Semiparametric analysis of regression models for longitudinal data

Gu, Xiecheng

2013

Gu, X. (2013). Semiparametric analysis of regression models for longitudinal data. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54876>

<https://doi.org/10.32657/10356/54876>

SEMIPARAMETRIC ANALYSIS OF REGRESSION MODELS FOR LONGITUDINAL DATA

NANYANG
TECHNOLOGICAL
UNIVERSITY

GU XIECHENG

School of Physical and Mathematical Sciences

Nanyang Technological University

2013

SEMIPARAMETRIC ANALYSIS OF REGRESSION MODELS FOR LONGITUDINAL DATA

GU XIECHENG

School of Physical and Mathematical Sciences
Nanyang Technological University

A thesis submitted to Nanyang Technological University in fulfillment of the
requirement for the degree of
Doctor of Philosophy

2013

Abstract

Longitudinal data are often observed in many areas of research in practice such as public health and economic studies, where each subject involved is measured repeatedly over time. The defining feature of longitudinal studies is that the individual subject remains independent of another but the repeated measurements within subject are correlated. Since that the underlying correlation is hardly captured leads to the infeasibility of commonly used Generalized Linear Models (GLM), two types of models for different targets of inference, the marginal models and mixed effects models have been well developed in the past 30 years. In this thesis, we investigate new methods, extending these two models to deal with more practical and complicated cases.

We propose a penalized quadratic inference functions (QIF) approach in Chapter 2 and apply it to analyze the longitudinal data with large number of covariates. Particularly for variable selection, we develop a new QIF incorporating the regularization technique and a quadratic penalty based on SCAD. Theoretical properties are well derived under the scenario of diverging number of covariates. Extensive simulation studies have been conducted to assess the performance of our proposed

modeling approaches in comparison with the standard penalized GEE and QIF.

Modeling more complex longitudinal data with mismeasured covariates, in Chapter 3 we consider partially linear mixed effects models, in which the regression predictor is naturally set to incorporate a combination of linear and nonlinear effects to improve model fitting efficiency. To deal with measurement error in covariates, a two-step estimation procedure has been developed: we perform the standard multiple imputation for mismeasured covariates based on a measurement error model at the first step; with the nonlinear component approximated by smoothing techniques, we then estimate regression coefficients using the penalized quasi-likelihood at the second step. We further develop an iterative algorithm for implementation and establish the asymptotic properties for the resulting estimators. Simulation studies are carried out to evaluate the performance of the method. Finally, the proposed method is applied to a real data set from a randomized controlled trial of home-based intervention program for seniors.

This thesis is concluded with some remarks and the future research topics are discussed in the last chapter.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Xiang Liming, for her sustained support and excellent guidance during my PHD study at Nanyang Technological University. Her rigorous academic attitude and wealth knowledge of statistics always impress me and motivate me to enhance myself in future.

Secondly, I am grateful to my friends in statistics group for their selfless help and encouragement. I would like to give my deepest thanks to NTU for offering me a great opportunity to pursue my PHD, as well as the faculty and staffs of the Division of Mathematical Sciences, who taught me a lot and gave me kindly assistance in these years.

Finally, I want to express my great gratitude to my family who always stand firmly behind me during my PHD study. Specially appreciate my parents for supporting my decision in each moment. Thank my wife, Hongyi, for giving me unbounded help and courage to get over all the difficulties and challenges.

Contents

List of Figures	v
List of Tables	vii
1 Introduction	1
1.1 Longitudinal data analysis	1
1.2 Review of related methods	2
1.2.1 Generalized Estimating Equations	3
1.2.2 Linear Mixed Models	5
1.2.3 Generalized Linear Mixed Models	7
1.2.4 Generalized Partially Linear Mixed Models	9
1.3 PANS data set	10
1.4 Challenge problems	10
1.5 Outline	14
2 Regularized quadratic inference function with quadratic penalty	15
2.1 Introduction	15
2.1.1 Quadratic inference functions	15
2.1.2 High-dimensional longitudinal data and variable selection .	16
2.2 Model specification	19
2.3 Methodology	22

CONTENTS

2.3.1	Regularization	22
2.3.2	Penalization under the high-dimensional settings	23
2.3.3	Quadratic SCAD with a new iterative algorithm	24
2.4	Asymptotic inference	27
2.5	Simulation studies	32
2.6	Chapter summary	39
2.7	Appendix 1	40
3	Covariate measurement error in GPLMM	51
3.1	Introduction	51
3.2	Model specification	53
3.3	MIPQL estimation	56
3.3.1	Multiple imputation algorithm	56
3.3.2	Local smoother and penalized quasi-likelihood	59
3.4	Asymptotic inference	63
3.5	Simulation study	67
3.6	Application to PANS data	73
3.7	Chapter summary	77
3.8	Appendix 2	80
4	Conclusions and future work	95
4.1	Concluding remarks	95
4.2	Future work	97
	References	99

List of Figures

- 3.1 Trace plots of $\hat{\beta}^*$, $\hat{\eta}^*$, $\hat{\mu}_x^*$, $\hat{\sigma}^*$, $\hat{\sigma}_u^*$ and $\hat{\sigma}_x^*$ in the case that $N = 100$, $T = 5$, $\beta = 1$, $\eta = 0.5$, $\sigma = 1$ and $\sigma_b = 1$. The blue and red lines represents two iterative chains with two different sets of starting values, respectively. 69
- 3.2 True curve (solid line) and estimated curve (dotted line) values of nonlinear function $\theta(Z)$ calculated from 500 simulations. In (a): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (b): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$ 70
- 3.3 Estimated variance components σ and σ_b for 4 scenarios, based on 500 simulations. In (a): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (b): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$; In (c): $N = 100$, $T = 5$, $\beta = 1$, $\eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (d): $N = 100$, $T = 5$, $\beta = 1$, $\eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$ 71
- 3.4 nonparametric estimate of unknown nonlinear $\theta(Z)$. The estimated value is drawn in solid line with its 95% confidence interval drawn between two dotted lines. 78

LIST OF FIGURES

List of Tables

2.1	Comparison of quadratic SCAD and SCAD with simulated $AR(1)$ normal data when the model correlation is correctly specified, based on 1000 replications.	34
2.2	Comparison of quadratic SCAD and SCAD with simulated <i>Exchangeable</i> normal data when the model correlation is misspecified as $AR(1)$, based on 1000 replications.	34
2.3	The performance of the proposed RQIF with/without penalty, and the penalized GEE. For normal distribution with $AR(1)$ correlation structure, $\rho = 0.7$. QIF.r&p and QIF.r represent regularized QIF with and without proposed quadratic SCAD penalty, respectively, while GEE.p stands for penalized GEE.	36
2.4	The result analysis for the normal responses after 1000 simulations, using regularized QIF with quadratic SCAD penalty.	37
2.5	Variable selection for exchangeable Poisson data using the regularized QIF with penalty.	39
3.1	The mean, sample standard error (s.s.e) and estimated standard error (e.s.e) of the model parameter estimator $\hat{\beta}$ based on 500 simulations using proposed MIPQL method.	72
3.2	Description of baseline characteristics for our PANS data study. .	74

LIST OF TABLES

3.3	Inference for the PANS data and the performance of our proposed MIPQL compared with other existing approaches.	76
-----	---	----

Abbreviations

- **GLM:** *Generalized Linear Models*
- **GEE:** *Generalized Estimating Equations*
- **GLS:** *Generalized Least Squares*
- **MLE:** *Maximum Likelihood Estimator*
- **LMM:** *Linear Mixed Models*
- **GLMM:** *Generalized Linear Mixed Models*
- **GPLM:** *Generalized Partially Linear Models*
- **EM:** *Expectation-Maximization*
- **GPLMM:** *Generalized Partially Linear Mixed Models*
- **MI:** *Multiple Imputation*
- **RC:** *Regression Calibration*
- **CC:** *Covariates Calibration*
- **QIF:** *Quadratic Inference Function*
- **SCAD:** *Smoothly Clipped Absolute Deviation*
- **GMM:** *Generalized Method of Moments*
- **MIPQL:** *Multi-Imputed Penalized Quasi Likelihood*
- **SIMEX:** *Simulation-extrapolation*
- **MCMC:** *Markov chain Monte Carlo*
- **PQL:** *Penalized Quasi Likelihood*
- **CV:** *Cross Validation*
- **EBBS:** *Empirical Bias Bandwidth Selection*
- **MCEM:** *Monte Carlo Expectation-Maximization*
- **MCNR:** *Monte Carlo Newton-Raphson*

LIST OF TABLES

1 Introduction

1.1 Longitudinal data analysis

Longitudinal data have repeated measurements on every subject over a designed experiment period. Diggle et al. (2002) and Fitzmaurice et al. (2004) summarized various features and the fundamental structure of longitudinal data, and discussed important modeling techniques and their asymptotic inference in details.

The goal of longitudinal study is to evaluate the change in responses made by the effects of involving covariates over time. An important feature of longitudinal data is that the within-subject structure for each independent subject, also includes the non-zero correlation and except independence case, such correlation ρ is usually hard to identify. Hence, the linear model often used in the univariate analysis is no longer suitable as it does not account for the covariance structure that may exist among repeated measure within subjects. Another needed model specification is that the observations on one subject are very similar to another, in terms of the within-subject correlation. We assume the repeated observations on all the subjects in population share the same time-varying association. The number of repeated measurements for each subject may differ from one to another, subject to the study or experiment requirements. We refer the data sample with equal number of measurements within subjects as balanced data and unbalanced data otherwise. In general, the balanced design provides more robust estimators

1. INTRODUCTION

compared with the unbalanced design where for example, the marginal modeling approach may yield inconsistent estimators for covariance components of regression estimates such that the confidence interval estimation and other hypothesis testing regarding regression parameters fail subsequently.

The longitudinal data are often seen in economics, health and medical studies, where the target factors vary along the time line and the association among different measurements on the same subject can not be ignored. In the following subsections, we give a brief review of the most commonly used methods for handling longitudinal data and describe two real data examples that are used in this thesis.

1.2 Review of related methods

Generalized linear models (GLM) provide a useful class of tools for modeling various types of univariate data. A straightforward application of GLM to longitudinal data is not appropriate, due to the lack of independence among repeated measures on the same subject.

Many methods have been developed over past decades, accounting for the within-subject correlation under different assumptions. Classified by the distinct interpretations on regression parameters, the general extended methods are often referred to as population-averaged models and subject-specific models, specifically known as marginal models and mixed effect models. The two types of models are distinguished by the different targets of inference: the focus of marginal model is on estimating the main effects and variance structure of the overall population, while the mixed effect models concentrate on the interpretation of individual

subjects. Note that in the linear mixed models, the regression parameters for the conditional means also have interpretation in terms of the population means, as same as the marginal models. But in general, two models are not interconvertible.

1.2.1 Generalized Estimating Equations

Marginal models do not require distribution assumptions for the observations, but only a correct structure for the marginal mean of the response. This leads to a robust method known as generalized estimating equations (GEE) first proposed by Liang & Zeger (1986). The basic idea of GEE raised up from the quasi-likelihood initially set up by Wedderburn (1974), who stated that when applying the GLM to repeated measure data, only the first two moment structures of responses were required in order to avoid strict distribution assumptions. The GEE approach can provide a consistent estimator in modeling longitudinal data even when the correlation assumption is misspecified. Note that for the case of marginal linear model, the generalized least squares (GLS) can be regarded as a special case of the GEE method.

Suppose that we have a vector of longitudinal observations for subject i , $Y_i = (Y_{i1}, \dots, Y_{iT})^T$. We assume that the mean of Y_{ij} , $\mu_{ij} = E(Y_{ij})$ depends on covariates X_{ij} through an inverse link function $h(\cdot)$,

$$\mu_{ij} = h(X_{ij}^T \beta)$$

. The form of the GEE is then given by

$$\sum_{i=1}^N D_i^T \tilde{V}_i^{-1} (Y_i - \mu_i) = 0_p, \quad (1.2.1)$$

1. INTRODUCTION

where $\mu_i = (\mu_{i1}, \dots, \mu_{iT})^T$ is the marginal mean vector of the response and $D_i = \partial\mu_i/\partial\beta^T$ denotes the 1st derivative of μ_i with respect to p-dimension parameter vector β . In the equations (1.2.1), $\tilde{V}_i = A_i^{1/2}R(\rho)A_i^{1/2}$ is known as a working covariance matrix where A_i denotes $T \times T$ marginal variance matrix with diagonal elements $a_{jj}^2 = \text{var}(Y_{ij})$ and the working correlation $R(\rho)$ is assumed to be determined by the nuisance parameter ρ . Solving equations (1.2.1) for β gives the GEE estimate $\hat{\beta}$.

The key features of the GEE methodology include that the mean response and with-subject correlation are modeled separately by introducing a working correlation matrix R , and the asymptotic inference on the regression estimator $\hat{\beta}$ holds by constructing the so-called sandwich estimator of covariance even if the assumed correlation structure R is incorrect.

In recent decades, many outstanding works have been made in the GEE model. Prentice & Zhao (1991) proposed ad hoc estimating equations to solve covariance components and regression parameters jointly. Distinguished from the standard GEE, their joint GEE is referred as GEE2. Ye & Pan (2006) used modified Cholesky decomposition technique to decompose the underlying within-subject covariance matrices and then modeled the mean and covariance structures by GEE, motivated by the similar work in the context of GLM done by Pourahmadi (1999). Further Leng et al. (2010) considered the same decomposition skills in partially linear model which is more flexible in practice. When the specifications of the link function and variance-covariance structure are unknown, a nonparametric data-driven approach, such as kernel smoothing, can be implemented to approximate them and the regression parameters can be obtained via the estimated estimating equations (Chiou & Muller, 2005).

In summary, the GEE methodology enjoys a number of appealing properties for the regression analysis. First, the GEE estimator $\hat{\beta}$ is almost efficient as the optimal Maximum Likelihood Estimator (MLE). Secondly, it performs strongly against misspecification of the within-subject correlation. GEE is able to deal with discrete data in marginal model when the multivariate likelihood is hardly derived and the maximum likelihood estimation MLE fails to estimate the regression parameter β . Finally it is convenient computationally.

On the other hands, the GEE method has limitations. Its assumptions for correlations and variances sometimes are mathematically impossible, no matter how the data is generated. Consequently, sandwich estimation in turn inflates standard errors of GEE estimators. Without a full specified model, the GEE lacks effective ways for choosing the best model and for conducting goodness-of-fit tests. For more model specifications and applications, please refer to Hardin & Hilbe (2002).

An alternative way to account for the within-subject correlation is by incorporating random effects into generalized linear models. That is, a subset of the regression parameters is allowed to randomly vary from one subject to another. We give a review of linear mixed models (LMM) first and then discuss about generalized linear mixed models (GLMM) due to its conceptual and analytic complexity.

1.2.2 Linear Mixed Models

In LMM, without loss of generality, for each repeated response Y_{ij} , a normal distribution given random effect vector b_i is assumed with constant conditional

1. INTRODUCTION

variance σ_b^2 . The important feature of mixed effect models is the conditional independence assumption, that is, given b_i , Y_{ij} are independent of each other. Usually, random effects b_i are also assumed to be normally distributed. Hence, the identity link function gives the conditional mean:

$$E(Y_{ij}|b_i) = X_{ij}^T\beta + B_{ij}^Tb_i, \quad (1.2.2)$$

where X_{ij} and B_{ij} are corresponding covariates vectors for fixed and random effects. b_i are also assumed to be independent from one subject to another. The LMM can be represented as follows,

$$Y_{ij} = X_{ij}^T\beta + B_{ij}^Tb_i + e_{ij}, \quad (1.2.3)$$

where e_{ij} is random error term having normal distribution with zero mean and variance σ_e^2 . The simple form of LMM leads to the marginal mean response over all subjects,

$$E(Y_{ij}) = E(E(Y_{ij}|b_i)) = X_{ij}^T\beta. \quad (1.2.4)$$

Thus the regression parameters β in the subject-specific model (1.2.2) also have population-averaged interpretations on how the changes of related covariates influence the marginal mean response.

For longitudinal data, Laird & Ware (1982) summarized the estimation and inference in random-effects models under known or unknown covariance components. The Expectation-Maximization (EM) algorithm was introduced to estimate random effect distribution. The details of Newton-Raphson and EM algorithms for LMM were discussed further in the work of Lindstrom & Bates

(1988). They also proposed a two-step algorithm to fit repeated measures data in nonlinear mixed models, by constructing pseudo data (Lindstrom & Bates, 1990). In more complicated cases, say, when the random-effect model may have time-varying coefficients, the penalized spline method was incorporated in order to estimate both population-averaged and subject-specific model factors (Chen & Wang, 2011). LMM has widely used in many scientific applications as an analytical tool. For example in biology, it provided convenient way to test the significance of genes from cDNA microarray expression data (Wolfinger et al., 2001).

1.2.3 Generalized Linear Mixed Models

An extension of LMM is known as Generalized Linear Mixed Models (GLMM) (McCulloch et al. 2001). The random effects is involved into GLM to handle the data with repeated measures and more types of distribution are allowed for handling discrete or non-negative responses. Moreover, in contrast to LMM, the inverse link function $h(\cdot)$ in GLMM is not just simply identity for normal responses but describes more complicated relationship between mean and predictors,

$$E(Y_{ij}|b_i) = h(X_{ij}^T\beta + B_{ij}^Tb_i). \quad (1.2.5)$$

Other model specifications are analogous to LMM. Unlike marginal modeling in which only first two moments structures are specified, the joint distribution of Y_{ij} and b_i can be fully specified in the context of GLMM. That is, $f(Y_{ij}, b_i) = f(Y_{ij}|b_i)f(b_i)$. Furthermore, $f(Y_{ij}|b_i)$ is assumed to have an exponential family distribution.

1. INTRODUCTION

The interpretations of the fixed effects β have different meanings from population-averaged models, GEE. They represent the influence of covariates on a specific subject in terms of the change in a subject's transformed mean response due to the change of related covariates.

Although the regression parameters in the special case of GLMM, LMM also have population-averaged interpretation as mentioned above, in general, the non-linear inverse link function generally does not lead to the same conclusion,

$$E(Y_{ij}) = E(h(X_{ij}^T\beta + B_{ij}^Tb_i)) \neq E(h(X_{ij}^T\beta)). \quad (1.2.6)$$

Thus the interpretations of β are not generally identical in the marginal and conditional modelings and yields different targets of inference. One of the advantages of the GLMM approach is to assume that the within-subject correlation is explained by introducing random effects b . Under the conditional independence assumption, the joint distribution of responses Y and b is easily derived and follows with the statistical inference based on the likelihood function. Therefore we are able to obtain the MLE of fixed effects β and covariance components. Another appealing features of GLMM is that even the distribution of random effects is misspecified, the estimates of β are still sufficiently unbiased (Fitzmaurice et al., 2004). Due to its structural complexity, the exact inference is hardly given and the great effort made by Breslow & Clayton (1993) leads us to some approximated inference.

1.2.4 Generalized Partially Linear Mixed Models

As a class of more flexible and compatible tools for modeling complex longitudinal or clustered data, Generalized partially linear mixed models (GPLMM) has been drawn many attentions in the recent decades. Carroll et al. (1997) considered a single-index model with random effects and proposed the kernel smoothing method in nonparametric estimation of unknown nonlinear function. If there exists varying coefficients in longitudinal data, the double penalized quasi-likelihood approach was adopted to jointly estimate variance components and varying coefficients in mixed model (Zhang 2004). In addition to (1.2.5), we can define the relationship between responses and partially linear predictors as follows

$$E(Y_{ij}|b_i) = h(X_{ij}^T\beta + B_{ij}^Tb_i + \theta(Z_{ij})), \quad (1.2.7)$$

where Z_{ij} denote the covariates related to nonlinear function $\theta(\cdot)$.

Similar to the GLMM, random effects are specified in the GPLMM to incorporate potential dependence among repeated observations under the same subject. In the meantime, the GPLMM allow some covariates to be linear and others (e.g., time-varying covariates) to be nonlinear, which naturally inherit advantages of generalized linear models and nonparametric models. Excellent applications of the GPLMM in analyzing longitudinal data include the study of major risk factors of coronary heart disease mortality rates in men (Carroll et al., 1997) and semi-parametric analysis of the relation between virologic and immunologic responses in AIDS clinical trials (Liang, 2009).

1. INTRODUCTION

1.3 PANS data set

This example data set was provided by Burke et al. (2012), who conducted a 6-month trial to investigate the impact of the physical activities and nutrition intakes by instruction for seniors (PANS) on their central obesity. The two indicators for central obesity are recorded: the body mass index (BMI) and waist-to-hip ratio (WHR). In 2010, totally 478 participants were recruited and randomly allocated to the intervention (248 peoples) and control (230 peoples) groups. About three-quarter of them (375 peoples) were available for the following statistical analysis. The initial data were collected by participants on their own at baseline and post-time after half a year. The self-completion reports submitted to investigator included many demographic and lifestyle characteristics, such as age, sex, working and smoking statuses. The participants also reported how much time they spent on physical exercises every week, classified by intensity of activities. For example, the participants were required to record how long they sat in the room and went out for walking per week.

1.4 Challenge problems

The two main challenges that we encounter and would like to overcome in Chapter 2 and 3 are (1) when the number of covariates is relatively large comparing with the sample size in longitudinal data, the existing GEE-type method is not able to accurately choose the most important variables and efficiently estimate the corresponding coefficients of interest at the same time; (2) It is common that the practical longitudinal data is observed with measurement error, so a

hierarchical model which takes the measurement error into account and remains powerful performance on consistency and other asymptotic properties of resulting estimators is strongly needed but difficult to specify.

The GEE approach which leads to consistent estimators of regression parameters may cause loss of efficiency unless the within-subject correlation is correctly specified, since the asymptotic variance of $\hat{\beta}$ does not attain to the Cramér-Rao lower bound. Many works have been done to handle this problem. For example, Qu et al. (2000) improved the performance of the GEE by using quadratic inference functions and the evidence that QIF outperforms GEE in terms of efficiency was support by the simulation study therein; Leung et al. (2009) proposed a hybrid method to enhance the efficiency by combining equations based on different working correlations; Wang et al. (2009) discussed two kinds of empirical likelihood based methods which lead to more efficient estimators in the context of GEE approach, and so forth. Another drawback of the GEE approach is found in the estimation of correlation parameter ρ . Shown by Crowder (1995), there are some cases that the estimator of ρ does not exist. Particularly, GEE as an extension of GLM to deal with the multivariate discrete data, may provide less convinced estimator when the dispersion parameter ϕ , representing the common phenomenon overdispersion in discrete case, is sufficiently large (Cox 1983).

With high-dimensional covariates, the problem of applying GEE-type method relies on the accuracy and efficiency of the variable selection. In our preliminary experiment in Chapter 2, the existing penalized approaches do not yield satisfying result, combined with conventional QIF. Besides, QIF is able to remarkably improve the efficiency of the resulting estimators but relatively large number of covariates may lead it to some singular form. The excellent works recently done

1. INTRODUCTION

by Wang (2011) and earlier Fan & Peng (2004) assessed the performance of GEE with relatively small number of covariates but did not further deal with the case that the number of covariates is comparable to the sample size in the simulation therein, which probably is the more common problem in variable selection in practice.

For both population-averaged and subject-specific models specified above, a common assumption is that the variable components are linearly independent and combined in the predictor. This linear combination is inconsistent with the fact we usually encounter in the medical and related studies. For example, in the infectious disease study (Diggle et al. 2002), it was suggested that the ages of patients be treated via an unknown nonlinear function. Also in CD4 count data study, the length of treatment time was considered to have impact on the mean response through a nonlinear relationship (Wu 2002). Therefore, if a subset of covariates is partially combined into predictor, the estimation procedure of GLMM becomes more complicated as the nonlinear components have to be approximated first. Many asymptotic inference problems may raise when considering the nonparametric function (Lin & Carroll 2001).

Another challenge often encountered in many empirical studies is the measurement error in the data. it commonly arises from instrumental error, biological variation, or errors in self-report questionnaire-based data. When the explanatory variable is measured with error, it is found that the corresponding estimated effect in the regression model will be attenuated eventually. Particularly, Carroll et al. (1995) interpreted the negative influence of the measurement error in the covariates such as blood pressure in the Framingham Heart Study data (Kannel et al. 1986), and discussed the sources of error for covariate CD4 count in the

AIDS study of Tsiatis et al. (1995). Furthermore, Kipnis et al. (2003) found that measurement error in dietary intake assessment by using the Food Frequency Questionnaires (FFQ) could result in severe attenuation in the estimated effect of disease relative risk in a biomarker study. Cotton et al. (2005) noticed that measurement error could lead to inaccurate diagnosis of appropriate educational placement for children with reading difficulties. Using conditional expectation method to surrogate the mismeasured ambient concentrations, the recent environment research by Wannemuehler et al. (2009) shows that the measurement error can substantially affect the assessment of the association between pollutant exposure and a health outcome.

Various methods for dealing with error-prone variables have been studied in the literature. Schafer (1987) provided an EM algorithm to approximately estimate the coefficients when the covariate has error in a logistic regression. Cook & Stefanski (1994) and Stefanski & Cook (1995) introduced a simulation-extrapolation (SIMEX) for parametric measurement error models where the variance of error is at least well estimated. A good review of existing methods can be found in the work of Carroll et al. (1995) especially for linear models. In the GLMM with error in one of predictors, Wang et al. (1998) implemented the SIMEX method in GLMM with error in one of the predictor to correct the estimation bias. Further, Lin & Carroll (2000) adopted SIMEX procedure when the predictor is of unknown form. In the same situation that the predictor is hardly identifiable, Ma & Carroll (2006) constructed a local estimating equations to deal with the error in predictor semiparametrically. Another recent work by Thomas et al. (2011) imputed the values of the true covariates by matching the recommended four moments.

1. INTRODUCTION

In the subsequent chapters, we aim to develop new and improved methods for tackling these complex longitudinal data problems.

1.5 Outline

The rest of this thesis is organized as follows. In Chapter 2, we propose a more efficient method to deal with high-dimension data, based on the marginal modeling. A new quadratic penalty for variable selection is defined and plays quite well in the simulation section. The asymptotic inference is concretely derived under the case of diverging number of covariates. In Chapter 3, we turn to mixed effects model where random effects are introduced in order to handle the underlying within-subject correlation. We consider partially linear predictor in mixed model to improve the efficiency and accuracy of model fitting. Particularly, we describe and discuss how to perform subject-specific model when the data are measured with error. A penalized quasi-likelihood algorithm is then specified, with multiple imputation procedure in advance. To evaluate the performance of our proposed method, a small simulation study and the example of PANS data are illustrated in the end. We conclude this thesis and address the future work in Chapter 4.

2 Regularized quadratic inference function with quadratic penalty

2.1 Introduction

2.1.1 Quadratic inference functions

To enhance the performance of the estimating equations method, Qu et al. (2000) introduced the quadratic inference function (QIF), which is equally or more efficient than GEE and has useful properties such as robustness, a goodness-of-fit test and model selection. The QIF approach ensures the existence of optimal estimator $\hat{\beta}$ no matter whether the correlation structure is correctly specified or not. Based on the idea of the work done by Hansen (1982), it only requires to directly solve a one-dimensional minimization problem rather than a multivariate correlation problem in the standard GEE. From the form of the QIF, it shows that theoretically QIF is able to embody all the possibilities of working correlations. This forms its core to improve the efficiency of resulting estimates by combining appropriate working structures as many as possible so as to approach the true structure as closed as possible.

However, definitely it is impractical to get as many working matrices into the QIF as we expect. Suppose that we have N individuals, with q covariates and r

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

working correlations assumed. The fact that $rq > N$ causes the instability of the QIF in term of the weighting covariance matrix, that is, it leads the empirical matrix to be singular. In this case, the QIF becomes invalid. Although it can still be used to obtain fairly good estimators, their efficiency is not guaranteed. The invalidity of the QIF may also be found in high-dimensional problems where usually q is set to be quite large.

2.1.2 High-dimensional longitudinal data and variable selection

Some of the recent research in clinical studies is aimed at small-sample, high-dimensional data sets. Spellman et al. (1998) identified nearly 800 genes which vary during the yeast cell cycle. Golub et al. (1999) data set consist of 72 patients with leukemia and 7129 gene expressions. Similarly, Alon et al.(1999) collected 40 tumour and 22 normal colon tissues and investigated more than 2000 gene expressions. The study of Chiaretti et al. (2004) contains microarray expressions for 128 patients with either T-cell or B-cell type leukemia. It is often that not all of the predictor variables may contribute equally well to the explanation of the outcome. Some of them may even not contribute at all. Thus it is important to select only those variables which are informative in order to obtain best model, in terms of accuracy and smaller number of variables.

When the number of parameters is fixed, Xie & Yang (2003) investigated the performance of the GEE when the cluster size was supposed to go to infinity. On the study of high-dimensional longitudinal data, Fan & Peng (2004) introduce the non-concave penalized likelihood when the number of parameters q diverges at

the rate much slower than N . The purpose of their study is to find a well-defined penalty that can be used to estimate the coefficients and select the influential variables simultaneously. Furthermore, under the similar ratio of q and N , Wang (2011) analyzed the binary data with diverging q using GEE and Wang et al. (2012) provided a penalized GEE method for variable selection based in a high-dimensional setting.

Since the QIF has naturally advantage on the efficiency and optimality of the consistent estimator $\hat{\beta}$, in comparison with the GEE-type method, our aim in this chapter is to develop a new QIF inference procedure based on the longitudinal data with a large number of covariates, especially for the situations where the number of parameters is comparable to the sample size. In particular, our work is influenced by Warton (2011) who studied a regularity approach on GEE to overcome the problem that the usual sandwich estimator becomes singular when the cluster size exceeds number of clusters. In the work of Warton (2008), the ridge regularization is shown as a specified penalty added to the likelihood function. This inspires us to develop a new QIF with regularity, which is naturally embedding a regularity term into the QIF and treating it as an underlying penalty whose form unfortunately is unknown. Penalization methods are commonly used to deal with high-dimensional data to find the most influential coefficients. Many penalized regression based inference have been developed. Recently Qu & Li (2006) proposed a ridge-type penalty to QIF for their generalized varying coefficient model. However the purpose of their method is to select the new spline parameter γ rather than the regression covariates, whereafter Bai et al. (2009) extended their work and used the same penalty in the QIF to select the best spline approximation to the unknown link function. The widely used SCAD

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

penalty suggested by Fan & Li (2001) has excellent statistical properties, however during our simulation study, it shows that this penalty is no longer suitable for variable selection using QIF when number of initial covariates is sufficiently large. Depending on the structure of QIF as well as SCAD, we then develop a new penalty named the quadratic SCAD, which performs well regarding both parameter estimation and variable selection. Under the QIF model framework, it enables us to find the most useful covariates without loss of coefficients consistency. Furthermore, since the SCAD is not twice differentiable and has to be approximated by its reduced form, we develop a new iterative algorithm for implementation. The idea of our proposed algorithm comes from the MM method proposed by Hunter & Li (2005) and the unified algorithm by Fan & Li (2001).

In our early simulation studies, we find that although SCAD performs quite well for many cases (mainly $rq < N$), it fails to provide strong result in the cases that $rq > N$ or, rq and N are close. The standard penalty always requires large regularity parameter during the iteration step in order to guarantee the correct variable selection using QIF. However this inevitably leads to a significant bias between the estimators and the underlying true coefficients. In short, the use of standard SCAD in QIF under high-dimensional setting does not provide desired strong power to simultaneously handle the estimation and variable selection processes. These force us to seek for a stronger penalty that can penalized QIF even RQIF without loss of accuracy and efficiency, when the number of covariates is relatively large.

The rest of this chapter is organized as follows. The QIF model is first specified and discussed in Section 2.2. In Section 2.3, the regularization technique is introduced into the QIF to overcome its infeasibility in high-dimensional settings,

followed by a new quadratic penalty function presented in details. Asymptotic properties are given in Section 2.4 when number of covariates is diverging. We assess the performance of our proposed method and compare it with standard GEE through simulations in Section 2.5. Section 2.6 provides some further discussion. Technical details are given in the last section.

2.2 Model specification

We focus on the generalized linear model defined as follows:

- Let Y_{ij} be longitudinal response with $E(Y_{ij}) = \mu_{ij}$ and $Var(Y_{ij}) = v(\mu_{ij})$, where $\{Y_{ij}, i = 1, \dots, N, j = 1, \dots, T\}$ are supposed to be correlated within each subject in terms of R_i . Let $Y_i = (Y_{i1}, \dots, Y_{iT})$, for $i = 1, \dots, N$, representing T -vector of dependent variables, independent over all subjects. For simplicity, R_i is assumed to be identical for all subjects in practice.
- The linear predictor ζ on the mean: $\zeta_{ij}^\mu(\beta) = x_{ij}^T \beta$, where x_{ij}^T is the j th row vector of the $T \times q$ design matrix $\mathbf{X}_i$ and $\beta = (\beta_1, \dots, \beta_q)^T$ the vector of the regression coefficients.
- The corresponding inverse link function h :

$$\mu_{ij} = E(Y_{ij}|\beta) = h(\zeta_{ij}^\mu(\beta)) \quad (2.2.1)$$

Remark 2.2.1. ζ above represents the linear predictor on the mean. and $h(\cdot)$ denotes the inverse link function associating the ζ with the mean of response μ . In the context of Generalized linear models (GLM), we always assume that the

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

response y comes from some distribution belonging to the exponential family,

$$f(y; \theta, \phi) = \exp\{[y\theta - \gamma(\theta)]/\phi + \tau(y, \phi)\},$$

where θ denotes the canonical parameter, ϕ is a scale parameter, and γ and τ are known functions. Then the canonical link function expressing θ in terms of μ , and the corresponding canonical inverse link $h(\theta) = \mu$. Besides, many other commonly used link functions can be used, provided that it can be easy to match the domain of the link function to the range of the mean of distribution function.

Instead of solving such a linear combination problem, Qu et al. (2000) provided an alternative way by introducing QIF. First, an extended score function is defined as

$$g(\beta) = \frac{1}{N} \sum_{i=1}^N g_i(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N D_i^T A_i^{-1/2} M_1 A_i^{-1/2} (y_i - \mu_i) \\ \sum_{i=1}^N D_i^T A_i^{-1/2} M_2 A_i^{-1/2} (y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^N D_i^T A_i^{-1/2} M_r A_i^{-1/2} (y_i - \mu_i) \end{pmatrix}, \quad (2.2.2)$$

where M_i is the inverse of R_i . Note that when $r = 1$, g is reduced to the standard GEE, otherwise, it contains more estimating equations than the number of parameters q . We can estimate β by applying the generalized method of moments (GMM) developed by Hansen (1982), which optimizes the linear combination of the elements in g by solving the following problem,

$$\hat{\beta} = \arg \min_{\beta} N g^T G^{-1} g, \quad (2.2.3)$$

where the optimal choice of the $rp \times rp$ weighting matrix G is the variance of g and can be consistently approximated by

$$G_N = \frac{1}{N^2} \sum_{i=1}^N g_i(\beta) g_i^T(\beta). \quad (2.2.4)$$

Let $C_N(\beta) = (g_1(\beta), \dots, g_N(\beta))$. Then G_N can also be expressed as

$$G_N = \frac{1}{N^2} C_N(\beta) C_N^T(\beta). \quad (2.2.5)$$

Hence we can define the QIF by using G_N as

$$Q_N = N g^T G_N^{-1} g. \quad (2.2.6)$$

(2.2.6) has the analogous form of the commonly used Wald's statistic but with higher dimension of the estimating equation g .

To implement the Newton-Raphson algorithm for $\hat{\beta}$, Q_N provides consistent approximation of 1st and 2nd derivatives with respect to β :

$$\dot{Q}_N(\beta) = 2N \dot{g}^T G_N^{-1} g, \quad (2.2.7)$$

$$\ddot{Q}_N(\beta) = 2N \dot{g}^T G_N^{-1} \dot{g}, \quad (2.2.8)$$

where $\dot{g} = \partial g / \partial \beta$, is $rq \times q$ matrix.

Using QIF method is not compulsory to get exactly the explicit solution of the linear combination of g_i but instead, the optimization is reached by minimizing the quadratic inference function Q_N . Since it considers more possibilities of working correlation, it definitely outperforms the GEE in terms of efficiency, especially

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

when the true correlation is unknown and R_i is misspecified.

2.3 Methodology

2.3.1 Regularization

In (2.2.4) and (2.2.5), we notice that G_N is invertible if $rq \leq N$ and Q_N turns to be invalid when $rq > N$. Therefore, direct application of the QIF is not appropriate in this case.

To solve this singular problem of the covariance matrix in the likelihood function and further in the sandwich estimator of QIF, we assume, in a similar manner of the ridge regularization provided by Warton (2008, 2011), that

$$G_{N,\lambda_1} = \lambda_1 G_N + (1 - \lambda_1) G_{\bar{R}}, \quad (2.3.1)$$

for $rq > N$, where $G_{\bar{R}}$ is supposed to be some working covariance matrix only determined by the specified combined choices $\bar{R} \subset \{R_1, \dots, R_r\}$. As a special case,

$$G_{N,\lambda_1} = \lambda_1 G_N + \frac{1}{N^2} (1 - \lambda_1) I \quad (2.3.2)$$

for independence working correlation I . $G_{\bar{R}}$ can also be specified the diagonal matrix $diag(G_N)$ generated from the corresponding diagonal elements of G_N .

The regularity terms in (2.3.1) are nonsingular, so the corresponding G_{N,λ_1} can be used as a substitute for the weighting matrix G_N in (2.2.6). The regularized

QIF (RQ) is then defined as

$$RQ_{N,\lambda_1} = Ng^T G_{N,\lambda_1}^{-1} g. \quad (2.3.3)$$

Now as G_N is shrunk toward the specified matrix $G_{\bar{R}}$, the instability problem of the estimator of G is somehow resolved.

2.3.2 Penalization under the high-dimensional settings

Now denote the total number of parameters by q , and $\beta = (\beta_1, \dots, \beta_q)^T$. The extended design matrix X_i for each i belongs to $R^{T \times q}$. We define that

$$\Omega = \{i = 1, \dots, q\}, \quad (2.3.4)$$

$$\Lambda = \{i = 1, \dots, q \mid \beta_i \neq 0\}. \quad (2.3.5)$$

Let p be the dimension of Λ and $q \rightarrow \infty$ so that $q \gg p$. It becomes easier to trigger the singularity conditions in the QIF that $rq > N$. Besides estimating β'_i s from Λ , we have to exclude the zero coefficients β'_i s from $\Omega - \Lambda$ at the final selected model.

In order to estimate the non-zero coefficients and proceed the variable selection at the same time, we introduce the penalization into the RQ in (2.3.3) and form the following penalized RQ (PRQ):

$$PRQ(\beta) = Ng(\beta)^T G_{N,\lambda_1}^{-1} g(\beta) + K_N P_{\lambda_2}(|\beta|), \quad (2.3.6)$$

where K_N is a scale for the penalty term and usually equals N for simplicity.

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

The penalty function $P_{\lambda_2}(|\beta|) = \sum_{j=1}^q p_{\lambda_2}(|\beta_j|)$. For each penalty term $p_{\lambda_2}(\theta)$, we first adopt the SCAD penalty, though other forms of penalty can be chosen alternatively. The 1st derivative of SCAD penalty must satisfy

$$p'_{\lambda_2}(\theta) = \lambda_2 \left\{ I(\theta \leq \lambda_2) + \frac{(a\lambda_2 - \theta)_+}{(a-1)\lambda_2} I(\theta > \lambda_2) \right\} \quad (2.3.7)$$

for $\theta \geq 0$ and $a > 2$. Compared with other choices of penalty such as *lasso* and *ridge*, the SCAD penalty in penalized likelihood yields the estimator $\hat{\beta}$ to have three good properties: unbiasedness, sparsity and continuity. See Fan & Li (2001) for details. When the sample size is very large, unfortunately, it is found that SCAD penalty is not sufficiently strong to penalize the regularized QIF proposed. The drawback of SCAD and the proposed new penalty will be discussed in the following sections.

2.3.3 Quadratic SCAD with a new iterative algorithm

For non-concave penalized likelihood, Fan & Li (2001) developed a unified algorithm to deal with the penalty term that is not twice differentiable at the origin. In particular, the local quadratic approximation (LQA) instead of was applied to take over the real derivatives. Suppose an initial value β_0 is given to be closed to the consistent estimator $\hat{\beta}$, the derivatives of the penalty function can be then approximated by

$$p'_{\lambda}(|\beta_j|) \approx \frac{p'_{\lambda}(|\beta_{0j}|)}{|\beta_{0j}|} \beta_j, \quad (2.3.8)$$

where β_{0j} is the j th element of β_0 and required to be non-zero. For $\beta_{0j} = 0$, we add a sufficiently small $\epsilon > 0$ to the denominator and the approximation should

be rewritten as

$$p'_\lambda(|\beta_j|) \approx \frac{p'_\lambda(|\beta_{0j}|)}{|\beta_{0j}| + \epsilon} \beta_j. \quad (2.3.9)$$

Before deriving the algorithm of the PRQ with quadratic SCAD penalty, we first review a simple example mentioned in Fan & Li (2001), where $\beta^{(k)}$ in the k th step of the iteration procedure should satisfy that

$$\beta^{(k)} = \beta^{(k-1)} - [\ddot{l}(\beta^{(k-1)}) + K_N \Sigma(\beta^{(k-1)}, \hat{\lambda})]^{-1} [\dot{l}(\beta^{(k-1)}) + K_N \Sigma(\beta^{(k-1)}, \hat{\lambda}) \beta^{(k-1)}], \quad (2.3.10)$$

where l denotes some loss function,

$$\begin{aligned} \dot{l}(\beta^{(k-1)}) &= \frac{\partial l(\beta^{(k-1)})}{\partial \beta}, \\ \ddot{l}(\beta^{(k-1)}) &= \frac{\partial^2 l(\beta^{(k-1)})}{\partial \beta \partial \beta^T}, \end{aligned}$$

and Σ is a diagonal matrix with elements of standard penalty *SCAD*,

$$\Sigma(\beta^{(k-1)}, \hat{\lambda}) = \text{diag} \left\{ \frac{p'_\lambda(|\beta_1^{(k-1)}|)}{|\beta_1^{(k-1)}| + \epsilon}, \dots, \frac{p'_\lambda(|\beta_q^{(k-1)}|)}{|\beta_q^{(k-1)}| + \epsilon} \right\}. \quad (2.3.11)$$

The new penalty for QIF and the regularized QIF is based on the relatively similar structures of GEE and QIF. The QIF derived from the quasi-score functions can be approximately considered as the quadratic GEE. Therefore we define the new penalty as the quadratic SCAD, $qS_\lambda(\theta)$, whose 1st derivative should satisfy

$$qS'_\lambda(\theta) = \lambda^2 \left\{ I(\theta \leq \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\}^2 / (|\theta| + \epsilon). \quad (2.3.12)$$

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

Approximated by LQA method, as same as (3.5.1) and (2.3.9), the quadratic SCAD penalty can be rewritten as

$$qS'_{\lambda_2}(|\beta_j|) \approx \frac{qS'_{\lambda}(|\beta_{0j}|)}{|\beta_{0j}| + \epsilon} \beta_j. \quad (2.3.13)$$

Solving the $PRQ(\beta)$ with quadratic SCAD penalty gives the following iterative algorithm. Specifically, in the k th step of iteration, the estimator denoted by $\beta^{(k)} = \beta^{(k)}(\lambda_1, \lambda_2)$ is updated by:

$$\begin{aligned} \beta^{(k)} = & \beta^{(k-1)} - [R\ddot{Q}IF(\beta^{(k-1)}, \hat{\lambda}_1) + K_N(\Sigma(\beta^{(k-1)}, \hat{\lambda}_2))^2]^{-1} \\ & [R\dot{Q}IF(\beta^{(k-1)}, \hat{\lambda}_1) + K_N(\Sigma(\beta^{(k-1)}, \hat{\lambda}_2))^2 \beta^{(k-1)}], \end{aligned} \quad (2.3.14)$$

where $R\dot{Q}IF$ and $R\ddot{Q}IF$ denote the 1st and 2nd derivatives of the regularized QIF. This new penalty $qS(\theta)$ shows outstanding performance under both low and high-dimensional settings in the following simulations. Given $\epsilon = 10^{-6}$, and selected tuning parameter λ_1 as well as λ_2 , the computation process stops until the tolerance is reached and the resulting estimator is denoted by $\hat{\beta}$. For variable selection, the estimated coefficients are penalized to be zero when the values of them are below the threshold value 10^{-4} .

There are some existing methods for choosing λ_1 and λ_2 . Usually, the optimal estimators of such tuning parameters can be obtained by minimizing the targeted loss function, for example, the RQ in our cases. Using w -fold cross-validation, we first obtain $\hat{\lambda}_1$ by

$$\hat{\lambda}_1 = \arg \min_{\lambda_1} \sum_{j=1}^w RQ(\hat{\beta}^{/j}, D_j, \lambda_1) \quad (2.3.15)$$

where D and D_j denote the full and test data in fold j , $j = 1, \dots, w$. $\hat{\beta}^{/j}$ is

obtained by using training data $D - D_j$. λ_2 can be estimated in a similar way by minimizing

$$\sum_{j=1}^w \{RQ(\hat{\beta}^{/j}, D_j, \hat{\lambda}_1) + K_n P(|\hat{\beta}|^{/j}, \lambda_2)\} \quad (2.3.16)$$

In practice, another tuning parameter a in the SCAD penalty is often assigned to be 3.7 as suggested by Fan & Li (2001).

2.4 Asymptotic inference

In this section we investigate the asymptotic properties for the regularized QIF proposed with penalty when the number of initial covariates is diverging. We show the consistency and asymptotic normality of the resulting estimators. Recalling from (2.3.4), we know that each entry β_i of unknown parameter vector β is allocated into two subgroups: Λ for the true coefficients and $\Omega - \Lambda$ for the unused zero coefficients, with dimension p and $q - p$ respectively. First of all, we assume that

(A1) $q^4 = o_p(N)$. β lies in a compact set $\mathcal{B}$ and the true parameter β_0 also in the interior of $\mathcal{B}$.

From (A1), we know that $p^5 = o_p(N)$. Throughout this section, G_{N, λ_1} for RQ_{N, λ_1} is assumed to be in form of (2.3.2) without loss of generality. Before investigating the asymptotics of $\hat{\beta}$, we take a look on the regularity parameter λ_1 . Assuming that

(A2) $\lambda_{\min}(N^2 G_N) \longrightarrow \infty$ as $N \longrightarrow \infty$, where $\lambda_{\min}$ finds the smallest eigenvalue of the objective matrix.

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

We then have the following lemma,

Lemma 2.4.1. *If $N \rightarrow \infty$ so $Nw \rightarrow \infty$, then $\hat{\lambda}_1$ obtained from (2.3.15) tends to 1 in probability, that is,*

$$\hat{\lambda}_1 \xrightarrow{P} 1. \quad (2.4.1)$$

This lemma tells us that when sample size increases, the optimal $\hat{\lambda}_1$ is very closed to 1, yielding $G_{N,\lambda_1} \rightarrow G_N$ and $Q_N \rightarrow RQ_{N,\lambda_1}$. Therefore, it can be proved that the estimator $\hat{\beta}$ obtained from the regularized QIF still remains consistent. Suppose that (A 2) holds and for the discussion below, we let $\hat{\lambda}_1 \rightarrow 1$ and furthermore, use λ_1 replacing $\hat{\lambda}_1$ for simplicity. We will establish our theorems, under the following regularity conditions:

$$(A3) \quad \|g\| = O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right) \text{ and } \|\dot{g}^T G_{N,\lambda_1}^{-1} g\| = O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right).$$

$$(A4) \quad \text{All the entries of } \dot{G}_{N,\lambda_1}^{-1} \text{ and } \ddot{G}_{N,\lambda_1}^{-1} \text{ are } O_p(1).$$

$$(A5) \quad \text{All the entries of } \dot{g}^T \dot{G}_{N,\lambda_1}^{-1} \text{ and } \ddot{g}^T G_{N,\lambda_1}^{-1} \text{ are } O_p(1).$$

$$(A6) \quad RQ_{N,\lambda_1} \text{ is thrice differentiable and all the third derivatives of } RQ_{N,\lambda_1} \text{ are } O_p(N).$$

$$(A7) \quad \text{All the entries of } \dot{g}^T G_{N,\lambda_1}^{-1} \dot{g} \text{ and } \dot{g}^T G_{N,\lambda_1}^{-1} \text{ are } O_p(1).$$

Note that in the regularity conditions above, G_{N,λ_1} can be relaxed to G_N , indicated by the property of λ_1 from Lemma 2.4.1.

To show the consistency of $\hat{\beta}$, first we show the consistency of the 1st and 2nd derivatives of RQ_{N,λ_1} in the lemma below:

Lemma 2.4.2.

$$\frac{1}{N} \|\dot{R}Q_{N,\lambda_1} - 2N\dot{g}^T G_{N,\lambda_1}^{-1} g\| = o_p(1) \quad (2.4.2)$$

$$\frac{1}{N} \|\ddot{R}Q_{N,\lambda_1} - 2N\ddot{g}^T G_{N,\lambda_1}^{-1} \dot{g}\| = o_p(1) \quad (2.4.3)$$

It follows from this lemma that

Theorem 2.4.1. ($\sqrt{N/q}$ consistency) *Suppose that all the regularity conditions above hold, then $\hat{\beta} = \arg \min_{\beta} RQ_{N,\lambda_1}(\beta)$ satisfies*

$$\|\hat{\beta} - \beta_0\| = O_p(\sqrt{q/N}). \quad (2.4.4)$$

When $\hat{\beta} = \arg \min_{\beta} PRQ_{N,\lambda_1,\lambda_2}(\beta) = \arg \min_{\beta} \{RQ_{N,\lambda_1}(\beta) + K_n P_{\Lambda_2}(|\beta|)\}$, we need further technical assumptions to show the consistency.

(A8) for $\beta \in \mathcal{B}$, $P_{\lambda_2}(|\beta|) \geq 0$ and are twice differentiable.

$$(A9) \sup_{j \in \Lambda} \dot{p}_{\lambda_2}(|\beta_{0j}|) = O\left(q^{-\frac{1}{2}} N^{\frac{1}{2}} K_N^{-1}\right).$$

$$(A10) \sup_{j \in \Lambda} \ddot{p}_{\lambda_2}(|\beta_{0j}|) = O\left(q^{-\frac{1}{2}} N^{-1} K_N\right).$$

With the penalty p_{λ_2} , Theorem 2.4.1 is then extended to

Theorem 2.4.2. ($\sqrt{N/q}$ consistency) *Suppose that all the regularity conditions above hold, then there exists $\hat{\beta} = \arg \min_{\beta} PRQ_{N,\lambda_1,\lambda_2}(\beta)$, such that*

$$\|\hat{\beta} - \beta_0\| = O_p(\sqrt{q/N}). \quad (2.4.5)$$

Detailed proofs of Theorem 2.4.1 and 2.4.2 are given in the Appendix 1.

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

From (A1), the converge rate in (2.4.5) is known as $o_p(1)$. In this chapter, the consistency when $q \rightarrow \infty$ is not expected, as implied by Warton (2008). In particular, when qr is proportional to N , or q is proportional to N by fixing r , say, $q = O(N)$, the estimator of covariance matrix is less accurate, being far away from identity I_{rq} , because as z tends to 1, the ratio of the largest and smallest eigenvalues of G_N becomes very large. In practice, our new algorithm is still available for dealing with high-dimensional data, though it may be computationally intensive.

From Theorem 2.4.2, the two lemmas can be derived below.

Lemma 2.4.3. (*Sensitivity*) *If $\hat{\beta}$ is defined as in Theorem 2.4.2, then*

$$P(\exists j \in \Lambda, \hat{\beta}_j = 0) \rightarrow 0 \quad (2.4.6)$$

Lemma 2.4.4. (*Sparsity*) *If $\hat{\beta}$ is defined as in Theorem 2.4.2 and*

(A11) *as $\lambda_2 \rightarrow 0$ and $\sqrt{Nq}K_N^{-1}\lambda_2^{-1} \rightarrow 0$,*

then

$$P(\hat{\beta}_{\Omega-\Lambda} = \beta_{0(\Omega-\Lambda)} = 0_{q-p}) \rightarrow 1. \quad (2.4.7)$$

These two lemmas together with Theorem 2.4.3 below guarantee the oracle property of $\hat{\beta}$, indicating that the correct subset model is identified.

Under Lemma 2.4.3–2.4.4, we build up the asymptotic normality for the proposed penalized RQ_{N,λ_1} estimators. The following three more technical assumptions are required:

(A12) For penalty function $p_{\lambda_2,j} = p_{\lambda_2}(\beta_j)$, there exist $c_1, c_2 > 0$, such that

$$\|\ddot{p}_{\lambda_2,j}(t_1) - \ddot{p}_{\lambda_2,j}(t_2)\| \leq q^{-1}c_2\|t_1 - t_2\|, \text{ where } t_1, t_2 > c_1\lambda_2.$$

(A13) $E\|S_{N,\Lambda}^{-\frac{1}{2}}\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1}\|^2 = o_p(q^2)$, where $\dot{g}_\Lambda = \frac{\partial g(\beta_0)}{\partial \beta_\Lambda}$ and $H_{N,\Lambda} = \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} \dot{g}_\Lambda$. $S_{N,\Lambda}$ is defined as

$$S_{N,\Lambda} = \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} G G_{N,\lambda_1}^{-1} \dot{g}_\Lambda, \quad (2.4.8)$$

where G can be replaced by G_N which will remain the asymptotic normality of the corresponding estimator $\hat{\beta}$.

(A14) for every $i \in \Omega$, $Cov(g_i(\beta_0))$ exists and the entries of $Cov(g_i(\beta_0))$ are all $O_p(1)$.

Theorem 2.4.3. *Suppose conditions (A1)–(A14) hold and let*

$b_N = \frac{K_N}{2N} \left\{ H_{N,\Lambda} + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta_{0\Lambda}) \right\}^{-1} \dot{P}_{\lambda_2,\Lambda}(\beta_{0\Lambda})$. Then as $N \rightarrow \infty$, we have

$$S_{N,\Lambda}^{-\frac{1}{2}} \left\{ H_{N,\Lambda} + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta_{0\Lambda}) \right\} (\hat{\beta}_\Lambda - \beta_{0\Lambda} + b_N) \xrightarrow{d} \mathcal{N}(0, I_p). \quad (2.4.9)$$

The proof of Theorem 2.4.3 is provided in the Appendix 1.

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

2.5 Simulation studies

In this section, we conduct two simulation examples to assess the performance of the proposed method.

Example 1. In the first example, without loss of generality, we simply simulated longitudinal data from a normal distribution in our experiment. The true underlying model is assumed to be

$$Y_i = \mathbf{X}_i\beta + \epsilon_i, \quad (2.5.1)$$

where T -vector Y_i denotes the responses in each subject i and $\mathbf{X}_i$ is $T \times q$ covariates matrix whose first column elements are all equal to 1 as the intercept and the rest column elements independently drawn from a multivariate normal distribution with mean $(0.1, 0.2, \dots, 0.1 * T)$ and identity covariance matrix. The correlation within each subject is determined by the noise ϵ_i in (2.5.1), which comes from a mean zero normal distribution with a correlated variance structure.

Specifically throughout this example, we considered the number of non-zero coefficients in the true model, $p = 3$, and the true value of coefficients in the underlying model $\beta = (3, 2, 2, 0, \dots, 0)^T$ where the 1st entry denotes the intercept. For comparison, the subject number N and subject size T were different values while the number of coefficients, q was set to be relatively large. The $AR(1)$ structure was selected in the initial data generating step. Note that the inverse of $AR(1)$ correlation matrix can be decomposed into three matrices (Qu et al., 2000 and Sutradhar & Kumar, 2003), therefore the number r of M_i used in QIF was set to 3, such that $rq > N$ held for singularity condition. M_i 's had the same

structures as the decomposition matrices mentioned above. Alternatively other forms of the working matrix like the *exchangeable* can be easily applied after the same setting steps as well. We set the correlation parameter $\rho = 0.7$ represent a strong within-cluster association in the simulated data.

According to the iterative algorithm in Section 2.3, the initial values β_0 plays a significant role in the estimation procedure. To make our simulation more efficient, we used the screening method proposed by Fan & Lv (2008) to obtain the starting values. Briefly speaking, the idea of this method is to select the most related covariates corresponding to the response in terms of deviance or the value of the coefficient estimator, incorporated with the SCAD penalization. Details and further discussions on this topic can be referred to Fan & Lv (2008).

To clearly demonstrate the effects of the model selection proposed, we defined the following terms,

- (CS) average number of truly non-zero coefficients estimated as non-zero,
- (IS) average number of truly non-zero coefficients estimated as zero,
- (CD) average number of truly zero coefficients estimated as zero,
- (ID) average number of truly zero coefficients estimated as non-zero,
- (EXACT) proportion of exactly selecting the true model with non-zero coefficients.

After 1000 repeated runs, results are reported in the Tables 2.1—2.4.

Table 2.1 summarizes results of both quadratic SCAD and SCAD for simulated $AR(1)$ normal data when the correlation structure within each subject is correctly specified. In addition, we considered the misspecified model with *exchangeable*

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

Table 2.1: Comparison of quadratic SCAD and SCAD with simulated $AR(1)$ normal data when the model correlation is correctly specified, based on 1000 replications.

Quadratic SCAD					
N	T	q	CS	IS	EXACT
80	5	50	3	0.122	0.896
100	5	60	3	0.126	0.891
SCAD					
N	T	q	CS	IS	EXACT
80	5	50	3	37.367	0.159
100	5	60	3	49.511	0.079

Table 2.2: Comparison of quadratic SCAD and SCAD with simulated *Exchangeable* normal data when the model correlation is misspecified as $AR(1)$, based on 1000 replications.

Quadratic SCAD					
N	T	q	CS	IS	EXACT
80	5	50	3	0.110	0.902
100	5	60	3	0.123	0.893
SCAD					
N	T	q	CS	IS	EXACT
80	5	50	3	37.788	0.139
100	5	60	3	48.939	0.084

correlation and concluded results in Table 2.2. It is evidently shown that the SCAD performs much worse than the quadratic SCAD in terms of **IS** and **EXACT**. The SCAD fails to refine the underlying true model in all cases listed in both tables. On the contrary, the quadratic SCAD which strongly performs when number of covariates is sufficiently large has high rate in **EXACT**.

In Table 2.3 we summarize results of the proposed RQIF with quadratic penalty and its comparison with the standard GEE with SCAD penalty in the high-dimension situation.

Here the tuning parameter λ_2 in both penalties in RQIF and GEE were computed respectively and independently. As expected, without any penalty, the RQIF could not deal with variable selection, regardless of the sample size and the relative dimension of variables. However, as we add quadratic SCAD penalty into RQIF, it correctly excludes the zero coefficients from the suggested model with a quite high rate of exact selection. In the cases of small sample size under singularity, when $N \leq 100$, the RQIF also exhibits better performance than the GEE in terms of **IS** and **CD**. Using the proposed method, the rate of exact selection becomes very high when the sample size is sufficiently large. Note that in the simulation, if q is too large, the **EXACT** shows an abnormally low level. For example, when $N = 100$, $T = 5$ and $q = 80$, **EXACT** is 0.654 for RQIF. These unusual results may be caused by the insufficient observations. In general, our proposed RQIF performs competently in the high-dimensional variable selection, although it does not outperforms the penalized GEE in terms of **EXACT** because of regularization for the singular structure in standard QIF triggered by relatively large number of covariates. But its performance on **IS** and **CD** is comparable to the penalized GEE. Regarding our primary aim to perform model

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

Table 2.3: The performance of the proposed RQIF with/without penalty, and the penalized GEE. For normal distribution with $AR(1)$ correlation structure, $\rho = 0.7$. QIF.r&p and QIF.r represent regularized QIF with and without proposed quadratic SCAD penalty, respectively, while GEE.p stands for penalized GEE.

	N	T	q	CS	IS	CD	ID	EXACT
QIF.r&p	40	5	30	3	0.062	26.938	0	0.942
	40	10	30	3	0.318	26.682	0	0.766
	80	5	50	3	0.122	46.878	0	0.896
	80	10	50	3	0.286	46.714	0	0.820
	100	5	60	3	0.126	56.874	0	0.891
	100	10	60	3	0.305	56.695	0	0.705
	200	5	110	3	0.159	106.841	0	0.877
	200	10	110	3	0.258	106.742	0	0.853
QIF.r	40	5	30	3	26.986	0.014	0	0
	40	10	30	3	26.974	0.026	0	0
	80	5	50	3	46.962	0.038	0	0
	80	10	50	3	46.903	0.097	0	0
	100	5	60	3	56.935	0.065	0	0
	100	10	60	3	56.873	0.127	0	0
	200	5	110	3	106.772	0.228	0	0
	200	10	110	3	106.591	0.409	0	0
GEE.p	40	5	30	3	0.496	26.504	0	0.893
	40	10	30	3	0.609	26.391	0	0.923
	80	5	50	3	0.318	46.682	0	0.915
	80	10	50	3	0.093	46.907	0	0.975
	100	5	60	3	0.312	56.688	0	0.926
	100	10	60	3	0.132	56.868	0	0.979
	200	5	110	3	0.316	106.684	0	0.905
	200	10	110	3	0.041	106.959	0	0.985

2.5 Simulation studies

Table 2.4: The result analysis for the normal responses after 1000 simulations, using regularized QIF with quadratic SCAD penalty.

N=400, T=5, q=50			
	$\beta_1=3$	$\beta_2=2$	$\beta_3=2$
Mean	2.998	1.999	1.998
Std.err	0.048	0.017	0.017
Sandwich	0.032	0.012	0.012
CP	0.808	0.853	0.843
N=800, T=5, q=50			
	$\beta_1=3$	$\beta_2=2$	$\beta_3=2$
Mean	2.999	1.999	2.000
Std.err	0.031	0.011	0.011
Sandwich	0.025	0.010	0.010
CP	0.902	0.903	0.907
N=1200, T=5, q=50			
	$\beta_1=3$	$\beta_2=2$	$\beta_3=2$
Mean	2.999	1.999	2.000
Std.err	0.024	0.009	0.009
Sandwich	0.021	0.008	0.008
CP	0.916	0.916	0.931

Mean denotes the average estimate value of coefficients; Std.err denotes the sample standard error; Sandwich denotes the standard error estimated by the sandwich formula; CP stands for the coverage probability of 95% confidence interval.

selection and estimation simultaneously and the fact that QIF provides more efficient estimator in regression parameters than GEE, it is worth proposing new regularized QIF to adapt the high-dimensional context.

To further investigate the performance of RQIF for the data with large sample size, we considered $N = 400, 800, 1200$, $T = 5$ and $q = 50$. The results are concluded in Table 2.4. We observe that the coverage property **CP** grows as sample size increases. When N is sufficiently large, the sandwich estimate is close to the sample standard deviation and **CP** is close to the nominal 95% in-

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

dicating the accuracy of the proposed sandwich estimate in Section 2.4. Besides, the coefficient estimation performs very well with small bias.

Example 2. Another simulation study was conducted on correlated Poisson count data. Differing from the continuous case, in this example, we used the *exchangeable* within-cluster correlation. Elements in each column of $T \times q$ covariates matrix X_i has same value generated from a standard uniform distribution. We adopted the binomial thinning approach developed by Steutel et al. (1983) and McKenzie (1988) to simulate correlated Poisson data. Given the structure of x_{ij} and correlation matrix, we generated the count response y_{ij} by satisfying the relation below:

$$y_{ij} = \sqrt{\rho} * y_{i0} + e_{ij}, \quad (2.5.2)$$

where $\rho \in [0, 1]$ and conditional on y_{i0} , $\sqrt{\rho} * y_{i0}$ is a random variable with $Binomial(y_{i0}, \sqrt{\rho})$. From the mechanism of generating X_i , we obtained that all elements of μ_i , μ_{ij} 's are same, that is $\mu_{ij} \equiv \mu_{i1}$ for all $j \geq 2$. So the starting value y_{i0} was given from the Poisson distribution with mean μ_{i1} defined as

$$\mu_{i1} = \exp(x_{i1}^T \beta), \quad (2.5.3)$$

and so e_{ij} was drawn from Poisson distribution with mean $(1 - \sqrt{\rho})\mu_{i1}$. Since the marginal distribution of $\sqrt{\rho} * y_{i0}$ is Poisson with mean $\sqrt{\rho}\mu_{i1}$, it follows that y_{ij} has the same distribution as y_{i0} and $\text{corr}(y_{ij}, y_{i,j'}) = \rho$ where $j \neq j'$.

The same screening method as in example 1 was still applied to select the initial values of β in the iteration. We considered the true value $\beta = (1, 1, 1, 1, 1)$ with $p = 5$. Large number of subject $N = 400, 800$ is preferred in this example.

We varied the subject size T from 5 to 10 to investigate the estimation results. The number of coefficients $q = 50$ and the correlation parameter ρ was set to 0.7.

Variable selection results based on 1000 replications are summarized in Table 2.5. As we expected, the RQIF has no problem in dealing with the Poisson distributed data with relatively high-dimensional variables. As sample size becomes sufficiently large, all the terms related to variable selection process, such as **IS** and **EXACT** rate, show greatly strong performance.

Table 2.5: Variable selection for exchangeable Poisson data using the regularized QIF with penalty.

N	T	q	CS	IS	CD	ID	EXACT
400	5	50	5	0	45	0	1
400	10	50	5	0	45	0	1
800	5	50	5	0	45	0	1
800	10	50	5	0	45	0	1

2.6 Chapter summary

For the high-dimensional longitudinal data, the proposed regularized QIF, RQ_{N,λ_1} can solve the problem of invalidation when the weighting structure in the QIF turns to be singular. By the asymptotic built in Section 2.4, the RQ_{N,λ_1} inherits the statistical properties of the QIF fairly well. From our simulation studies, we found that on the premise of consistency, the performance of the resulting estimator is mainly influenced by penalty function rather than the strength of tuning parameter. This may be caused by the strictly quadratic structure of the original QIF. Another influential factor is the rate of q tending to infinity. We observe that when q is so large that rq exceeds N , the asymptotics may not hold

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

any more. Therefore, the consistency and normality properties are limited in the case of diverging number of covariates. Actually, when $rq > N$, it has been proved that the covariance estimator in QIF would be inaccurate and differs from the true covariance matrix. That is why the initial underlying model should be penalized first.

2.7 Appendix 1

Proof of Lemma 2.4.1.

Without loss of generality, let $w = 1$. Then

$$\hat{\lambda}_1 = \arg \min_{\lambda_1} RQ(\hat{\beta}, D, \lambda_1). \quad (2.7.1)$$

We know that G_N is a symmetric matrix so it can be decomposed into PVP^{-1} , where V is a diagonal matrix which consists of the eigenvalues of G_N and P the corresponding unitary matrix whose columns are the eigenvectors of G_N , thus

$$\begin{aligned} RQ_{\lambda_1} &= Ng^T G_{N,\lambda_1}^{-1} g \\ &= Ng^T \left(\lambda_1 G_N + \frac{1}{N^2} (1 - \lambda_1) I \right)^{-1} g \\ &= Ng^T \left(\lambda_1 PVP^{-1} + \frac{1}{N^2} (1 - \lambda_1) I \right)^{-1} g \\ &= Ng^T P \left(\lambda_1 V + \frac{1}{N^2} (1 - \lambda_1) I \right)^{-1} P^{-1} g \\ &= N \sum_{i=1}^{rq} z_i^2 \left(\lambda_1 v_{ii} + \frac{1}{N^2} (1 - \lambda_1) \right)^{-1}, \end{aligned}$$

where z_i is the entry of the vector $Z = P^{-1}g$ and v_{ii} the diagonal element of V . Due to (A2), $v_{ii} \rightarrow \infty$ as $N \rightarrow \infty$. Therefore, when $\hat{\lambda}_1$ is very closed to 1, RQ_{λ_1} attains the minimum value. $\square$

Proof of Lemma 2.4.2.

Since

$$\dot{R}Q_{N,\lambda_1} = 2N\dot{g}^T G_{N,\lambda_1}^{-1}g + Ng^T \dot{G}_{N,\lambda_1}^{-1}g,$$

where according to (A1), (A3) and (A4), $\|g^T \dot{G}_{N,\lambda_1}^{-1}g\| = O_p\left(q^{\frac{5}{2}}N^{-1}\right) = o_p(1)$, the first part now is proved.

For the second part,

$$\ddot{R}Q_{N,\lambda_1} = 2N\ddot{g}^T G_{N,\lambda_1}^{-1}g + 4N\dot{g}^T \dot{G}_{N,\lambda_1}^{-1}g + 2N\dot{g}^T G_{N,\lambda_1}^{-1}\dot{g} + Ng^T \ddot{G}_{N,\lambda_1}^{-1}g.$$

Conditions (A1), (A3), (A4) and (A5) indicate that

$$\|\ddot{g}^T G_{N,\lambda_1}^{-1}g\| \leq \|\ddot{g}^T G_{N,\lambda_1}^{-1}\| \|g\| = O_p\left(q^2 N^{-\frac{1}{2}}\right) = o_p(1),$$

$$\|\dot{g}^T \dot{G}_{N,\lambda_1}^{-1}g\| = O_p\left(q^2 N^{-\frac{1}{2}}\right) = o_p(1),$$

$$\|g^T \ddot{G}_{N,\lambda_1}^{-1}g\| = O_p(q^3 N^{-1}) = o_p(1).$$

$\square$

Proof of Theorem 2.4.1.

It is sufficient to prove that for any $\epsilon > 0$, there exists some large constant $\mathcal{C}_\epsilon$ such that a local minimizer $\hat{\beta}$ lies in the interior of the ball $\left\{\beta_0 + q^{\frac{1}{2}}N^{-\frac{1}{2}}u, \|u\| \leq \mathcal{C}_\epsilon\right\}$.

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

Then we define

$$H(u) = RQ_{N,\lambda_1}(\beta_0 + q^{\frac{1}{2}}N^{-\frac{1}{2}}u) - RQ_{N,\lambda_1}(\beta_0).$$

By Taylor expansion,

$$\begin{aligned} RQ_{N,\lambda_1}(\beta_0 + q^{\frac{1}{2}}N^{-\frac{1}{2}}u) &= RQ_{N,\lambda_1}(\beta_0) + q^{\frac{1}{2}}N^{-\frac{1}{2}}u^T \dot{R}Q_{N,\lambda_1}(\beta_0) \\ &+ \frac{1}{2}qN^{-1}u^T \ddot{R}Q_{N,\lambda_1}(\beta_0)u \\ &+ \frac{1}{6}q^{\frac{3}{2}}N^{-\frac{3}{2}} \left(\frac{\partial}{\partial \beta}(u^T \ddot{R}Q_{N,\lambda_1}(\beta^*)u) \right)^T u, \end{aligned}$$

where β^* is between β_0 and $\beta_0 + q^{\frac{1}{2}}N^{-\frac{1}{2}}u$. We denote the last three terms on the right side by (a), (b) and (c) in order, then by (A3) and Lemma 2.4.2,

$$\begin{aligned} \|(a)\| &= q^{\frac{1}{2}}N^{-\frac{1}{2}}u^T(2N\dot{g}^T G_{N,\lambda_1}^{-1}g + O_p(1)) \\ &= O_p(q)\|u\| + O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right)\|u\|, \\ \|(b)\| &= qu^T \dot{g}^T G_{N,\lambda_1}^{-1} \dot{g}u + o_p(q)\|u\|^2. \end{aligned}$$

By (A5) and (A6),

$$\begin{aligned} \|(c)\| &\leq \frac{1}{6}q^{\frac{3}{2}}N^{-\frac{3}{2}}\|u\| \left(\sum_{i=1}^q \sum_{j=1}^q \sum_{k=1}^q \frac{\partial^3 RQ_{N,\lambda_1}(\beta^*)}{\partial \beta_i \partial \beta_j \partial \beta_k} \|u\|^2 \right)^{\frac{1}{2}} \\ &= q^{\frac{3}{2}}N^{-\frac{3}{2}}\|u\|^2 O_p\left(q^{\frac{3}{2}}N^{\frac{1}{2}}\right) \\ &= O_p(q^3 N^{-1})\|u\|^2 \\ &= o_p(q)\|u\|^2. \end{aligned}$$

As a result, we have

$$\begin{aligned} H(u) &= (a) + (b) + (c) \\ &= O_p(q)\|u\| + O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right)\|u\| + o_p(q)\|u\|^2 + qu^T(\dot{g}^T G_{N,\lambda_1}^{-1} \dot{g})u. \end{aligned}$$

By (A7), $H(u)$ is dominated by the positive term $qu^T(\dot{g}^T G_{N,\lambda_1}^{-1} \dot{g})u$, thus $H(u) > 0$ and there at least exists a local minimizer $\hat{\beta}$ which lies in the ball, as $N \rightarrow \infty$ and sufficiently large $\mathcal{C}_\epsilon$. $\square$

Proof of Theorem 2.4.2.

To prove Theorem 2.4.2, we just need to further discuss the penalty term $K_N P_{\lambda_2}(|\beta|)$, based on the proof of Theorem 2.4.1. Let

$$L(u) = P_{\lambda_2}(|\beta_0 + q^{\frac{1}{2}}N^{-\frac{1}{2}}u|) - P_{\lambda_2}(|\beta_0|).$$

From the conclusion of Theorem 2.4.1, it is sufficient to show that $L(u)$ is also dominated by the same term. By Taylor expansion, for each $j \in \Lambda$,

$$\begin{aligned} p_{\lambda_2}(|\beta_{0j} + q^{\frac{1}{2}}N^{-\frac{1}{2}}u_j|) &= p_{\lambda_2}(|\beta_{0j}|) + q^{\frac{1}{2}}N^{-\frac{1}{2}}\dot{p}_{\lambda_2}(|\beta_{0j}|) + \frac{1}{2}qN^{-1}\ddot{p}_{\lambda_2}(|\beta_{0j}^*|) \\ &= p_{\lambda_2}(|\beta_{0j}|) + q^{\frac{1}{2}}N^{-\frac{1}{2}}\dot{p}_{\lambda_2}(|\beta_{0j}|) + \frac{1}{2}qN^{-1}\ddot{p}_{\lambda_2}(|\beta_{0j}|)(1 + o(1)) \\ &= p_{\lambda_2}(|\beta_{0j}|) + q^{\frac{1}{2}}N^{-\frac{1}{2}}\dot{p}_{\lambda_2}(|\beta_{0j}|) + \frac{1}{2}qN^{-1}\ddot{p}_{\lambda_2}(|\beta_{0j}|) + o(qN^{-1}). \end{aligned}$$

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

Denote the 2nd and 3rd terms on the right side by d and e . By (A8)—(A10),

$$K_N(d) = O_p(q)\|u\|$$

$$K_N(e) = o_p(q)\|u\|^2.$$

Therefore, we have $H(u) + L(u) = (a) + (b) + (c) + K_N(d) + K_N(e)$ and definitely it is dominated by $qu^T(\dot{g}^T G_{N,\lambda_1}^{-1} \dot{g})u$. Finally, it is easy to see that there exists a local minimizer $\hat{\beta}$ in the interior of the ball. The proof is finished. $\square$

Proof of Lemma 2.4.3 .

Define $\delta = \liminf_{j \in \Lambda} |\beta_{0j}|$ and let $\epsilon = \frac{1}{2}\delta$, then

$$\begin{aligned} P(\exists j \in \Lambda, \hat{\beta}_j = 0) &\leq P(\exists j \in \Lambda, |\hat{\beta}_j - \beta_{0j}| > \epsilon) \\ &\leq P(\|\hat{\beta} - \beta_0\|^2 > \epsilon^2) \\ &= O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right). \end{aligned}$$

The last equation above is directly derived from Theorem 2.4.2. $\square$

Proof of Lemma 2.4.4.

From Theorem 2.4.2, it is already known that

$$\|\hat{\beta} - \beta_0\| = O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right)$$

and by Lemma 2.4.3,

$$\|\hat{\beta}_\Lambda - \beta_{0\Lambda}\| = O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right).$$

Then it is equivalent to show that $\exists k_0 > 0$ such that

$$PRQ_{N,\lambda_1,\lambda_2}(\hat{\beta}_\Lambda, 0) = \arg \min\{PRQ_{N,\lambda_1,\lambda_2}(\hat{\beta}_\Lambda, \beta_{\Omega-\Lambda}), \|\beta_j\| \leq k_0 q^{\frac{1}{2}}N^{-\frac{1}{2}}\}.$$

Next applying Taylor expansion, we have

$$\begin{aligned} \frac{\partial PRQ_{N,\lambda_1,\lambda_2}(\beta)}{\partial \beta_i} &= \frac{\partial RQ_{N,\lambda_1,\lambda_2}(\beta_0)}{\partial \beta_i} \\ &+ \sum_{j=1}^q \frac{\partial^2 RQ_{N,\lambda_1,\lambda_2}(\beta_0)}{\partial \beta_i \partial \beta_j} (\beta_j - \beta_{0j}) \\ &+ \frac{1}{2} \sum_{j=1}^q \sum_{k=1}^q \frac{\partial^3 RQ_{N,\lambda_1,\lambda_2}(\beta_0)}{\partial \beta_i \partial \beta_j \partial \beta_k} (\beta_j - \beta_{0j})(\beta_k - \beta_{0k}) \\ &+ K_N \dot{p}_{\lambda_2}(|\beta_i|). \end{aligned}$$

We assign the terms on the right side as (a), (b), (c) and (d), respectively. By (A3), (A7) and Lemma 2.4.2,

$$\|(a)\| = \|2N\dot{g}^T G_{N,\lambda_1}^{-1} g\| + O_p(1) = O_p(q^{\frac{1}{2}}N^{\frac{1}{2}}).$$

Also by Theorem 2.4.2, we obtain that

$$\|(b)\| = O_p(q^{\frac{1}{2}}N^{\frac{1}{2}}).$$

Similarly, due to (A6) and Theorem 2.4.2,

$$\|(c)\| = O(q^2)O_p(N)O_p(qN^{-1}) = O_p(q^3) = O_p(q^{\frac{1}{2}}N^{\frac{1}{2}}).$$

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

At last, by (A9) and (A11)

$$\begin{aligned}\frac{\partial PRQ_{N,\lambda_1,\lambda_2}(\beta)}{\partial \beta_i} &= O_p(q^{\frac{1}{2}}N^{\frac{1}{2}}) + (d) \\ &= K_N \lambda_2 \{ \lambda_2^{-1} \dot{p}_{\lambda_2}(\beta_i) \text{sgn}(\beta_i) + O_p(q^{\frac{1}{2}}N^{\frac{1}{2}}) K_N^{-1} \lambda_2^{-1} \},\end{aligned}$$

then it is easy to see that $\text{sgn}(\beta_i)$ determines the sign of $\frac{\partial PRQ_{N,\lambda_1,\lambda_2}(\beta)}{\partial \beta_i}$. The proof is finished. $\square$

Proof of Theorem 2.4.3.

By the preceding Lemma 2.4.2, it is already known that

$$\frac{1}{N} \dot{R}Q_{N,\lambda_1}(\beta_0) = 2\dot{g}(\beta_0)^T G_{N,\lambda_1}^{-1}(\beta_0) g(\beta_0) + o_p(1),$$

so for $i \in \Lambda$, we have

$$\frac{1}{N} \dot{R}Q_{N,\lambda_1}(\beta_{0\Lambda}) = 2\dot{g}_\Lambda(\beta_0)^T G_{N,\lambda_1}^{-1}(\beta_0) g(\beta_0) + o_p(1).$$

Similarly,

$$\frac{1}{N} \ddot{R}Q_{N,\lambda_1}(\beta_{0\Lambda}) = 2H_{N,\Lambda}(\beta_{0\Lambda}) + o_p(1).$$

Since $\hat{\beta}_\Lambda$ is the solution of $\dot{R}Q_{N,\lambda_1}(\beta_\Lambda) + K_N \dot{P}_{\lambda_2,\Lambda}(\beta_\Lambda)$, by Taylor expansion,

$$\begin{aligned}0 &= \frac{\partial RQ_{N,\lambda_1}(\beta_0)}{\partial \beta_\Lambda} + \frac{\partial^2 RQ_{N,\lambda_1}(\beta_0)}{\partial \beta_\Lambda \partial \beta_\Lambda^T} (\hat{\beta}_\Lambda - \beta_{0\Lambda}) \\ &+ (\hat{\beta}_\Lambda - \beta_{0\Lambda})^T \ddot{R}Q_{N,\lambda_1}(\beta^*) (\hat{\beta}_\Lambda - \beta_{0\Lambda}) + K_N \dot{P}_{\lambda_2,\Lambda}(\beta_0) \\ &+ K_N \ddot{P}_{\lambda_2,\Lambda}(\beta^{**}) (\hat{\beta}_\Lambda - \beta_{0\Lambda})\end{aligned}$$

where β^* and β^{**} are between $\hat{\beta}_\Lambda$ and $\beta_{0\Lambda}$.

First, denote the third term on the right side of the equation above by A . According to (A6),

$$\|(A)\| = O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right) O_p\left(q^{\frac{3}{2}}N\right) O_p\left(q^{\frac{1}{2}}N^{-\frac{1}{2}}\right) = o_p\left(N^{-\frac{1}{2}}\right).$$

Then

$$\begin{aligned} 0 &= \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} g(\beta_0) + H_{N,\Lambda}(\beta_0)(\hat{\beta}_\Lambda - \beta_{0\Lambda}) \\ &+ \frac{K_N}{2N} \dot{P}_{\lambda_2,\Lambda}(\beta_0) + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta^{**})(\hat{\beta}_\Lambda - \beta_{0\Lambda}) \\ &+ o_p\left(N^{-\frac{1}{2}}\right), \end{aligned}$$

that is,

$$\begin{aligned} -\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} g(\beta_0) &= \left(H_{N,\Lambda}(\beta_0) + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta^{**}) \right) (\hat{\beta}_\Lambda - \beta_{0\Lambda}) \\ &+ \frac{K_N}{2N} \dot{P}_{\lambda_2,\Lambda}(\beta_0) + o_p\left(N^{-\frac{1}{2}}\right). \end{aligned}$$

From (A12), we obtain that

$$\begin{aligned} -\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} g(\beta_0) &= \left(H_{N,\Lambda}(\beta_0) + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta_0) \right) (\hat{\beta}_\Lambda - \beta_{0\Lambda}) \\ &+ \frac{K_N}{2N} \dot{P}_{\lambda_2,\Lambda}(\beta_0) + o_p\left(N^{-\frac{1}{2}}\right) \\ &= \left(H_{N,\Lambda}(\beta_0) + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta_0) \right) (\hat{\beta}_\Lambda - \beta_{0\Lambda} + b_N) + o_p\left(N^{-\frac{1}{2}}\right) \end{aligned}$$

So

$$\begin{aligned} -S_{N,\Lambda}^{-\frac{1}{2}} \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} g(\beta_0) &= -S_{N,\Lambda}^{-\frac{1}{2}} \left(H_{N,\Lambda}(\beta_0) + \frac{K_N}{2N} \ddot{P}_{\lambda_2,\Lambda}(\beta_0) \right) (\hat{\beta}_\Lambda - \beta_{0\Lambda} + b_N) \\ &+ o_p\left(N^{-\frac{1}{2}} S_{N,\Lambda}^{-\frac{1}{2}}\right). \end{aligned}$$

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

By (A7), $o_p\left(N^{-\frac{1}{2}}S_{N,\Lambda}^{-\frac{1}{2}}\right) = o_p(1)$. Next, since

$$S_{N,\Lambda}^{-\frac{1}{2}}\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1}g(\beta_0) = \frac{1}{N} \sum_{i=1}^N S_{N,\Lambda}^{-\frac{1}{2}}\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1}g_i(\beta_0),$$

let $D_i = \frac{1}{N}S_{N,\Lambda}^{-\frac{1}{2}}\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1}g_i(\beta_0)$, we only need to prove that

$$\sum_{i=1}^N D_i \xrightarrow{d} \mathcal{N}(0, I_p)$$

Note that for $\forall \epsilon > 0$,

$$\sum_{i=1}^N E\{\|D_i\|^2 I(\|D_i\| > \epsilon)\} \leq \sum_{i=1}^N \sqrt{E\|D_i\|^4 P(\|D_i\| > \epsilon)}$$

and by Chebyshev's inequality,

$$P(\|D_i\| > \epsilon) \leq \epsilon^{-2} E\|D_i\|^2.$$

Besides, by (A 13) and (A14)

$$\begin{aligned} E\|D_i\|^2 &= E\left\|\frac{1}{N}S_{N,\Lambda}^{-\frac{1}{2}}\dot{g}_\Lambda^T G_{N,\lambda_1}^{-1}g_i(\beta_0)\right\|^2 \\ &\leq \frac{1}{N^2}O_p(q^2)O_p(q) \\ &= O_p\left(\frac{q^3}{N^2}\right). \end{aligned}$$

In the same way, we have

$$E\|D_i\|^4 = O_p\left(\frac{q^6}{N^4}\right).$$

Therefore,

$$\begin{aligned} \sum_{i=1}^N E\{\|D_i\|^2 I(\|D_i\| > \epsilon)\} &\leq \sum_{i=1}^N \sqrt{O_p\left(\frac{q^3}{N^2}\right) O_p\left(\frac{q^6}{N^4}\right)} \\ &= o_p(1). \end{aligned}$$

By Linderberg-Feller theorem, $\sum_{i=1}^N D_i$ is asymptotically normal with mean zero and covariance

$$\begin{aligned} Cov\left(\sum_{i=1}^N D_i\right) &= \sum_{i=1}^N Cov(D_i) \\ &= \frac{1}{N^2} \sum_{i=1}^N E\left(S_{N,\Lambda}^{-\frac{1}{2}} \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} g_i(\beta_0) g_i^T(\beta_0) G_{N,\lambda_1}^{-1} \dot{g}_\Lambda S_{N,\Lambda}^{-\frac{1}{2}}\right) \\ &= E\left(S_{N,\Lambda}^{-\frac{1}{2}} \dot{g}_\Lambda^T G_{N,\lambda_1}^{-1} G G_{N,\lambda_1}^{-1} \dot{g}_\Lambda S_{N,\Lambda}^{-\frac{1}{2}}\right) \\ &= I_p. \end{aligned}$$

The proof of Theorem 2.4.3 is completed. □

2. REGULARIZED QUADRATIC INFERENCE FUNCTION WITH QUADRATIC PENALTY

3 Covariate measurement error in GPLMM

3.1 Introduction

Longitudinal data arising in medical, health and epidemiological studies are commonly hampered in the presence of measurement errors, where covariates are not measured precisely due to the careless mistakes made by the data collectors, misusing the measurement equipment or technical difficulties. Ignoring measurement errors among covariates often results in biased estimates, and hence yielding invalid inference. Therefore it is important to take measurement errors into consideration when conducting inference based on such data.

Liang & Ren (2005) and later Liang (2009), discussed the GPLM and GPLMM when data were clustered and partially measured with error. However, they treated the error-prone covariates either based on the individual measurement error model using the validation set information only or assuming that the exact error structure was already known. In this chapter, denoting the response matrix by $\mathbf{Y}$, the true covariates matrix by $\mathbf{X}$ and observed covariates matrix by $\mathbf{W}$, we present a more general approach accounting for both $\mathbf{Y}-\mathbf{X}$ and $\mathbf{W}-\mathbf{X}$ structures and thus utilizing information of entire data set to improve the error correction procedure and thus achieve greater efficiency. We incorporate the

3. COVARIATE MEASUREMENT ERROR IN GPLMM

multiple imputation (MI) with the GPLMM to impute the pseudo completely correct data sets for the model fitting procedure. As proposed by Rubin (1976), as a useful MCMC method, the MI was developed to fill in the missing data in the sample and make inference on the multiple imputation sets using the standard statistical model. It is appropriate for measurement error problems where a validation set exists. Yucel & Zaslavsky (2005) adopted the imputation method to analyze the administrative data with mismeasured binary treatment outcomes. Cole et al. (2006) designed a hypothetical study to demonstrate the performance of the MI and compared it with the regression calibration (RC). Guo & Little (2011) discussed the heteroscedastic measurement error in covariates in the normal linear model. Their simulation showed that the MI method led to convinced results compared with other calibration methods.

Because of the complexity of the GPLMM the posterior distribution of $\mathbf{X}$ given other available information has no closed form. So, the conditional expectation based on this posterior is hard to calculate. We adopt a multiple imputation procedure which provides the exactly approximated values of mismeasured covariates and can be implemented by some MCMC algorithm like the random walk when the target distribution is implicitly given. As pointed out by Wang & Robins (1998) and Robins & Wang (2000), the traditional MI method usually underestimates variances and fails to construct their appropriate confidence intervals. Putting the imputed complete data sets into the GPLMM, we then develop a new approach to consistently estimate the regression coefficients under the conditional independence structure.

The rest sections of this chapter are organized as follows. In Section 3.2,

we introduce the standard imputation procedure and the PQL principle for the GPLMM. The model specification is also given in details in the same section. A two-step semiparametric method MIPQL (Multi-Imputed Penalized Quasi Likelihood) is developed in Section 3.3. Asymptotic properties of proposed method are discussed in Section 3.4. Results of simulation studies and an application to real data analysis are provided in Section 3.5—3.6. The discussions on the future research are concluded at the end of this chapter. All the proofs are elaborated in the Appendix 2.

3.2 Model specification

Consider a longitudinal data with W_{ije} being the single covariate subject to error distributed as $\mathcal{N}(\mu_w, \sigma_w^2)$, the corresponding true covariate X_{ije} following $\mathcal{N}(\mu_x, \sigma_x^2)$ and the random error U_{ije} , independent of X_{ije} , being assumed to be normal with mean 0 and variance σ_u^2 , where $i = 1, \dots, N$, $j = 1, \dots, T$ and $1 \leq e \leq p$ indicates the mismeasured element in p -dimensional vector of covariates X_{ij} . From commonly used measurement error model $W_{ije} = X_{ije} + U_{ije}$, it is easy to have the following relationships: $\mu_w = \mu_x$, and $\sigma_w^2 = \sigma_x^2 + \sigma_u^2$.

Normally, we can have many covariates with mismeasured observations, and they can be interpreted by several measurement error models as above. For simplicity, in the rest of this section, we consider the measurement error model that involving single covariate only.

More generally we assume that

$$W_{ije} = X_{ije}\eta + U_{ije}, \tag{3.2.1}$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

where parameter η adjusts the underlying relationship between W_{ije} and X_{ije} . The use of η takes into account both cases that the variation of W_{ije} is greater and less than that of X_{ije} .

The main analysis model in this chapter is to fit the conditional expectation of longitudinal response Y , on partially linear predictor with mixed effects, $X^T\beta + B^Tb + \theta(Z)$ through a link function, where X and Z are vectors of covariates for linear and nonlinear terms, while B represents vector of covariates related to the random effects b , and $\theta(\cdot)$ denotes unknown function. The detailed estimation method and model specification for such a complicated model are presented in the next section. In the presence of measurement error, the complete data set actually consists of (Y, X, W, B, Z) , where a subset of covariates in X is unfortunately measured with error and observed as W instead. Furthermore, W are supposed to have no relation to the outcomes Y at all. That is, we assume the error term U should determine W only.

To implement multiple imputation method to generate the "correct" complete data set, first of all we need to extract the subset from the original data set (Y, X, W, B, Z) . Define it as the validation set in our study and denote it by $(Y^v, X^v, W^v, B^v, Z^v)$. Other observed data in the remaining data set are viewed as $(Y^e, X^e, W^e, B^e, Z^e)$. Note that X^e is unobserved and in fact does not exist. Under the longitudinal setting, we let N_v and N denote the number of subjects in the validation and whole set, respectively. T denotes the total number of repeated times in the experiment.

The key of imputation is to draw unknown X from its posterior distribution given (Y, W, B, Z) by in advance generating the model parameter β , random effects b and nonlinear component $\theta(Z)$ given the artificial complete data. The

random effects b is set to account for the underlying correlation within each subject and is distributed with some known density function. Usually it is assumed to be normal. After some calculation, the posterior distribution of X given the remaining data, random effects and corresponding parameters is proportional to

$$\begin{aligned} p(X|Y, W, B, Z, b) &\propto p(X, Y, W, b|B, Z) \\ &= p(Y|X, B, b, Z)p(W|X)p(X). \end{aligned} \quad (3.2.2)$$

Once we have successfully generated M imputed complete values, $\{X_{(m)}^\epsilon, m = 1, \dots, M\}$, the mismeasured covariates X can be replaced and then the estimation procedure can be performed in the usual way. To simplify the notations, in the following we use X^ϵ to surrogate $X_{(m)}^\epsilon$. Now we have longitudinal data observations $\{Y_{ij}, X_{ij}^\epsilon, B_{ij}, Z_{ij}, j = 1, \dots, T\}$, for $i = 1, \dots, N$, where response Y_{ij} , covariates X_{ij}^ϵ and Z_{ij} , and B_{ij} are obtained from the i th subject at the j th repeated time. Consider the responses Y 's to be independent between subjects but apparently correlated during measurement period. Given a q -dimensional vector of random effects b_i , the conditional expectation of y_{ij} is defined as

$$\mu_{ij}^b = E(Y_{ij}|b_i) = h(X_{ij}^{\epsilon T}\beta + B_{ij}^T b_i + \theta(Z_{ij})), \quad (3.2.3)$$

where β is a p -dimensional vector of fixed effects and $h(\cdot)$ is the inverse link function. b_i are independently followed $\mathcal{N}(0, C)$ with positive definite covariance matrix C . Define μ^b as the conditional mean of Y , the conditional variance v^b is assumed to be the function of μ^b , $v(\mu^b)$. We then develop the estimation procedure with MI in the parametric and nonparametric steps.

3. COVARIATE MEASUREMENT ERROR IN GPLMM

In the next section, we take the partially mixed linear model for example to illustrate our proposed algorithm. That is the analysis model for each subject i is

$$Y_i = \mathbf{X}_i\beta + \mathbf{B}_ib_i + \theta(Z_i) + e_i, \quad (3.2.4)$$

where p -dimensional vector β denotes the regression parameters of interest, $Y_i = (Y_{i1}, \dots, Y_{iT})^T$, $T \times p$ matrix $\mathbf{X}_i = (X_{i1}, \dots, X_{iT})^T$ and $T \times q$ matrix $\mathbf{B}_i = (B_{i1}, \dots, B_{iT})^T$. Z_i denotes T -dimensional vector of covariate. T -dimensional error term $e_i \sim \mathcal{N}(0, \sigma^2 I)$ and all the elements in q -dimensional random effect vector b_i follows $\mathcal{N}(0, \sigma_b^2)$.

3.3 MIPQL estimation

We propose a multiple imputation estimation procedure to generate X^ϵ according to model (3.2.3). Without loss of generality, we describe our imputation algorithm steps using the simpler model (3.2.4). The steps are described as follows. In order to simplify notation, we replace X_{ije} by X_i in this section. So does W_{ije} .

3.3.1 Multiple imputation algorithm

Step 1. Perform initial least squares estimation of η and σ_u^2 , denoted resulting estimators by $\hat{\eta}$ and $\hat{\sigma}_u^2$, based on the measurement error model (3.2.1),

$$\hat{\eta} = \arg \min_{\alpha} \frac{1}{N_v T - 1} \sum_{k=1}^{N_v} \|W_k^v - X_k^v \eta\|^2$$

$$\hat{\sigma}_u^2 = \frac{1}{N_v T - 1} \sum_{k=1}^{N_v} \|W_k^v - X_k^v \hat{\eta}\|^2.$$

Note that as we mentioned in Section 3.2, here X_k^v and W_k^v represent the true and mismeasured observations of single covariate X for the subject k in the validation set, respectively. Let $X^v = (X_1^v, \dots, X_{N_v}^v)^T$, the initial imputed parameters $\hat{\eta}^*$ and $\hat{\sigma}_u^{*2}$ are then drawn from

$$\begin{aligned}\hat{\eta}^* &\sim \mathcal{N}(\hat{\eta}, [X^{vT} X^v]^{-1} \hat{\sigma}_u^{*2}) \\ \hat{\sigma}_u^{*2} &\sim (N_v T - 1) \hat{\sigma}_u^2 / \chi_{N_v T - 1}^2.\end{aligned}$$

Given the validation data, $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\sigma}_b^2$ can be obtained by the PQL estimation step derived in the next section. In the same way the initial imputed $\hat{\beta}^*$, $\hat{\sigma}^{*2}$ and $\hat{\sigma}_b^{*2}$ are drawn from

$$\begin{aligned}\hat{\beta}^* &\sim \mathcal{N}(\hat{\beta}, [X^{vT} X^v]^{-1} \hat{\sigma}^{*2}) \\ \hat{\sigma}^{*2} &\sim N_v T \hat{\sigma}^2 / \chi_{N_v T}^2 \\ \hat{\sigma}_b^{*2} &\sim N_v \hat{\sigma}_b^2 / \chi_{N_v}^2.\end{aligned}$$

Based on the marginal distribution of X^v , the general moment method gives the estimates $\hat{\mu}_x$ and $\hat{\sigma}_x$. The yielding imputed parameters $\hat{\mu}_x^*$ and $\hat{\sigma}_x^{*2}$ are generated from

$$\begin{aligned}\hat{\mu}_x^* &\sim \mathcal{N}(\hat{\mu}_x, \hat{\sigma}_x^{*2} / (N_v T - 1)) \\ \hat{\sigma}_x^{*2} &\sim (N_v T - 1) \hat{\sigma}_x^2 / \chi_{N_v T - 1}^2.\end{aligned}$$

Step 2. For each subject $i = N_v + 1, N_v + 2, \dots, N$, the imputed value X_i^* of X_i

3. COVARIATE MEASUREMENT ERROR IN GPLMM

is generated from

$$X_i^* \sim p(Y_i | X_i, B_i, \hat{b}_i, \hat{\theta}(Z_i))p(W_i | X_i)p(X_i). \quad (3.3.1)$$

where $\hat{b}_i$ and $\hat{\theta}(Z_i)$ are also estimated by the PQL estimation method. It is worth pointing out that because X^e is unknown, in this initial step we take the value of W^e instead to evaluate the posterior distribution. Hence the imputed complete values X^e are denoted by (X^v, X^*) .

Step 3. With X^e , repeat **(Step 1)**—**(Step 2)**, substituting X^e and N for X^v and N^v . Consequently, the imputed value X_i^* of X_i^e for $i = N_v + 1, N_v + 2, \dots, N$, is updated by

$$X_i^* \sim p(Y_i | X_i^e, B_i, \hat{b}_i, \hat{\theta}(Z_i))p(W_i | X_i^e)p(X_i^e), \quad (3.3.2)$$

and the imputed complete values X^e is expressed as (X^v, X^*) .

Step 4. We run **Step 3** until the result converges.

Remark 3.3.1. To monitor convergence of the model parameters, we can use the method proposed by Gelman & Rubin (1992) by estimating the factor by which the scale of the current distributions for $\hat{\eta}^*$ and $\hat{\sigma}^{*2}$ decline to 1 as $n \rightarrow \infty$. It is evident that the further iterations could improve the estimating performance about the target distribution if we observe high potential scale reduction. In practice, we observe $\hat{\eta}^*$ and $\hat{\sigma}^{*2}$ converge to a stable distribution after sufficiently long burn-in period. The M imputed complete data sets then are drawn by selecting every d th iteration to reduce the underlying correlation among the imputed values as much as possible. Computationally, we can use $d = 100$ or $d = 200$ to avoid such related imputed sets.

Remark 3.3.2. In **Step 2** and corresponding **Step 3**, it is required to generate X^* as the imputed values of X^e from its underlying posterior distribution. However, this procedure is not as straightforward as the model parameters imputation, when only implicit form of posterior density is known. The use of random-walk Metropolis algorithm makes it easy to impute X^e . The target distribution $\pi(X)$ is the posterior distribution of X and an appropriate proposal distribution F is given to generate X^* . Depending on the experiment conditions, the density F is usually randomly chosen by the users, although the optimal one could improve the algorithm efficiency significantly. In order to reduce the computation complexity, let F be the normal density with mean that the current values of X^* , and σ_p^2 scaled sampling variance of the target density. In the d th iteration, $X^{*(d)}$ is thereby drawn from $F(X^{*(d)}, X^{*(d-1)}) \sim \mathcal{N}(X^{*(d-1)}, \sigma_p^2)$. We define the acceptance probability as $\min \left[1, \frac{\pi(X^{*(d)})F(X^{*(d)}, X^{*(d-1)})}{\pi(X^{*(d-1)})F(X^{*(d-1)}, X^{*(d)})} \right]$, with which we accept $X^{*(d)}$ as the new imputed values, otherwise set $X^{*(d)} = X^{*(d-1)}$.

3.3.2 Local smoother and penalized quasi-likelihood

For more general model GPLMM (3.2.3), we develop estimation procedure as follows, which consists mainly of three parts: (1) estimation of nonlinear component $\theta(\cdot)$; (2) PQL estimation of β and b ; (3) estimation of the variance component.

Estimation of $\theta(z)$

By Taylor expansion, $\theta(z)$ is approximated in the small neighborhood of z_0 at its 1st order as

$$\theta(z) \approx \theta(z_0) + \theta^{(1)}(z_0)(z - z_0) + \dots + \frac{\theta^{(s)}}{s!}(z - z_0)^s, \quad (3.3.3)$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

where $\theta(z_0)$ and $\theta^{(1)}(z_0)$ can be substituted by the local smoothing parameters α_0 and α_1 . The choices for local smoother are suggested to be constant, linear or quadratic. With $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_s)^T$, the conditional mean structure (3.2.3) can be then estimated by

$$\hat{\mu}_{ij}^b = h(X_{ij}^{\epsilon T} \beta + B_{ij}^T b_i + \Theta_{ij}^T \alpha). \quad (3.3.4)$$

Θ_{ij} is a local smoothing vector $(1, (Z_{ij} - Z_0)/h, \dots, [(Z_{ij} - Z_0)/h]^s)^T$ where h is the bandwidth. The common choice of s is 0, 1 or 2, which results in average, linear or quadratic kernel estimator. And now v^b can be estimated by $\hat{v}^b = v(\hat{\mu}^b)$.

PQL estimation for β and b

Given imputed values X^ϵ , we fit the model (3.3.4) using *PQL* principle proposed by Breslow & Clayton (1993). Based on the quasi-likelihood, we can define two conditional scores and jointly solve them to obtain the required estimators.

Suppose that we have $\hat{\beta}$ and $\hat{b}_i$ given fixed z_0 , then for each z_0 , α is evaluated by the kernel score functions:

$$\frac{1}{N} \sum_{i=1}^N U_Z^i(\alpha; X^\epsilon) = \frac{1}{N} \sum_{i=1}^N \Theta_s(\mathbf{Z}_i) \Delta_i K_h^{1/2}(\mathbf{Z}_i) V_i^{-1} K_h^{1/2}(\mathbf{Z}_i) (Y_i - \hat{\mu}^b\{\mathbf{Z}_i\}) = \mathbf{0}, \quad (3.3.5)$$

where we define $s \times T$ matrix $\Theta_s(\mathbf{Z}_i)$ as $(\Theta_{i1}, \dots, \Theta_{iT})$ and diagonal matrix $\Delta_i = [\hat{\mu}_i^b(\cdot)]'$. V_i is conditional covariance matrix whose j th diagonal elements are $\hat{v}_{ij}^b$. $K_h(\cdot) = K(\cdot/h)/h$ where K is optional symmetric kernel density function with unknown bandwidth parameter h . h can be estimated by the given data. Although it is difficult to identify V_i in practice, we can apply working correlation

structure R_i without loss of consistency. Therefore it turns out that V_i is approximated by $A_i^{1/2} R_i A_i^{1/2}$ where A_i is the diagonal conditional variance matrix with elements $v(\hat{\mu}_i^b)$. The terms in (3.3.5) is evaluated at the initial values Z_0 , with X^ϵ and given $\hat{\beta}$ and $\hat{b}$. The yielding estimate of α , $\hat{\alpha}$ is denoted as $\hat{\alpha}(\hat{\beta}, \hat{b}, Z_0)$. After going through all the values of Z in the sample, we can estimate $\theta(Z_{ij})$ by $\hat{\theta}(Z_{ij}) = \hat{\alpha}_0(Z_{ij}; \hat{\beta}, \hat{b})$. We use $\hat{\alpha}_0(Z_i; \hat{\beta}, \hat{b})$ to represent T -dimensional vector with j th entry $\hat{\alpha}_0(Z_{ij}; \hat{\beta}, \hat{b})$ for the following derivation.

Remark 3.3.3. Bandwidth selection is an important step in the local smoothing. A traditional way to solve h is cross-validation (CV):

$$h = \arg \min_h \sum_{i=1}^N \sum_{j=1}^T \frac{[Y_{ij} - \tilde{\mu}_{-i,j}^b(Z_{ij})]^2}{v[\tilde{\mu}_{-i,j}^b(Z_{ij})]}$$

where $\tilde{\mu}_{-i,j}^b$ denotes the estimate of $\hat{\mu}_{ij}^b$ using all the data except subject i . Known as a computationally intensive approach, CV is able to select a better h when sample size increases, but usually is difficult to implement. Another approach is the empirical bias bandwidth selection (EBBS) method raised by Ruppert (1997). The central idea is to compute empirical mean squared error (EMSE) of $\hat{\theta}(Z_{ij}; h)$ which is estimated by fitting a polynomial regression:

$$EMSE[\hat{\theta}(Z_{ij}; h)] = \{bias[\hat{\theta}(Z_{ij}; h)]\}^2 + var[\hat{\theta}(Z_{ij}; h)]$$

where $bias[\hat{\theta}(Z_{ij}; h)]$ can be approximated by $\hat{c}_1 h^{p+1} + \dots + \hat{c}_t h^{p+t}$ and the estimate of $var[\hat{\theta}(Z_{ij}; h)]$ can be chosen from the local score functions (3.3.5). Concretely, we establish user-supplied finite set $\mathcal{H}$ for h from which we choose h to minimize $EMSE[\hat{\theta}(Z_{ij}; h)]$ for each value of Z . More details are given in Ruppert

3. COVARIATE MEASUREMENT ERROR IN GPLMM

(1997). Suggested by Lin & Carroll (2001), The chosen h often works quite well in practical problem if h is proportional to $N^{-1/3}$.

Next we consider the PQL estimation for β and b . Using *PQL* principle, the estimators $\hat{\beta}$ and each $\hat{b}_i$ satisfy that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N U_X^i(\beta; X^\epsilon) &= \frac{1}{N} \sum_i^N \tilde{\mathbf{X}}_i^{\epsilon T} \Delta_i V^{-1}(\mathbf{X}_i^\epsilon) (Y_i - \hat{\mu}^b\{\mathbf{X}_i^\epsilon\}) = \mathbf{0}, \\ U_B(b_i; X^\epsilon) &= \tilde{\mathbf{B}}_i^T \Delta_i V^{-1}(\mathbf{B}_i) (Y_i - \hat{\mu}^b\{\mathbf{B}_i\}) - C^{-1}b_i = \mathbf{0}, \end{aligned}$$

where $\tilde{\mathbf{X}}_i^\epsilon = \mathbf{X}_i^\epsilon + \partial \hat{\alpha}_0(\mathbf{Z}_i; \beta, b) / \partial \beta^T$ and $\tilde{\mathbf{B}}_i = \mathbf{B}_i + \partial \hat{\alpha}_0(\mathbf{Z}_i; \beta, b) / \partial b_i^T$ as the estimate $\hat{\alpha}_0$ is based on the estimate of β and b .

Before undertaking the main estimating steps, we independently generate M imputed complete data sets $\{X_{(m)}^\epsilon, m = 1, \dots, M\}$, so the preceding estimating equations for $\hat{\alpha}$, $\hat{\beta}$ and $\hat{b}_i$ can be transformed as

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M U_Z^i(\alpha; X_{(m)}^\epsilon) = \mathbf{0}, \quad (3.3.6)$$

$$\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M U_X^i(\beta; X_{(m)}^\epsilon) = \mathbf{0}, \quad (3.3.7)$$

$$\frac{1}{M} \sum_{m=1}^M U_B^i(b_i; X_{(m)}^\epsilon) = \mathbf{0}. \quad (3.3.8)$$

Unless the inverse link $h(\cdot)$ is identical such that the corresponding estimating equations can be solved explicitly, some numerical methods, like Newton-Raphson or Fisher scoring algorithm, must be applied to get the solutions $\hat{\alpha}$, $\hat{\beta}$ and $\hat{b}$.

Estimation of variance components

Denote the current resulting estimate of b in the iterative k th step of the Newton-Raphson algorithm by $\hat{\alpha}^{(k)}$, $\hat{\beta}^{(k)}$ and $\hat{b}^{(k)}$.

Following the EM estimation used in GLMM (Lindstrom & Bates, 1988 and Laird et al., 1987), we have the updated estimation of $C^{(k+1)}$ as

$$\hat{C}^{(k+1)} = \hat{C}^{(k)} + \frac{1}{N} \sum_{i=1}^N [\hat{b}_i^{(k)} \hat{b}_i^{(k)T} - \hat{C}^{(k)} \mathbf{B}_i^T (\mathbf{P}_i^{(k)})^{-1} \mathbf{B}_i \hat{C}^{(k)}] \quad (3.3.9)$$

for ML estimates and

$$\hat{C}^{(k+1)} = \hat{C}^{(k)} + \frac{1}{N} \sum_{i=1}^N [\hat{b}_i^{(k)} \hat{b}_i^{(k)T} - \hat{C}^{(k)} \mathbf{B}_i^T \mathbf{Q}_i^{(k)} \mathbf{B}_i \hat{C}^{(k)}] \quad (3.3.10)$$

for REML estimates where $\mathbf{P}_i^{(k)}$ denotes the variance of Y_i evaluated at $\hat{\alpha}^{(k)}$, $\hat{\beta}^{(k)}$ and $\hat{b}^{(k)}$. $\mathbf{Q}_i^{(k)} = (\mathbf{P}_i^{(k)})^{-1} - (\mathbf{P}_i^{(k)})^{-1} \mathbf{X}_i \left(\sum_{i=1}^N \mathbf{X}_i^T (\mathbf{P}_i^{(k)})^{-1} \mathbf{X}_i \right)^{-1} \mathbf{X}_i^T (\mathbf{P}_i^{(k)})^{-1}$. Note that X_i should be replaced by X_m^ϵ in our proposed algorithm for each imputed complete set. For the details, please see Appendix 2.

Compared with the ML estimates, EM for REML estimates always leads to smaller estimated variance components (Laird & Ware, 1982).

3.4 Asymptotic inference

From this section, the local linear kernel, namely $s = 1$, accounts for inferring the asymptotic properties of $\hat{\theta}(Z_{ij})$. We investigate the asymptotic bias and variance of $\hat{\theta}(Z_{ij})$, detailing the consistency and asymptotic normality theory of estimator $\hat{\beta}$. The previous elaborate work in generalized nonlinear or partially linear model

3. COVARIATE MEASUREMENT ERROR IN GPLMM

without random effects, has been well developed by Lin & Carroll (2000, 2001). Zhou & Liang (2009) studied the similar semiparametric model with varying coefficients and inferred the asymptotics on both parametric and nonparametric components based on covariates calibration (CC) that error-prone covariates are sufficiently estimated by other auxiliary variables. In addition to imputing data values alone, which always be seen in the Monte Carlo studies like MCEM, MI allows one further step to impute parameters in order to fully account for the posterior information behind the data structure. Thus, the calibration methods like RC and CC, can be somewhat considered as the special cases in which the posterior distribution of model parameters are noninformative.

At preliminary stage, we assume that each subject i is measured finite T times and $f_j(\cdot)$ is the marginal density of Z_{ij} . The (j, k) th element of V^{-1} is denoted by w^{jk} . Define $E_c(l) = \{c_K(l), \dots, c_K(l+s)\}^T$ and $E_s(c)$ the matrix whose (j, k) th element is $c_K(j+k-2)$, with $c_K(r) = \int z^r K(z)dz$. Furthermore, let $E_s(\gamma)$ denote the matrix whose (j, k) th element is $\gamma_K(j+k-2)$, with $\gamma_K(r) = \int z^r K^2(z)dz$. Suppose K to be symmetric about 0, so we have $c_K(r) = 0$ if r is odd. Some notations needs to be clarified at the beginning. In the following discussions, the low subscript (m) indicates that the term is computed in the m th imputed complete data set. $\{(m), i\}$ and $\{(m), i, j\}$ represent the i th subject and its j th repeated observation, respectively.

We then conclude the asymptotics of the nonparametric component $\hat{\theta}(z)$ in the following theorems:

Theorem 3.4.1. *Let $\hat{\theta}(z)$ be the solution of (3.3.5) with respect to $\hat{\beta}$ and $\hat{b}$. $Nh \rightarrow \infty$ as $N \rightarrow \infty$, $h \rightarrow 0$. If $\hat{\beta}$ and $\hat{b}$ are both $\sqrt{N}$ consistent then the*

3.4 Asymptotic inference

asymptotic bias of $\hat{\theta}(z)$ given linear kernel K with $s = 1$ is

$$\text{bias}[\hat{\theta}(z)] \approx h^2 \theta^{(2)}(z) / 2c_K(2) \quad (3.4.1)$$

Under conditional independence assumption that $R = I$, the asymptotic conditional variance of $\hat{\theta}(z)$ is

$$\text{var}^b[\hat{\theta}(z)] = \frac{M}{Nh} \left(\sum_{m=1}^M \hat{\mu}_{(m)}^{b(1)}(\theta(z)) \sum_{j=1}^T f_j(z) w_{(m)}^{jj}(z) \right)^{-1} \gamma_K(0), \quad (3.4.2)$$

where $\hat{\mu}_{(m)}^{b(1)}(\theta(z))$ is the 1st derivative of $\hat{\mu}_{(m)}^b$.

In practice, once C is evaluated by the method in Section 3.3, and the marginal variance of $\hat{\theta}(z)$ can be estimated by empirical MCMC method. Detailed proof is given in Appendix 2.

We summarize the conditional inference of $\hat{\beta}$ and its the marginal consistency and asymptotic normality in the following theorems.

Theorem 3.4.2. *Suppose that for conditionally independent correlation R , given the estimate $\hat{\theta}(z)$ and $\hat{b}$, as $N \rightarrow \infty$, by Law of Large Number we have*

$$\sqrt{N} \|\hat{\beta} - \beta\| \rightarrow O_p(1). \quad (3.4.3)$$

And $\sqrt{N}(\hat{\beta} - \beta)$ asymptotically tends to conditionally normal distribution with mean 0 and covariance V_β^b which is

$$V_\beta^b = D_{(M)}^{-1} E \left\{ \frac{1}{M} \sum_{m=1}^M G_{(m)} \Omega G_{(m)}^T + \frac{1}{M} \sum_{m=1}^M G_{(m)} \Delta_{(m),i} \beta \text{cov}(U^*) \beta^T \Delta_{(m),i} G_{(m)}^T \right\} D_{(M)}^{-1}, \quad (3.4.4)$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

where $G_{(m)} = \tilde{X}_{(m)}^{\epsilon T} \Delta_{(m)} V^{-1}(X_{(m)}^{\epsilon})$, $D_{(M)} = \frac{1}{M} \sum_{m=1}^M E[\tilde{X}_{(m)}^{\epsilon T} \Delta_{(m)} V^{-1}(X_{(m)}^{\epsilon}) \Delta_{(m)} \tilde{X}_{(m)}^{\epsilon}]$, and Ω is the conditional covariance matrix of Y .

The asymptotic variance V_{β} can be evaluated at the current estimates $\hat{\beta}$ and $\hat{b}$. Let $\hat{G}_{(m)}$ and $\hat{D}_M$ denote the corresponding estimated value of $G_{(m)}$ and $D_{(m)}$ and we have

$$\begin{aligned} \hat{G}_{(m),i} &= \tilde{\mathbf{X}}_{(m),i}^{\epsilon T} \Delta(\mathbf{X}_{(m),i}^{\epsilon}; \hat{\beta}, \hat{b}_i) V^{-1}(\mathbf{X}_{(m),i}^{\epsilon}; \hat{\beta}, \hat{b}_i) \\ \hat{D}_{(M)} &= \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{M} \sum_{m=1}^M \tilde{X}_{(m),i}^{\epsilon T} \Delta(\mathbf{X}_{(m),i}^{\epsilon}; \hat{\beta}, \hat{b}_i) V^{-1}(\mathbf{X}_{(m),i}^{\epsilon}; \hat{\beta}, \hat{b}_i) \Delta(\mathbf{X}_{(m),i}^{\epsilon}; \hat{\beta}, \hat{b}_i) \tilde{\mathbf{X}}_{(m),i}^{\epsilon} \right\}. \end{aligned}$$

Until now the model discussed is supposed to be conditionally independent, the only one that needs to be estimated in Ω is the diagonal elements. The estimated asymptotic variance $\hat{V}_{\beta}^b$ is given by

$$\hat{V}_{\beta}^b = \hat{D}_{(M)}^{-1} + \hat{D}_{(M)}^{-1} \frac{1}{N} \left\{ \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \hat{G}_{(m),i} \Delta_{(m),i} \hat{\beta}_e \widehat{cov}(U_{(m)}^{\star}) \hat{\beta}_e \Delta_{(m),i} \hat{G}_{(m),i}^T \right\} \hat{D}_{(M)}^{-1}$$

where $\widehat{cov}(U_{(m)}^{\star})$ is estimated by the variation difference between $X_{(m)}^{\epsilon}$ and W . $\hat{\beta}_e$ denotes the elements of $\hat{\beta}$ corresponding to the error-prone covariate X^{ϵ} in covariates vector X .

Theorem 3.4.2 shows that $\hat{\beta}$ has strong asymptotic performance conditional on b . The marginal distribution then can be calculated by the total expectation and variance formulas. We have that as $N \rightarrow \infty$, $\sqrt{N}(\hat{\beta} - \beta)$ is still normally

distributed with mean zero and covariance $\hat{V}_\beta$

$$\begin{aligned} \hat{V}_\beta = & E_b[\hat{D}_{(M)}^{-1}] + \\ & E_b[\hat{D}_{(M)}^{-1}] \frac{1}{N} \left\{ \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M E_b[\hat{G}_{(m),i}] \Delta_{(m),i} \beta_e \text{cov}(U_{(m)}^*) \beta_e \Delta_{(m),i} E_b[\hat{G}_{(m),i}]^T \right\} \\ & \times E_b[\hat{D}_{(M)}^{-1}]. \end{aligned} \quad (3.4.5)$$

Calculating $\hat{V}_\beta$ is actually very complicated. In practice, a common solution is to use a first-order approximation following the work of Breslow & Clayton (1993) and Liang (2009). We also use this technique to construct the approximated variance of Y_i in Appendix 2. Particularly, when the responses Y are assumed to be simply normal, we obtain that

$$E_b[\hat{G}_{(m),i}] = \tilde{\mathbf{X}}_{(m),i}^\epsilon (\Omega + \mathbf{B}_i C \mathbf{B}_i^T)^{-1} \quad (3.4.6)$$

$$E_b[\hat{D}_{(M)}^{-1}] = N \left\{ \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \tilde{\mathbf{X}}_{(m),i}^{\epsilon T} (\Omega + \mathbf{B}_i C \mathbf{B}_i^T)^{-1} \tilde{\mathbf{X}}_{(m),i}^\epsilon \right\}^{-1}. \quad (3.4.7)$$

3.5 Simulation study

We conducted a simulation study to assess the performance of the proposed methodology and investigate the impact of ignoring measurement error on estimation. In the following simulation, we considered a hierarchical model with identity link:

$$\mu_{ij}^b = E(y_{ij}|b_i) = x_{ij}\beta + r_{ij}b_i + \sin(\pi z_{ij}), \quad (3.5.1)$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

for $i = 1, \dots, N$ and $j = 1, \dots, T$, where x_{ij} was generated from normal $(0, 2^2)$, while r_{ij} and z_{ij} were both from uniform distribution over $(-1, 1)$ and $(0, 1)$ respectively. A random effect b_i was used, $b_i \sim \mathcal{N}(0, \sigma_b^2)$. In all simulation, we considered relatively small and large sample sizes $N = 30$ and 100 , $T = 5$. The true model parameter $\beta = 1$ or 2 . Under model (3.5.1), response data y_{ij} were then generated from $\mathcal{N}(\mu_{ij}^b, \sigma^2)$. The two variance components σ and σ_b were set as $(1, 1)$ or $(0.5, 0.5)$.

With measurement error, the observed value w_{ij} was generated from $w_{ij} = x_{ij}\eta + u_{ij}$, where $u_{ij} \sim \mathcal{N}(0, 0.3^2)$ and $\eta = 1$ or 0.5 . The size of validation set N_v was set to be approximately $N/3$ to meet the need of imputation procedure. Hence x_{ij} was supposed to be correctly recorded only in the first N_v subjects. For the rest of subjects, mismeasured covariate w_{ij} were observed.

The values X^ϵ of X were first imputed by the iterative procedure described in Section 3.2 then substituted into the analysis model specified in the same section. Normally, for the purpose of generating appropriate X^ϵ it is better to choose the length of the initial burn-in period as long as possible. However for the sake of the computation convenience, a too long burn-in period does not worth being initialized at the beginning. Investigating the method performance by many repeated experiments, we used a moderate length depending on the structure of measurement error model itself. In our simulations, the first 2000 imputing iterations were dropped in advance, followed by $M = 3$ imputed sets drawn for every 100th iteration afterwards. The starting values for iterative computation was obtained from the validation set.

For the purpose of illustrating the convergence of MI process, we consider the case of $N = 100$, $T = 5$, $\beta = 1$, $\eta = 0.5$, $\sigma = 1$ and $\sigma_b = 1$. The trace plots

Figure 3.1: Trace plots of $\hat{\beta}^*$, $\hat{\eta}^*$, $\hat{\mu}_x^*$, $\hat{\sigma}^*$, $\hat{\sigma}_u^*$ and $\hat{\sigma}_x^*$ in the case that $N = 100$, $T = 5$, $\beta = 1$, $\eta = 0.5$, $\sigma = 1$ and $\sigma_b = 1$. The blue and red lines represents two iterative chains with two different sets of starting values, respectively.

for visually inspecting the iterative imputations are shown in Figure 3.1. Two different sets of initial values are used for comparison. The length of simulated period is set to 5000. It is found that our proposed algorithm for MI almost con-

3. COVARIATE MEASUREMENT ERROR IN GPLMM

verges after sufficiently long iterative period. Therefore it is feasible to generate the imputed data sets by discarding the first 2000 iterations.

Figure 3.2: True curve (solid line) and estimated curve (dotted line) values of nonlinear function $\theta(Z)$ calculated from 500 simulations. In (a): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (b): $N = 30$, $T = 5$, $\beta = 1$, $\eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$.

We ran 500 simulations and utilized the two-step algorithm to fit the GPLMM with error-prone covariates. The results are reported in Table 3.1 based on the proposed method for imputed X^ϵ and the naive method with values of X and W . If measurement error is relatively minor, for example, when $\eta = 1$, the three methods provide fairly comparable results in estimating β in terms of bias and standard error. Our approach can calibrate the bias significantly for the underlying partially linear model, especially for cases ($\eta = 0.5$) where the observed W appears to be very different from the true X . As expected, the bias of the proposed estimator decreases with increased sample size. Even for relatively small sample cases, for instance $N = 30$ and $\beta = 1$, $\eta = 0.5$ (large measurement error), our proposed method gave $\beta = 1.011$ which is much closer to the result 1.001 with

the true covariate, while the naive method led to considerably biased estimate for β . The estimated standard error for estimating β coincides with the sample standard error in all scenarios. This indicates that the variance estimation formula given in Theorem (3.4.2) performs quite well. When comparing to the estimation with the true X , it is not surprising that accounting for measurement error in the proposed estimation procedure loss some efficiency in return.

Figure 3.3: Estimated variance components σ and σ_b for 4 scenarios, based on 500 simulations. In (a): $N = 30, T = 5, \beta = 1, \eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (b): $N = 30, T = 5, \beta = 1, \eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$; In (c): $N = 100, T = 5, \beta = 1, \eta = 1$, and the true variance components $\sigma = 1$ and $\sigma_b = 1$; In (d): $N = 100, T = 5, \beta = 1, \eta = 0.5$, and the true variance components $\sigma = 0.5$ and $\sigma_b = 0.5$.

We used local smoothing approach to estimate nonlinear component $\theta(Z)$. To evaluate the performance of the nonparametric estimate $\hat{\theta}(Z)$, we show kernel-

3. COVARIATE MEASUREMENT ERROR IN GPLMM

based estimators in Figure 3.2 for two small sample cases. It can be seen that the kernel-based estimate is slightly apart from the true $\theta(Z)$ on the upper boundary. Fortunately, this slight difference did not provide any significant bias in the resulting estimate $\hat{\beta}$ in the simulation.

Table 3.1: The mean, sample standard error (s.s.e) and estimated standard error (e.s.e) of the model parameter estimator $\hat{\beta}$ based on 500 simulations using proposed MIPQL method.

N	β	η	σ	σ_b	Method	Mean	S.s.e	E.s.e
30	1	1	1	1	TRUE	1.002	0.045	0.043
					NAIVE	0.979	0.046	0.045
					MIPQL	0.984	0.046	0.045
30	1	0.5	0.5	0.5	TRUE	1.001	0.023	0.022
					NAIVE	1.835	0.064	0.062
					MIPQL	1.011	0.041	0.034
30	2	1	1	1	TRUE	2.002	0.045	0.043
					NAIVE	1.957	0.051	0.049
					MIPQL	1.969	0.050	0.050
30	2	0.5	0.5	0.5	TRUE	2.001	0.023	0.022
					NAIVE	3.668	0.104	0.100
					MIPQL	2.016	0.060	0.060
100	1	1	1	1	TRUE	1.000	0.024	0.024
					NAIVE	0.978	0.025	0.025
					MIPQL	0.985	0.024	0.025
100	1	0.5	0.5	0.5	TRUE	1.000	0.012	0.012
					NAIVE	1.834	0.033	0.034
					MIPQL	1.006	0.023	0.018
100	2	1	1	1	TRUE	2.000	0.024	0.024
					NAIVE	1.956	0.027	0.027
					MIPQL	1.970	0.026	0.027
100	2	0.5	0.5	0.5	TRUE	2.000	0.012	0.012
					NAIVE	3.669	0.053	0.055
					MIPQL	2.006	0.034	0.032

Figure 3.3 shows the performance of estimated variance components $\hat{\sigma}$ and $\hat{\sigma}_b$

over 500 replications. We considered four scenarios for small and large samples. Compared with (a), (b) gives less convincing estimates $\hat{\sigma}$ and $\hat{\sigma}_b$ (0.431 and 0.526). When sample size increases, the variances of $\hat{\sigma}$ and $\hat{\sigma}_b$ decrease.

3.6 Application to PANS data

As an illustration, we apply our proposed method to analyze the PANS data (Burke et al., 2012). Totally 375 individuals from Australia were included in the study. The baseline characteristics of participants are shown in Table 3.2. The intervention and control groups were similar in terms of demographic and lifestyle factors. In summary, nearly half of the recruited seniors were men, the averaged age was 66 years overall, almost all of them received secondary education and above, most of them had a relationship and majority of them were in struggle with their financial status. Meanwhile, no significant differences in alcohol drinking and smoking were found between these two groups.

It is of interest to explore the effect of weekly weak exercise on the change of **WHR**. We considered the **WHR** be the response, and the covariates included in the longitudinal model are the participants gender **Sex**, the walking exercise time **Walk** per week, **Group** indicating whether in control or intervention and **time** whose value equals 0 at baseline stage and 1 otherwise. The other covariate involved into an unknown nonlinear function $\theta(\cdot)$, is baseline measurement **Age**. The response and the covariates are postulated by the hierarchical model:

$$\begin{aligned} \log(WHR_{ij}) = & Sex_i\beta_1 + Walk_{ij}\beta_2 + Group_i\beta_3 + Group_i * time_{ij}\beta_4 + \\ & \theta(Age_i) + b_{1i} + time_{ij}b_{2i} + \varepsilon_{ij}, \end{aligned} \quad (3.6.1)$$

Table 3.2: Description of baseline characteristics for our PANS data study.

Characteristic	Overall(n=375)	Treatment (n=176)	Control (n=199)	P-value
Age (median, IQR)	66.0(63.0-68.0)	66.0(63.0-68.0)	66.0(63.0-68.5)	0.883
Male (%)	194(51.7)	93(52.8)	101(50.8)	0.756
Education level (%):				0.492
primary school	24(6.4)	8(4.5)	16(8.0)	
secondary school	174(46.4)	83(47.2)	91(45.7)	
diploma	105(28.0)	48(27.3)	57(28.6)	
university	72(19.2)	37(21.0)	35(17.6)	
Relationship: with partner (%)	287(76.5)	128(72.7)	159(79.9)	0.113
Working (%)	157(40.7)	77(43.8)	80(40.2)	0.530
Financial struggle (%):				0.963
never	49(13.1)	24(13.6)	25(12.6)	
sometimes	246(65.6)	115(65.3)	131(65.8)	
always	80(21.3)	37(21.0)	43(21.6)	
Alcohol drinker: yes (%)	253(67.5)	116(65.9)	137(68.8)	0.582
Smoking (%):				0.274
never	191(50.9)	97(55.1)	94(47.2)	
former	163(43.5)	69(39.2)	94(47.2)	
current	16(6.0)	10(5.7)	11(5.5)	

where $i = 1, \dots, 375$, $j = 1, 2$, and the within-subject correlation in each individual i arises from common random intercept b_{1i} and random time slope b_{2i} . ε_{ij} defined as overall model error, was assumed to be homoscedastic. The log-transformation of **WHR** is necessary due to the positive values in the sample. The mixed model (3.6.1) can be interpreted in terms of treatment status and group allocation: for the control group at baseline, $\log(WHR_{ij}) = Sex_i\beta_1 + Walk_{ij}\beta_2 + \theta(Age_i) + b_{1i} + \varepsilon_{ij}$; for the control group at post period, $\log(WHR_{ij}) = Sex_i\beta_1 + Walk_{ij}\beta_2 + \theta(Age_i) + b_{1i} + time_{ij}b_{2i} + \varepsilon_{ij}$; for intervention group at baseline, $\log(WHR_{ij}) = Sex_i\beta_1 + Walk_{ij}\beta_2 + Group_i\beta_3 + \theta(Age_i) + b_{1i} + \varepsilon_{ij}$; for intervention group at post period, $\log(WHR_{ij}) = Sex_i\beta_1 + Walk_{ij}\beta_2 + Group_i\beta_3 + Group_i * time_{ij}\beta_4 + \theta(Age_i) + b_{1i} + time_{ij}b_{2i} + \varepsilon_{ij}$. The fixed intercept was not introduced into the model because its effect is absorbed by the nonparametric fitting procedure. As we mentioned, the unknown function of **Age** was approximated by a local linear kernel. The subject-specific interpretation of regression parameter β is of our major interest. For example, β_4 alone indicates the difference between the changes of **WHR** in control and intervention groups over the six-month period. The negative values of β_4 shows that the treatment indeed has positive impact on decreasing **WHR**. In PANS study although the participants received supports more or less in both intervention and control groups, we believe the education background, one of significant covariates in the early work of Burke et al. (2012), may affect the initial data collection by the participants themselves. Walking exercise time was collected based on subjects' self reports, which may be inaccurately measured. So we considered **Walk** as a covariate subject to error.

Walk taken by the participants with low education performs less appealing in our preliminary study, in contrast to the observations taken by high-educated

3. COVARIATE MEASUREMENT ERROR IN GPLMM

seniors, under the same model (3.6.1) above. We chose $N_v = 177$ out of $N = 375$ participants as the validation objectives, corresponding to high-educated subjects. The proposed PQL method was applied based on (1) the original data set ignoring the measurement error, (2) data in validation set only, and (3) data accounting for measurement error using MI approach. In order to explain the necessity of introducing nonlinear term in (3.6.1), a generalized linear mixed model was also adopted to analyze the original data using function *glmer()* in R package *lme4*. The result is summarized in Table 3.3.

Table 3.3: Inference for the PANS data and the performance of our proposed MIPQL compared with other existing approaches.

Parameter		Method			
		Naive	Validation	MIPQL	GLMER
Sex	Estimate	0.134	0.132	0.135	0.133
	Std.err	0.005	0.008	0.006	0.007
Walk	Estimate	-1.895	-2.708	-2.305	-1.137
	Std.err	0.454	0.828	0.603	0.435
Group	Estimate	0.010	0.015	0.009	0.010
	Std.err	0.006	0.010	0.007	0.007
Group $\times$ time	Estimate	-0.015	-0.016	-0.015	-0.017
	Std.err	0.008	0.011	0.008	0.004
σ_b					
Intercept	Estimate	0.011	0.021	0.024	0.062
Time	Estimate	0.017	0.030	0.032	0.035
Correlation	Estimate	-0.701	-0.356	-0.200	-0.303
σ	Estimate	0.069	0.071	0.074	0.032

Table 3.3 shows that walking exercise has positive effect on reducing the waist-to-hip ratio as we expected. Regardless of the error in covariates and comparing $\hat{\beta}_2$ in **Naive** and **GLMER**, we found that introducing the nonlinear function θ

leads to remarkable difference in the estimates of model parameters. The **Naive** method shows more significant influence on **WHR** than the **GLMER** per unit walking time. The comparison between **Naive** and **Validation**, especially the difference in β_2 (-1.895, -2.708), indicates the error in recording walking time affects the model fitting dramatically.

Further, we calculate the sum of squared residuals (SSR), and get 3.551 for **Naive** and 3.606 for **GLMER**. This somehow supports the use of nonlinear function θ in the predictor. The curve shown in Figure 3.4 exhibits clear fluctuations as **Age** increases, confirming the existence of nonlinear components in (3.6.1). The estimate of **Walk** in **MIPQL** was found distinguished from that in **Naive** and much closer to the estimate of the true coefficient (**Validation**), whilst the other covariates have the nearly same effects on the response **WHR**. Especially, the **Group*time** interaction term remains significant after correcting the measurement error on **Walk**. The intervention is still effective in reducing **WHR** among recruited seniors. We noted that the significance level for **Walk** is not improved in **MIPQL**. This is not surprising as the error correction approaches always attain a trade-off bound between bias and variation (Carroll et al. 1995).

3.7 Chapter summary

In this chapter, we propose a multiple imputation method for analyzing longitudinal data with measurement error in covariates under generalized partial linear mixed models, a class of flexible tools for modeling complex longitudinal or clustered data. A two-step estimation procedure is developed: we perform the standard MI for mismeasured covariates based on a measurement error model at the

3. COVARIATE MEASUREMENT ERROR IN GPLMM

Figure 3.4: nonparametric estimate of unknown nonlinear $\theta(Z)$. The estimated value is drawn in solid line with its 95% confidence interval drawn between two dotted lines.

first step; with the nonlinear component approximated by smoothing techniques, we then estimate regression coefficients using the penalized quasi-likelihood at the second step.

The use of random effects in the model can remarkably enhance the model flexibility to cover various problems. The mixed model can explain the variation between and within clusters by assigning the corresponding covariates in the model itself, while a user-driven working association (except unstructured assumption) in GEE has less interpretation with the data. The PANS data we

analyzed in this chapter may contain false observations in walking time covariates. Accounting for error correction, we fitted a GPLMM with proposed MIPQL, which performs better than the naive model as expected. The walking exercise was found that it has positive impact on the seniors with obesity problem.

The introduction of nonlinear term in the predictor extends our mixed effect model to more general scenarios. The model fitting is no longer limited by the existing link functions. In our study, the case we concern was specific link relationship with unknown partially linear predictor. In more complicated case where the underlying link is not easy to identify in practice, the nonparametric approximation also can be implemented in both link and predictor simultaneously. The data-driven kernel method is very effective to evaluate the unknown function of some baseline characteristics. The treatment time, for example, in most health studies are not treated to be linear absolutely. In our simulation and data example, the nonlinear term $\theta(\cdot)$ is approximated by local smoother with relatively low order.

Multiple imputation can completely utilize the information provided by the data itself. When the measurement error is found in the covariates, the imputation method becomes more reasonable and stronger than other error correction methods, in which the unobserved X is simply surrogated by the conditional expectation of it given W . If we regard the error-prone covariates to be missing, our semiparametric algorithm to generate X^ϵ is very similar to the MCEM and MCNR approaches (McCulloch 1997), except that we do not only take into account the information stored in the response and covariates, but also investigate the background of parameters specified by the model itself. Although doing this could make the whole imputation procedure more intensive, it is worth involving

3. COVARIATE MEASUREMENT ERROR IN GPLMM

the parameter imputing steps, because their estimator should also be considered as random variable in the modeling process.

3.8 Appendix 2

Derivation of posterior distribution for model parameters in Multiple imputation algorithm

The derivation for the algorithm steps in Section 3.3 is not difficult but complicated. The details of imputation process for $\hat{\eta}^*$ and $\hat{\sigma}_u^{*2}$ in **Step 1** are shown below. Other steps can be derived in the similar way.

First, let $W^v = (W_1^v, \dots, W_{N_v}^v)^T$ and we rewrite (3.2.1) as

$$W^v = X^v \eta + U^v$$

The posterior distribution of η is expressed as

$$f(\eta|X^v, W^v) = \frac{f(\eta, W^v|X^v)}{\int f(\eta, W^v|X^v) d\eta}$$

Assuming the marginal distribution of η is constant, that is, $f(\eta) = c$, we have $f(\eta|X^v, W^v)$ is proportional to $f(W^v|\eta, X^v)f(\eta)$ which represents multivariate normal distribution with mean $X^v \eta$ and covariance matrix whose diagonal components are all σ_u^2 .

Since $\|W^v - X^v\eta\|^2 = \|W^v - X^v\hat{\eta}\|^2 + \|X^v(\hat{\eta} - \eta)\|^2$,

$$\begin{aligned} \int f(\eta, W^v|X^v)d\eta &= c \int \frac{1}{(2\pi\sigma_u^2)^{N_v/2}} \exp\left(\frac{\|W^v - X^v\hat{\eta}\|^2 + \|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right) d\eta \\ &= c \frac{\exp\left(\frac{\|W^v - X^v\hat{\eta}\|^2}{-2\sigma_u^2}\right)}{(2\pi\sigma_u^2)^{N_v/2}} \int \exp\left(\frac{\|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right) d\eta \\ &= \frac{c(2\pi\sigma_u^2)^{1/2} \exp\left(\frac{\|W^v - X^v\hat{\eta}\|^2}{-2\sigma_u^2}\right)}{(2\pi\sigma_u^2)^{N_v/2} \|X^v\|} \int \frac{\|X^v\|}{(2\pi\sigma_u^2)^{1/2}} \exp\left(\frac{\|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right) d\eta. \end{aligned}$$

Then,

$$f(\eta|X^v, W^v) = \frac{\|X^v\|}{(2\pi\sigma_u^2)^{1/2}} \exp\left(\frac{\|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right) \int \frac{\|X^v\|}{(2\pi\sigma_u^2)^{1/2}} \exp\left(\frac{\|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right) d\eta.$$

By $\int f(\eta|X^v, W^v)d\eta = 1$, we conclude that

$$f(\eta|X^v, W^v) = \frac{\|X^v\|}{(2\pi\sigma_u^2)^{1/2}} \exp\left(\frac{\|X^v(\hat{\eta} - \eta)\|^2}{-2\sigma_u^2}\right).$$

Therefore, $\eta|X^v, W^v \sim \mathcal{N}(\hat{\eta}, [X^{vT}X^v]^{-1}\sigma_u^2)$. Further, since $\frac{(N_vT-1)\sigma_u^2}{\sigma_u^2} \sim \chi_{N_vT-1}^2$, it follows that $\sigma_u^2|X^v, W^v \sim (N_vT-1)\hat{\sigma}_u^2/\chi_{N_vT-1}^2$. Thereafter we can impute $\hat{\eta}^*$ and $\hat{\sigma}_u^{*2}$ from the posterior distributions above.

Derivation of estimator $\hat{C}$

As discussed by Breslow & Clayton (1993) for Generalized Linear Mixed Models (GLMM), an appropriate approximation often makes the computation less intensive but reasonable. We have defined the conditional variance v^b where the random effects b is known. However, in Section 3.4 where we introduce the asymptotic behavior of $\hat{\beta}$ and a powerful method to estimate the variance com-

3. COVARIATE MEASUREMENT ERROR IN GPLMM

ponents, the marginal variance of Y is needed but hardly to obtain when the relation between conditional mean and predictor is not simply linear.

In the case of (3.3.4), Y_i is often assumed to be

$$Y_i \approx \hat{\mu}_i^b + \varepsilon_i$$

where $\text{var}(\varepsilon_i) = v(\hat{\mu}^b)$. We further approximate it at its 1st order with respect to $\mathbf{B}_i b_i$ by

$$Y_i \approx h(\mathbf{X}_i^\epsilon \beta + \Theta_i \alpha) + h^{(1)}(\mathbf{X}_i^\epsilon \beta + \Theta_i \alpha) \mathbf{B}_i b_i + \varepsilon_i,$$

and the approximated marginal variance is written as

$$\mathbf{P}_i = \text{var}(Y_i) = v(\hat{\mu}_i^b) + L_i \mathbf{B}_i C \mathbf{B}_i^T L_i, \quad (3.8.1)$$

where $\hat{\mu}_i^b = h(\mathbf{X}_i^\epsilon \beta + \Theta_i \alpha)$ and matrix $L_i = h^{(1)}(\mathbf{X}_i^\epsilon \beta + \Theta_i \alpha)$.

During the estimation procedure, the unknown variance components C and R_i can be estimated separately. Generally, R_i should be defined as conditional independence given b_i . The model correlation within each subject i is assumed to be interpreted by the random effects b_i only. The asymptotic inference discussed below shows that being conditionally independent guarantees the consistency of our estimators under some regularity assumptions. The covariance matrix C can be estimated by *EM* algorithm, or alternatively by profile penalized likelihood. According to the method first proposed by Laird & Ware (1982) via conventional maximum likelihood, $\hat{C}$ is of the form:

$$\hat{C} = \frac{1}{N} \sum_{i=1}^N b_i b_i^T = t/N. \quad (3.8.2)$$

From the results of Lindstrom & Bates (1988) and Laird et al. (1987), given the current estimate $\hat{C}$, we let $\hat{b}_i$ be $\hat{b}_i(\hat{C})$, and the estimate of b_i , and $\hat{t}$ denote the corresponding estimate of t , which satisfies

$$\begin{aligned}\hat{t} &= E\left\{\sum_{i=1}^N b_i b_i^T | y_i, \hat{\beta}, \hat{C}\right\} \\ &= \sum_{i=1}^N \{\hat{b}_i(\hat{C}) \hat{b}_i(\hat{C})^T + \text{var}(b_i | y_i, \hat{\beta}, \hat{C})\}.\end{aligned}\tag{3.8.3}$$

To summarize, with the reasonable initial value of $\hat{C}$, first we calculate $\hat{t}$ via its conditional expectation (E-step), then update the value of $\hat{C}$ by ML estimates (M-step). Run iterations between these two steps until convergence.

In mixed-effect model context, REML estimates are usually more reasonable than ML estimates, accounting for the influence of fixed effects $\hat{\beta}$ on the estimation of b . By integrating β out of the likelihood, the M-step does not change but the E-step for REML estimates becomes

$$\begin{aligned}\hat{t} &= E\left\{\sum_{i=1}^N b_i b_i^T | y_i, \hat{C}\right\} \\ &= \sum_{i=1}^N \{\hat{b}_i(\hat{C}) \hat{b}_i(\hat{C})^T + \text{var}(b_i | y_i, \hat{C})\}.\end{aligned}\tag{3.8.4}$$

Proof of Theorem 3.4.1.

Given that $\hat{\beta}$ and $\hat{b}$ are consistent, the similar asymptotics have been developed by Lin & Carroll (2000, 2001) for generalized linear models with nonlinear and partially linear predictors. Now assuming that $\sqrt{N}\|\hat{\beta} - \beta\| = O_p(1)$ and $\sqrt{N}\|\hat{b} -$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

$b\| = O_p(1)$ and using Taylor expansion, we have

$$\begin{aligned}\sqrt{Nh}[\hat{\theta}(Z; \hat{\beta}, \hat{b}) - \theta(Z)] &= \sqrt{Nh}[\hat{\theta}(Z; \hat{\beta}, \hat{b}) - \hat{\theta}(Z; \beta, b)] + \sqrt{Nh}[\hat{\theta}(Z; \beta, b) - \theta(Z)] \\ &= \sqrt{h} \left[\frac{\partial \hat{\theta}(Z; \beta, b)}{\partial \beta^T} \right] [\sqrt{N}(\hat{\beta} - \beta)] + \sqrt{h} \left[\frac{\partial \hat{\theta}(Z; \beta, b)}{\partial b^T} \right] [\sqrt{N}(\hat{b} - b)] \\ &\quad + \sqrt{Nh}[\hat{\theta}(Z; \beta, b) - \theta(Z)] + o_p(1).\end{aligned}\tag{3.8.5}$$

From (3.3.6), $\hat{\theta}(Z; \beta, b) - E[\hat{\theta}(Z; \beta, b)] = e^T B_{NM}^{-1} A_{NM} + o_p(N^{-1/2})$ where

$$\begin{aligned}e &= (1, 0)^T \\ B_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Theta_s(\mathbf{Z}_i) \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) \Delta_{(m),i} \Theta_s^T(\mathbf{Z}_i), \\ A_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Theta_s(\mathbf{Z}_i) \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) [Y_i - \hat{\mu}_{(m)}^b(\mathbf{Z}_i)].\end{aligned}$$

Let $AA_{NM} = \partial A_{NM} / \partial \beta^T$, so

$$AA_{NM} = -\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Theta_s(\mathbf{Z}_i) \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) \Delta_{(m),i} \tilde{\mathbf{X}}_{(m),i}^\epsilon.$$

It follows that $\partial \hat{\theta}(Z; \beta, b) / \partial \beta^T = e^T B_{NM}^{-1} AA_{NM} + o_p(N^{-1/2}) = O_p(1)$, which implies

$$\sqrt{Nh}[\hat{\theta}(Z; \hat{\beta}, \hat{b}) - \theta(Z)] = \sqrt{Nh}[\hat{\theta}(Z; \beta, b) - \theta(Z)] + o_p(1).\tag{3.8.6}$$

Consequently, $\hat{\theta}(Z; \hat{\beta}, \hat{b})$ has the same asymptotic distribution as $\hat{\theta}(Z; \beta, b)$.

B_{NM} and A_{NM} are evaluated at β and b . Further, as $N \rightarrow \infty$ with respect to

subscript i , B_{NM} tends to

$$B_M = \frac{1}{M} \sum_{m=1}^M E\{\Theta_s(Z)\Delta_{(m)}K_h^{1/2}(Z)V_{(m)}^{-1}(Z)K_h^{1/2}(Z)\Delta_{(m)}\Theta_s^T(Z)\}$$

whose (l_1, l_2) th element is

$$\begin{aligned} & \frac{1}{M} \sum_{m=1}^M E\left\{\sum_{j=1}^T \sum_{j'=1}^T \left(\frac{Z_j - Z_0}{h}\right)^{l_1-1} \left(\frac{Z_{j'} - Z_0}{h}\right)^{l_2-1} \right. \\ & \quad \left. \times \delta_{j|(m)} \delta_{j'|(m)} w_{(m)}^{jj'} K_h^{1/2}(Z_j - Z_0) K_h^{1/2}(Z_{j'} - Z_0)\right\}, \end{aligned}$$

where $\delta_{j|(m)}$ denotes the j th element of Δ_i in m th imputed set.

Denoting $r_j = (Z_j - Z_0)/h$, we derive that

$$\begin{aligned} B_M[l_1, l_2] &= \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \int [\delta_{j|(m)}]^2 w_{(m)}^{jj} f_j(Z_0 + r_j h) r_j^{l_1+l_2-2} K(r_j) dr_j + o(h) \\ &= \frac{1}{M} \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \sum_{j=1}^T w_{(m)}^{jj} f_j(Z_0) c_K(l_1 + l_2 - 2) + o(h). \end{aligned}$$

Followed by $E_1(c) = I$, $B_M = \frac{1}{M} \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \sum_{j=1}^T w^{jj} f_j(Z_0) I + o(h)$.

The same idea is applied to calculate $E(A_{NM})$ where the expectation is taken over Y . First, $\lim_{N \rightarrow \infty} A_{NM} = A_M$ where

$$A_M = \frac{1}{M} \sum_{m=1}^M E\{\Theta_s(Z)\Delta_{(m)}K_h^{1/2}(Z)V_{(m)}^{-1}(Z)K_h^{1/2}(Z)[Y - \hat{\mu}_{(m)}^b(Z)]\}.$$

According to the result of Carroll et al. (1998), the l th element of $E(A_M)$ is

3. COVARIATE MEASUREMENT ERROR IN GPLMM

$$\begin{aligned}
E(A_M)[l] &= \frac{1}{M} \sum_{m=1}^M E \left\{ \sum_{j=1}^T \sum_{j'=1}^T \left(\frac{Z_j - Z_0}{h} \right)^{l-1} \delta_{j|(m)} \delta_{j'|(m)} w_{(m)}^{jj'} K_h^{1/2}(Z_j - Z_0) K_h^{1/2}(Z_{j'} - Z_0) \right. \\
&\quad \times \left[\frac{h^2 \theta^{(2)}(Z_0)}{2!} \left(\frac{Z_{j'} - Z_0}{h} \right)^2 + \frac{h^3 \theta^{(3)}(Z_0)}{3!} \left(\frac{Z_{j'} - Z_0}{h} \right)^3 \right] \Big\} + o(h^3) \\
&= \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \int [\delta_{j|(m)}]^2 w_{(m)}^{jj} f_j(Z_0 + r_j h) \\
&\quad \times \left[\frac{h^2 \theta^{(2)}(Z_0)}{2!} r_j^{l+1} + \frac{h^3 \theta^{(3)}(Z_0)}{3!} r_j^{l+2} \right] dr_j + o(h^3) \\
&= \frac{1}{M} \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \\
&\quad \times \left[\frac{h^2 \theta^{(2)}(Z_0)}{2!} \sum_{j=1}^T w_{(m)}^{jj} f_j(Z_0) c_K(l+1) + \frac{h^3 \theta^{(3)}(Z_0)}{3!} \sum_{j=1}^T w_{(m)}^{jj} f_j(Z_0) c_K(l+2) \right] \\
&\quad + o(h^3). \tag{3.8.7}
\end{aligned}$$

It follows that

$$\begin{aligned}
E(A_M) &= \frac{1}{M} \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \\
&\quad \times \left[\frac{h^2 \theta^{(2)}(Z_0)}{2!} \sum_{j=1}^T w_{(m)}^{jj} f_j(Z_0) E_c(2) + \frac{h^3 \theta^{(3)}(Z_0)}{3!} \sum_{j=1}^T w_{(m)}^{jj} f_j(Z_0) E_c(3) \right] + o(h^3).
\end{aligned}$$

By definition, $E_c(2) = (c_K(2), c_K(3))^T = (c_K(2), 0)^T$ and $E_c(3) = (c_K(3), c_K(4))^T = (0, c_K(4))^T$. We also know that $B_M[1, 2] = 0$. Finally we have that

$$bias[\hat{\theta}(z)] = e^T B_M^{-1} E(A_M) \approx \frac{h^2 \theta^{(2)}(z)}{2} c_K(2).$$

Next, recall that $\hat{\theta}(Z; \beta, b) - E[\hat{\theta}(Z; \beta, b)] = e^T B_M^{-1} A_{NM} + o_p(N^{-1/2})$, then the

asymptotic variance of $\hat{\theta}(Z; \beta, b)$ is

$$\text{var}(\hat{\theta}(Z; \beta, b)) = e^T B_M^{-1} \text{cov}(A_{NM}) B_M^{-1} e,$$

where

$$\begin{aligned} \text{cov}(A_{NM}) &= \sum_{m=1}^M E\{\Theta_s(Z) \Delta_{(m)} K_h^{1/2}(Z) V_{(m)}^{-1}(Z) K_h^{1/2}(Z) E\{[Y - \hat{\mu}_{(m)}^b(Z)][Y - \hat{\mu}_{(m)}^b(Z)]^T\} \\ &\quad \times K_h^{1/2}(Z) V_{(m)}^{-1}(Z) K_h^{1/2}(Z) \Delta_{(m)} \Theta_s^T(Z)\} / NM + o\{(Nh)^{-1}\} \\ &= \frac{1}{NM} \sum_{m=1}^M E\{\Theta_s(Z) \Delta_{(m)} K_h^{1/2}(Z) V_{(m)}^{-1}(Z) K_h^{1/2}(Z) \Sigma K_h^{1/2}(Z) \\ &\quad \times K_h^{1/2}(Z) V_{(m)}^{-1}(Z) K_h^{1/2}(Z) \Delta_{(m)} \Theta_s^T(Z)\} + o\{(Nh)^{-1}\}. \end{aligned}$$

Let σ_{jk} be the (j, k) th element of Σ , and we can write the (l_1, l_2) th element of $\text{cov}(A_{NM})$ is

$$\begin{aligned} \text{cov}(A_{NM})[l_1, l_2] &= \frac{1}{NM} \sum_{m=1}^M E\left\{ \sum_{j=1}^T \sum_{j'=1}^T \sum_{s=1}^T \sum_{s'=1}^T \delta_{j|(m)} \delta_{s|(m)} w_{(m)}^{jj'} \sigma_{js} w_{(m)}^{ss'} \right. \\ &\quad \times K_h^{1/2}(Z_j - Z_0) K_h^{1/2}(Z_{j'} - Z_0) K_h^{1/2}(Z_s - Z_0) K_h^{1/2}(Z_{s'} - Z_0) \\ &\quad \times \left(\frac{Z_j - Z_0}{h} \right)^{l_1-1} \left(\frac{Z_{s'} - Z_0}{h} \right)^{l_2-1} \Big\} \\ &= \frac{1}{NMh} \sum_{m=1}^M \sum_{j=1}^T \int [\delta_{j|(m)}]^2 w_{(m)}^{jj} \sigma_{jj} K^2(r_j) f_j(Z_0 + r_j h) \\ &\quad \times r_j^{l_1+l_2-2} dr_j + o\{(Nh)^{-1}\} \\ &= \frac{1}{NMh} \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \sum_{j=1}^T w_{(m)}^{jj} \sigma_{jj} f_j(Z_0) \gamma_K(l_1 + l_2 - 2) \\ &\quad + o\{(Nh)^{-1}\}. \end{aligned}$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

Now given $R_i = I$, $\text{var}(\hat{\theta}(Z; \beta, b))$ becomes

$$\text{var}(\hat{\theta}(Z; \beta, b)) = \frac{M}{Nh} \left\{ \sum_{m=1}^M [\delta_{(m)}(\theta(Z_0))]^2 \sum_{j=1}^T w_{(m)}^{jj} \sigma_{jj} f_j(Z_0) \right\}^{-1} \gamma_K(0).$$

We already have proved that $\hat{\theta}(Z; \hat{\beta}, \hat{b})$ has the same asymptotic distribution as $\hat{\theta}(Z; \beta, b)$, hence Theorem 3.4.1 holds. $\square$

Proof of Theorem 3.4.2.

Similarly, using Taylor expansion for (3.3.7) we have

$$\sqrt{N}(\hat{\beta} - \beta) = D_{NM}^{-1}(\sqrt{N}C_{NM}) + o_p(1) \quad (3.8.8)$$

where

$$\begin{aligned} D_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \tilde{\mathbf{X}}_{(m),i}^{\epsilon T} \Delta_{(m),i} V_{(m),i}^{-1} \Delta_{(m),i} \tilde{\mathbf{X}}_{(m),i}^{\epsilon} \\ C_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \tilde{\mathbf{X}}_{(m),i}^{\epsilon T} \Delta_{(m),i} V_{(m),i}^{-1} [Y_i - \mu^b(\mathbf{X}_{(m),i}^{\epsilon} \beta + \mathbf{B}_i b_i + \hat{\theta}(\mathbf{Z}_i; \beta, b))]. \end{aligned}$$

As $N \rightarrow \infty$, let D_{NM} approaches to $D_M = \frac{1}{M} \sum_{m=1}^M E[\tilde{X}_{(m)}^{\epsilon T} \Delta_{(m)} V_{(m)}^{-1} \Delta_{(m)} \tilde{X}_{(m)}^{\epsilon}]$.

In the imputation procedure, (3.2.2), (3.2.4) and (3.3.2) hold. Further, by defining

$G_{(m),i} = \tilde{\mathbf{X}}_{(m),i}^{\epsilon T} \Delta_{(m),i} V_{(m),i}^{-1}$, we can expand C_{NM} as

$$C_{NM} = C_{1NM} - C_{2NM} - C_{3NM} + o_p(1), \quad (3.8.9)$$

where

$$\begin{aligned}
 C_{1NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} [Y_i - \mu^b(\mathbf{X}_i \beta + \mathbf{B}_i b_i + \theta(\mathbf{Z}_i; \beta, b))] \\
 C_{2NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} [\hat{\theta}(\mathbf{Z}_i; \beta, b) - \theta(\mathbf{Z}_i; \beta, b)] \\
 C_{3NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} [W_i(\hat{\eta}_{(m)}^\epsilon - \hat{\eta})\beta + (U_{(m),i}^\epsilon - U^*)\beta + o_p(1)]
 \end{aligned}$$

Before calculating C_{2NM} and C_{3NM} , we introduce definitions of B'_{NM} and A'_{NM} :

$$\begin{aligned}
 B'_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) \Delta_{(m),i}, \\
 A'_{NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) [Y_i - \mu_{(m)}^b(\mathbf{Z}_i)], \\
 A' A'_{NM} &= -\frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \Delta_{(m),i} K_h^{1/2}(\mathbf{Z}_i) V_{(m)}^{-1}(\mathbf{Z}_i) K_h^{1/2}(\mathbf{Z}_i) \Delta_{(m),i} \tilde{\mathbf{X}}_{(m),i}^\epsilon.
 \end{aligned}$$

Solving (3.3.5) and considering the estimate α_0 , we have that $e^T B_{NM}^{-1} A_{NM} = B_{NM}'^{-1} A'_{NM}$. By Theorem 3.4.1 then we have

$$\begin{aligned}
 C_{2NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T \tilde{X}_{(m),i,j}^\epsilon \delta_{i,j|(m)} w_{(m),i}^{jj'} \delta_{i,j'|(m)} \left\{ B_{NM}'^{-1} A'_{NM} + \frac{h^2 \theta^{(2)}(Z_{ij'})}{2} c_K(2) \right\} \\
 &\quad + o_p(1) \\
 &= \frac{1}{N} \sum_{i'=1}^N \frac{1}{M} \sum_{m'=1}^M \sum_{k'=1}^T \delta_{i',k'|(m)} w_{(m),i'}^{k'k'} \\
 &\quad \times \left\{ \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T \tilde{X}_{(m),i,j}^\epsilon \delta_{i,j|(m)} w_{(m),i}^{jj'} \delta_{i,j'|(m)} B_{NM}'^{-1}(Z_{ij'}) K_h(Z_{ij'} - Z_{i'k'}) \right\} \\
 &\quad \times (Y_{i'k'} - \mu_{(m),i',k'}^b)
 \end{aligned}$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

$$\begin{aligned}
& + \frac{h^2}{2} \left\{ \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T \tilde{X}_{(m),i,j}^\epsilon \delta_{i,j|(m)} w_{(m),i}^{jj'} \delta_{i,j'|(m)} \theta^{(2)}(Z_{ij'}) c_K(2) \right\} + o_p(1) \\
& = \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T H_{(m),i,j}((Y_{ij} - \mu_{(m),i,j}^b) \\
& + \frac{h^2}{2} \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T E \left\{ \tilde{X}_{(m),j}^\epsilon \delta_{j|(m)} w_{(m)}^{jj'} \delta_{j'|(m)} \theta^{(2)}(Z_{ij'}) c_K(2) \right\} + o_p(1) \\
& = \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i}((Y_i - \mu_{(m),i}^b) \\
& + \frac{h^2}{2} \frac{1}{M} \sum_{m=1}^M E \left\{ \tilde{X}_{(m)}^\epsilon \Delta_{(m)} V_{(m)}^{-1} \Delta_{(m)} \theta^{(2)}(Z) c_K(2) \right\} + o_p(1), \tag{3.8.10}
\end{aligned}$$

where following Lin & Carroll (2001), we know that

$$\begin{aligned}
H_{(m),i',k'} & = \delta_{i',k'|(m)} w_{(m),i'}^{k'k'} \\
& \times \left\{ \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T \tilde{X}_{(m),i,j}^\epsilon \delta_{i,j|(m)} w_{(m),i}^{jj'} \delta_{i,j'|(m)} B_{NM}'^{-1}(Z_{ij'}) K_h(Z_{ij'} - Z_{i'k'}) \right\} \\
& = \delta_{i',k'|(m)} w_{(m),i'}^{k'k'} E \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T \tilde{X}_{(m),j}^\epsilon \delta_{j|(m)} w_{(m)}^{jj'} \delta_{j'|(m)} B_{NM}'^{-1}(Z_{j'}) K_h(Z_{j'} - Z_{i'k'}) \right\} \\
& = \delta_{i',k'|(m)} w_{(m),i'}^{k'k'} E \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T \sum_{j'=1}^T E \left\{ \tilde{X}_{(m),j}^\epsilon \delta_{j|(m)} w_{(m)}^{jj'} \delta_{j'|(m)} | Z_{j'} = Z_{i'k'} \right\} \right\} \\
& \times B_{NM}'^{-1}(Z_{i'k'}) f_{k'}(Z_{i'k'}), \\
H_{(m),i} & = (H_{(m),i,1}, \dots, H_{(m),i,T})^T.
\end{aligned}$$

Let $e_{(m)}^\epsilon = \hat{\eta}_{(m)}^\epsilon - \hat{\eta}$, and the similar calculations shows

$$\begin{aligned}
 C_{3NM} &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} [W_i(\hat{\eta}_{(m)}^\epsilon - \hat{\eta})\beta + (U_{(m),i}^\epsilon - U^\star)\beta + o_p(1)] \\
 &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} W_i \beta e_{(m)}^\epsilon \\
 &\quad + \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} (U_{(m),i}^\epsilon - U^\star)\beta + o_p(1).
 \end{aligned}$$

Then,

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i}((Y_i - \mu_{(m),i}^b)) &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i}((Y_i - \mu_i^b)) \\
 &\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i} \Delta_{(m),i} W_i \beta e_{(m)}^\epsilon \\
 &\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i} \Delta_{(m),i} (U_{(m),i}^\epsilon - U^\star)\beta + o_p(1).
 \end{aligned}$$

Therefore, it is not hard to write that

$$\begin{aligned}
 C_{NM} &= C_{1NM} - C_{2NM} - C_{3NM} + o_p(1) \\
 &= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} (Y_i - \mu_i^b) - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i}((Y_i - \mu_i^b)) \\
 &\quad + \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i} \Delta_{(m),i} W_i \beta e_{(m)}^\epsilon \\
 &\quad + \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M H_{(m),i} \Delta_{(m),i} (U_{(m),i}^\epsilon - U^\star)\beta \\
 &\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} W_i \beta e_{(m)}^\epsilon - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} (U_{(m),i}^\epsilon - U^\star)\beta \\
 &\quad - \frac{h^2}{2} \frac{1}{M} \sum_{m=1}^M D_M d_{(m)}(\theta, \beta, b) + o_p(1)
 \end{aligned}$$

3. COVARIATE MEASUREMENT ERROR IN GPLMM

$$\begin{aligned}
&= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i})(Y_i - \mu_i^b) \\
&\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i}) \Delta_{(m),i} W_i \beta e_{(m)}^\epsilon \\
&\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i}) \Delta_{(m),i} (U_{(m)}^\epsilon - U^\star) \beta \\
&\quad - \frac{h^2}{2} \frac{1}{M} \sum_{m=1}^M D_M d_{(m)}(\theta, \beta, b) \\
&= \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i})(Y_i - \mu_i^b) \\
&\quad - \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i}) \Delta_{(m),i} (U_{(m)}^\epsilon - U^\star) \beta \\
&\quad - \frac{1}{M} \sum_{m=1}^M D_M c_{(m)}(e_{(m)}^\epsilon, \beta, b) \\
&\quad - \frac{h^2}{2} \frac{1}{M} \sum_{m=1}^M D_M d_{(m)}(\theta, \beta, b),
\end{aligned}$$

where $D_M c_{(m)}(e_{(m)}^\epsilon, \beta, b) = E \left\{ (G_{(m)} - H_{(m)}) \Delta_{(m)} W \beta e_{(m)}^\epsilon \right\}$ and $D_M d_{(m)}(\theta, \beta, b) = E \left\{ \tilde{X}_{(m)}^\epsilon \Delta_{(m)} V_{(m)}^{-1} \Delta_{(m)} \theta^{(2)}(Z) c_K(2) \right\}$. From (3.8.8) it follows that

$$\begin{aligned}
\sqrt{N}(\hat{\beta} - \beta) &= D_M^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i})(Y_i - \mu_i^b) \\
&\quad - D_M^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M (G_{(m),i} - H_{(m),i}) \Delta_{(m),i} (U_{(m)}^\epsilon - U^\star) \beta \\
&\quad - \frac{\sqrt{N}}{M} \sum_{m=1}^M c_{(m)}(e_{(m)}^\epsilon, \beta, b) - \frac{\sqrt{N} h^2}{2M} \sum_{m=1}^M d_{(m)}(\theta, \beta, b) + o_p(1)
\end{aligned}$$

Under the conditional independence assumption and let $f_j(\cdot) = f(\cdot)$, that is $R_i =$

I , we have

$$\frac{1}{M} \sum_{m=1}^M d_{(m)}(\theta, \beta, b) = D_M^{-1} E \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^T E[\tilde{X}_{(m),j}^\epsilon (\delta_{j|(m)})^2 w_{(m)}^{jj} | Z_j] \theta^{(2)}(Z) c_K(2) \right\}$$

. Recalling the relation between $\tilde{X}_{(m),i,j}^\epsilon$ and $X_{(m),i,j}^\epsilon$, we further have

$$\begin{aligned} \tilde{X}_{(m),i,j}^\epsilon &= X_{(m),i,j}^\epsilon - \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j'=1}^T E[(\delta_{j'|(m)})^2 w_{(m)}^{j'j'} | Z'_j = Z_{ij}] \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j'=1}^T E[(\delta_{j'|(m)})^2 w_{(m)}^{j'j'} X_{(m),j'}^T | Z'_j = Z_{ij}] \right\} \\ \tilde{X}_{(m),i,j}^\epsilon &= X_{(m),i,j}^\epsilon - \left\{ \frac{1}{M} \sum_{m=1}^M E[(\delta_{i,j|(m)})^2 w_{(m),i}^{jj} | Z_{ij}] \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{M} \sum_{m=1}^M E[(\delta_{i,j|(m)})^2 w_{(m),i}^{jj} X_{(m),i,j}^T | Z_{ij}] \right\}. \end{aligned}$$

Thus it is easy to show that

$$E\left[\frac{1}{M} \sum_{m=1}^M \tilde{X}_{(m),j}^\epsilon (\delta_{j|(m)})^2 w_{(m)}^{jj} | Z_j\right] = 0.$$

So

$$\frac{1}{M} \sum_{m=1}^M d_{(m)}(\theta, \beta, b) = 0.$$

Since we know that $H_{(m),i}$ has the same term $E[\frac{1}{M} \sum_{m=1}^M \tilde{X}_{(m),j}^\epsilon (\delta_{j|(m)})^2 w_{(m)}^{jj} | Z_j]$, $H_{(m),i} = 0$. Given that, $D_M c_{(m)}(e_{(m)}^\epsilon, \beta, b)$ becomes $E\left\{G_{(m)} \Delta_{(m)} W \beta e_{(m)}^\epsilon\right\}$ and $\frac{1}{M} \sum_{m=1}^M c_{(m)}(e_{(m)}^\epsilon, \beta, b) = 0$. Consequently, under the conditional independence

3. COVARIATE MEASUREMENT ERROR IN GPLMM

assumption,

$$\begin{aligned}\sqrt{N}(\hat{\beta} - \beta) &= D_M^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} (Y_i - \mu_i^b) \\ &\quad - D_M^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{M} \sum_{m=1}^M G_{(m),i} \Delta_{(m),i} (U_{(m)}^\epsilon - U^\star) \beta + o_p(1).\end{aligned}\tag{3.8.11}$$

Given $U^\star$, each $U_{(m)}^\epsilon$ has variance $(1 - \lambda\eta^2\sigma_x^{\star 2})$, which can be easily estimated from the sample variances of X_m^ϵ and W . The consistency can be directly implied by (3.8.11). By the central limit theorem, the asymptotic distribution of $\hat{\beta}$ follows immediately. $\square$

4 Conclusions and future work

4.1 Concluding remarks

In this thesis, two kinds of statistical models for longitudinal data are discussed and extended to cover more practical cases.

In Chapter 2, a much more efficient marginal modeling approach is proposed and combined with regularization technique to handle the high-dimension data in the sense that the number of regression covariates q is relatively large. In the numerical study, we compared regularized QIF with proposed quadratic penalty, with standard penalized GEE. The result shows that it performs competently in terms of variable selection. In both continuous and discrete cases, our quadratic SCAD penalty is able to select the true non-zero regression covariates, with regularized QIF. Being different from the moment technique used in the GEE, another important finding in the parameter estimation is that the correlation parameter ρ do always exist, solved by minimizing an optimal least squares problem.

The limitation of our proposed new QIF approach for variable selection in longitudinal data can be found in the simulation study. Due to its complex structure in which the regularization term is introduced, it does not outperform the standard penalized GEE with simpler form, which has no singularity problem in their variance structure. However, as our primary goal of study is to efficiently estimate the coefficients of interest and select the most important variables si-

4. CONCLUSIONS AND FUTURE WORK

multaneously, it is worth proposing a variable selection method in the spirit of quadratic inference function. Besides, we suggest to apply paralleling computation method to decrease the computation time spending on the iterative algorithm introduced in this chapter, since the major case we consider is that the number of covariates is comparable to the large sample size.

In Chapter 3, we fit the GPLMM to longitudinal data and implement imputation algorithm to correct underlying measurement error in covariates. The partially linear predictor gives more efficient estimator and better model interpretation by incorporating nonlinear function of some subset of covariates. The computation complexity exponentially increases as we must approximate this unknown function in advance. One practical way to shorten the estimation process, is to transform the generalized linear model to nonlinear least squares problem. It can obtain approximately unbiased estimator of fixed effect covariates, although such action may not yield exactly asymptotic inference. In PANS example, the self-reported data may have high probability on recording wrong continuous observations such as time on weekly exercises so the bias correction step must be conducted before continuing the standard statistical analysis. The analytical results of both simulation and example studies show the necessity to deal with error-in-prone variables and coincide with our derived asymptotic properties.

However, the proposed modeling fitting procedure needs a heavy computation intensity, which makes it hard to implement when the sample size may become very large in practice. That is because the multiple imputation model requires a sufficient long burn-in period to ensure the convergence and select no less than 3 imputed data sets for the further steps. Although in data analysis, a powerful *R* package (“Parallel”) is used accordingly for parallel computing, the iterative

algorithm remains relatively slow.

4.2 Future work

Beyond the scope of this thesis, we are interested in further exploring some other aspects in variable selection and hypothesis testing.

In this thesis, the variable selection procedure takes its effects on the linear components in the regression predictor only. If the predictor is partially combined with some unknown function, like the way we specify in Chapter 3, what we concern about thereafter is how to select which subset of covariates are linear or nonlinear. The excellent work of Zhang et al. (2011) provided a structure selection procedure called LAND (linear and nonlinear discoverer) to distinguish the nonlinear term from other linear components. The extension of their selection approach to mixed effect model is feasible. An major advantage of QIF is that the inference function has an explicit asymptotic form, which allows us to test whether coefficients are significant in the model or not. Further, it also enables us to conduct goodness-of-fit tests for checking model assumptions. In our future research, we would like to construct a hypothesis test for the regularized QIF model when the number of covariates is relatively large to the sample size. Qu & Li (2006) applied QIF approach for the varying-coefficients model in the longitudinal data, in which a subset of variables is associated with the time variable. It can be considered as GPLM where the varying coefficients are treated as unknown nonlinear function of time variable. Therefore, another interesting topic in future is to introduce unknown nonlinear term in the model linear predictor in QIF. The hypothesis testing can also be conducted but probably very complicated.

4. CONCLUSIONS AND FUTURE WORK

Our research on measurement error in covariates assumes constant variance for the measurement error. But often such variance varies as the underlying true value of covariates changes, as discussed in the work of Guo & Little (2011), who proposed a MI approach to cope with the problem that covariates have heteroscedastic measurement error. We can extend our measurement error model in the same way to cover more cases. Then the hypothesis testing for heteroscedastic error can be accordingly developed.

In the GLMM context, the hypothesis testing can be directly made as the likelihood-based inference is given. However, asymptotic inference therein is greatly affected by the specified distribution of random effects, so the likelihood ratio test may fail when the random effect is misspecified. As a result, the issue of test for random effect have drawn many attention recently. Huang (2009, 2010) diagnosed the misspecification on random effect by constructing coarsened data from the initial data set. For the linear mixed model, Wu & Li (2012) in their unpublished work proposed a moment-based test statistic to examine whether the covariates related to random effects is significant or not. We can learn and apply their excellent methods for our GPLMM as the only difference is the nonlinear term in the model predictor.

References

- [1] ALON, U., BARKAI, N., NOTTERMAN, D. A., GISH, K., YBARRA, S., MACK, D., AND LEVINE, A. J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences* 96, 12 (1999), 6745–6750.
- [2] BAI, Y., FUNG, W. K., AND ZHU, Z. Y. Penalized quadratic inference functions for single-index models with longitudinal data. *Journal of Multivariate Analysis* 100, 1 (2009), 152–361.
- [3] BRESLOW, N. Extra-poisson variation in log-linear models. *Journal of Applied Statistics* 33, 1 (1984), 38–44.
- [4] BRESLOW, N. E., AND CLAYTON, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 421 (1993), 9–25.
- [5] BURKE, L., LEE, A. H., PASALICH, M., JANCEY, J., KERR, D., AND HOWAT, P. Effects of a physical activity and nutrition program for seniors on body mass index and waist-to-hip ratio: A randomised controlled trial. *Preventive Medicine* 54, 6 (2012), 397–401.
- [6] CARROLL, R. J., FAN, J. Q., GIJBELS, I., AND WAND, M. P. Generalized partially linear single-index models. *Journal of the American Statistical Association* 92, 438 (1997), 477–489.

REFERENCES

- [7] CARROLL, R. J., RUPPERT, D., AND STEFANSKI, L. A. *Measurement Error in Nonlinear Models*. Chapman & Hall, London, 1995.
- [8] CHEN, H. H., AND WANG, Y. J. A penalized spline approach to functional mixed effects model analysis. *Biometrics* 67, 3 (2011), 861–870.
- [9] CHIARETTI, S., LI, X. C., GENTLEMAN, R., VITALE, A., VIGNETTI, M., MANDELLI, F., RITZ, J., AND FOA, R. Gene expression profile of adult t-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood* 103, 7 (2004), 2771–2778.
- [10] CHIOU, J. M., AND MULLER, H. G. Estimated estimating equations: semi-parametric inference for clustered and longitudinal data. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 67, 4 (2005), 531–553.
- [11] COLE, S. R., CHU, H., AND GREENLAND, S. Multiple-imputation for measurement-error correction. *Int J Epidemiol* 35, 4 (2006), 1074–81.
- [12] CONSUL, P. C., AND JAIN, G. C. A generalization of the poisson distribution. *Technometrics* 15, 4 (1973), 791–799.
- [13] COOK, J. R., AND STEFANSKI, L. A. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89, 428 (1994), 1314–1328.
- [14] COTTON, S. M., CREWTER, D. P., AND CREWTER, S. G. Measurement error: implications for diagnosis and discrepancy models of developmental dyslexia. *Dyslexia* 11, 3 (2005), 186–202.
- [15] COX, D. R. Some remarks on overdispersion. *Biometrika* 70, 1 (1983), 269–274.

REFERENCES

- [16] CROWDER, M. On the use of a working correlation matrix in using generalized linear-models for repeated-measures. *Biometrika* 82, 2 (1995), 407–410.
- [17] DIGGLE, P. J., HEAGERTY, P., LIANG, K. Y., AND ZEGER, S. L. *Analysis of longitudinal data*. Oxford University Press, New York, 2002.
- [18] EFRON, B. Double exponential-families and their use in generalized linear-regression. *Journal of the American Statistical Association* 81, 395 (1986), 709–721.
- [19] FAHRMEIR, L., AND KAUFMANN, H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* 13, 1 (1985), 342–368.
- [20] FAN, J. Q., AND LI, R. Z. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 456 (2001), 1348–1360.
- [21] FAN, J. Q., AND LV, J. C. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 70 (2008), 849–883.
- [22] FAN, J. Q., AND PENG, H. Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics* 32, 3 (2004), 928–961.
- [23] FITZMAURICE, G. M., LAIRD, N. M., AND WARE, J. H. *Applied Longitudinal Analysis*. John Wiley & Sons, New York, 2004.
- [24] FLEISS, J. L. *Statistical methods for rates and proportions*. J. Wiley & Sons, New York, 1981.

REFERENCES

- [25] GELFAND, A. E., AND DALAL, S. R. A note on overdispersed exponential families. *Biometrika* 77, 1 (1990), 55–64.
- [26] GELMAN, A., AND RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 4 (1992), 457–511.
- [27] GOLUB, T. R., SLONIM, D. K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J. P., COLLIER, H., LOH, M. L., DOWNING, J. R., CALIGIURI, M. A., BLOOMFIELD, C. D., AND LANDER, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 5439 (1999), 531–537.
- [28] GUO, Y., AND LITTLE, R. J. Regression analysis with covariates that have heteroscedastic measurement error. *Stat Med* 30, 18 (2011), 2278–2294.
- [29] HANSEN, L. P. Large sample properties of generalized-method of moments estimators. *Econometrica* 50, 4 (1982), 1029–1054.
- [30] HARDIN, J. W., AND HILBE, J. M. *Generalized estimating equations*. Chapman & Hall, London, 2002.
- [31] HUANG, X. Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. *Biometrics* 65, 2 (2009), 361–368.
- [32] HUANG, X. Detecting random-effects model misspecification via coarsened data. *Computational Statistics and Data Analysis* 55, 1 (2011), 703–714.
- [33] HUNTER, D. R., AND LI, R. Variable selection using mm algorithms. *Ann Stat* 33, 4 (2005), 1617–1642.
- [34] JOWAHEER, V., AND SUTRADHAR, B. C. Analysing longitudinal count data with overdispersion. *Biometrika* 89, 2 (2002), 389–399.

REFERENCES

- [35] KANNEL, W. B., NEATON, J. D., WENTWORTH, D., THOMAS, H. E., STAMLER, J., HULLEY, S. B., AND KJELSBERG, M. O. Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for mrfit. *American Heart Journal* 112, 4 (1986), 825–836.
- [36] KIPNIS, V., SUBAR, A. F., MIDTHUNE, D., AND FREEDMAN, L. S. Structure of dietary measurement error: results of the open biomarker study. *American Journal of Epidemiology* 158, 1 (2003), 14–21.
- [37] LAIRD, N., LANGE, N., AND STRAM, D. Maximum-likelihood computations with repeated measures - application of the em-algorithm. *Journal of the American Statistical Association* 82, 397 (1987), 97–105.
- [38] LAIRD, N. M., AND WARE, J. H. Random-effects models for longitudinal data. *Biometrics* 38, 4 (1982), 963–974.
- [39] LAWLESS, J. F. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics* 15, 3 (1987), 209–225.
- [40] LENG, C. L., ZHANG, W. P., AND PAN, J. X. Semiparametric mean covariance regression analysis for longitudinal data. *Journal of the American Statistical Association* 105, 489 (2010), 181–193.
- [41] LEUNG, H. Y., WANG, Y. G., AND ZHU, M. Efficient parameter estimation in longitudinal data analysis using hybrid gee method. *Biostatistics* 10, 3 (2009), 436–445.
- [42] LIANG, H. Generalized partially linear mixed-effects models incorporating mis-measured covariates. *Annals of the Institute of Statistical Mathematics* 61, 1 (2009), 27–46.

REFERENCES

- [43] LIANG, H., AND REN, H. B. Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics* 14, 1 (2005), 237–250.
- [44] LIANG, K. Y., AND ZEGER, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 1 (1986), 13–22.
- [45] LIN, X. H., AND CARROLL, R. J. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* 95, 450 (2000), 520–534.
- [46] LIN, X. H., AND CARROLL, R. J. Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* 96, 455 (2001), 1045–1056.
- [47] LINDSTROM, M. J., AND BATES, D. M. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* 83, 404 (1988), 1014–1022.
- [48] LINDSTROM, M. J., AND BATES, D. M. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46, 3 (1990), 673–687.
- [49] MA, Y. Y., AND CARROLL, R. J. Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association* 101, 476 (2006), 1465–1474.
- [50] MCCULLOCH, C. E. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 92, 437 (1997), 162–170.
- [51] MCCULLOCH, C. E., SEARLE, S. R., AND NEUHAUS, J. M. *Generalized, linear, and mixed models*. John Wiley & Sons, New York, 2001.

REFERENCES

- [52] MCKENZIE, E. Some arma models for dependent sequences of poisson counts. *Advances in Applied Probability* 20, 4 (1988), 822–835.
- [53] POURAHMADI, M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* 86, 3 (1999), 677–690.
- [54] PRENTICE, R. L., AND ZHAO, L. P. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 47, 3 (1991), 825–839.
- [55] QU, A., AND LI, R. Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics* 62, 2 (2006), 379–391.
- [56] QU, A., AND LINDSAY, B. G. Building adaptive estimating equations when inverse of covariance estimation is difficult. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 65 (2003), 127–142.
- [57] QU, A., LINDSAY, B. G., AND LI, B. Improving generalised estimating equations using quadratic inference functions. *Biometrika* 87, 4 (2000), 823–836.
- [58] ROBINS, J. M., AND WANG, N. S. Inference for imputation estimators. *Biometrika* 87, 1 (2000), 113–124.
- [59] RUBIN, D. B. Inference and missing data. *Biometrika* 63, 3 (1976), 581–590.
- [60] RUPPERT, D. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association* 92, 439 (1997), 1049–1062.
- [61] SCHAFER, D. W. Covariate measurement error in generalized linear-models. *Biometrika* 74, 2 (1987), 385–391.

REFERENCES

- [62] SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D., AND FUTCHER, B. Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of Cell* 9, 12 (1998), 3273–3297.
- [63] STEFANSKI, L. A., AND COOK, J. R. Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association* 90, 432 (1995), 1247–1256.
- [64] STEUTEL, F. W., VERVAAT, W., AND WOLFE, S. J. Integer-valued branching processes with immigration. *Advances in Applied Probability* 15, 4 (1983), 713–725.
- [65] SUTRADHAR, B. C., AND KUMAR, P. The inversion of correlation matrix for $ma(1)$ process. *Applied Mathematics Letters* 16, 3 (2003), 317–321.
- [66] THOMAS, L., STEFANSKI, L., AND DAVIDIAN, M. A moment-adjusted imputation method for measurement error models. *Biometrics* 67, 4 (2011), 1461–70.
- [67] TSIATIS, A. A., DEGRUTTOLA, V., AND WULFSOHN, M. S. Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and cd4 counts in patients with aids. *Journal of American Statistics Association* 90, 429 (1995), 27–37.
- [68] WANG, L. Gee analysis of clustered binary data with diverging number of covariates. *Annals of Statistics* 39, 1 (2011), 389–417.
- [69] WANG, L., ZHOU, J. H., AND QU, A. N. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* 68, 2 (2012), 353–360.

REFERENCES

- [70] WANG, N., AND ROBINS, J. M. Large-sample theory for parametric multiple imputation procedures. *Biometrika* 85, 4 (1998), 935–948.
- [71] WANG, N. Y., LIN, X. H., GUTIERREZ, R. G., AND CARROLL, R. J. Bias analysis and simex approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association* 93, 441 (1998), 249–261.
- [72] WANG, S., QIAN, L., AND CARROLL, R. J. Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika* 97, 1 (2010), 79–93.
- [73] WANNEMUEHLER, K. A., LYLES, R. H., WALLER, L. A., HOEKSTRA, R. M., KLEIN, M., AND TOLBERT, P. A. A conditional expectation approach for associating ambient air pollutant exposures with health outcomes. *Environmetrics* 20, 7 (2009), 877–894.
- [74] WARTON, D. I. Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association* 103, 481 (2008), 340–349.
- [75] WARTON, D. I. Regularized sandwich estimators for analysis of high-dimensional data using generalized estimating equations. *Biometrics* 67, 1 (2011), 116–123.
- [76] WEDDERBURN, R. W. M. Quasi-likelihood functions, generalized linear-models, and gauss-newton method. *Biometrika* 61, 3 (1974), 439–447.
- [77] WOLFINGER, R. D., GIBSON, G., WOLFINGER, E. D., BENNETT, L., HAMADEH, H., BUSHEL, P., AFSHARI, C., AND PAULES, R. S. Assessing gene significance from cdna microarray expression data via mixed models. *Journal of Computational Biology* 8, 6 (2001), 625–637.

REFERENCES

- [78] WU, J. H., AND LI, G. D. Moment-based tests for random effects in panel data models. *Transcript* (2012).
- [79] WU, L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to aids studies. *Journal of the American Statistical Association* 97, 460 (2002), 955–964.
- [80] XIE, M. G., AND YANG, Y. N. Asymptotics for generalized estimating equations with large cluster sizes. *Annals of Statistics* 31, 1 (2003), 310–347.
- [81] YE, H. J., AND PAN, J. X. Modelling of covariance structures in generalized estimating equations for longitudinal data. *Biometrika* 93, 4 (2006), 927–941.
- [82] YUCEL, R. M., AND ZASLAVSKY, A. M. Imputation of binary treatment variables with measurement error in administrative data. *Journal of the American Statistical Association* 100, 472 (2005), 1123–1132.
- [83] ZHANG, D. W. Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics* 60, 1 (2004), 8–15.
- [84] ZHANG, H. H., CHENG, G., AND LIU, Y. F. Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association* 106, 495 (2011), 1099–1112.
- [85] ZHOU, Y., AND LIANG, H. Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *Ann Stat* 37, 1 (2009), 427–458.