)*</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 3.1: Quantile-Quantile plot under AR(1) structure without overdispersion.

Figure 3.2: Power of the hypothesis test with overdispersion parameter from 0 to 0.01 under AR(1), compound symmetry, and independent structures.
Chapter 4

Adjusting for Baseline Outcomes in Comparing the Effects of Randomized Treatments

4.1 Introduction

In randomized clinical trials, an outcome of particular interest is generally measured from subjects prior to treatment. The outcome of disease at baseline is often considered as a decision support tool in the treatment assignment (Elkin et al., 1995; Eberhard et al., 2000; Kirsch et al., 2008; Lazar et al., 2010; Bower et al., 2013). For example, a randomized controlled trial study in which low-income young minority women suffering from depression was conducted in Washington, D.C. from March 1997 to May 2002. Depression scores from subjects in the study were measured at baseline before one of three treatments, antidepressant medication, psychotherapy, and referral to community mental health services, was randomly assigned. Siddique et al. (2012) reported that the severity of depression at baseline influenced on the efficacy of treatments. In particular, antidepressant medication was superior to psychotherapy for women with moderate level of
depression at baseline, whereas both treatments were comparable for patients with severe depression at the beginning of study.

Clinicians regularly monitor the patient’s condition over time to evaluate the effectiveness of treatments. In the aforementioned study, the depression score was measured monthly over a period of a half year. Miranda et al. (2003) fitted a linear regression model for repeatedly measured data and concluded that medication and psychotherapy performed similar, but were effective compared to referral to community. The results did not account for the baseline outcome in comparing the treatment effects. Kim et al. (2018) developed an initial severity-dependent model to demonstrate how a degree of severity at baseline has an influence on that of subsequent treatment. The model was on the basis of varying coefficient models (Hastie and Tibshirani, 1993; Huang et al., 2002; Fan and Zhang, 2008), and thereby enjoyed the flexibility of a nonparametric model by allowing coefficients to vary across baseline levels. However, efficient statistical inferences about comparison of treatment effects have not been fully studied. Moreover, the linearity assumption between measurement times and outcomes might lead to an imprecise trajectory of outcomes at a certain baseline level.

In this paper, we propose a bivariate varying coefficient model for longitudinal data. Suppose that \( y_{i0} \) and \( Y_{ij} \) are outcomes of subject \( i \) measured at baseline and time \( t_{ij} \), \( j = 1, \ldots, n_i \), respectively. The conditional mean of \( Y_{ij} \) is modeled as:

\[
E(Y_{ij}) = Z_i \alpha(y_{i0}, t_{ij}) + H_{ij} \gamma(y_{i0}), \quad i = 1, \ldots, n, \tag{4.1}
\]

where \( n \) is the number of subjects, \( Z_i \) is a \( q \)-dimensional row vector indicating the assigned treatment group, \( \alpha(y_{i0}, t_{ij}) = [\alpha_1(y_{i0}, t_{ij}), \ldots, \alpha_q(y_{i0}, t_{ij})]' \) is a \( q \)-dimensional vector of bivariate varying coefficients represented by the smooth regression functions of baseline and time, \( H_{ij} \) is a \( m \)-dimensional row vector of controlling variables such as gender and race, and \( \gamma(y_{i0}) \) is the corresponding coefficient vector that can vary with baseline levels.
the aforementioned trial, if \( y_0 \) is the baseline outcome of interest, three treatment effects at the level of \( y_0 \) are evaluated by comparing \( \alpha_1(y_0, t_{ij}), \alpha_2(y_0, t_{ij}), \text{ and } \alpha_3(y_0, t_{ij}). \) If the mean trajectory of outcomes of the \( k \)th treatment group at a different level of baseline is of interest, we can assess \( \alpha_k(y_{0i}, t_{ij}) \) across baseline outcomes. Therefore, efficient estimation and statistical inference about \( \alpha(y_{0i}, t_{ij}) \) in model (4.1) will be addressed.

In this article, the \( k \)th bivariate varying coefficient function in \( \alpha(y_0, t) \) is approximated by a linear combination of basis functions, \( B_{k1}(t), \ldots, B_{kp_{y0}}(t), \) as \( \alpha_k(y_0, t) \approx \sum_{i=0}^{p_{y0}} \theta_{ki}(y_0) B_{ki}(t), \) where \( \theta_{k1}(y_0), \ldots, \theta_{kp_{y0}}(y_0) \) are smooth functions of \( y_0 \) and \( p_{y0} \) is the number of basis functions for the \( k \)th coefficient, which allows to vary with a value of \( y_0 \). Various sets of basis functions can be considered such as Fourier basis functions, splines, or polynomials. Although the proposed approach is not restricted to one specific choice of basis functions, considering the nature of the aforementioned study in which each subject was measured monthly up to seven times, we suppose that polynomial functions with a sufficiently high order are guaranteed to provide arbitrarily good fits to the data. Therefore, for a given value of \( y_0 \), \( \alpha_k(y_0, t) \) is modeled as \( \alpha_k(y_0, t) = \sum_{i=0}^{p_{y0}} \theta_{ki}(y_0) t^i. \)

The choice of the optimal polynomial order plays an important role in balancing between the complexity of the model and lack of fit of the data. We propose a model selection procedure based on the Bayesian information criterion and an empirical log-likelihood ratio (Owen, 1988; Qin and Lawless, 1994). The proposed method adaptively identifies the optimality of a polynomial at each baseline level. After the parsimonious model is selected, it is of particular interest to assess whether the treatment effects are identical at each baseline level. Therefore, we develop a hypothesis test to check the therapeutic effects by comparing the mean trajectories with respect to treatment groups. We propose an empirical log-likelihood ratio test statistic that asymptotically follows the chi-square distribution under the null hypothesis. We remark that the polynomial basis function approximation offers practical convenience in conducting these statistical inference about the bivariate
varying coefficients and model selection. Our numerical studies also indicate that the proposed model selection chooses the true degree of polynomial with a high frequency, and the proposed test detects the difference between the mean trajectories of outcomes successfully in all cases under consideration.

In longitudinal studies, repeated measurements within the subject are likely to be correlated, and estimation efficiency can be improved by incorporating the within-subject correlation. Qu et al. (2000) extended the inverse of a working correlation structure to a set of predetermined matrices and developed the quadratic inference functions. This method does not require estimation of the additional nuisance parameters associated with the working correlation structure. However, the quadratic inference functions might be inefficient in a highly over-identified case, because estimation of a variance matrix of many estimating equations is required, yet could be unreliable. Alternatively, the empirical log-likelihood is constructed on the basis of the set of the over-identified estimating equations and yields a more efficient estimator than the one ignoring the within-subject correlation.

The paper is organized as follows. In Section 4.2, we provide a complete procedure including estimation of the varying coefficients, selection of the optimal degree of a polynomial, and comparison of mean trajectories among treatment groups. Section 4.3 confirms the effectiveness of the proposed procedure in finite samples. In Section 4.4, we illustrate the analysis of the aforementioned depression data and explore the dynamic change of treatment effects across baseline levels. The theoretical proofs and regularity conditions are placed in Section 4.5.
4.2 Methodologies

4.2.1 Estimation of Varying Coefficients

Model (4.1) under the polynomial basis function approximation can be represented in vector or matrix notation as:

$$E(Y_i) = G_i(y_{i0})\Theta(y_{i0})Z_i + H_i\gamma(y_{i0}), \quad (4.2)$$

where $Y_i = \{Y_{i1}, \ldots, Y_{in_i}\}'$ is a vector of $n_i$ outcomes measured from subject $i$ at times $t_{i1}, \ldots, t_{in_i}$, $G_i(y_{i0}) = \{G_{i1}(y_{i0})', \ldots, G_{in_i}(y_{i0})'\}'$ is an $n_i \times (p_{y0} + 1)$-dimensional matrix having $G_{ij}(y_{i0}) = (1, t_{ij}, t_{ij}^2, \ldots, t_{ij}^{p_{y0}})$, $\Theta(y_{i0}) = \{\Theta_1(y_{i0}), \ldots, \Theta_q(y_{i0})\}$ with $\Theta_k(y_{i0}) = \{\theta_{k0}(y_{i0}), \ldots, \theta_{kp_{y0}}(y_{i0})\}'$ is a $(p_{y0} + 1) \times q$-dimensional matrix of smooth regression functions of $y_{i0}$, and $H_i = \{H_{i1}', \ldots, H_{in_i}'\}'$.

To obtain the efficient estimation of parameters $\Theta(y_{i0})$ and $\gamma(y_{i0})$ in model (4.2), we first specify a $n_i \times (q(p_{y0} + 1) + m)$-dimensional matrix $X_i = \{Z_i' \otimes G_i(y_{i0}), H_i\}$ and combine parameters as $\beta = \{\Theta_1(y_{i0})', \ldots, \Theta_q(y_{i0})', \gamma(y_{i0})'\}'$, then model (4.2) can be expressed as $E(Y_i) = X_i\beta$, where $\otimes$ is a Kronecker product. To estimate $\beta$ at a certain baseline level $y_0$, we extend the generalized estimating equations (Liang and Zeger, 1986) by incorporating the within-subject correlation

$$\sum_{i=1}^n X_i' A_i^{-1/2} R_i(\alpha)^{-1} A_i^{-1/2} (Y_i - X_i\beta) K_i(y_0) = 0, \quad (4.3)$$

where $A_i$ is a diagonal marginal variance matrix of $Y_i$, $R_i(\alpha)$ is a working correlation structure of $Y_i$ with a few nuisance parameters of $\alpha$, and $K_i(y_0) = K \{(y_{i0} - y_0)/h\}$ is a kernel function with bandwidth $h$. Note that the estimation of $\alpha$ in $R_i(\alpha)$ is required prior to solving (4.3), while it can be inefficient if the assumed working correlation structure is misspecified.
To tackle this problem, Qu et al. (2000) employed a class of basis matrices to represent the inverse of $R_i(\alpha)$ in (4.3) as $R_i(\alpha)^{-1} = \sum_{s=1}^{d} \eta_s U_{is}$, where $U_{is}$ is the so-called basis matrix and $\eta_s$ is an unknown constant. The set of basis matrices can be predetermined by the assumed working correlation structure. For instance, if the working correlation structure is compound symmetry, then $R_i(\alpha)^{-1} = \eta_1 U_{i1} + \eta_2 U_{i2}$, where $U_{i1}$ is an identity matrix, and $U_{i2}$ is a matrix with 0 on the diagonal and 1 elsewhere. If $R_i(\alpha)$ corresponds to a first-order autoregressive structure, denoted by AR(1), then $R_i(\alpha)^{-1} = \eta_1 U_{i1} + \eta_2 U_{i2} + \eta_3 U_{i3}$, where $U_{i1}$ is an identity matrix, $U_{i2}$ is a matrix with 1 on the sub-diagonals and 0 elsewhere, and $U_{i3}$ is a matrix with 1 on the corners $(1,1)$ and $(t_i, t_i)$ and 0 elsewhere.

By replacing $R_i(\alpha)^{-1}$ in (4.3) with the basis matrices, the equation (4.3) can be extended as $\sum_{i=1}^{n} g_i(\beta) = 0$ having

$$g_i(\beta) = \begin{pmatrix}
X_i' A_i^{-1/2} U_{i1} A_i^{-1/2} (Y_i - X_i \beta) K_i(y_0)
\vdots
X_i' A_i^{-1/2} U_{id} A_i^{-1/2} (Y_i - X_i \beta) K_i(y_0)
\end{pmatrix}. \tag{4.4}$$

We denote that $\beta(y_0)$ is the true parameter value of $\beta$ at a baseline outcome of interest $y_0$. Under $E[g_i{\beta(y_0)}] = 0$, we employ the empirical log-likelihood function

$$L \{\beta(y_0)\} = \sup \left\{ \sum_{i=1}^{n} \log(w_i) \bigg| \sum_{i=1}^{n} w_i g_i(\beta) = 0, \sum_{i=1}^{n} w_i = 1, 0 \leq w_i \leq 1 \right\}. \tag{4.5}$$

where $w_i$ denotes a point mass assigned to the observations from the $i$th subject. The proposed empirical log-likelihood estimator can be obtained as $\hat{\beta}(y_0) = \arg\max_{\beta} L \{\beta(y_0)\}$. Note that the proposed approach incorporates the within-subject correlation information, while it does not need to estimate the nuisance parameter $\alpha$. Moreover, the proposed estimator $\hat{\beta}(y_0)$ enjoys the asymptotic properties under the regularity conditions in Section 4.5, as the following theorem. We first define $g_i{\beta(y_0)} = h_i{\beta(y_0)} K_i(y_0)$, and $\Sigma =$
\[(\Phi'\Phi)^{-1}\text{ with } \Phi = -E[\partial h_i\{\beta(y_0)\}/\partial \beta(y_0)] \text{ and } V = E\left[h_i\{\beta(y_0)\}h_i\{\beta(y_0)\}^T\right].\]

**Theorem 4.1.** Under the regularity conditions in Section 4.5 and the assumption that \(nh \to \infty\) and \(nh^5 \to 0\), the estimator of \(\beta\) holds \(\sqrt{nh}\{\hat{\beta}(y_0) - \beta(y_0)\} \overset{d}{\to} N\{0, \psi \Sigma\}\), where \(\psi = \int K^2(u)du/f(y_0)\). Moreover, \(\Sigma_i - \Sigma\) is positive semi-definite such that \(\nu^T \Sigma_i \nu \leq \nu^T \Sigma \nu\) for any nonzero constant vector of \(\nu\), where the asymptotic covariance matrix of \(\hat{\beta}(y_0)\) under the independent working correlation structure is \(\psi \Sigma_i\).

Theorem 4.1 confirms that the estimator is asymptotically normally distributed with mean \(\beta(y_0)\) and covariance matrix \(\psi \Sigma\). When undersmoothing is conducted, i.e., \(nh^5 \to 0\), the proposed estimator \(\hat{\beta}(y_0)\) is consistent with \(\beta(y_0)\), while if it is not undersmoothing, i.e., \(nh^5 < \infty\), there exists a bias term of \(\hat{\beta}(y_0)\). Moreover, the asymptotic estimation efficiency can be improved by incorporating the within-subject correlation structure, even though the working correlation structure is specified incorrectly. The proof of Theorem 4.1 is provided in Section 4.5.

### 4.2.2 Choice of the Most Parsimonious Model

The most parsimonious polynomial regression model might differ at various baseline outcomes, which motivates us to select an optimal polynomial order for a given baseline level, denoted by \(p_{y_0}\). Bayesian information criterion is a model selection procedure, which enables us to identify the best polynomial regression model with an optimal polynomial order that neither overfits nor underfits the data. Let \(s\) be the largest polynomial order we considered, and it is enough to capture the true polynomial order \(p_{y_0}\), i.e., \(p_{y_0} \leq s\). We compare all models with polynomial order from 0 to \(s\), and index these candidate models by \(p\), where \(p = 0, \ldots, s\). For each candidate model, we estimate the parameters \(\Theta(y_0)\) and \(\gamma(y_0)\) in model (4.2). Then we construct the \((s+1) \times q\) matrix \(\hat{\Theta}_p(y_0) = \{\hat{\Theta}_{1p}(y_0), \ldots, \hat{\Theta}_{qp}(y_0)\}\) and \(m\)-dimensional vector \(\hat{\gamma}_p(y_0)\), where the first \((p+1)\) rows in
\( \hat{\Theta}_p(y_0) \) are \( \hat{\Theta}(y_0) \) obtained from the candidate model with a polynomial order \( p \) and all remaining items are 0. With \( \hat{\beta}_p(y_0) = \{ \hat{\Theta}_1(y_0)', \ldots, \hat{\Theta}_p(y_0)', \hat{\gamma}_p(y_0)' \}' \), we employ the Bayesian information criterion based on the empirical loglikelihood ratio as

\[
ELBIC_p = -2L \{ \hat{\beta}_p(y_0) \} + df_p \log(n\h).
\]  

(4.6)

where \( df_p \) is the number of non-zero coefficients in \( \hat{\beta}_p(y_0) \). The optimal polynomial order for a given baseline level is the model has the smallest \( ELBIC \), i.e., \( \hat{p} = \text{argmin}_p ELBIC_p \).

**Theorem 4.2.** Under the regularity conditions in Section 4.5 and the assumption that \( nh \to \infty \) and \( n\h^5 \to 0 \), with probability approaching to one, \( P(ELBIC_p > ELBIC_{p_0}) \to 1 \) for all \( p \neq \hat{p} \).

Theorem 4.2 ensures that the proposed selection procedure identifies the true order of the polynomial regression model consistently. Note that all curves in model (4.2) are constant over time if \( \hat{p} = 0 \). The proof of Theorem 4.2 is provided in Section 4.5.

In addition, we adopt the iterative method based on \( ELBIC \) in (4.6) to determine the optimal bandwidth parameter \( h \) in (4.3). We first employ a candidate model with a polynomial order \( p \), and obtain the optimal \( h \) by minimizing \( ELBIC \), denoted by \( h_p \). Based on bandwidth \( h_p \) and apply the above mentioned method to determine the best polynomial order \( \hat{p} \) by comparing \( ELBIC \) among all candidate models. If \( \hat{p} = p \), the process stops with the optimal polynomial order \( p \) and optimal bandwidth \( h_p \); while if \( \hat{p} \neq p \), then repeating this procedure by using the model with the current order \( \hat{p} \). Therefore, after several iteration, we may get the optimal polynomial order \( \hat{p} \) and the best bandwidth \( h_{\hat{p}} \) together through \( ELBIC \).
4.2.3 Hypothesis Test for Comparing Treatment Effects

For the proposed nonparametric polynomial regression model, one of the important studies is to test whether treatment effects are the same on the response variable. Therefore, we develop a hypothesis test for

\[ H_0 : \Theta_1(y_0) = \Theta_2(y_0) = \ldots = \Theta_q(y_0). \]

Since the empirical log-likelihood function in (4.5) plays a similar role as minus twice the log-likelihood ratio test function, we propose a test statistic based on \( L \{ \beta(y_0) \} \). We construct an empirical log-likelihood ratio test statistic as

\[ T(y_0) = -2 \left[ L \{ \tilde{\beta}(y_0) \} - L \{ \hat{\beta}(y_0) \} \right]. \quad (4.7) \]

where \( \tilde{\beta}(y_0) \) and \( \hat{\beta}(y_0) \) are the estimators obtained under the null and alternative hypothesis, respectively. The following theorem shows that the proposed test statistic has the same chi-square asymptotic property as the log-likelihood ratio test, while it does not need to specify the likelihood function.

**Theorem 4.3.** Under the regularity conditions in Section 4.5 and the assumption that \( nh^5 \to 0 \), \( T(y_0) \xrightarrow{d} \chi^2_{(p_{y_0} + 1)(q - 1)} \) under \( H_0 \) as \( nh \to \infty \).

This theorem illustrates that the proposed test statistic with undersmoothing asymptotically follows the chi-square distribution with \((p_{y_0} + 1)(q - 1)\) degrees of freedom under the null hypothesis. We note that the proposed test statistic can be readily extended to test the equality of some curves by setting the parameters of curves of particular interest as equal for the null hypothesis. The proof of Theorem 4.3 is provided in Section 4.5.
4.3 Simulation Studies

In this section, we evaluate the proposed approach through the following simulation studies. Two treatment groups are considered and each consists of 300 subjects. Every subject is repeatedly measured seven times, including a baseline outcome and 6 follow-up outcomes. We set an unbalanced case with 15% of follow-up outcomes are missing. The true polynomial order is 2, and we generate the correlated continuous response from the following model

\[
Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 6 \\ 1 & 2^2 & \cdots & 6^2 \end{pmatrix} \begin{pmatrix} \theta_{10} & \theta_{20} \\ \theta_{11} & \theta_{21} \\ \theta_{12} & \theta_{22} \end{pmatrix} \begin{pmatrix} 1_{300} & 0 \\ 0 & 1_{300} \end{pmatrix} + H\gamma + \epsilon
\]

where

\[
\theta_{10} = \cos((y_{i0} + \pi)/4), \quad \theta_{11} = \sin(y_{i0}/2\pi), \quad \theta_{20} = (y_{i0} + 1.3)/4.5, \quad \theta_{21} = (2 - y_{i0})/6.3, \quad \theta_{12} = \theta_{22} = \sin(2\pi y_{i0})/20 + 0.5, \quad \gamma = 0.05y_{i0} + 0.25,
\]

the covariate \( H \) is generated from an standard normal distribution, the random error \( \epsilon \) is a 6 \times 600-dimensional matrix, it follows a multivariate normal distribution under a compound symmetry correlation structure with a correlation coefficient of 0.8, and \( 1_{300} \) is a 300-dimensional one vector. In addition, we generate an equal number of baseline values independently for each treatment group by using the uniform distribution on range \((c, c + 1)\) with \( c = 0, 1, 2, 3, 4 \). We also remark that the effects of the two treatment groups are identical at the baseline level of one, since two polynomial curves are exactly same at \( y_0 = 1 \). To implement the proposed approach, the standard Gaussian kernel with an optimal bandwidth \( h = 0.4 \) is employed in the estimating equations.
We estimate the parameters through the proposed approach under the three working correlation structure: compound symmetry, AR(1), and independent. The mean squared errors of the estimators are computed at \( y_0 = 1, 2, 3, 4 \) from 500 simulation runs, defined as \( MSE(\hat{\beta}_\delta) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\beta}_\delta^{(i)} - \beta_\delta)^2 \), where \( \hat{\beta}_\delta^{(i)} \) is the estimator from the \( i \)th simulation and \( \beta_\delta \) is the true parameter with \( \delta = 1, \ldots, 6 \). Table 4.1 reports the relative efficiency of the proposed estimator under the compound symmetry or AR(1) working correlation structure to the one under the independent structure, defined as the ratio of the mean squared errors. A large value of relative efficiency indicates that the proposed estimator is more efficient. According to the results in Table 4.1, the estimators obtained by incorporating a within-subject correlation structure is more efficient than the one assuming working independently, since the relative efficiencies are greater than one in all cases, even though the working correlation structure is misspecified.

Furthermore, we apply the proposed selection procedure to chose an optimal polynomial order for each simulation study. Over 98% of times, the \textit{ELBIC} identifies that the optimal order of the polynomial is 2 at all baseline levels under consideration. This ensures that the proposed model selection approach accurately captures the true polynomial order. In addition, we also assess the equality of the polynomial curves at each baseline level. The proposed test successfully detects the identical and the difference between the two polynomial curves, since the rejection rate are 5.6% and 5.2% at \( y_0 = 1 \) for compound symmetry and AR(1) working correlation structure, and it always be 100% at \( y_0 = 2, 3, 4 \), respectively. Moreover, we investigate the Theorem 4.3 at \( y_0 = 1 \) as shown in Figure 4.1, the quantile-quantile plots confirm that the proposed test statistic follows the chi-square distribution with three degrees of freedom under the null hypothesis.
4.4 Application to Depression Study

In this section, we illustrate the proposed procedure through the depression data for low-income young minority women studied in Washington, D.C. from March 1997 to May 2002 (Miranda et al., 2003). A total of 267 women with major depression were enrolled in the study. They were randomly assigned to one of the three treatment groups: an antidepressant medication intervention (medication, \( n = 88 \)), a psychotherapy intervention (psychotherapy, \( n = 90 \)), and referral to community mental health services (community, \( n = 89 \)). The study recorded a Hamilton depression rating scale for each patient to measure their depression severity, a higher score indicates a more severe depression. To assess the effects of treatments on depression, the Hamilton depression rating scale was measured monthly from baseline severity through 6 months. In addition, other characteristics of patients were recorded at the first visit. More details can be found in Miranda’s paper. The objective of this study is to identify the most effective treatment for each patient based on the severity of depression at their baseline while adjusting for other characteristics. Therefore, we treat the depression score measured during 6 months follow-up as the response variable \( (Y) \), and the centralized age \( (\text{c age}, \text{subtract average age 29}) \) and employment status \( (\text{employ}, 0=\text{unemployed}, 1=\text{employed}) \) as covariates in the model. Accordingly, the polynomial regression model at a certain baseline severity \( y_0 \) is specified as

\[
E(Y) = \begin{pmatrix}
1 & 1 & \ldots & 1^{p_0} \\
1 & 2 & \ldots & 2^{p_0} \\
\vdots & \vdots & \ddots & \vdots \\
1 & 6 & \ldots & 6^{p_0}
\end{pmatrix}
\begin{pmatrix}
\theta_{10} & \theta_{20} & \theta_{30} \\
\theta_{11} & \theta_{21} & \theta_{31} \\
\vdots & \vdots & \vdots \\
\theta_{1p_0} & \theta_{2p_0} & \theta_{3p_0}
\end{pmatrix}
\begin{pmatrix}
1_{88} & 0 & 0 \\
0 & 1_{90} & 0 \\
0 & 0 & 1_{89}
\end{pmatrix}
+ \text{c age} \gamma_1 + \text{employ} \gamma_2.
\] (4.9)
Before we evaluate the model (4.9), we first choose the optimal polynomial order $p$ at each baseline severity level using the proposed model selection approach in Section 4.2.2 under the $AR(1)$ correlation structure. The $ELBIC$ confirms that the best polynomial order is zero when the baseline level is less than or equal to thirteen, indicating that these three treatments are not beneficial for patients when their degree of depression is mild, perhaps, choosing other therapies might be effective for them. The quadratic polynomial curve, $p_{y_0} = 2$, is the optimal choice for $y_0 = 14$ to 20, and the linear regression, $p_{y_0} = 1$, is adopted for $y_0 = 21$ to 23. Note that the selection results are unreliable when the severity at baseline is greater than or equal to twenty-four, because there is not enough data for modeling. Therefore, the following model evaluations focus on $y_0 = 14$ to 23.

We then estimate the coefficients in (4.9) using the proposed estimation approach under $AR(1)$ correlation structure. For the baseline from 14 to 20, the adopted regression models include the quadratic polynomial coefficients, which makes it difficult to assess the change of treatment effects over time through coefficient values. Therefore, we plot the polynomial curves of the three treatments at $y_0 = 14$ to 20 for those patients whose age is around 29 and do not have a job, as shown in Figure 4.2. At each baseline severity, both medication and psychotherapy effectively decrease the depression scores over time, while the reduction rates are declining because the predicted depression score eventually tends to be a lower value. The community treatment curve remains steady over time, this indicates that this treatment is not beneficial for patients. Compared with psychotherapy, when the severity at baseline is relatively low such as $y_0 = 14$, the effect of medication is similar to that of the psychotherapy, while its efficacy tends to be better than the psychotherapy as the baseline increases, i.e., $y_0 = 20$. Figure 4.2 also indicates that the ultimate predicted depression score for medication is close to 8 for all levels of baseline, whereas the endpoint of predicted score increases as the baseline increases for psychotherapy. Note that the depression score 8 is considered to be almost no depression
(American Psychiatric Association, 2009; National Collaborating Centre for Mental Health, 2010). Moreover, Table 4.2 reports all estimated coefficients obtained from the linear regression model for $y_0 = 21$ to $23$, respectively. The signs of slope coefficients are all negative, indicating that all therapies are intended to reduce the depression scores over time. Within each treatment group, they have a similar trend over the baseline outcome, since the slope coefficient has not much change as the baseline level increases. On the other hand, comparison of three therapies, medication is the most effective with a larger slope. In addition, Table 4.3 reports the estimated coefficients of age and employment status at each baseline level. The results confirm that age has an effect on depression in moderate patients but insignificant in severe patients; employment status is mostly and statistically significant on depression scores.

We further investigate whether three treatments are identical at each baseline severity of interest using the proposed hypothesis test in Section 4.2.3, the results are shown in Table 4.4. We conclude at a significance level of 0.05 that the proposed test rejects the null hypothesis at all baseline levels of interest, note that the p-value is close to 0.05 at $y_0 = 23$. This is consistent with the results of Figure 4.2 and Table 4.2, due to medication and community treatment effects are significantly different. Moreover, we construct a hypothesis test in the pairwise curves to explore which treatment has obvious difference. The results indicate that psychotherapy is as beneficial as medication when severity at baseline is moderate, while it is similar to that in the community for serious patients; the medication therapy is always more beneficial regardless of the baseline severity of depression.
4.5 Proofs of Theorems

The following regularity conditions are required to establish the asymptotic properties of the proposed estimators:

1. $\beta(\cdot)$, $V$, and $\Phi$ are twice continuously differentiable in a neighborhood of $y_0$.

2. The density function $f(\cdot)$ of $y_0$ is bounded. Assuming that $f(\cdot)$ is positive and twice continuously differentiable in a neighborhood of $y_0$.

3. $K(\cdot)$ is symmetric with bounded support, has bounded derivative, and $\int K(u)du = 1$.

Proof of Theorem 4.1. Recall that $g_i \{\beta(y_0)\} = h_i \{\beta(y_0)\} K_i(y_0)$, and $h_i \{\beta(y_0)\}$ can be decomposed as $h_i \{\beta(y_0)\} = h_{i1} + h_{i2}$ with

$$h_{i1} = \begin{pmatrix} X'_i A^{-1/2}_i U_{i1} A^{-1/2}_i X_i \{\beta - \beta(y_0)\} \\ \vdots \\ X'_i A^{-1/2}_i U_{id} A^{-1/2}_i X_i \{\beta - \beta(y_0)\} \end{pmatrix},$$

and

$$h_{i2} = \begin{pmatrix} X'_i A^{-1/2}_i U_{i1} A^{-1/2}_i \epsilon_i \\ \vdots \\ X'_i A^{-1/2}_i U_{id} A^{-1/2}_i \epsilon_i \end{pmatrix},$$

where $\epsilon_i$ is an error term with mean 0. Based on the proof of Lemma 1 in Kim et al. (2018), $1/nh \sum_{i=1}^n \partial g_i \{\beta(y_0)\} / \partial \beta(y_0) \overset{D}{\to} f(y_0)\Phi$ and $1/nh \sum_{i=1}^n g_i \{\beta(y_0)\} g_i \{\beta(y_0)\}' \overset{D}{\to}$
\( \lambda f(y_0)V \), and according to the proof of Theorem 1 in Qin and Lawless (1994), we have

\[
\sqrt{n} h \left\{ \hat{\beta}(y_0) - \beta(y_0) \right\} = \left\{ f(y_0) \Phi' \left( \frac{1}{\lambda f(y_0)} V^{-1} f(y_0) \Phi \right) \right\}^{-1} f(y_0) \Phi' \left( \frac{1}{\lambda f(y_0)} V^{-1} \frac{1}{\sqrt{n} h} \sum_{i=1}^{n} g_i \{ \beta(y_0) \} \right) \\
= \left\{ \Phi' V^{-1} \Phi \right\}^{-1} \Phi' V^{-1} \frac{1}{f(y_0) \sqrt{n} h} \sum_{i=1}^{n} g_i \{ \beta(y_0) \},
\]

where \( \lambda = \int K^2(u) du \). Since \( h_i \{ \beta(y_0) \} = h_{i1} + h_{i2} \), we get

\[
\frac{1}{\sqrt{n} h} \sum_{i=1}^{n} g_i \{ \beta(y_0) \} = \frac{1}{\sqrt{n} h} \sum_{i=1}^{n} h_{i1} K_i(y_0) + \frac{1}{\sqrt{n} h} \sum_{i=1}^{n} h_{i2} K_i(y_0). \tag{A.1}
\]

By the Taylor's expansion, we have

\[
\frac{1}{\sqrt{n} h} \sum_{i=1}^{n} h_{i1} K_i(y_0) \tag{A.2}
= \sqrt{n} h^5 \left[ \{ f(y_0) + f(y_0) \} \Phi \beta(y_0) + f(y_0) \Phi \beta(y_0) + \frac{f(y_0)}{2} \Phi \beta(y_0) \right] \int u^2 K(u) du + o_p(1).
\]

Under the assumption that \( nh^5 \to 0 \), then (A.2) converges to 0 in probability. Under \( E(\epsilon_i|X_i) = 0 \) and the Lemma 1 in Kim et al (2018), the second term in (A.1) converges to a normal distribution with mean 0 and covariance matrix \( \lambda f(y_0)V \), such that

\[
\frac{1}{\sqrt{n} h} \sum_{i=1}^{n} h_{i2} K_i(y_0) \xrightarrow{d} N(0, \lambda f(y_0)V).
\]

Therefore, the desired result in Theorem 4.1 follows

\[
\sqrt{n} h \left\{ \hat{\beta}(y_0) - \beta(y_0) \right\} \xrightarrow{d} N\left\{ 0, \lambda/f(y_0) \Sigma \right\}.
\]

Now, we proof \( \Sigma_i - \Sigma \) is positive semi-definite. Recall that \( \Sigma = \left\{ \Phi' V^{-1} \Phi \right\}^{-1} \) with \( \Phi = -E[\partial_i \{ \beta(y_0) \} / \partial \beta(y_0)] \) and \( V = E \left[ h_i \{ \beta(y_0) \} h_i \{ \beta(y_0) \}' \right] \), and the corresponding
score vector is formulated as

\[
g_i(\beta) = \begin{pmatrix}
X_i' A_i^{-1/2} U_i A_i^{-1/2} (Y_i - X_i \beta) K_i(y_0) \\
X_i' A_i^{-1/2} U_i A_i^{-1/2} (Y_i - X_i \beta) K_i(y_0)
\end{pmatrix} = \begin{pmatrix}
h_{ii}(\beta) K_i(y_0) \\
h_{IR}(\beta) K_i(y_0)
\end{pmatrix}. \tag{A.3}
\]

With \( \Phi_i = -E [\partial h_{ii} \{ \beta(y_0) \} / \partial \beta(y_0)] \) and \( \Lambda_i = E \left[ h_{ii} \{ \beta(y_0) \} h_{ii} \{ \beta(y_0) \}' \right] \), we can formulate \( \Sigma_i = (\Phi_i' \Lambda_i^{-1} \Phi_i)^{-1} \). To differentiate the contributions of \( h_{ii}(\beta) \) and \( h_{IR}(\beta) \), we orthogonalize them as \( h^o_{ii}(\beta) = h_{IR}(\beta) - \Lambda_i^{-1} h_{ii}(\beta) \), where \( \Lambda_i = E \left[ h_{ii} \{ \beta(y_0) \} h_{ii} \{ \beta(y_0) \}' \right] \), which satisfies \( E \left[ h^o_{ii} \{ \beta(y_0) \} h_{ii} \{ \beta(y_0) \}' \right] = 0 \). By replacing \( h_{IR}(\beta) \) in (A.3) with \( h^o_{ii}(\beta) \), we obtain

\[
\Sigma^{-1} = \Phi' \Lambda^{-1} \Phi = (\Phi_i, \Phi^o) \begin{pmatrix}
\Lambda_i^{-1} & 0 \\
0 & \Lambda^o_i^{-1}
\end{pmatrix} \begin{pmatrix}
\Phi_i \\
\Phi^o
\end{pmatrix} = \Sigma_i^{-1} + \Phi^o \Lambda^o_i^{-1} \Phi^o, \tag{A.4}
\]

where

\[
\Phi^o = \Phi_R - \Lambda Ri \Lambda_i^{-1} \Phi_i,
\]

\[
\Lambda^o = E \left[ h^o_{ii} \{ \beta(y_0) \} h^o_{ii} \{ \beta(y_0) \}' \right]
\]

with \( \Phi_R = E [\partial h_{IR} \{ \beta(y_0) \} / \partial \beta(y_0)] \). Since \( \Lambda^o \) in (A.4) is positive semi-definite, it consequently follows that \( \Sigma^{-1} \geq \Sigma_i^{-1} \).

\[\square\]

**Proof of Theorem 4.2.** Based on the proof of Lemma 1 in Qin and Lawless (1994), we have

\[
\lambda(\beta) = \left\{ \frac{1}{nh} \sum_{i=1}^n g_i(\beta) g_i(\beta)' \right\}^{-1} \frac{1}{nh} \sum_{i=1}^n g_i(\beta) + o_p(1), \tag{A.5}
\]

where it satisfies \((nh)^{-1} \sum_{i=1}^n g_i(\beta)/\{1 + \lambda(\beta)' g_i(\beta)\} = 0\). uniformly on \( \beta \) for \( ||\beta -
\[ \beta(y_0) = O((nh)^{-1/2}). \] It follows from (A.5) and Taylor expansion that

\[
-2L \{ \beta(y_0) \} = -2 \sum_{i=1}^{n} \log(nw_i) = 2 \sum_{i=1}^{n} \log \left\{ 1 + \lambda(\beta) g_i(\beta) \right\}
\]

\[
= 2 \sum_{i=1}^{n} \lambda(\beta) g_i(\beta) - \sum_{i=1}^{n} \left\{ \lambda(\beta) g_i(\beta) \right\}^2 + o_p(1)
\]

\[
= nh \left\{ \frac{1}{nh} \sum_{i=1}^{n} g_i(\beta) \right\}' \left\{ \frac{1}{nh} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)' \right\}^{-1} \left\{ \frac{1}{nh} \sum_{i=1}^{n} g_i(\beta) \right\} + o_p(1)
\]

\[
= nh \bar{g}(\beta)' \bar{C}(\beta)^{-1} \bar{g}(\beta) + o_p(1).
\]

By letting \( p \) be a finite integer and to prove that \( P(ELBI_{C_p} \leq ELBI_{C_{p_0}}) \to 0 \) for all \( p \neq \beta \). The following proof has two steps, similar to Nishii (1984) and Wang and Qu (2009). Firstly, when \( p < p_{y_0} \), it follows from the Uniform Law of Large Numbers, continuous mapping theorem, \( E \left[ \bar{g} \{ \beta_p(y_0) \} \right] \neq 0 \), and a positive definiteness of \( \Sigma \{ \beta_p(y_0) \} \) that

\[
\frac{1}{nh} ELBI_{C_p} = \bar{g} \{ \beta_p(y_0) \}' \bar{C} \{ \beta_p(y_0) \}^{-1} \bar{g} \{ \beta_p(y_0) \} + o_p(1) + \frac{1}{nh} df_p \log(nh)
\]

\[
\Rightarrow E \left[ \bar{g} \{ \beta_p(y_0) \} \right]' \Sigma \{ \beta_p(y_0) \}^{-1} E \left[ \bar{g} \{ \beta_p(y_0) \} \right] > 0. \quad (A.6)
\]

From (A.6) thus \( P(ELBI_{C_p} > ELBI_{C_{p_0}}) \to 1 \) for all \( p < p_{y_0} \). Secondly, when \( p > p_{y_0} \), with \( E \left[ \bar{g} \{ \beta_p(y_0) \} \right] = 0 \) and \( \bar{g} \{ \beta_p(y_0) \} = O_p((nh)^{-1/2}) \) that

\[
\frac{1}{nh} ELBI_{C_p} = \bar{g} \{ \beta_p(y_0) \}' \bar{C} \{ \beta_p(y_0) \}^{-1} \bar{g} \{ \beta_p(y_0) \} + o_p(1) + \frac{1}{nh} df_p \log(nh)
\]

\[
= O_p(n^{-1}) + O(\log(nh)/nh) = o_p(1). \quad (A.7)
\]

Therefore, it is sufficient to

\[
P(ELBI_{C_p} \leq ELBI_{C_{p_0}}) = P(-2L \{ \beta_{p_0}(y_0) \} + 2L \{ \beta_p(y_0) \} \geq (df_p - df_{p_0}) \log(nh)) \to 0.
\]
Theorem 4.3 and (A.7) confirm that $-2L \{ \beta_{p_0}(y_0) \} + 2L \{ \beta_p(y_0) \}$ has an asymptotic chi-square distribution, i.e., $-2L \{ \beta_{p_0}(y_0) \} + 2L \{ \beta_p(y_0) \} = O_p(1)$, while $df_p - df_{p_0}$ is positive with $\log(nh)$ divergent as $nh \to \infty$.

Proof of Theorem 4.3. Recall the null hypothesis is specified as

$$H_0 : \Theta_{ij}(y_0) = \ldots = \Theta_{iq}(y_0) \text{ for all } i = 1, \ldots, p_{y_0} + 1.$$ 

where $\Theta_j(y_0) = \{ \Theta_{2j}(y_0), \ldots, \Theta_{(p_{y_0}+1)j}(y_0) \}^T$ for $j = 1, \ldots, q - 1$. To contrast $\Theta_{ij}(y_0)$'s for $i = 1, \ldots, p_{y_0} + 1$ and $j = 1, \ldots, q - 1$, we denote $\Theta_{ij}^*(y_0) = \Theta_{ij}(y_0) - \Theta_{iq}(y_0)$, and the null hypothesis can be rewritten as

$$H_0 : \Theta_{ij}^*(y_0) = 0 \text{ for all } i = 1, \ldots, p_{y_0} + 1, j = 1, \ldots, q - 1.$$ 

We let $\beta^* = (\beta_1^*, \beta_2^*)^T$, where $\beta_1^*$ is a vector of $(p_{y_0} + 1)(q - 1) \Theta_{ij}^*(y_0)$'s, and $\beta_2^*$ is a vector including $p_{y_0} \Theta_{iq}(y_0)$'s and $m \gamma(y_0)$'s, respectively. Then, the parameter $\beta$ can be denoted by $\beta = C\beta^*$, where $C$ is a $(q(p_{y_0} + 1) + m) \times (q(p_{y_0} + 1) + m)$ matrix. The first $q(p_{y_0} + 1) \times q(p_{y_0} + 1)$ matrix of $C$ can be expressed as $I_{p_{y_0}+1} \otimes D$ with a $(p_{y_0}+1) \times (p_{y_0}+1)$ identity matrix $I_{p_{y_0}+1}$ and a square matrix $D$ of order $q$ with 1 on the diagonal and the last column, and 0 elsewhere. For the rest part of $C$, 1 on the diagonal and 0 elsewhere. An estimator of $\beta^*$ is accordingly obtained by maximizing $L^* \{ \beta^*(y_0) \}$, and

$$g_i^*(\beta^*) = \left( X_i' C A_i^{-1/2} U_i A_i^{-1/2} (Y_i - X_i C \beta^*) K_i(y_0) \right)^T = h_i^*(\beta^*) K_i(y_0).$$
Similar to Theorem 4.1, the estimator \( \hat{\beta}^*(y_0) = \{\hat{\beta}^*_1(y_0), \hat{\beta}^*_2(y_0)\}' \) satisfies

\[
\sqrt{nh} \begin{pmatrix}
\hat{\beta}^*_1(y_0) - \beta^*_1(y_0) \\
\hat{\beta}^*_2(y_0) - \beta^*_2(y_0)
\end{pmatrix} \xrightarrow{d} N(0, \psi^{*^{-1}}),
\]

where \( \beta^*(y_0) = \{\beta^*_1(y_0), \beta^*_2(y_0)\}' \) is the true value of \( \beta^* = (\beta_1^*, \beta_2^*)' \) and \( \Omega^* = \Phi^* V^{*-1} \Phi^* \) with \( \Phi^* = -E [\partial h_i^* \{\beta^*(y_0)\} / \partial \beta^*(y_0)] \) and \( V^* = E \left[ h_i^* \{\beta^*(y_0)\} h_i^* \{\beta^*(y_0)\}' \right]. \)

Under the null hypothesis, the true parameter of \( \beta^* \) and its estimator are specified as \( \beta^*(y_0) = \{0, \beta^*_2(y_0)\}' \) and \( \tilde{\beta}^* = (0, \tilde{\beta}^*_2)' \), respectively. Then the test statistic can be rewritten as

\[
T(y_0) = -2 \left[ L^* \{\tilde{\beta}^*(y_0)\} - L^* \{\hat{\beta}^*(y_0)\} \right].
\]

From similar arguments as in the proof of Corollary 5 in Qin and Lawless (1994), we obtain

\[
T(y_0) = -2 \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^n g_i^* \{\beta^*(y_0)\} \right]' V^{*-1/2} (\Phi^* \Omega^{*-1} \Phi^* - \Phi_1^* \Omega_1^{*-1} \Phi_1^*') V^{*-1/2} \left[ \frac{1}{\sqrt{nh}} \sum_{i=1}^n g_i^* \{\beta^*(y_0)\} \right] + o_p(1)
\]

where \( \Phi^* = (\Phi_1^*, \Phi_2^*) \) and \( \Phi_1^* \) corresponds to \( \beta_1^* \), and \( \Omega_1^* = \Phi_1^* V^{*-1} \Phi_1^* \). Therefore, since

\[
\Phi^* \Omega^{*-1} \Phi^* \geq \begin{pmatrix} \Phi_1^* & \Phi_2^* \end{pmatrix} \begin{pmatrix} \Omega_1^{*-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Phi_1^* \\ \Phi_2^* \end{pmatrix} = \Phi_1^* \Omega_1^{*-1} \Phi_1^*,
\]

that the distribution holds \( T(y_0) \xrightarrow{d} \chi^2_{(p_0 + 1)(q - 1)}. \)
Table 4.1: Relative efficiency of proposed estimators obtained from quadratic regression under compound symmetry and AR(1) structures.

<table>
<thead>
<tr>
<th></th>
<th>compound symmetry</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_0 = 1$ $y_0 = 2$ $y_0 = 3$ $y_0 = 4$</td>
<td>$y_0 = 1$ $y_0 = 2$ $y_0 = 3$ $y_0 = 4$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.166 1.186 1.108 1.192</td>
<td>1.062 1.034 1.037 1.145</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.316 1.500 1.233 1.399</td>
<td>1.140 1.169 1.132 1.217</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>1.250 1.442 1.229 1.284</td>
<td>1.097 1.113 1.126 1.152</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>1.075 1.316 1.113 1.158</td>
<td>1.074 1.093 1.012 1.115</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>1.271 1.637 1.511 1.193</td>
<td>1.125 1.272 1.232 1.377</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>1.310 1.584 1.665 1.199</td>
<td>1.112 1.331 1.344 1.288</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>2.490 2.644 2.660 2.748</td>
<td>2.220 2.657 2.551 2.262</td>
</tr>
</tbody>
</table>
Table 4.2: Estimated coefficients of depression data adopting linear regression under AR(1) structure.

<table>
<thead>
<tr>
<th></th>
<th>Medication</th>
<th></th>
<th>Psychotherapy</th>
<th></th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>Linear</td>
<td>Intercept</td>
<td>Linear</td>
<td>Intercept</td>
</tr>
<tr>
<td>$\gamma_0 = 21$</td>
<td>14.986</td>
<td>-1.060</td>
<td>14.409</td>
<td>-0.378</td>
<td>13.684</td>
</tr>
<tr>
<td>$\gamma_0 = 22$</td>
<td>15.366</td>
<td>-1.125</td>
<td>15.041</td>
<td>-0.520</td>
<td>14.042</td>
</tr>
<tr>
<td>$\gamma_0 = 23$</td>
<td>16.027</td>
<td>-1.164</td>
<td>15.736</td>
<td>-0.667</td>
<td>14.958</td>
</tr>
</tbody>
</table>

Table 4.3: Estimated covariate coefficients of depression data.

<table>
<thead>
<tr>
<th>$\gamma_0 = 14$</th>
<th>$\gamma_0 = 15$</th>
<th>$\gamma_0 = 16$</th>
<th>$\gamma_0 = 17$</th>
<th>$\gamma_0 = 18$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cage</td>
<td>-0.067</td>
<td>-0.072</td>
<td>-0.074</td>
<td>-0.072</td>
</tr>
<tr>
<td>Employ</td>
<td>-0.079</td>
<td>0.084</td>
<td>0.169</td>
<td>0.169</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\gamma_0 = 19$</th>
<th>$\gamma_0 = 20$</th>
<th>$\gamma_0 = 21$</th>
<th>$\gamma_0 = 22$</th>
<th>$\gamma_0 = 23$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cage</td>
<td>-0.064</td>
<td>-0.050</td>
<td>-0.017</td>
<td>-0.019</td>
</tr>
<tr>
<td>Employ</td>
<td>0.234</td>
<td>0.040</td>
<td>-0.078</td>
<td>-0.123</td>
</tr>
</tbody>
</table>
Table 4.4: Test statistic (T), p-value and degrees of freedom (df) of equality test of three treatments: medication (M), psychotherapy (P) and community (C).

<table>
<thead>
<tr>
<th>T</th>
<th>p-value</th>
<th>df</th>
<th>T</th>
<th>p-value</th>
<th>df</th>
<th>T</th>
<th>p-value</th>
<th>df</th>
<th>T</th>
<th>p-value</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.400</td>
<td>0.004</td>
<td>6</td>
<td>1.455</td>
<td>0.693</td>
<td>3</td>
<td>9.039</td>
<td>0.029</td>
<td>3</td>
<td>18.273</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>23.651</td>
<td>0.001</td>
<td>6</td>
<td>2.315</td>
<td>0.510</td>
<td>3</td>
<td>8.443</td>
<td>0.038</td>
<td>3</td>
<td>22.012</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>23.870</td>
<td>0.001</td>
<td>6</td>
<td>3.225</td>
<td>0.358</td>
<td>3</td>
<td>7.529</td>
<td>0.057</td>
<td>3</td>
<td>22.306</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>21.658</td>
<td>0.001</td>
<td>6</td>
<td>3.927</td>
<td>0.269</td>
<td>3</td>
<td>5.785</td>
<td>0.123</td>
<td>3</td>
<td>21.308</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>19.913</td>
<td>0.003</td>
<td>6</td>
<td>5.183</td>
<td>0.159</td>
<td>3</td>
<td>4.482</td>
<td>0.214</td>
<td>3</td>
<td>19.564</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>18.364</td>
<td>0.005</td>
<td>6</td>
<td>4.919</td>
<td>0.178</td>
<td>3</td>
<td>3.682</td>
<td>0.298</td>
<td>3</td>
<td>18.465</td>
<td>&lt;0.001</td>
<td>3</td>
</tr>
<tr>
<td>15.940</td>
<td>0.014</td>
<td>6</td>
<td>5.564</td>
<td>0.135</td>
<td>3</td>
<td>2.606</td>
<td>0.456</td>
<td>3</td>
<td>14.419</td>
<td>0.002</td>
<td>3</td>
</tr>
<tr>
<td>13.731</td>
<td>0.008</td>
<td>4</td>
<td>8.325</td>
<td>0.016</td>
<td>2</td>
<td>0.439</td>
<td>0.803</td>
<td>2</td>
<td>10.187</td>
<td>0.006</td>
<td>2</td>
</tr>
<tr>
<td>11.343</td>
<td>0.023</td>
<td>4</td>
<td>5.397</td>
<td>0.067</td>
<td>2</td>
<td>0.523</td>
<td>0.770</td>
<td>2</td>
<td>9.357</td>
<td>0.009</td>
<td>2</td>
</tr>
<tr>
<td>9.488</td>
<td>0.050</td>
<td>4</td>
<td>3.343</td>
<td>0.188</td>
<td>2</td>
<td>1.117</td>
<td>0.572</td>
<td>2</td>
<td>8.689</td>
<td>0.013</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 4.1: Quantile-Quantile plots for chi-square distribution with three degrees of freedom versus proposed test statistic testing equality of polynomial curves at $y_0 = 1$ under compound symmetry and AR(1) structures.
Figure 4.2: Fitted quadratic polynomial curves of medication (solid curve), psychotherapy (dashed curve) and community (dotted curve).
References


Kim, S., Cho, H. and Zhang, X. (2018). Initial severity-dependent longitudinal model with application to a randomized controlled trial of women with depression.


Multivariate Analysis **111**, 241–255.


Appendix A

HSIRB Approval Letter
Date: December 12, 2017

To: Hyunkeun Ryan Cho, Principal Investigator
   Xiaomeng Niu, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: Approval not needed for HSIRB Project Number 17-12-15

This letter will serve as confirmation that your project titled “Statistical Models for Correlated Data” has been reviewed by the Western Michigan University Institutional Review Board (WMU IRB). Based on that review, the WMU IRB has determined that approval is not required for you to conduct this project because you are not collecting personal identifiable (private) information about individual and your scope of work does not meet the Federal definition of human subject.

45 CFR 46.102 (f) Human Subject

(f) Human subject means a living individual about whom an investigator (whether professional or student) conducting research obtains

(1) Data through intervention or interaction with the individual, or
(2) Identifiable private information.

Intervention includes both physical procedures by which data are gathered (for example, venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes. Interaction includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.

“About whom” – a human subject research project requires the data received from the living individual to be about the person.

Thank you for your concerns about protecting the rights and welfare of human subjects.

A copy of your protocol and a copy of this letter will be maintained in the HSIRB files.