ABSTRACT

LI, ERNING. Estimation for Generalized Linear Models When Covariates Are Subject-specific Parameters in a Mixed Model for Longitudinal Measurements. (Under the direction of Marie Davidian and Daowen Zhang.)

In many studies, a primary endpoint and longitudinal measures of a continuous response are collected for each participant along with other covariates, and the association between the primary endpoint and features of the longitudinal profiles is of interest. One challenge is that the features of the longitudinal profiles are observed only through the longitudinal measurements, which are subject to measurement error and other variation.

A relevant framework assumes that the longitudinal data follow a linear mixed model whose random effects are covariates in a generalized linear model for the primary endpoint. Naive implementation by imputing subject-specific effects from individual regression fits yields biased inference, and several methods for reducing this bias have been proposed, which include regression calibration, pseudo-expected estimating equations, and refined regression calibration (for logistic primary model only). However, these methods require a parametric (normality) assumption on the random effects, which may be unrealistic.

Adapting a strategy of Stefanski and Carroll (1987 Biometrika 74:703-716), we propose a conditional estimation approach in which estimators for the generalized linear model parameters require no assumptions on the random effects and yield consistent inference regardless of the true distribution of random effects. This approach is straightforward and fast to implement. However, this approach can not give insight
into the features of the random effects and inference on the longitudinal data process
is not available as it “conditions away” the random effects. Furthermore, conditional
estimation approach might be less efficient relative to full likelihood approach.

In order to estimate the distribution of random effects and to improve efficiency,
we further propose a semiparametric full likelihood approach which approximates
the random effects distribution by the seminonparametric density representation of
Gallant and Nychka (1987 Econometrica 55:363-390). This approach requires only
the very mild assumption that the random effects have a “smooth” but unspecified
density. When the primary endpoint is normally distributed given the random effects,
implementation of this approach is straightforward via optimization techniques. EM
algorithm is used for implementation of this approach when the primary endpoint
does not follow a normal model given the random effects and it involves increased
computational burden.

The performance of both approaches is demonstrated via simulation. Results show
that, in contrast to methods predicated on a parametric (normality) assumption for
the random effects, the approaches yield valid inferences under departures from this
assumption and are competitive when the assumption holds. The semiparametric full
likelihood approach shows the potential of efficiency gains over other methods. We
also illustrate the performance of the approaches by application to a study of bone
mineral density and longitudinal progesterone levels in peri-menopausal women. Data
analysis results obtained from the approaches offer the analyst assurance of credible
estimation of the relationship.
ESTIMATION FOR GENERALIZED LINEAR MODELS WHEN COVARIATES ARE SUBJECT-SPECIFIC PARAMETERS IN A MIXED MODEL FOR LONGITUDINAL MEASUREMENTS

by

ERNING LI

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial satisfaction of the requirements for the Degree of Doctor of Philosophy

STATISTICS

Raleigh

2004

Approved By:

Professor Marie Davidian
Chair of Advisory Committee

Professor Daowen Zhang
Co-Chair of Advisory Committee

Professor Anastasios A. Tsiatis

Professor Leonard A. Stefanski

Professor Kevin Gross
To My Family
Biography

LI, ERNING was born and raised in Wuhan, China. She earned her B.S. and Ph.D. in Electrical Engineering from Huazhong University of Science and Technology in China in 1992 and 1998, respectively. She received her M.Phil. in Electrical Engineering from Hong Kong Polytechnic University in 1998. She worked as an assistant professor in the Department of Electrical Engineering at Huazhong University of Science and Technology for one year before entering the graduate program in the Department of Statistics at North Carolina State University in 2000. She received her M.S. in Statistics in 2001 and continued to pursue her Ph.D. in Statistics at North Carolina State University. After graduation, she is joining the Department of Statistics at Texas A&M University as an assistant professor in fall of 2004.
Acknowledgement

I can not thank my advisors, Dr. Marie Davidian and Dr. Daowen Zhang, enough for their priceless guidance, encouragement and support throughout my dissertation research. It truly has been a pleasure and honor to work with them. I would also like to offer heartfelt thanks to my other advisory committee members, Dr. Anastasios A. Tsiatis, Dr. Leonard A. Stefanski, and Dr. Kevin Gross, for their support, valuable comments, and review of this work. I especially admire all my advisory committee members’ intuition, wisdom, endless knowledge, and enthusiasm.

It has been a pleasant experience to study in this department. I would like to express my appreciation to all faculty, staff, and fellow graduate students for their help.

Certainly, I owe a debt of gratitude to my family, especially my husband. Their love is always essential to my success.
Contents

List of Figures vii
List of Tables viii

1 Introduction and Literature Review 1
   1.1 Introduction ...................................... 1
   1.2 Joint Model ........................................ 3
   1.3 Brief Review of Literature ........................ 5
   1.4 Proposed Methods .................................. 7

2 Conditional Estimation Approach 10
   2.1 Introduction ....................................... 10
   2.2 Methods ............................................. 11
      2.2.1 Sufficiency Score .............................. 11
      2.2.2 Conditional Estimator .......................... 13
      2.2.3 Inference ....................................... 15
   2.3 Special Cases ....................................... 15
      2.3.1 Normal Primary Model ......................... 16
      2.3.2 Logistic Primary Model ....................... 17
   2.4 Technical Details .................................. 18
      2.4.1 Updating Scheme ............................... 18
      2.4.2 Empirical Sandwich Estimator ................. 19

3 Empirical Results For Conditional Estimation Approach 21
   3.1 Introduction ....................................... 21
   3.2 Simulation Studies .................................. 22
   3.3 Application to SWAN ............................... 24
   3.4 Summary ............................................ 26

4 Semiparametric Likelihood Approach 30
   4.1 Introduction ....................................... 30
4.2 SNP Density Approximation ........................................... 32
4.3 Estimation ............................................................... 34
  4.3.1 Likelihood Function ........................................... 34
  4.3.2 Normal Primary Model .......................................... 35
  4.3.3 EM Algorithm for Non-normal Primary Model ............... 36
4.4 Selection of the Tuning Parameter $K$ .......................... 41
4.5 Technical Details ...................................................... 42
  4.5.1 Computation of $E\{s(X_i)|Y_i, W_i; \hat{\Theta}^{(k)}\}$ for Logistic Primary Model 42
  4.5.2 Computation of Empirical Bayes Estimates .................. 44
  4.5.3 Computation of Observed-data Log-likelihood ............. 44
  4.5.4 Computation of the Observed Information .................. 45

5 Empirical Results For Semiparametric Likelihood Approach 46
  5.1 Introduction ......................................................... 46
  5.2 Simulation Studies .................................................. 47
    5.2.1 Simulation Studies Based on SWAN Design ............... 47
    5.2.2 Simulation Studies Under Generic Scenarios ............. 49
  5.3 Application to SWAN ............................................... 53
  5.4 Summary ........................................................... 56

6 Discussion 68

Bibliography 73
List of Figures

1.1  a. Natural logarithm of progesterone levels measured in urine (PDG) for four randomly selected women plotted against days in standardized menstrual cycle. Individual ordinary least squares fits of the piecewise linear model are superimposed. b. Estimated joint density of woman-specific intercepts (level up to day 14) and slopes (rate of change for days 14–21 and negative rate of change for days 21–28) in a linear mixed model for log PDG using the approach of Zhang and Davidian (2001).

5.1 Estimated densities for the bimodal random effects simulation. a. True density of $X_i$. b. Monte Carlo average of estimated densities of $X_i$ preferred by HQ. c and d. True marginal densities (solid line), Monte Carlo average of estimated marginal densities preferred by AIC (long-dashed line), HQ (dashed line) and BIC (dotted line). e and f. Estimated marginal densities preferred by HQ for 200 data sets.

5.2 Fit to the SWAN data using SNP approach. a. Estimated density of $X_i$. b. Contour plot of the density in a with empirical Bayes estimates superimposed (Contours are 10, 25, 50, 75, 90, and 95%). c and d. Corresponding estimated marginal densities for $K = 0$ (dotted line), $K = 1$ (dashed line), and $K = 2$ (solid line) with the histogram of empirical Bayes estimates of the intercept and slope superimposed, respectively.

5.3 Empirical Bayes estimates of the slope plotted against observed covariates in the SWAN data. a. Body mass index. b. Age. c. Ethnicity.
List of Tables

3.1 Simulation results for four underlying random effect distributions for $P(Y_i = 1|X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1}$, with true values $\beta_0 = -2.5$, $\beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T$, $\sigma^2_u = 0.5$, $n = 500$, RB, relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score. .............................................................. 28

3.2 Parameter estimates for the SWAN data analysis via several methods; abbreviations for methods are as in Table 3.1. Estimated standard errors are in parentheses below each estimate. Standard errors for estimates of $\sigma^2_u$ are multiplied by 10. For the naive method, $\sigma^2_u$ was estimated by mean square of pooled individual least squares residuals. 29

5.1 Simulation results for SWAN design. True parameter values are in parentheses below the parameters. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score; SNP, proposed semiparametric likelihood method, where $K = 0$ denotes normal random effects, and HQ denotes the estimates preferred by HQ. MEAN, Monte Carlo average; RB, relative bias (%); MSE, mean square error; CP, Monte Carlo coverage probability of 95% Wald confidence interval. .......................... 58
5.2 Simulation results for normal and bimodal underlying random effect distributions for $P(Y_i = 1 | X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1}$, with true values $\beta_0 = -2.5$, $\beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T$, $\sigma_u^2 = 0.5$, $n = 500$. RB, relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval; RE, relative efficiency, i.e., Monte Carlo mean square errors for CS divided by that for the indicated estimator. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score; SNP, proposed semiparametric method, where $K = 0$ denotes normal random effects, and AIC, HQ, and BIC are the estimates preferred by AIC, HQ, and BIC, respectively.

5.3 Simulation results for skewed and heavy-tailed underlying random effect distributions for $P(Y_i = 1 | X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1}$, with true values $\beta_0 = -2.5$, $\beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T$, $\sigma_u^2 = 0.5$, $n = 500$; abbreviations for estimates and methods are as in Table 5.2.

5.4 Percentage of simulated data sets with $K = 0, 1, 2$ preferred by each information criterion.

5.5 Parameter estimates and $p$-values for the SWAN data analysis via proposed semiparametric likelihood (SNP) method. Estimated standard errors are in parentheses below each estimate. Standard errors for estimates of $\sigma_u^2$ are multiplied by 10.

5.6 Estimates and standard errors of the parameters in the SNP representation for the SWAN data analysis via proposed semiparametric likelihood method.

5.7 Model selection criteria for $K = 0, 1, 2$ for the SWAN data analysis via proposed semiparametric likelihood method. Smaller values are preferred.
Chapter 1

Introduction and Literature Review

1.1 Introduction

In longitudinal studies, a primary endpoint and longitudinal measurements of a continuous response are often collected for each participant, along with other covariate information. Elucidating the relationship between the primary endpoint and features of the longitudinal profiles may be the main objective. The primary endpoint can be continuous or discrete, which is usually placed in the framework of the generalized linear model. Knowledge of the true features of the longitudinal profiles is needed for implementation. However, one challenge is that the features of the longitudinal profiles can not be observed directly, but they are observed only through the longitudinal measurements, and it is well known that longitudinal data are collected intermittently and are subject to measurement error and other variation. A further objective may be to characterize the longitudinal data process and estimate
the features of the longitudinal profiles.

For example, in the Study of Women’s Health Across the Nation (SWAN, Sowers et al., 2003), investigators wished to understand the association between osteopenia, characterized by bone mineral density (BMD) at or below the 33rd percentile, and features of hormonal patterns over the menstrual cycle in women who were transitioning to menopause. For one cycle, longitudinal progesterone levels derived from urine (PDG) and a single total hip BMD measurement were obtained from 624 women. Because cycle lengths vary among women, it is common (e.g., Zhang et al., 1998) to standardize lengths to a reference of 28 days. Figure 1.1a shows the log PDG level plotted against days in the standardized menstrual cycle for four randomly chosen women in SWAN, in which profiles of log PDG versus standardized cycle time are well approximated by a straight line for the first 14 days, with an approximately symmetric rise and fall over days 14 through 28. Assessing whether log PDG level over the initial 14 days or the later rate of rise and fall are associated with osteopenia would provide insight on hormonal mechanisms that could guide interventions to maintain bone mass and reduce risk of osteoporotic fracture in these women. However, these features are observed only through the log PDG measurements, which are subject to assay error and other variation. Study on the underlying distribution of the features of log PDG, which could offer the potential for guiding subject-matter information in these women, may also be a focus. This example is used in the sequel to illustrate the data analytic objectives and to demonstrate the performance of proposed methods.
1.2 Joint Model

A general framework to model the relationship between the primary endpoint and the features of the longitudinal profiles is to postulate a joint model for the primary endpoint and longitudinal data in which both incorporate the underlying subject-specific random effects. Specifically, the longitudinal data are assumed to follow a linear mixed model whose random effects (characterizing individual longitudinal inherent trajectories) are covariates in a generalized linear model for the primary endpoint.

A popular model for the longitudinal data is a linear mixed model (e.g., Laird and Ware, 1982). Longitudinal measurements of a continuous response for subject $i$, $W_i = (W_{i1}, \ldots, W_{im_i})^T$ at times $t_{i1}, \ldots, t_{im_i}$, are assumed to follow the linear mixed model

\[
W_i = D_i X_i + U_i, \tag{1.1}
\]

where $D_i$ is a full-rank $(m_i \times q)$ design matrix for subject $i$; and $U_i = (U_{i1}, \ldots, U_{im_i})^T$ are within-subject errors reflecting uncertainty in measuring $W_i$, independently and identically distributed with mean zero and variance $\sigma_u^2$, i.e., $U_i \sim \mathcal{N}(0, \sigma_u^2 I_{m_i})$ for $(k \times k)$ identity matrix $I_k$. The $X_i$ are $(q \times 1)$ random effects representing unobserved subject-specific features of the longitudinal profiles. For example, $D_i$ with rows $(1, t_{ij})$, $j = 1, \ldots, m_i$, and $X_i = (X_{1i}, X_{2i})^T$ yields a linear random coefficient model for the longitudinal data with subject-specific intercept $X_{1i}$ and slope $X_{2i}$. $U_i$ are independent of $X_i$. $\sigma_u^2$ is a parameter to be estimated.

A relevant framework for the primary endpoint is a generalized linear model (McCullagh and Nelder, 1989) with random effects as covariates. The class of generalized
linear models may be considered as extending the usual classical linear model to handle special features of different kinds of data. Regression models based on many distributions such as normal, Poisson, Binomial, Inverse Gaussian, and Gamma, etc. in a general class may be considered within the common framework of generalized linear models. Let $Y_i$ denote the observed primary response for subject $i, i = 1, \ldots, n$, and $C_i$ denote a $p$-vector of explanatory variables that can be observed. The primary endpoint and longitudinal profile features are assumed to be related via a generalized linear model in canonical form; i.e., the conditional density of $Y_i$ given $X_i$ (and $C_i$; conditioning on $C_i$ is suppressed throughout) is

$$f(Y_i|X_i; \beta, \rho) = \exp \left\{ \frac{Y_i(\beta_0^T C_i + \beta_1^T X_i) - b(\beta_0^T C_i + \beta_1^T X_i)}{a(\rho)} + c(Y_i, \rho) \right\}, \quad (1.2)$$

where $\beta = (\beta_0^T, \beta_1^T)^T$ are regression parameters, $\rho$ is a dispersion parameter, and $a(\cdot), b(\cdot)$, and $c(\cdot, \cdot)$ are known functions. $\beta$ and $\rho$ are parameters of interest.

All variables are independent across $i$. We assume that $Y_i$ and $W_i$ are conditionally independent given $X_i$ (e.g., Carroll et al, 1995, sec. 1.6). It is often reasonable to assume the within-subject errors $U_i$ are normally distributed, perhaps on a transformed scale.

For SWAN, a piecewise linear model for log PDG with random effects representing baseline-to-14 days level and rate of symmetric rise/fall for days 14–28 captures the salient features of woman-specific profiles, and, for the primary endpoint, presence/absence of osteopenia, a logistic model incorporating these random effects and other covariates is a natural choice.
1.3 Brief Review of Literature

When all of the covariates in a generalized linear model are observed directly, maximum likelihood and iteratively reweighted least squares techniques can be implemented to make inference on parameters of interest. However, in the joint model described in Section 1.2, the random effects which are covariates in the generalized linear model can not be observed directly. Some approaches to inference in this joint model have been proposed.

Implementation of this joint model which substitutes subject-specific ordinary least squares estimates of the random effects in the primary generalized linear model is regarded as naive approach (N). This method can lead to biased and misleading inference on parameters of the generalized linear model as discussed by Wang, Wang, and Wang (2000).

Viewing this as a measurement error problem, Wang et al. (2000) considered regression calibration (RC; Carroll, Ruppert, and Stefanski, 1995, Chapter 3), where the random effects are replaced in the primary model by “estimated” best linear unbiased predictors from the fit of the mixed model, which reduces but does not eliminate bias.

Wang et al. (2000) also proposed a pseudo-expected estimating equation (EEE) approach based on the average of conditional expectations of the primary-model score given the observed data, which requires numerical integration to compute the conditional expectations. To circumvent this integration for logistic regression, they developed an approximate pseudo-EEE that may be viewed as “refined” RC (RR).

These procedures depend on the usual normality assumptions on random effects
and intra-subject errors. In the joint model described in Section 1.2, interest may focus on estimating $\theta = (\beta_0^T, \beta_1^T, \rho, \sigma_u^2)^T$. Although normality of the within-subject errors $U_i$ may be often reasonable, perhaps on a transformed scale, the usual assumption of normal $X_i$ may be a poor representation of the population variation in longitudinal profiles and obscure important features of among-individual variation. Several authors (e.g., Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001) have shown that its violation can compromise inference, raising this concern for the joint model over the validity of inferences both on fixed and random effects if the normality assumption on the random effects is violated. Consequences for inference on parameters of interest if methods depending on this assumption are used when it is violated are not known.

For SWAN, we fitted the above mixed model for log PDG with pre-14-day “intercept” and post-14-day “rise/fall slope” random effects by the approach of Zhang and Davidian (2001), which assumes the random effects have a “smooth” but unspecified density that is estimated along with other model parameters. Flexibility of the density estimate is controlled by a tuning parameter $K$, where $K = 0$ corresponds to normality, chosen via standard model selection criteria, allowing assessment of departure from normality. All criteria reject $K = 0$, the normality assumption. Figure 1.1b shows the estimated density, providing visual evidence of a departure from normality, particularly for the “slope,” which indicates that the normality assumption for log PDG random effects may be inappropriate, raising concern that inference on population effects and individual effects would be compromised if the foregoing methods were applied.
1.4 Proposed Methods

The usual assumption of normal $X_i$ may be a poor representation of the population variation in longitudinal profiles. Because of these considerations, methods for the joint model that do not require or can relax the parametric random effects assumption and the investigation of sensitivity of inference to violations of the normality assumption are needed. This motivates us to propose two approaches.

In Chapter 2, we develop a conditional estimation approach adapting the strategy of Stefanski and Carroll (1987) and derive two score functions, a sufficiency score (SS) and a conditional score (CS). The methods “condition away” the random effects, so no distributional assumption on the random effects is required. We assess the finite sample properties of this approach via simulation studies and by application to the SWAN data in Chapter 3.

The conditional estimation approach is straightforward to implement and yields valid inference regardless of the underlying random effects distribution. However, a drawback is that it makes inefficient use of the information in longitudinal data. Moreover, because the random effects are “conditioned away” from consideration, it provides neither insight into the underlying features of the random effects, nor inference on the longitudinal data process. Therefore, we further propose an alternative method which is a semiparametric likelihood approach with relaxed distributional assumption on random effects. We approximate the random effects distribution using the seminonparametric (SNP) density representation of Gallant and Nychka (1987), which only requires the assumption that the random effects have a “smooth” but unspecified density, and use a full likelihood procedure to obtain estimators for both the
joint model parameters and the random effects distribution. Because the approach is full likelihood based, it may have the potential for improved efficiency over the conditional estimation approach if the “smoothness” assumption is valid. In addition, since it is a likelihood procedure, standard information criteria may be used for model selection. It is a semiparametric approach in that the random effects distribution is unspecified. This semiparametric likelihood approach is described in Chapter 4 and its performance is demonstrated via simulation studies and by application to the SWAN data in Chapter 5.
Figure 1.1: **a.** Natural logarithm of progesterone levels measured in urine (PDG) for four randomly selected women plotted against days in standardized menstrual cycle. Individual ordinary least squares fits of the piecewise linear model are superimposed. **b.** Estimated joint density of woman-specific intercepts (level up to day 14) and slopes (rate of change for days 14–21 and negative rate of change for days 21–28) in a linear mixed model for log PDG using the approach of Zhang and Davidian (2001).
Chapter 2

Conditional Estimation Approach

2.1 Introduction

In the last chapter, a joint model framework has been introduced to characterize the association between a primary endpoint and features of longitudinal profiles of a continuous response. The longitudinal measurements are assumed to follow a mixed model whose subject-specific random effects can capture the underlying features of longitudinal profiles and are considered as explanatory variables in the generalized linear model for the primary endpoint. The challenge is the underlying random effects can not be observed directly. Existing methods in literature require a parametric (normality) assumption on the random effects, which may be unrealistic. These considerations suggest that methods for the joint model that do not require a parametric random effects assumption are needed.

For a generalized linear model with normal covariate measurement error, Stefanski and Carroll (1987) derived estimating equations by conditioning on certain sufficient statistics that are unbiased regardless of the true covariate distribution. Tsiatis and
Davidian (2001) adapted this to the proportional hazards model with longitudinal covariates and demonstrated consistency of inferences on hazard parameters across a range of true distributions.

In this chapter, adapting the strategy of Stefanski and Carroll (1987), we propose estimators for inference in the joint model that require no assumptions on the random effects distribution. We derive the proposed estimators in Section 2.2, and we explicate popular special cases in Section 2.3. The methods are illustrated via simulation and by application to the SWAN data in Chapter 3.

2.2 Methods

2.2.1 Sufficiency Score

Because $f(Y_i | X_i; \beta, \rho)$ is an exponential family, re-arranging (1.2) shows that $S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_1 / a(\rho)$ is a sufficient and complete “statistic” for $X_i$ when all parameters are known. Let $E_i$, a $(m_i \times (m_i - q))$ matrix, be orthogonal to $D_i$ (i.e., $D_i^T E_i = 0$) with $E_i^T E_i = I_{m_i-q}$, so $E_i$ is determined by $D_i$. Let $\tilde{X}_i^{(ls)} = (D_i^T D_i)^{-1} D_i^T W_i$, the least squares estimator for $X_i$, and $R_i = E_i^T W_i$, a “residual-type” vector. It is straightforward to deduce that $\tilde{X}_i^{(ls)} | X_i \sim \mathcal{N}\{X_i, \sigma_u^2 (D_i^T D_i)^{-1}\}$, $R_i | X_i \sim \mathcal{N}(0, \sigma_u^2 I)$, and $R_i$ and $\tilde{X}_i^{(ls)}$ are independent given $X_i$. Thus, $R_i$ and $S_i$ are independent given $X_i$. As the transformation from $W_i$ to $(\tilde{X}_i^{(ls)}, R_i)$ is free of unknown parameters, $f(Y_i, W_i | X_i; \theta) \propto f(Y_i | X_i; \beta, \rho) f(\tilde{X}_i^{(ls)} | X_i; \sigma_u^2) f(R_i | X_i; \sigma_u^2)$, where $\theta = (\beta_0^T, \beta_1^T, \rho, \sigma_u^2)^T$. Also, because the Jacobian of the transformation from $(Y_i, \tilde{X}_i^{(ls)}, R_i)$ to $(Y_i, S_i, R_i)$ does not involve unknown parameters, the conditional
distribution of data \((Y_i, W_i)\) given \(S_i\) is

\[
f(Y_i, W_i | X_i, S_i; \theta) = f(Y_i, W_i | S_i; \theta) \propto f(Y_i | S_i)f(R_i). \tag{2.1}
\]

By arguments similar to those leading to (2.5) of Stefanski and Carroll (1987), we have that

\[
f(Y_i | S_i) = \exp \left[ Y_i \eta - Y_i^2 \sigma_u^2 \beta_1^T (D_i^T D_i)^{-1} \beta_1 / \{2a^2(\rho)\} + c(Y_i, \rho) - \log \{H(\eta_i, \beta_1, \rho, \sigma_u^2)\} \right],
\]

where

\[
H(\eta_i, \beta_1, \rho, \sigma_u^2) = \int \exp \{y\eta_i - y^2 \sigma_u^2 \beta_1^T (D_i^T D_i)^{-1} \beta_1 / \{2a^2(\rho)\} + c(y, \rho)\} dm(y),
\]

\(m(y)\) is the dominating measure for \(Y_i\), and

\[
\eta_i = \{\beta_0^T C_i + S_i^T (D_i^T D_i)^{-1} \beta_1\} / a(\rho).
\]

Moreover, because \(f(R_i)\) is proportional to the likelihood of \(\sigma_u^2\) for data \(R_i\),

\[
f(R_i) \propto (2\pi \sigma_u^2)^{-\{m_i - q\}/2} \exp \{-R_i^T R_i / (2\sigma_u^2)\}.
\]

For simplicity, \(R_i\) may be represented by the degenerate version

\[
R_i = \{I - D_i (D_i^T D_i)^{-1} D_i^T\} W_i,
\]

the least squares residuals. Substituting in (2.1), we obtain an expression proportional to the conditional density \(f(Y_i, W_i | S_i; \theta)\) and hence the log conditional likelihood for \(\theta\).

Following Stefanski and Carroll (1987), \(f(Y_i | S_i)\) is also an exponential family; thus \(E(Y_i | S_i) = \partial / \partial \eta_i (\log H)\) evaluated at \(\eta_i = \{\beta_0^T C_i + S_i^T (D_i^T D_i)^{-1} \beta_1\} / a(\rho)\) and
\[ S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_1 / a(\rho). \] Thus, we can construct an estimating equation for \( \theta \) that is unbiased regardless of the distribution of \( X_i \) of the form

\[
\sum_{i=1}^{n} \psi_S(Y_i, W_i | S_i; \theta) = 0,
\]

where \( \psi_S(Y_i, W_i | S_i; \theta) = \partial / \partial \theta \{ \log f(Y_i, W_i | S_i) \} \) evaluated at \( S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_1 / a(\rho). \) In particular, the sufficiency score function \( \psi_S(Y_i, W_i | S_i; \theta) \) is

\[
\begin{pmatrix}
\frac{Y_i - E(Y_i | S_i)}{a(\rho)} \\
\frac{a^2(\rho)}{\{Y_i - E(Y_i | S_i)\}\{D_i^T D_i\}^{-1} S_i} - \frac{\{Y_i^2 - E(Y_i^2 | S_i)\}\{D_i^T D_i\}^{-1} \sigma_u^2 \beta_1}{a^2(\rho)} \\
\frac{\{Y_i^2 - E(Y_i^2 | S_i)\}\{D_i^T D_i\}^{-1} \sigma_u^2 \beta_1}{a^2(\rho)} + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \right | S_i) \\
\frac{m_i - q}{2 \sigma_u^2} + \frac{R_i^T R_i}{2 \sigma_u^4} - \frac{\{Y_i^2 - E(Y_i^2 | S_i)\}\beta_1^T (D_i^T D_i)^{-1} \beta_1}{2 a^2(\rho)}
\end{pmatrix}
\]

evaluated at \( S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_1 / a(\rho). \) Following Stefanski and Carroll (1987), we call the solution to such an estimating equation \( \hat{\theta}_S \) a sufficiency estimator.

### 2.2.2 Conditional Estimator

Alternatively, we consider the conditional score approach (McCullagh and Nelder, 1989, sec. 7.2.2) based on the bias-corrected score function. Treating \( X_i \) as a nuisance parameter, the bias-corrected score,

\[
\partial / \partial \theta \log \{ f(Y_i, W_i | X_i; \theta) \} - E[\partial / \partial \theta \log \{ f(Y_i, W_i | X_i; \theta) \} | S_i],
\]
is

\[
\begin{pmatrix}
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} X_i \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T X_i \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T X_i \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
- \frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T X_i \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
- \frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T X_i \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right)
\end{pmatrix}
\]
evaluated at \( S_i = D_i^T W_i + Y_i \sigma^2_u \beta_1 / a(\rho) \). As this involves the unknown \( X_i \), we again follow Stefanski and Carroll (1987) and replace \( X_i \) by a \( q \)-dimensional function \( t(S_i) \), which yields the conditional score function \( \psi_C(Y_i, W_i|S_i; \theta) \) given by

\[
\begin{pmatrix}
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} C_i \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} t(S_i) \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T t(S_i) \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
\frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T t(S_i) \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
- \frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T t(S_i) \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right) \\
- \frac{\{Y_i - E(Y_i|S_i)\}}{a(\rho)} \left( \beta_0^T C_i + \beta_1^T t(S_i) \right) a'(\rho) + \frac{\partial c(Y_i; \rho)}{\partial \rho} - E \left( \frac{\partial c(Y_i, \rho)}{\partial \rho} \bigg| S_i \right)
\end{pmatrix}
\]
evaluated at \( S_i = D_i^T W_i + Y_i \sigma^2_u \beta_1 / a(\rho) \). As long as \( t(S_i) \) depends on \((Y_i, W_i)\) only through \( S_i \), \( \psi_C(Y_i, W_i|S_i; \theta) \) has mean zero; thus, we may form an unbiased estimating equation

\[
\sum_{i=1}^{n} \psi_C(Y_i, W_i|S_i; \theta) = 0
\]
for \( \theta \). As in Stefanski and Carroll (1987), we call \( \hat{\theta}_C \) solving this equation a conditional estimator. Because \( \tilde{X}_i^{(ls)} \) is unbiased for \( X_i \) and \( S_i \) is complete and sufficient for \( X_i \), a natural choice is

\[
t(S_i) = E(\tilde{X}_i^{(ls)}|S_i) = (D_i^T D_i)^{-1}\{S_i - E(Y_i|S_i)\sigma^2_u \beta_1 / a(\rho)\},
\]
a uniformly minimum variance unbiased estimator (UMVUE) for $X_i$.

2.2.3 Inference

Both the sufficiency and conditional score estimators effectively “condition away” dependence on $X_i$ so require no assumption on the distribution of $X_i$. Both estimating equations are unbiased. In general, $\psi_S$ and $\psi_C$ differ regardless of the choice of $t(S_i)$. Solution of either equation may be implemented via the Newton-Raphson algorithm without much computational burden. Similar to Stefanski and Carroll (1987) and Tsiatis and Davidian (2001), the equations may have multiple roots, not all of which are consistent for $\theta$. We have found that by using a naive estimator as the starting value, as in Tsiatis and Davidian (2001), a consistent solution may be identified, demonstrated in Section 3.2. Moreover, as both are M-estimators, under regularity conditions, such consistent solutions should be asymptotically normal, and standard errors for either estimator may thus be obtained by using the usual empirical sandwich approach (e.g., Carroll, Ruppert, and Stefanski, 1995, sec. A.3.1). We show that this technique yields reliable inferences in Section 3.2.

2.3 Special Cases

We elucidate the form of the sufficiency and conditional scores in two popular cases.
2.3.1 Normal Primary Model

Suppose \( Y_i | X_i \sim \mathcal{N}(\beta_0^T C_i + \beta_1^T X_i, \rho) \). Then the dominating measure \( m(\cdot) \) for \( Y_i \) is Lebesgue measure; and, in (1.2), \( a(\rho) = \rho, b(\gamma) = \gamma^2/2, \) and \( c(y, \rho) = -\{y^2/\rho + \log(2\pi\rho)\}/2. \) Hence \( \eta_i = \{\beta_0^T C_i + S_i^T (D_i^T D_i)^{-1} \beta_1\}/\rho. \) Substituting in the expression for \( f(Y_i | S_i) \) in Section 2.2.1, it follows that \( Y_i | S_i \sim \mathcal{N}(\mu_i, v_i), \) where \( \mu_i = \{\beta_0^T C_i + S_i^T (D_i^T D_i)^{-1} \beta_1\}/\{1 + \sigma_u^2 \beta_1^T (D_i^T D_i)^{-1} \beta_1/\rho\}, \) and \( v_i = \rho/\{1 + \sigma_u^2 \beta_1^T (D_i^T D_i)^{-1} \beta_1/\rho\}, \) so that \( E(Y_i | S_i) = \mu_i \) and \( E(Y_i^2 | S_i) = \mu_i^2 + v_i. \) Substituting these results, \( a(\rho) = \rho, \) and \( a'(\rho) = 1 \) in \( \psi_S(Y_i, W_i | S_i; \theta) \) yields the sufficiency score function for \( \theta, \) i.e.,

\[
\psi_S(Y_i, W_i | S_i; \theta) = \left( \begin{array}{c}
\frac{(Y_i - \mu_i)C_i}{(Y_i - \mu_i)(D_i^T D_i)^{-1} S_i} - \frac{(Y_i^2 - \mu_i^2)(D_i^T D_i)^{-1} \sigma_u^2 \beta_1}{\rho^2} + \frac{(D_i^T D_i)^{-1} \sigma_u^2 \beta_1}{\rho}
- \frac{(Y_i - \mu_i)\{\beta_0^T C_i + S_i^T (D_i^T D_i)^{-1} \beta_1\}}{\rho^2} + (Y_i^2 - \mu_i^2) \left\{ \frac{1}{2\rho^2} + \frac{\sigma_u^2 \beta_1^T (D_i^T D_i)^{-1} \beta_1}{\rho^3} \right\}
\end{array} \right)
\]

evaluated at \( S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_1/\rho. \)

Similarly, substitution in \( \psi_C(Y_i, W_i | S_i; \theta) \) with \( t(S_i) = (D_i^T D_i)^{-1}(S_i - \mu_i \sigma_u^2 \beta_1/\rho), \)
the UMVUE for $X_i$, gives the conditional score function $\psi_C(Y_i, W_i|S_i; \theta) =$

\[
\begin{pmatrix}
(Y_i - \mu_i)C_i \\
\frac{\rho}{(Y_i - \mu_i)(D_i^TD_i)^{-1}(S_i - \mu_i \sigma_u^2 \beta_2 \rho)} \\
-\frac{(Y_i - \mu_i)\{\beta_0^T C_i + \beta_1^T (D_i^TD_i)^{-1}(S_i - \mu_i \sigma_u^2 \beta_1 \rho)\}}{\rho^2} \frac{Y_i^2 - \mu_i^2}{2\rho^2} \\
\frac{1}{\rho^2} \frac{2\{\rho + \sigma_u^2 \beta_1^T (D_i^TD_i)^{-1} \beta_1\}}{\rho^2} \\
-\frac{m_i - q}{2\sigma_u^2} + \frac{R_i^TR_i}{2\sigma_u^4} - \frac{(Y_i - \mu_i)\mu_i \beta_1^T (D_i^TD_i)^{-1} \beta_1}{\rho^2} + \frac{(Y_i^2 - \mu_i^2)\beta_1^T (D_i^TD_i)^{-1} \beta_1}{2\rho^2} \\
-\frac{\beta_1^T (D_i^TD_i)^{-1} \beta_1}{2\rho^2} \frac{2\{\rho + \sigma_u^2 \beta_1^T (D_i^TD_i)^{-1} \beta_1\}}{\rho^2}
\end{pmatrix}
\]

evaluated at $S_i = D_i^TW_i + Y_i \sigma_u^2 \beta_1 / \rho$.

### 2.3.2 Logistic Primary Model

Suppose $Y_i$ is binary (=0,1) and $P(Y_i = 1|X_i) = \left[1 + \exp\{-\beta_0^T C_i + \beta_1^T X_i\}\right]^{-1}$.

Then $a(\rho) = 1$, $\rho = 1/n$, $b(\gamma) = \log(1 + e^\gamma)$, and $c(y, \rho) = \log \left(\begin{pmatrix}n\end{pmatrix}\right)$ in (1.2); $m(\cdot)$ is counting measure on (0,1); and $\eta_i = \beta_0^T C_i + S_i^T (D_i^TD_i)^{-1} \beta_1$. Substitution in $f(Y_i|S_i)$ in Section 2.2.1 yields $P(Y_i = 1|S_i) = \left(1 + \exp\{-\beta_0^T C_i + (S_i - \sigma_u^2 \beta_1 / 2)^T (D_i^TD_i)^{-1} \beta_1\}\right)^{-1} = \mu_i$, say, which is also a logistic model. Hence, $E(Y_i|S_i) = E(Y_i^2|S_i) = \mu_i$. As $\rho$ is known, the third entry in each of $\psi_S(Y_i, W_i|S_i; \theta)$ and $\psi_C(Y_i, W_i|S_i; \theta)$ may be ignored; combining with the foregoing results, we obtain

\[
\psi_S(Y_i, W_i|S_i; \theta) = \begin{pmatrix}
(Y_i - \mu_i)C_i \\
(Y_i - \mu_i)(D_i^TD_i)^{-1}(S_i - \sigma_u^2 \beta_1) \\
-\frac{m_i - q}{2\sigma_u^2} + \frac{R_i^TR_i}{2\sigma_u^4} - \frac{(Y_i - \mu_i)\beta_1^T (D_i^TD_i)^{-1} \beta_1}{2}
\end{pmatrix}
\]
and

$$\psi_C(Y_i, W_i | S_i; \theta) = \begin{pmatrix} (Y_i - \mu_i)C_i \\ (Y_i - \mu_i)(D_i^T D_i)^{-1}(S_i - \mu_i\sigma_u^2\beta_1) \\ -\frac{m_i - q}{2\sigma_u^2} + \frac{R_i^T R_i}{2\sigma_u^4} + (Y_i - \mu_i)\beta_1^T(D_i^T D_i)^{-1}\beta_1\left(\frac{1}{2} - \mu_i\right) \end{pmatrix}$$

with $t(S_i) = (D_i^T D_i)^{-1}(S_i - \mu_i\sigma_u^2\beta_1)$, both evaluated at $S_i = D_i^T W_i + Y_i\sigma_u^2\beta_1$.

### 2.4 Technical Details

#### 2.4.1 Updating Scheme

Let $\hat{\theta}$ denote either $\hat{\theta}_S$ or $\hat{\theta}_C$. Based on the Newton-Raphson algorithm, the updating scheme for the estimators at the $(k + 1)$th iteration is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \left\{ \sum_{i=1}^{n} \frac{\partial}{\partial \theta^T} \psi(Y_i, W_i | S_i; \hat{\theta}^{(k)}) \right\}^{-1} \sum_{i=1}^{n} \psi(Y_i, W_i | S_i; \hat{\theta}^{(k)})$$

For the logistic primary model described in Section 2.3.2, the derivative of the sufficiency score is

$$\frac{\partial}{\partial \theta^T} \psi_S(Y_i, W_i | S_i; \theta) = \begin{pmatrix} \frac{\partial \psi_S}{\partial \mu} \\ \frac{\partial \psi_S}{\partial \beta_0} \\ \frac{\partial \psi_S}{\partial \beta_1} \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{12} & J_{22} & J_{23} \\ J_{13} & J_{23} & J_{33} \end{pmatrix}$$

where

$$J_{11} = \mu_i(\mu_i - 1)C_i C_i^T,$$

$$J_{12} = \mu_i(\mu_i - 1)C_i ((D_i^T D_i)^{-1}(S_i - \sigma_u^2\beta_1))^T,$$

$$J_{13} = -\mu_i(\mu_i - 1)\beta_1^T(D_i^T D_i)^{-1}\beta_1 C_i/2,$$

$$J_{22} = \mu_i(\mu_i - 1)(D_i^T D_i)^{-1}(S_i - \sigma_u^2\beta_1)(S_i - \sigma_u^2\beta_1)^T(D_i^T D_i)^{-1} - (Y_i - \mu_i)\sigma_u^2(D_i^T D_i)^{-1},$$
2.4.2 Empirical Sandwich Estimator

\[ J_{23} = -\mu_i(\mu_i - 1)\beta_1^T(D_i^T D_i)^{-1}\beta_1(D_i^T D_i)^{-1}(S_i - \sigma^2 u_1) / 2 - (Y_i - \mu_i)(D_i^T D_i)^{-1}\beta_1, \]
\[ J_{33} = (m_i - q)/(2\sigma^4 u) - R_i^T R_i/\sigma^6 u + \mu_i(\mu_i - 1)\{\beta_1^T(D_i^T D_i)^{-1}\beta_1\}^2 / 4, \]

and the derivative of the conditional score is

\[
\frac{\partial}{\partial \theta^T} \psi_C(Y_i, W_i|S_i; \theta) = \begin{pmatrix}
\frac{\partial \psi_C}{\partial \beta_0^T} \\
\frac{\partial \psi_C}{\partial \beta_1^T} \\
\frac{\partial \psi_C}{\partial \sigma^2_u} 
\end{pmatrix} = \begin{pmatrix}
J_{11} & J_{12} & J_{13} \\
J_{21} & J_{22} & J_{23} \\
J_{31} & J_{32} & J_{33}
\end{pmatrix}
\]

where

\[ J_{11} = \mu_i(\mu_i - 1)C_i C_i^T, \]
\[ J_{12} = \mu_i(\mu_i - 1)C_i((D_i^T D_i)^{-1}(S_i - \sigma^2 u_1)) C_i^T, \]
\[ J_{13} = -\mu_i(\mu_i - 1)\beta_1^T(D_i^T D_i)^{-1}\beta_1 C_i/2, \]
\[ J_{21} = \mu_i(\mu_i - 1)(D_i^T D_i)^{-1}\{S_i + (Y_i - 2\mu_i)\sigma^2 u_1\} C_i^T, \]
\[ J_{22} = \mu_i(\mu_i - 1)(D_i^T D_i)^{-1}\{S_i + (Y_i - 2\mu_i)\sigma^2 u_1\}(S_i - \sigma^2 u_1)^T(D_i^T D_i)^{-1} \]
\[ -\mu_i(Y_i - \mu_i)\sigma^2 u(D_i^T D_i)^{-1}, \]
\[ J_{23} = -\mu_i(\mu_i - 1)\beta_1^T(D_i^T D_i)^{-1}\beta_1(D_i^T D_i)^{-1}\{S_i + (Y_i - 2\mu_i)\sigma^2 u_1\} / 2 \]
\[ -\mu_i(Y_i - \mu_i)(D_i^T D_i)^{-1}\beta_1, \]
\[ J_{31} = \mu_i(\mu_i - 1)(Y_i - 2\mu_i + 1/2)\beta_1^T(D_i^T D_i)^{-1}\beta_1 C_i^T, \]
\[ J_{32} = \mu_i(\mu_i - 1)(Y_i - 2\mu_i + 1/2)\beta_1^T(D_i^T D_i)^{-1}\beta_1(S_i - \sigma^2 u_1)^T(D_i^T D_i)^{-1} \]
\[ + (Y_i - \mu_i)(1 - 2\mu_i)\beta_1^T(D_i^T D_i)^{-1}, \]
\[ J_{33} = (m_i - q)/(2\sigma^4 u) - R_i^T R_i/\sigma^6 u - \mu_i(\mu_i - 1)(Y_i - 2\mu_i + 1/2)\{\beta_1^T(D_i^T D_i)^{-1}\beta_1\}^2 / 2. \]

2.4.2 Empirical Sandwich Estimator

To obtain standard errors for the estimators, the usual empirical sandwich estimators may be used because both sufficiency and conditional estimators are M-estimators. As an M-estimator, \( \hat{\theta} \sim N(\theta_0, V(\theta_0)/n) \) as \( n \to \infty \), where \( \theta_0 \) is the
true parameter vector and

\[
V(\theta_0) = A^{-1}(\theta_0)B(\theta_0)\{A^{-1}(\theta_0)\}^T,
\]

\[
A(\theta_0) = E\left\{-\frac{\partial}{\partial \theta^T} \psi(Y_i, W_i|S_i; \theta_0)\right\},
\]

\[
B(\theta_0) = E\left\{\psi(Y_i, W_i|S_i; \theta_0)\psi^T(Y_i, W_i|S_i; \theta_0)\right\}.
\]

The empirical sandwich estimator of \(V(\theta_0)\) is

\[
V_n(Y, W|S; \hat{\theta}) = A_n^{-1}(Y, W|S; \hat{\theta})B_n(Y, W|S; \hat{\theta}) \left\{A_n^{-1}(Y, W|S; \hat{\theta})\right\}^T,
\]

where

\[
A_n(Y, W|S; \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{-\frac{\partial}{\partial \theta^T} \psi(Y_i, W_i|S_i; \hat{\theta})\right\},
\]

\[
B_n(Y, W|S; \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, W_i|S_i; \hat{\theta})\psi^T(Y_i, W_i|S_i; \hat{\theta}).
\]

Here, the derivatives of the score functions differ from those in the updating scheme.

In the updating scheme, the derivatives of the score functions are derived with \(S_i\) held fixed, while for the sandwich estimator, the derivatives of the score functions are derived with respect to the parameters after the scores are evaluated at \(S_i = D_i^T W_i + Y_i \sigma_u^2 \beta_i / a(\rho)\).
Chapter 3

Empirical Results For Conditional Estimation Approach

3.1 Introduction

In the last chapter, a conditional estimation approach has been proposed. The proposed sufficiency and conditional estimators are shown to be consistent and asymptotically normal and to yield unbiased inference on parameters of the joint model. An attractive feature of the conditional estimation approach is that no assumption is needed to be made on the distribution of the random effects. Moreover, the estimators are straightforward to compute via Newton-Raphson algorithm.

In this chapter, we illustrate the finite sample performance and the utility of the conditional estimation approach. In Section 3.2, the performance of the approach is demonstrated via simulation. The utility of the approach is shown via application to the SWAN data in Section 3.3.
3.2 Simulation Studies

We carried out several simulations for the logistic primary model

\[ P(Y_i = 1|X_i) = \left[ 1 + \exp\{-\langle \beta_0 + \beta_1^T X_i \rangle \} \right]^{-1} \]

with \( X_i = (X_{1i}, X_{2i}) \) and \( \beta_1 = (\beta_{11}, \beta_{12})^T \). In all cases, \( E(X_{1i}) = E(X_{2i}) = 0.5 \), \( \text{var}(X_{1i}) = 1.0 \), \( \text{var}(X_{2i}) = 0.64 \), \( \text{cov}(X_{1i}, X_{2i}) = -0.2 \), \( \beta_0 = -2.5 \), and \( \beta_1 = (3.0, 2.0)^T \); longitudinal data were generated according to

\[ W_{ij} = X_{1i} + X_{2i}t_{ij} + U_{ij}, \]

where \( t_{ij} \sim N(j - 1, 0.01) \) for \( i = 1, \ldots, n \), \( j = 1, \ldots, m_i \), and \( U_{ij} \sim N(0, \sigma_u^2) \). In the simulations reported here, \( n = 500 \); \( \sigma_u^2 = 0.5 \), and \( m_i \leq 5 \); five \( W_{ij} \) were generated for each subject \( i \), with each arbitrarily missing with probability 0.05 to create imbalance. Each scenario below involved 1000 Monte Carlo data sets. Viewing this as a measurement error problem, the magnitude of the measurement error may be gauged informally by the “reliability ratio” defined by \( \lambda_k = \text{var}(X_{ki})/E\{\text{var}(\hat{X}_{ki})\} \) for \( k = 1, 2 \). For all scenarios, \( (\lambda_1, \lambda_2) = (0.72, 0.92) \), suggesting moderate (mild) measurement error associated with \( \hat{X}_{1i} \) (\( \hat{X}_{2i} \)).

To investigate performance under a range of true \( X_i \) distributions, we considered four scenarios: (a) \( X_i \) bivariate normal; (b) \( X_i \) distributed as a bimodal 50-50 mixture of normals; (c) \( X_i \) following the bivariate skew-normal (Azzalini and Dalla Valle, 1996), with coefficients of skewness \(-0.10 \) and \( 0.85 \) for \( X_{1i} \) and \( X_{2i} \), respectively; and (d) \( X_i \) bivariate \( t \) with 5 degrees of freedom (Lange, Little, and Taylor, 1989). In each case, the distribution was scaled to have the moments above. For each data set under each scenario, \( \theta = (\beta_0, \beta_{11}, \beta_{12}, \sigma_u^2)^T \) was estimated five ways: (i) by the naive
approach in Section 1.1 (N); (ii) using regression calibration (RC) (Wang et al., 2000, sec. 4.1); (iii) by refined regression calibration (Wang et al., 2000, sec. 4.2), which applies a probit approximation to the logistic to approximate the pseudo-EEE (RR); (iv) via the sufficiency score method (SS); and (v) via the conditional score method (CS). Methods (ii) and (iii) require normality of $X_i$, so their performance under violations of this assumption is of interest. Standard errors for (i) were obtained from usual logistic regression formulæ, ignoring least-squares imputation of $X_i$; those for (ii)–(v) were derived from the empirical sandwich approach. 95% Wald confidence intervals were constructed using the standard normal critical value of 1.96. The naive estimate was used as the starting value for (iii)–(v).

From Table 3.1, in all cases, the naive estimator shows considerable bias, while bias of the proposed estimators (iv) and (v) is negligible. The RC estimator is also unacceptably biased except for bimodal $X_i$; for normal $X_i$, this estimator shows appreciable bias, verifying concerns of Wang et al. (2000). Under normality, the RR and proposed estimators exhibit negligible bias and similar performance. However, under departures from normality, RC and RR can show degradation of performance. Under bimodality, although the RC estimator is remarkably unbiased for $\beta_0$ and $\beta_{11}$, coverages are well below the nominal level. Interestingly, here, the RR estimator (iii) shows considerable bias, and standard errors vastly under-represent true sampling variation. For scenario (c), with only mild skewness in the $X_{2i}$ dimension, the RR estimator achieves good performance; however, for the “heavy-tailed” bivariate $t$, although bias of the RR estimator is not large (although larger than that of SS and CS), inferences are flawed, with coverage falling far short of nominal.
We obtained similar results for other choices of $n$, $\max m_i$, and $(\beta_0, \beta_1)$; for larger $\sigma_\theta^2$, bias and failure to achieve nominal coverage of the naive, RC, and RR estimators worsen and are less pronounced for smaller $\sigma_\theta^2$. The sufficiency and conditional score estimators always show negligible bias and attain nominal coverage regardless of scenario. In all cases, the starting value strategy encountered no problems identifying the apparent consistent solution.

Overall, under the nonlinear (in $X_i$) logistic model, RC and RR show poor performance. On the other hand, for the linear (in $X_i$), normal primary model with $\rho$ estimated, for which there is no RR estimator, simulations not reported show that while the naive estimator is still biased, the RC estimator shows little bias in all cases and achieves slightly better performance than SS and CS under normality but is inefficient otherwise. For a related linear model, results of Buonaccorsi, Demidenko, and Tosteson (2000) suggest that, under normal random effects, RC is a pseudo-likelihood estimator for $\beta$, as also noted by Wang et al. (2000), so should be consistent and nearly efficient. Buonaccorsi et al. (2000) further show for their model that the RC estimator is consistent but with likely loss of precision, coinciding with our simulation results and suggesting that a similar property holds here.

### 3.3 Application to SWAN

In addition to osteopenia status, where $Y_i = 1$ denotes absence of osteopenia (BMD above the 33rd percentile) and $Y_i = 0$ denotes presence for woman $i$, $i = 1, \ldots, n = 624$, a number of baseline covariates and longitudinal log PDG measurements were collected; numbers of longitudinal measurements $m_i$ ranged from 8 to
25. For woman $i$, let $Z_{\text{BMI},i}$ denote body mass index (BMI, kg/m$^2$), $Z_{\text{age},i}$ denote age (years), and $Z_{B,i}$, $Z_{W,i}$ and $Z_{C,i}$ denote ethnicity indicator variables = 1 as woman $i$ is Black, White, or Chinese; $Z_{B,i} = Z_{W,i} = Z_{C,i} = 0$ indicates woman $i$ is Japanese. Let $C_i = (1, Z_{\text{BMI},i}, Z_{\text{age},i}, Z_{B,i}, Z_{W,i}, Z_{C,i})^T$.

From Figure 1.1a, a reasonable representation of woman $i$’s log PDG profile is

$$W_{ij} = X_{1i} + X_{2i}(t_{ij} - 1.4)_+ - 2X_{2i}(t_{ij} - 2.1)_+ + U_{ij},$$

where $u_+ = u$ if $u > 0$ and 0 otherwise, time is in units of 10 days for computational stability, $U_{ij}$ is a $\mathcal{N}(0, \sigma_u^2)$ measurement error, $X_{1i}$ is the woman-specific “true” (error-free) log PDG up to day 14, and $X_{2i}$ is 10× the woman-specific “slope” of the symmetric rise (days 14–21) and fall (days 21–28) depicted in Figure 1.1a. To describe the relationship between propensity for osteopenia and $C_i$ and $X_i = (X_{1i}, X_{2i})^T$, we considered the primary logistic regression model

$$P(Y_i = 1|X_i) = \frac{1}{1 + \exp\{-(\beta_0^T C_i + \beta_1^T X_i)\]}^{-1},$$

where $\beta_0 = (\beta_{0,\text{BMI}}, \beta_{0,\text{age}}, \beta_{0,B}, \beta_{0,W}, \beta_{0,C})^T$ and $\beta_1 = (\beta_{11}, \beta_{12})^T$.

Table 3.2 shows results from fitting the joint model by the methods in Section 3.2. Inferences on the coefficients of baseline covariates are similar across all methods and indicate strong association between absence of osteopenia and BMI after taking into account age, ethnicity, and hormone profile, consistent with observations in previous studies. Inferences on the relationship between osteopenia and pre-14-day log PDG level and post-14-day rise/fall are also similar, although point estimates are more disparate across methods. The proposed estimators suggest that absence of osteopenia may be positively associated with post-14-day “slope” of log PDG profile, after taking into account other factors.
The consistency across methods here may not be unexpected. The estimated “reliability ratio” defined in Section 3.2 is \( \lambda = (0.88, 0.84) \) for these data, so the magnitude of “measurement error” in both components is not great. From Figure 1.1, although bimodal, the estimated density for \( X_i \) does not deviate considerably from normality; coefficients of skewness based on the density estimate are \(( -0.13, -0.87)\) for \( (X_{1i}, X_{2i})^T \), comparable in magnitude to the bivariate skew-normal scenario (c) in Section 3.2, where RR, SS, and CS performed similarly. To investigate further, we carried out a simulation based on the estimates and design for the data, which will be discussed in Section 5.2.1, and found that all methods yielded similar inferences. Evidently, the design and variation/correlation configuration for these data leads to only minimal differences among the procedures. Nonetheless, use of the proposed methods, which, unlike the competitors are consistent under any conditions, offers the analyst assurance of credible estimation of the relationship.

### 3.4 Summary

The relationship between a primary endpoint and features of longitudinal profiles of a continuous response is often of interest, and a relevant framework is that of a generalized linear model with covariates that are subject-specific random effects in a linear mixed model for the longitudinal measurements. Naive implementation by imputing subject-specific effects from individual regression fits yields biased inference, and several methods for reducing this bias have been proposed. These require a parametric (normality) assumption on the random effects, which may be unrealistic. Adapting a strategy of Stefanski and Carroll (1987), we propose estimators for the
generalized linear model parameters that require no assumptions on the random effects and yield consistent inference regardless of the true distribution. The methods are illustrated via simulation and by application to the SWAN data.

The methods have been implemented using Newton-Raphson algorithm for normal and logistic primary models; SAS macro code is available from the author.
Table 3.1: Simulation results for four underlying random effect distributions for 
\(P(Y_i = 1|X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1}\), with true values \(\beta_0 = -2.5\), 
\(\beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T\), \(\sigma_u^2 = 0.5\), \(n = 500\), RB, relative bias (%); SD, Monte 
Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo 
coverage probability of 95% Wald confidence interval. Estimators: N, naive; RC, 
regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, 
conditional score.

<table>
<thead>
<tr>
<th>Method</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a) Normal</td>
<td></td>
<td></td>
<td></td>
<td>(b) Bimodal mixture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>N</td>
<td>-25.3</td>
<td>0.22</td>
<td>0.22</td>
<td>0.20</td>
<td>-14.0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.3</td>
<td>0.23</td>
<td>0.23</td>
<td>0.51</td>
<td>-4.4</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>1.4</td>
<td>0.38</td>
<td>0.37</td>
<td>0.95</td>
<td>39.5</td>
<td>1.05</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>3.0</td>
<td>0.41</td>
<td>0.42</td>
<td>0.97</td>
<td>2.8</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>2.6</td>
<td>0.38</td>
<td>0.41</td>
<td>0.97</td>
<td>3.0</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>(\beta_{11})</td>
<td>N</td>
<td>-32.0</td>
<td>0.18</td>
<td>0.19</td>
<td>0.01</td>
<td>-18.3</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.0</td>
<td>0.23</td>
<td>0.24</td>
<td>0.41</td>
<td>0.1</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>2.1</td>
<td>0.44</td>
<td>0.43</td>
<td>0.96</td>
<td>46.3</td>
<td>1.32</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>3.7</td>
<td>0.49</td>
<td>0.50</td>
<td>0.97</td>
<td>2.5</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>3.2</td>
<td>0.44</td>
<td>0.50</td>
<td>0.96</td>
<td>2.6</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>(\beta_{12})</td>
<td>N</td>
<td>-14.5</td>
<td>0.20</td>
<td>0.20</td>
<td>0.67</td>
<td>-8.0</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.0</td>
<td>0.21</td>
<td>0.21</td>
<td>0.58</td>
<td>-11.6</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>1.7</td>
<td>0.30</td>
<td>0.30</td>
<td>0.97</td>
<td>28.6</td>
<td>0.74</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>3.1</td>
<td>0.32</td>
<td>0.33</td>
<td>0.98</td>
<td>2.6</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>2.8</td>
<td>0.30</td>
<td>0.32</td>
<td>0.97</td>
<td>2.7</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(c) Skewed</td>
<td></td>
<td></td>
<td></td>
<td>(d) Bivariate (t_5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>N</td>
<td>-25.0</td>
<td>0.21</td>
<td>0.22</td>
<td>0.20</td>
<td>-28.9</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.1</td>
<td>0.23</td>
<td>0.23</td>
<td>0.53</td>
<td>-21.3</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>1.8</td>
<td>0.40</td>
<td>0.37</td>
<td>0.95</td>
<td>-6.6</td>
<td>0.37</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>3.3</td>
<td>0.45</td>
<td>0.43</td>
<td>0.96</td>
<td>3.5</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>3.1</td>
<td>0.42</td>
<td>0.42</td>
<td>0.95</td>
<td>3.4</td>
<td>0.44</td>
<td>0.45</td>
</tr>
<tr>
<td>(\beta_{11})</td>
<td>N</td>
<td>-32.1</td>
<td>0.19</td>
<td>0.19</td>
<td>0.01</td>
<td>-36.5</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.1</td>
<td>0.24</td>
<td>0.24</td>
<td>0.40</td>
<td>-22.5</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>1.9</td>
<td>0.47</td>
<td>0.43</td>
<td>0.94</td>
<td>-7.9</td>
<td>0.41</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>3.4</td>
<td>0.56</td>
<td>0.51</td>
<td>0.95</td>
<td>3.9</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>3.2</td>
<td>0.51</td>
<td>0.51</td>
<td>0.94</td>
<td>3.7</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>(\beta_{12})</td>
<td>N</td>
<td>-14.5</td>
<td>0.21</td>
<td>0.20</td>
<td>0.68</td>
<td>-17.0</td>
<td>0.20</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>-17.2</td>
<td>0.21</td>
<td>0.21</td>
<td>0.60</td>
<td>-19.3</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>1.6</td>
<td>0.33</td>
<td>0.30</td>
<td>0.94</td>
<td>-4.4</td>
<td>0.30</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td>2.9</td>
<td>0.35</td>
<td>0.33</td>
<td>0.96</td>
<td>3.6</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>2.8</td>
<td>0.33</td>
<td>0.33</td>
<td>0.95</td>
<td>3.4</td>
<td>0.34</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Table 3.2: Parameter estimates for the SWAN data analysis via several methods; abbreviations for methods are as in Table 3.1. Estimated standard errors are in parentheses below each estimate. Standard errors for estimates of $\sigma_u^2$ are multiplied by 10. For the naive method, $\sigma_u^2$ was estimated by mean square of pooled individual least squares residuals.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>RC</th>
<th>RR</th>
<th>SS</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-7.12</td>
<td>0.00</td>
<td>-7.28</td>
<td>0.00</td>
<td>-7.32</td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td></td>
<td>(2.18)</td>
<td></td>
<td>(2.21)</td>
</tr>
<tr>
<td>$\beta_{0,BMI}$</td>
<td>0.29</td>
<td>0.00</td>
<td>0.29</td>
<td>0.00</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td>(0.03)</td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>$\beta_{0,age}$</td>
<td>0.01</td>
<td>0.75</td>
<td>0.02</td>
<td>0.73</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>$\beta_{0,B}$</td>
<td>0.54</td>
<td>0.17</td>
<td>0.54</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td></td>
<td>(0.39)</td>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>$\beta_{0,W}$</td>
<td>-0.21</td>
<td>0.43</td>
<td>-0.21</td>
<td>0.44</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td></td>
<td>(0.27)</td>
<td></td>
<td>(0.28)</td>
</tr>
<tr>
<td>$\beta_{0,C}$</td>
<td>-0.68</td>
<td>0.01</td>
<td>-0.68</td>
<td>0.01</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td></td>
<td>(0.25)</td>
<td></td>
<td>(0.25)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.16</td>
<td>0.37</td>
<td>0.13</td>
<td>0.56</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td></td>
<td>(0.22)</td>
<td></td>
<td>(0.22)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.11</td>
<td>0.08</td>
<td>0.14</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
<td>(0.08)</td>
<td></td>
<td>(0.08)</td>
</tr>
<tr>
<td>$\sigma_u^2$</td>
<td>0.33</td>
<td></td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.12)</td>
</tr>
</tbody>
</table>
Chapter 4

Semiparametric Likelihood Approach

4.1 Introduction

In the last chapter, it has been illustrated that inference on the parameters of interest if methods depending on the normal random effects assumption are used when it is violated can be biased or have poor coverage probability, raising concern that inference on population effects and individual effects would be compromised if the foregoing methods were applied. We have developed a conditional estimation approach in Chapter 2, in which a sufficiency score and a conditional score are derived for inference in the joint model, without imposing any assumptions on the random effects distribution. This conditional estimation approach yields valid inferences under departure from the normality assumption for the random effects and is competitive when the assumption holds. It is straightforward and fast to implement. However, this approach makes inefficient use of the longitudinal measurements; i.e., no insight into the features of the random effects is available as it completely “conditions away” the random effects. Moreover, the efficiency and validity of inference on individual
effects is often a concern, hence estimation of random effects distribution under less restrictive assumptions may provide considerable insight into critical features of inherent subject heterogeneity and potential exclusion of important covariates which result in nonnormal behavior of the random effects. Because of these considerations, methods for the joint model that can relax the parametric random effects assumption and provide insight into the underlying distribution of the random effects, and the investigation of sensitivity of inference to violations of the normality assumption are needed.

Gallant and Nychka (1987) proposed a seminonparametric (SNP) representation for approximation of densities of various shapes. Davidian and Gallant (1993) applied the SNP representation to nonlinear mixed models. Linear mixed models using the SNP representation for the random effects distribution has been studied by Zhang and Davidian (2001). Song, Davidian, and Tsiatis (2002) modeled longitudinal and time-to-event data jointly based on the SNP density approximation. In this paper, we relax the normal random effects assumption and propose a full likelihood approach to joint estimation of the joint model parameters and random effects distribution requiring only that the random effects have a distribution in a plausible class with “smooth” but unspecified densities based on the SNP representation. If the “smoothness” assumption is valid, we would expect efficiency gains over the conditional estimation approach.

In Section 4.2, we approximate the random effects distribution using SNP representation, and we derive the estimation procedure including the likelihood-based EM algorithm in Section 4.3. In Chapter 5, we will compare the proposed method with
the existing methods in literature via simulation to investigate their performance under violation of the routine normality assumption on random effects. We will apply the method to the SWAN data in Chapter 5.

4.2 SNP Density Approximation

For continuous responses, it is reasonable to suppose that the random effects are continuous, thus rather than assuming normality, we relax the usual normality assumption on the random effects and require only that $X_i$ have a density belonging to a class of smooth densities studied by Gallant and Nychka (1987). Zhang and Davidian (2001) discussed that this class contains a variety of densities with flexible shapes such as normal, skewed, multimodal, and heavy- or thin-tailed densities, and unusual behavior including kinks, jumps, or oscillation is ruled out since densities in this class are sufficiently differentiable. As shown by Gallant and Nychka (1987) and discussed in Davidian and Giltinan (1995, Chapter 7), truncated series expansion may be used to approximate the densities in this class, which results in a density estimator, and the density approximation and estimator are referred to as seminonparametric (SNP). For inference, we apply the SNP representation to the density of the random effects $X_i$ and represent them as

$$X_i = \mu + RZ_i \quad (4.1)$$

where $\mu$ is a $(q \times 1)$ vector of parameters, $R$ is a $(q \times q)$ lower triangular matrix, and $Z_i$ is a $(q \times 1)$ random vector with smooth density approximated by

$$h_K(z) = P^2_K(z)\varphi(z) = \left( \sum_{|\lambda| \leq K} a_\lambda z^\lambda \right)^2 \varphi(z), \quad (4.2)$$
where $\lambda = (\lambda_1, \ldots, \lambda_q)$ is a $q$-dimensional vector of non-negative integers, $z^\lambda$ is the monomial $z_1^{\lambda_1} \cdots z_q^{\lambda_q}$ of order $|\lambda| = \sum_{k=1}^q \lambda_k$, $\varphi(z)$ is density of $N(0, I_q)$, $P_K(z)$ is a $K$th polynomial. For example, when $K = 0$, we have $a_{00} = 1$ and $P_K(z) \equiv 1$, thus $X_1 \sim N(\mu, RR^T)$, i.e., the SNP density representation includes normal as a special case. When $K = 1$ and $q = 2$, we have $z = (z_1, z_2)^T$ and $P_K(z) = a_{00} + a_{10} z_1 + a_{01} z_2$.

When $K = 2$ and $q = 2$, we have $z = (z_1, z_2)^T$ and $P_K(z) = a_{00} + a_{10} z_1 + a_{01} z_2 + a_{20} z_1^2 + a_{11} z_1 z_2 + a_{02} z_2^2$.

Because $h_K(z)$ is a density, the coefficients $a_\lambda$ must be chosen to ensure

$$\int h_K(z) dz = 1.$$  

As shown by Zhang and Davidian (2001), this can be accomplished by requiring $E\{P_K^2(V)\} = 1$, where $V \sim N(0, I_q)$, which is equivalent to imposing $a^T A a = 1$ for positive definite matrix $A$. For example, with $q = 2$, the coefficients in $P_K(z)$ may be collected into $d = (K + 1)(K + 2)/2$-dimensional vector $a = (a_{00}, \ldots, a_{0k}, a_{10}, \ldots, a_{1(k-1)}, \ldots, a_{k0})^T$ and $A$ is a $(d \times d)$ positive definite matrix where its $(i, j)$th element is $E(V_i^{i_1 + j_1})E(V_i^{i_2 + j_2})$ where $V_k$ is the $k$th component of $V$, and $i_1(j_1), i_2(j_2)$ are the subscripts corresponding to the $i$th ($j$th) element of $a$. In general, these kinds of expectations may be calculated using standard recursive formulae similar to Johnson and Kotz (1994). Let $A = BB^T$ and $c = B^T a$, then the density constraint can be written as $c^T c = 1$. Because $-c$ (and hence $-a$) yields the same density for $z$, $c$ lies on a half-unit $d$-dimensional sphere. Zhang and Davidian (2001) thus suggested to represent $c$ using the polar coordinate transformation:

$$c_1 = \sin(\phi_1),$$  

$$c_2 = \cos(\phi_1) \sin(\phi_2),$$
\[
\cd_{d-1} = \cos(\phi_1) \cos(\phi_2) \ldots \cos(\phi_{d-2}) \sin(\phi_{d-1}),
\]
\[
\cd_d = \cos(\phi_1) \cos(\phi_2) \ldots \cos(\phi_{d-1}),
\]
where \(-\pi/2 < \phi_k \leq \pi/2\) for \(k = 1, 2, \ldots, d-1\), hence the constraint is automatically satisfied and it leads to easy and stable implementation.

The order \(K\) acts as a tuning parameter controlling the degree of flexibility of the density \(h_K(z)\) shape. As suggested by Zhang and Davidian (2001), the value of \(K\) required to approximate a variety of complicated shapes such as skewness and multimodality is generally small, say, no greater than two. Letting \(r = \text{vech}(R)\) which denotes the nonzero elements of \(R\) and \(\phi = (\phi_1, \ldots, \phi_{d-1})^T\), the parameters of interest for a given \(K\) are \(\theta = (\beta^T, \rho, \sigma^2_u, \mu^T, r^T, \phi^T)^T\).

\section*{4.3 Estimation}

\subsection*{4.3.1 Likelihood Function}

We observe endpoints \((Y, W)\), where \(Y = (Y_1, \ldots, Y_n)^T\) and \(W = (W_1^T, \ldots, W_n^T)^T\). The observed data log-likelihood is

\[
l(\theta; Y, W) = \sum_{i=1}^{n} \log f(Y_i, W_i; \theta) \\
= \sum_{i=1}^{n} \log \left\{ \int f(Y_i, W_i|z; \theta) P_K^2(z; \phi) \varphi(z) dz \right\}, \tag{4.3}
\]

where \(f(Y_i, W_i; \theta)\) is the density of \((Y_i, W_i)\) for subject \(i\), and \(f(Y_i, W_i|Z_i; \theta)\) is the conditional density of \((Y_i, W_i)\) given \(Z_i\) for subject \(i\).
4.3.2 Normal Primary Model

Notice that when the primary endpoints are normally distributed given \( \mathbf{X}_i \); i.e.,
\[
Y_i = \beta_0^T \mathbf{C}_i + \beta_1^T \mathbf{X}_i + \epsilon_i \quad \text{where } \epsilon_i \sim N(0, \rho), \ i = 1, \ldots, n,
\]
f(\( Y_i, \mathbf{W}_i | \mathbf{Z}_i; \theta \)) \( \varphi(\mathbf{Z}_i) \) can be viewed as the joint density of (\( Y_i, \mathbf{W}_i \)) and \( \mathbf{Z}_i \), when \( \mathbf{Z}_i \) is assumed to be standard \( q \)-variate normal. Let \( g(Y_i, \mathbf{W}_i; \theta) \) denote the density of (\( Y_i, \mathbf{W}_i \)) and \( g(\mathbf{Z}_i|Y_i, \mathbf{W}_i; \theta) \) denote the conditional density of \( \mathbf{Z}_i \) given (\( Y_i, \mathbf{W}_i \)) when distribution of \( \mathbf{Z}_i \) is standard normal, then we have
\[
f(Y_i, \mathbf{W}_i|\mathbf{Z}_i; \theta) \varphi(\mathbf{Z}_i) = g(Y_i, \mathbf{W}_i; \theta)g(\mathbf{Z}_i|Y_i, \mathbf{W}_i; \theta)
\]
and hence
\[
f(Y_i, \mathbf{W}_i; \theta) = g(Y_i, \mathbf{W}_i; \theta) \int P_{\mathbf{K}}^2(z) g(z|Y_i, \mathbf{W}_i; \theta) dz
\]
\[
= g(Y_i, \mathbf{W}_i; \theta) E_{\mathbf{Z}_i|Y_i, \mathbf{W}_i, \theta}(P_{\mathbf{K}}^2(\mathbf{Z}_i)), \quad (4.4)
\]
where \( E_{\mathbf{Z}_i|Y_i, \mathbf{W}_i, \theta}(\cdot) \) represents expectation with respect to the conditional density of \( \mathbf{Z}_i \) given (\( Y_i, \mathbf{W}_i \)) when \( \mathbf{Z}_i \) is distributed as standard normal under \( \theta \). It is straightforward to show that \( g(Y_i, \mathbf{W}_i; \theta) \) is the normal density with mean
\[
\eta_i = \begin{pmatrix} \beta_0^T \mathbf{C}_i + \beta_1^T \mu \\ D_i \mu \end{pmatrix}
\]
and variance-covariance matrix
\[
\Omega_i = \begin{pmatrix} \beta_0^T \mathbf{R} \mathbf{R}^T \beta_1 + \rho & \beta_1^T \mathbf{R} \mathbf{R}^T D_i^T \\ D_i \mathbf{R} \mathbf{R}^T \beta_1 & D_i \mathbf{R} \mathbf{R}^T D_i^T + \sigma_u^2 \mathbf{I}_{m_i} \end{pmatrix},
\]
and that \( g(\mathbf{Z}_i|Y_i, \mathbf{W}_i; \theta) \) is the normal density with mean
\[
\eta_i^* = \begin{pmatrix} \beta_1^T \mathbf{R} \\ D_i \mathbf{R} \end{pmatrix}^T \Omega_i^{-1} \left\{ \begin{array}{c} Y_i \\ W_i \end{array} \right\} - \eta_i
\]
and variance-covariance matrix

$$\Omega^*_i = I_q - \begin{pmatrix} \beta_1^T R \\ D, R \end{pmatrix}^T \Omega_i^{-1} \begin{pmatrix} \beta_1^T R \\ D, R \end{pmatrix}. $$

Hence the observed data log-likelihood for $\theta$ becomes

$$l(\theta; Y, W) = \sum_{i=1}^n \log \{g(Y_i, W_i; \theta)\} + \sum_{i=1}^n \log \left[ E_{Z_i|Y_i, W_i, \theta} \{P^2_K(Z_i)\} \right].$$

(4.5)

where the first term is the usual likelihood function for the normal model for $(Y_i, W_i)$, which has a closed form, and the second term involves calculation of moments of a normal random vector with mean $\eta_i^*$ and variance matrix $\Omega_i^*, i = 1, \ldots, n$, which also has a closed form and can be calculated using the series representation of the moment generating function of a normal random variable similar to Zhang and Davidian (2001). Therefore, the log-likelihood of the observed data for $\theta$ has a closed form expression when the primary endpoint $Y_i$ is normally distributed given $X_i$, and standard optimization techniques can be applied to maximized $l(\theta; Y, W)$ to obtain parameter estimates. When the random effects have moderate dimension $q$, it is straightforward to implement. There can be increased computational burden when $q$ gets large.

### 4.3.3 EM Algorithm for Non-normal Primary Model

When the primary endpoint is not normally distributed given $X_i$, however, the likelihood of the observed data does not have a closed form representation. In principle, the observed data likelihood may be maximized with respect to $\theta$ for a given $K$ by numerically carrying out integrations, but direct optimization for such a high-dimensional case is prone to numerical instability. Therefore, we use an EM algorithm
to implement the approach.

Treating the random effects $X = (X_1, \ldots, X_n)^T$ as missing data, from this perspective, the completed data likelihood is

$$L_c(\theta; Y, W, X) = \prod_{i=1}^{n} \left\{ f(Y_i|X_i; \beta, \rho) f(W_i|X_i; \sigma_u^2) f(X_i; \mu, r, \phi) \right\}, \quad (4.6)$$

where $f(Y_i|X_i; \beta, \rho)$ is the conditional density of $Y_i$ given $X_i$, which follows the generalized linear model as shown in Section 1.2; $f(W_i|X_i; \sigma_u^2)$ is the conditional density of $W_i$ given $X_i$, which is a normal distribution according to the linear mixed model specified in Section 1.2; $f(X_i; \mu, r, \phi)$ is the density of $X_i$ and based on the SNP density representation

$$f(X_i; \mu, r, \phi) = h_K \{ R^{-1}(X_i-\mu) \} |R|^{-1} = P^2_K \{ R^{-1}(X_i-\mu) \} \varphi \{ R^{-1}(X_i-\mu) \} |R|^{-1}. \quad (4.7)$$

The estimator $\hat{\theta}$ is obtained by iterating between an E-step and an M-step. Let $\hat{\theta}^{(k)}$ denote the estimate at the $k$th iteration. In the E-step, we compute

$$Q(\theta|\hat{\theta}^{(k)}) = E\{ \log L_c(\theta; Y, W, X) | Y, W, \hat{\theta}^{(k)} \},$$

i.e., the conditional expectation of the complete data log-likelihood given the observed data and current parameter estimates. In the M-step, we maximize $Q(\theta|\hat{\theta}^{(k)})$ with respect to $\theta$ to obtain $\hat{\theta}^{(k+1)}$; i.e., the parameter estimates are updated by maximizing the conditional expectation.

The estimate of $\sigma_u^2$ at the $(k+1)$th iteration can be shown to be

$$\hat{\sigma}_u^2^{(k+1)} = \frac{\sum_{i=1}^{n} E \left\{ (W_i - D_iX_i)^T(W_i - D_iX_i) | Y_i, W_i; \hat{\theta}^{(k)} \right\}}{\sum_{i=1}^{n} m_i}. \quad (4.7)$$
The \((k+1)\)th iterates for the other parameters are not available in closed forms. \(\hat{\beta}_0^{(k+1)}, \hat{\beta}_1^{(k+1)}\) and \(\hat{\rho}^{(k+1)}\) are obtained by maximizing
\[
\sum_{i=1}^{n} \left[ \frac{Y_i \beta_0^T C_i + Y_i \beta_1^T E(X_i|Y_i, W_i; \hat{\theta}^{(k)}) - E\{b(\beta_0^T C_i + \beta_1^T X_i)|Y_i, W_i; \hat{\theta}^{(k)}\} + c(Y_i, \rho)}{a(\rho)} \right].
\]
(4.8)

They can also be obtained by a one-step Newton-Raphson approximation similar to Wulfsohn and Tsiatis (1997). Letting \(\gamma = (\beta^T, \rho)^T\), then the updates is
\[
\hat{\gamma}^{(k+1)} = \hat{\gamma}^{(k)} + I^{-1}_{\hat{\gamma}} S_{\hat{\gamma}}^{(k)},
\]
where \(S_{\hat{\gamma}}^{(k)}\) and \(I_{\hat{\gamma}}^{(k)}\) are the score vector and the information matrix for \(\gamma\) at the \(k\)th iteration, respectively. For example, for logistic primary endpoint model
\[
P(Y_i = 1|X_i) = \left[ 1 + \exp\{- (\beta_0^T C_i + \beta_1^T X_i)\} \right]^{-1},
\]
we have \(\gamma = \beta\) and
\[
S_{\hat{\beta}}^{(k)} = \left( \begin{array}{c} \sum_{i=1}^{n} \{Y_i - E(v_i|Y_i, W_i; \hat{\theta}^{(k)})\} C_i \\ \sum_{i=1}^{n} \{Y_i E(X_i|Y_i, W_i; \hat{\theta}^{(k)}) - E(v_i X_i|Y_i, W_i; \hat{\theta}^{(k)})\} \end{array} \right)
\]
\[
I_{\hat{\beta}}^{(k)} = \left( \begin{array}{cc} \sum_{i=1}^{n} E\{(v_i - v_i^2)|Y_i, W_i; \hat{\theta}^{(k)}\} C_i C_i^T & \sum_{i=1}^{n} C_i^T E\{(v_i - v_i^2)X_i^T|Y_i, W_i; \hat{\theta}^{(k)}\} \\ \sum_{i=1}^{n} E\{(v_i - v_i^2)X_i|Y_i, W_i; \hat{\theta}^{(k)}\} C_i & \sum_{i=1}^{n} E\{(v_i - v_i^2)X_i X_i^T|Y_i, W_i; \hat{\theta}^{(k)}\} \end{array} \right)
\]

where
\[
v_i = [1 + \exp\{- (\beta_0^{(k)} C_i + \beta_1^{(k)} X_i)\}]^{-1}.
\]

Because the parameters in the SNP representation including \(\hat{\mu}^{(k+1)}, \hat{\tau}^{(k+1)}\), and \(\hat{\phi}^{(k+1)}\) are separated from the other parameters in the completed data likelihood,
their iterative updates are obtained using standard optimizer to maximize

\[ \sum_{i=1}^{n} \left\{ E \left[ \log |P_K(X_i - \mu)| \right] | Y_i, W_i; \hat{\theta}^{(k)} \right\} + E \left[ \log |P_K^{-1}(X_i - \mu)| | Y_i, W_i; \hat{\theta}^{(k)} \right] \]

\[ -n \log |R|, \]

When \( K = 0 \), the algorithm reduces to the case for assumed normal random effects.

For more flexible models, implementation for various choices of \( K \) involves only additional complication of maximizing the above function, which only introduces moderate increase in computational burden.

All the expressions for the iterative updates of parameter estimates involve intractable integrals of the form

\[ E \{ s(X_i) | Y_i, W_i; \hat{\theta}^{(k)} \} = \int s(x_i) f(x_i | Y_i, W_i; \hat{\theta}^{(k)}) dx_i, \]

where \( s(X_i) \) is a function of \( X_i \) and \( f(X_i | Y_i, W_i; \theta) \) is the conditional density of \( X_i \) given the observed data \((Y_i, W_i)\), which have to be computed numerically. Notice that

\[ f(X_i | Y_i, W_i; \hat{\theta}^{(k)}) = \frac{f(Y_i, W_i, X_i; \hat{\theta}^{(k)})}{\int f(Y_i, W_i, x_i; \hat{\theta}^{(k)}) dx_i}, \quad (4.9) \]

where \( f(Y_i, W_i, X_i; \theta) \) is the joint density of \((Y_i, W_i)\) and \( X_i \), and

\[ f(Y_i, W_i, X_i; \hat{\theta}^{(k)}) = f(Y_i | X_i; \hat{\beta}^{(k)}, \hat{\rho}^{(k)}) f(W_i | X_i; \hat{\sigma}_u^{(k)}) f(X_i; \hat{\mu}^{(k)}, \hat{r}^{(k)}, \hat{\phi}^{(k)}), \]

which has an explicit expression. For implementation, we use adaptive Gauss-Hermite quadrature (e.g. Abramowitz and Stegun, 1972, p.890) to compute the integrals numerically.

In practice, in order to have good initial values for \( \theta \) for a fixed \( K \) for the EM Algorithm, we use the estimates of \( \beta, \rho, \) and \( \sigma_u^2 \) obtained from the conditional score
method described in Chapter 2 since the conditional score method is fast and straightforward to implement and yields consistent estimators regardless of the underlying random effects distribution. For the starting values for the parameters in the SNP density representation, we apply the approach of Zhang and Davidian (2001) to obtain the SNP estimates of $\mu, r$ and $\phi$ based on the linear mixed model for the longitudinal data with its tuning parameter equal to the value of the fixed $K$.

Once the parameter estimates $\hat{\theta}$ for a given $K$ are obtained, standard errors may be derived based on the inverse of the observed information, i.e., the second derivative matrix of the observed data log-likelihood evaluated at $\hat{\theta}$, which may be carried out by numerical integration (e.g., Gauss-Hermite quadrature) or Monte Carlo approximation similar to Chen, Zhang, and Davidian (2002). Preliminary simulation studies show that these two computing methods yield very similar results. Several other possible methods include profile score (i.e., the derivative of profile log-likelihood) and nonparametric bootstrap as discussed in Song et al. (2002).

After the SNP estimates of $\theta$ are obtained, inference on individuals may be carried out using the empirical Bayes approach of Davidian and Gallant (1992), where the empirical Bayes estimates of subject-specific random effects $X_i$ are obtained by finding the individual posterior modes for $X_i$, i.e.,

$$\hat{X}_i = \arg \max_{X_i} f(X_i|y_i, W_i; \hat{\theta}),$$

for a given $K$ or a preferred $K$. Alternatively, $X_i$ may be estimated by calculating $E(X_i|y_i, W_i; \hat{\theta})$. 
4.4 Selection of the Tuning Parameter $K$

The foregoing estimation procedure is developed based on a given value of the tuning parameter $K$. When $K$ is greater than 0, the density approximation of the random effects based on the SNP representation departs from normality, and the flexibility of the random effects density is controlled by the choice of the value of $K$. One can carry out inference for a fixed $K > 0$ to accommodate the approximation of a wide variety of densities of the random effects. Following Davidian and Gallant (1993) and Zhang and Davidian (2001), we choose $K$ by objectively inspecting information criteria which have the form of a penalized log-likelihood of the observed data

$$-l(\theta; Y, W)/N + C(N)(p_{net}/N)$$

evaluated at $\hat{\theta}$, where $N = n + \sum_{i=1}^{n} m_i$ is the total number of observations in the data set, $p_{net}$ is the number of free parameters in $\theta$ excluding $K$, and $C(N)$ is the penalty factor. For different choice of $C(N)$ which compensates for small value of $-l(\theta; Y, W)$ achieved by over-parameterization, different information criterion is considered: $C(N) = 1$ for Akaike Information Criterion (AIC), $C(N) = \log(\log N)$ for Hannan-Quinn Criterion (HQ), and $C(N) = 0.5\log N$ for Schwartz Information Criterion (BIC). For a given criterion, the value of $K$ which maximizes the penalized log-likelihood is preferred. Visual inspection of the estimated random effects density for $K$ values preferred by the information criteria is recommended for finding the most appropriate $K$ value among those preferred, as advocated by Davidian and Gallant (1993). Eastwood and Gallant (1991) showed that the SNP estimators based on such adaptive truncation rules for choosing $K$ achieve asymptotic normality. The
performance of these information criteria are demonstrated via simulation in Chapter 5.

4.5 Technical Details

4.5.1 Computation of $E\{s(X_i)|Y_i, W_i; \hat{\theta}^{(k)}\}$ for Logistic Primary Model

In the EM algorithm described in Section 4.3.3, all the expressions for the iterative updates of parameter estimates involve intractable integrals of the form

$$E\{s(X_i)|Y_i, W_i; \hat{\theta}^{(k)}\} = \int s(x_i)f(x_i|Y_i, W_i; \hat{\theta}^{(k)})dx_i,$$

where $s(X_i)$ is a function of $X_i$, which have to be computed numerically. Notice that

$$f(x_i|Y_i, W_i; \hat{\theta}^{(k)}) = \frac{f(Y_i, W_i, X_i; \hat{\theta}^{(k)})}{\int f(Y_i, W_i, x_i; \hat{\theta}^{(k)})dx_i}.$$

For logistic primary endpoint model

$$P(Y_i = 1|X_i) = \left[ 1 + \exp\{-\beta_0^T C_i + \beta_1^T X_i\} \right]^{-1},$$

we have

$$f(Y_i, W_i, X_i; \theta)$$

$$= f(Y_i|X_i; \beta)f(W_i|X_i; \sigma_u^2)f(X_i; \mu, r, \phi)$$

$$= \frac{\exp\{\beta_0^T C_i + \beta_1^T X_i\}Y_i}{1 + \exp(\beta_0^T C_i + \beta_1^T X_i)} \left(2\pi\sigma_u^2\right)^{-m/2} \exp\left\{ -\frac{(W_i - D_i X_i)^T (W_i - D_i X_i)}{2\sigma_u^2} \right\}$$

$$P_K \{R^{-1}(X_i - \mu)\} \varphi\{R^{-1}(X_i - \mu)\} |R|^{-1}.$$

Notice that

$$\exp\left\{ -\frac{(W_i - D_i X_i)^T (W_i - D_i X_i)}{2\sigma_u^2} \right\} \varphi\{R^{-1}(X_i - \mu)\}$$
\[ \varphi \{ H_i (X_i - \eta_i) \}, \]

where \( H_i \) and \( \eta_i \) satisfy

\[ H_i^T H_i = D_i D_i^T / \sigma_u^2 + (R^{-1})^T (R^{-1}), \]

\[ \eta_i = (H_i^T H_i)^{-1} \{ D_i^T W_i / \sigma_u^2 + (R^{-1})^T (R^{-1}) \mu \}. \]

Letting

\[ G(X_i, Y; \theta^{(k)}) = \frac{\exp \{ (\beta_0^T C_i + \beta_1^T X_i) Y_i \} \; P_k^2 \{ \tilde{R}^{(k)} (X_i - \mu^{(k)}) \} \; \tilde{R}^{(k)} \} \]

\[ 1 + \exp (\beta_0^T C_i + \beta_1^T X_i) \]

it follows that

\[ f(X_i|Y_i, W_i; \hat{\theta}^{(k)}) = \frac{G(X_i, Y; \hat{\theta}^{(k)}) \varphi \{ \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)}) \} \; \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)})}{\int G(X_i, Y; \hat{\theta}^{(k)}) \varphi \{ \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)}) \} dx_i}. \]

Hence,

\[ E \{ s(X_i)|Y_i, W_i; \hat{\theta}^{(k)} \} = \frac{\int s(X_i) G(X_i, Y; \hat{\theta}^{(k)}) \varphi \{ \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)}) \} dx_i}{\int G(X_i, Y; \hat{\theta}^{(k)}) \varphi \{ \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)}) \} dx_i} \]

\[ = \frac{\int s(X_i(\alpha_i)) G(X_i(\alpha_i), Y; \hat{\theta}^{(k)}) \exp (-\alpha_i^T \alpha_i) d\alpha_i}{\int G(X_i(\alpha_i), Y; \hat{\theta}^{(k)}) \exp (-\alpha_i^T \alpha_i) d\alpha_i}, \]

where \( \alpha_i = \tilde{H}_i^{(k)} (X_i - \tilde{\eta}_i^{(k)}) / \sqrt{2}. \)

For two dimensional random effects, this conditional expection can be approximated by \( M \)-point Gauss-Hermite quadrature

\[ \sum_{j=1}^{M} \sum_{l=1}^{M} s \{ X_i(\alpha_j, \alpha_l) \} G \{ X_i(\alpha_j, \alpha_l), Y; \hat{\theta}^{(k)} \} \omega_j \omega_l \]

\[ \sum_{j=1}^{M} \sum_{l=1}^{M} G \{ X_i(\alpha_j, \alpha_l), Y; \hat{\theta}^{(k)} \} \omega_j \omega_l, \]

where \( \alpha_j (\alpha_l) \) are abscissas with corresponding weights \( \omega_j (\omega_l) \), and \( X_i(\alpha_j, \alpha_l) = (\frac{X_1(\alpha_l)}{X_2(\alpha_l)}) = \sqrt{2} (\tilde{H}_i^{(k)})^{-1} (\alpha_l) + \tilde{\eta}_i^{(k)}. \)
4.5.2 Computation of Empirical Bayes Estimates

As discussed in Section 4.3.3, inference on individual subjects may be carried out using the empirical Bayes approach of Davidian and Gallant (1992) after the SNP estimates of $\theta$, denoted by $\hat{\theta}$, are obtained. The empirical Bayes estimates of subject-specific random effects $X_i$ are obtained by finding the individual posterior modes for $X_i$, i.e.,

$$\hat{X}_i = \arg \max X_i \ f(X_i | Y_i, W_i; \hat{\theta}),$$

for a given $K$ or a preferred $K$.

Notice that

$$f(X_i | Y_i, W_i; \hat{\theta}) = \frac{f(Y_i, W_i, X_i; \hat{\theta})}{\int f(Y_i, W_i, x_i; \theta) dx_i},$$

where

$$f(Y_i, W_i, X_i; \hat{\theta}) = f(Y_i | X_i; \hat{\beta}, \hat{\rho}) f(W_i | X_i; \hat{\sigma}_u^2) f(X_i; \hat{\mu}, \hat{r}, \hat{\phi}).$$

The integration in the denominator of the function may be approximated using Gauss-Hermite quadrature similar to that discussed in Section 4.5.1. Then the empirical Bayes estimate for subject $i$, $\hat{X}_i$, may be found by finding the value of $X_i$ which maximizes the above function.

4.5.3 Computation of Observed-data Log-likelihood

To calculate information criteria including AIC, HQ and BIC, we need to compute the observed-data log-likelihood after obtaining the parameter estimates, which is

$$l(\hat{\theta}; Y, W) = \sum_{i=1}^{n} \log f(Y_i, W_i; \hat{\theta})$$

$$= \sum_{i=1}^{n} \log \left\{ \int f(Y_i | z_i; \hat{\theta}) f(W_i | z_i; \hat{\theta}) P_K(z_i; \hat{\phi}) \varphi(z_i) dz_i \right\}.$$
which can be computed numerically using Gauss-Hermite quadrature similar to that discussed in Section 4.5.1.

### 4.5.4 Computation of the Observed Information

The standard errors for parameter estimates obtained from the semiparametric likelihood approach are estimated based on the observed information matrix

$$
\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(Y_i, W_i; \hat{\theta}) \frac{\partial}{\partial \theta^T} \log f(Y_i, W_i; \hat{\theta}).
$$

With $X_i = \mu + RZ_i$ and regularity conditions,

$$
\frac{\partial}{\partial \theta} \log f(Y_i, W_i; \hat{\theta}) = \frac{\partial}{\partial \theta} \log \int f(Y_i | z_i; \hat{\theta}) f(W_i | z_i; \hat{\theta}) P_K^2(z_i; \hat{\phi}) \varphi(z_i) dz_i
$$

$$
= \frac{\int \frac{\partial}{\partial \theta} f(Y_i | z_i; \hat{\theta}) f(W_i | z_i; \hat{\theta}) P_K^2(z_i; \hat{\phi}) \varphi(z_i) dz_i}{\int f(Y_i | z_i; \hat{\theta}) f(W_i | z_i; \hat{\theta}) P_K^2(z_i; \hat{\phi}) \varphi(z_i) dz_i},
$$

which can be computed numerically using Gauss-Hermite quadrature similar to that discussed in Section 4.5.1.

Alternatively, $\frac{\partial}{\partial \theta} \log f(Y_i, W_i; \hat{\theta})$ for subject $i$ can be computed numerically using Monte Carlo approximation given by

$$
\frac{\sum_{l=1}^{L} \frac{\partial}{\partial \theta} f(Y_i | z_i^{(l)}; \hat{\theta}) f(W_i | z_i^{(l)}; \hat{\theta}) P_K^2(z_i^{(l)}; \hat{\phi})}{\sum_{l=1}^{L} f(Y_i | z_i^{(l)}; \hat{\theta}) f(W_i | z_i^{(l)}; \hat{\theta}) P_K^2(z_i^{(l)}; \hat{\phi})},
$$

where $z_i^{(1)}, \ldots, z_i^{(L)}$ are samples from standard normal distribution $\varphi(z_i)$ for large $L$.

In practice, we use $L = 10000$. 

Chapter 5

Empirical Results For Semiparametric Likelihood Approach

5.1 Introduction

In the last chapter, a full likelihood procedure using SNP density approximated is proposed for inference on parameters of the joint model and estimation of the random effects distribution. An attractive feature of the semiparametric likelihood is that only a mild assumption (smoothness) is needed to be made on the distribution of the random effects. If the smoothness assumption holds, we would expect efficiency gains over the conditional estimation approach. The approach can be implemented using standard optimization techniques for normal primary model or via an EM algorithm for non-normal primary model. Implementation via the EM algorithm involves numerical integration, introducing increased computational burden.

In this chapter, we illustrate the finite sample performance and the utility of the semiparametric likelihood approach. In Section 5.2, the performance of the approach is demonstrated via simulation. The utility of the approach is shown via application
to the SWAN data in Section 5.3.

5.2 Simulation Studies

5.2.1 Simulation Studies Based on SWAN Design

To evaluate the performance of the approach, we conducted a simulation study based roughly on the design of the SWAN data. For each of \( n = 624 \) women in a data set, longitudinal log PDG data \( (W_{ij}, j = 1, \ldots, m_i) \) and the propensity for osteopenia \( (Y_i) \) were generated using the same joint model as described in Section 3.3, i.e., the mixed model for woman \( i \)'s log PDG profile

\[
W_{ij} = X_{1i} + X_{2i}(t_{ij} - 1.4)_{+} - 2X_{2i}(t_{ij} - 2.1)_{+} + U_{ij},
\]

and the primary logistic regression model

\[
P(Y_i = 1|X_i) = \left[1 + \exp\{- (\beta_0^T C_i + \beta_1^T X_i)\}\right]^{-1}.
\]

The parameters used are the same as the parameter estimates from the SWAN data analysis using SNP method with \( K = 2 \): \( \beta_0 = (\beta_{0, BMI}, \beta_{0, BMI}, \beta_{0, age}, \beta_{0, B}, \beta_{0, W}, \beta_{0, C})^T = (-7.22, 0.29, 0.01, 0.54, -0.22, -0.68)^T \), \( \beta_1 = (\beta_{11}, \beta_{12})^T = (0.15, 0.13) \), and \( \sigma_u^2 = 0.33 \). A bimodal 79-21 mixture of normals was considered for the distribution of the bivariate random effects \( X_i = (X_{1i}, X_{2i})^T \), which has characteristics similar to the estimated density in SWAN: \( E(X_{1i}) = -0.48 \), \( E(X_{2i}) = 2.99 \), \( \text{var}(X_{1i}) = 0.29 \), \( \text{var}(X_{2i}) = 2.49 \), \( \text{cov}(X_{1i}, X_{2i}) = 0.17 \), and the same apparent bimodal structure.

For simplicity, the time points \( (t_{ij}, j = 1, \ldots, m_i) \) were limited to 5 for each woman (i.e., \( m_i = 5 \)) by taking the first time point, the last time point, and three approximately equally-spaced time points in between from all the time points for that
woman in the SWAN data. For woman \( i \), we took the same baseline covariates \( \mathbf{C}_i = (1, C_{\text{BMI},i}, C_{\text{age},i}, C_{\text{B},i}, C_{\text{W},i}, C_{\text{C},i})^T \) as those in the SWAN data. The “reliability ratio” defined in Section 5.3 is \((\lambda_1, \lambda_2) = (0.78, 0.75)\), reflecting mild measurement errors similar to those estimates in the SWAN data analysis.

For each of the 200 Monte Carlo data sets generated, \((\beta_0, \beta_{0,\text{BMI}}, \beta_{0,\text{age}}, \beta_{0,\text{B}}, \beta_{0,\text{W}}, \beta_{0,\text{C}}, \beta_{11}, \beta_{12})^T\) was estimated the six ways: (i) by the naive approach in Section 1.1 (N); (ii) using regression calibration (RC) (Wang et al., 2000, sec. 4.1); (iii) by refined regression calibration, which applies a probit approximation to the logistic to approximate the pseudo-EEE (RR) (Wang et al., 2000, sec. 4.2); (iv) via the sufficiency score method (SS) described in Chapter 2; (v) via the conditional score method (CS) described in Chapter 2; and (vi) via the proposed semiparametric full likelihood procedure described in Section 4.3 (SNP) with \( K = 0, 1, \) and 2. In method (vi), eight-point Gauss-Hermite quadrature was used to calculate intractable integrals. Preliminary studies revealed that increasing the number of abscissas beyond eight could not improve performance appreciably. Notice that method (vi) also provides estimates of the parameters in the SNP density representation. In method (vi), the fit preferred by each of AIC, HQ, and BIC was recorded. Standard errors for (i) were obtained from usual logistic regression formulae, ignoring least-squares imputation of \( \mathbf{X}_i \), and those for (ii)–(v) were derived from the empirical sandwich approach. The naive estimate was used as the starting value for (iii)–(v). The standard errors and the starting value for (vi) were obtained using methods discussed in Section 4.3. 95% Wald confidence intervals were constructed using the standard normal critical value of 1.96.
In Table 5.1, simulation results for naive estimator, RR estimator, CS estimator, SNP estimator with $K = 0$, and SNP estimator preferred by HQ are reported; results for RC estimator are almost identical to those for RR estimator, and results of SNP estimators preferred by the three information criteria are very similar. Over all 200 data sets, AIC selected $K = 0$ 2.5% of the time, $K = 1$ 36.5% of the time, and $K = 2$ 61% of the time; HQ selected $K = 0, 1$, and 2 4%, 42%, and 54% of the time, respectively; and BIC selected $K = 0, 1$, and 2 11%, 59.5%, and 29.5% of the time, respectively. For the proposed SNP estimators preferred by the AIC, BIC, and HQ, bias is negligible and coverage probabilities attain the nominal level. The RC, RR, SS and CS estimators exhibit similar good performance with negligible or small bias and coverages close to nominal level. The naive estimator shows appreciable bias and poor coverage probabilities for the estimation of the coefficients of the random effects. The SNP estimators preferred by information criteria also show slight efficiency gains over the competing methods. These simulation results coincide with what we observe in the SWAN data analysis.

5.2.2 Simulation Studies Under Generic Scenarios

To further examine the robustness and performance of the approach, we carried out simulations under a range of true random effects distributions. Similar to Section 3.2, reported here are the results for binary primary endpoint. Longitudinal data were generated according to the mixed model

$$W_{ij} = X_{1i} + X_{2i}t_{ij} + U_{ij},$$
where \( t_{ij} \sim N(j - 1, 0.01) \) for \( i = 1, \ldots, n \), \( j = 1, \ldots, m_i \), and \( U_{ij} \sim N(0, \sigma_u^2) \). The binary primary endpoint was generated through a logistic model

\[
P(Y_i = 1|\mathbf{X}_i) = \frac{1}{1 + \exp\{-(\beta_0 + \mathbf{\beta}_1^T \mathbf{X}_i)\}}
\]

with \( \mathbf{X}_i = (X_{1i}, X_{2i}) \) and \( \mathbf{\beta}_1 = (\beta_{11}, \beta_{12})^T \). In all cases, \( E(X_{1i}) = E(X_{2i}) = 0.5 \), \( \text{var}(X_{1i}) = 1.0 \), \( \text{var}(X_{2i}) = 0.64 \), \( \text{cov}(X_{1i}, X_{2i}) = -0.2 \), \( \beta_0 = -2.5 \), and \( \mathbf{\beta}_1 = (3.0, 2.0)^T \). In the simulations reported here, \( n = 500 \); \( \sigma_u^2 = 0.5 \), and \( m_i \leq 5 \); five \( W_{ij} \) were generated for each subject \( i \), with each arbitrarily missing with probability 0.05 to create imbalance. Each scenario below involved 200 Monte Carlo data sets. For all scenarios, the “reliability ratio” is \( (\lambda_1, \lambda_2) = (0.72, 0.92) \), suggesting moderate (mild) measurement error associated with \( \mathbf{\beta}^{X_1}_i \) (\( \mathbf{\beta}^{X_2}_i \)).

Four scenarios for the true \( \mathbf{X}_i \) distribution were considered: (a) \( \mathbf{X}_i \) bivariate normal, which is the usual assumption; (b) \( \mathbf{X}_i \) distributed as a bimodal 70-30 mixture of normals, representing densities with more than one mode; (c) \( \mathbf{X}_i \) following the bivariate skew-normal (Azzalini and Dalla Valle, 1996), with coefficients of skewness \(-0.10\) and \(0.85\) for \( X_{1i} \) and \( X_{2i} \), respectively, representing skewed densities; and (d) \( \mathbf{X}_i \) bivariate \( t \) with 5 degrees of freedom, representing densities with heavy-tails. In each case, the distribution was scaled to have the moments above. For each data set under each scenario, \( (\beta_0, \beta_{11}, \beta_{12}, \sigma_u^2)^T \) was estimated the same six ways as described in Section 5.2.1: (i) \( N \), (ii) RC, (iii) RR, (iv) SS, (v) CS, and (vi) SNP. For each of the 200 data sets simulated, the joint model was fit three times with the density of \( \mathbf{X}_i \) represented by the SNP approximation with \( K = 0, 1, 2 \), and the three information criteria were calculated for each fit. Since method (vi) is a full likelihood approach, whether it could gain efficiency over competing methods such as methods (iv) and
(v) is of interest. Methods (i), (ii) and (iii) require normality of $X_i$, so comparison of their performance with method (vi) under violations of the normality assumption is also of interest.

Results for estimation are given in Table 5.2 and Table 5.3. In all cases, for the proposed SNP estimators preferred by the AIC, BIC, and HQ, bias is negligible or small, standard errors track the Monte Carlo standard deviations, and coverage probabilities attain the nominal level regardless of the underlying random effects distribution. The naive estimator always shows considerable bias and coverage probabilities of the naive estimator are far below the nominal level. Under normality, the RR estimator, SS estimator, CS estimator, and proposed SNP estimators preferred by the information criteria exhibit negligible bias and attain nominal coverage. However, under departures from normality, RC and RR estimators can show degradation of performance. The RC estimator is unacceptably biased except for bimodal $X_i$; for normal $X_i$, this estimator shows appreciable bias, which verifies concerns of Wang et al. (2000) and is similar to the observations in Section 3.2. Coverage probabilities of the RC estimators are well below the nominal level. Interestingly, under bimodality, the RR estimator shows considerable bias, and standard errors under-represent empirical sampling variation. For the skewed scenario, with only mild skewness in the $X_{2i}$ dimension, the RR estimator achieves good performance; however, for the “heavy-tailed” bivariate $t$, although bias of the RR estimator is not large (although larger than that of SS, CS, and SNP estimators), inferences are flawed, with coverage falling far short of nominal. Among the existing methods, SS and CS estimators exhibit good performance with negligible or small bias and coverages close to nominal
level. However, the SNP estimators preferred by information criteria show efficiency gains over CS estimator (which performs the best among the existing methods) with relative efficiency ranging from 1.02 to 1.40 with respective to the CS estimator and even higher efficiency gain with respect to the SS estimator, which shows that the CS and SS estimators (where the random effects distribution is completely unrestricted) are inefficient relative to the SNP estimators preferred by information criteria and demonstrates the potential for improved precision offered by the full likelihood based approach with only requirement of the class of smooth densities when the true density is smooth.

The advantage of estimating the random effects distribution is again appreciated via the illustration of Figure 5.1: Figure 5.1a shows the true density used to generate data in simulation for the bimodal scenario (b), Figure 5.1b shows the Monte Carlo average of joint density estimate preferred by HQ, Figures 5.1c and 5.1d show the Monte Carlo average of estimated marginal densities preferred by each information criterion, and Figures 5.1e and 5.1f show the 200 individual density estimates preferred by HQ. The estimated densities are close to the corresponding true densities; the pattern of capturing the actual shape of the underlying random effects densities for other scenarios is similar. It illustrates that the SNP density representation is capable of approximating the true underlying density and supports the flexibility of the approach, with only a few of the 200 SNP fits not capturing the bimodality of the true density as shown in Figure 5.1e. It is not unreasonable that allowing the additional flexibility to represent more complex densities would cause occasional over-modeling; however, only a few of SNP fits preferred by HQ has a spurious extra mode or bump.
In practice, visual inspection of the estimated densities is recommended to take a pragmatic approach and to choose $K$ to avoid over-interpretation of the behavior.

Table 5.4 summarizes the good performance of the information criteria. Close to 100% of the time, the information criteria including AIC, HQ, and BIC correctly selected $K = 0$ under the normal random effects scenario and selected $K > 0$ under the nonnormal scenarios, with the BIC criterion being slightly conservative. $K$ only plays the role of a tuning parameter and SNP density with larger $K$ does not necessarily imply a greater number of modes. In practice, inspection of information criteria should be augmented with examination of plots of the estimated densities for model selection. Table 5.4 also demonstrate that AIC tends to prefer larger $K$, while BIC tends to prefer smaller $K$ with HQ intermediate, echoing behavior observed in Zhang and Davidian (2001) and in Section 5.2.1, which illustrates the ability of these information criteria to detect departures from normality. In all cases, the starting value strategy described in Section 4.3 encountered no problems identifying the apparent consistent solution.

5.3 Application to SWAN

We illustrate the proposed semiparametric likelihood method by applying it to the SWAN data. The joint model as described in Section 3.3 was fitted using the same six ways as described in Section 5.2.1: (i) N; (ii) RC; (iii) RR; (iv) SS; (v) CS; and (vi) SNP with $K = 0, 1,$ and 2. Estimated standard errors for parameters of interest were obtained using the same methods as described in Section 5.2.1.

Table 5.5 shows results by the proposed full likelihood approach for $K = 0, 1,$
and 2, where in the SNP representation, \( X_i = \mu + RZ_i \) with \( \mu = (\mu_1, \mu_2)^T \), \( r = (r_{11}, r_{21}, r_{22})^T \), and \( Z_i = (Z_{1i}, Z_{2i})^T \). The estimates of the parameters in the SNP density representation including \( \mu^T \), \( r^T \), and \( \phi^T \) are given in Table 5.6. Table 5.7 presents the performance of the information criteria for \( K = 0, 1, \) and 2.

As shown in Table 5.5, the data analysis results illustrate that SNP estimates have slightly smaller (or comparable) standard errors than those estimates obtained from the other approaches as shown in Table 3.2, reflecting slight gain in efficiency. All three information criteria prefer the SNP representation with \( K = 2 \) as shown in Table 5.7, although estimates for \( K = 2 \) are almost identical to \( K = 1 \) with a slight increase in log-likelihood, supporting the contention of a departure from normality as observed in Section 1.1. This also illustrates a useful feature of our approach as a model selection device because the SNP representation contains the normal, fitting the model with \( K = 0 \) and then increasing \( K \) provides both visual and empirical information with which to gauge the validity of the usual normality assumption.

Similar to the observations in Section 3.3, inferences on the coefficients of baseline covariates are similar across all methods and indicate strong association between absence of osteopenia and BMI after taking into account age, ethnicity, and hormone profile, consistent with observations in previous studies. Inferences on the relationship between osteopenia and pre-14-day log PDG level and post-14-day rise/fall are also similar, although point estimates are more disparate across methods. The proposed estimators suggest that absence of osteopenia may be positively associated with post-14-day “slope” of log PDG profile, after taking into account other factors. Age effect is insignificant due to limited age range in these women (43 ~ 53). These results are
consistent with epidemiology studies.

The consistent inferences across methods are not unexpected because of similar arguments to those in Section 3.3. The estimated “reliability ratio” is \( \lambda = (0.88, 0.84) \) for the SWAN data, so the magnitude of “measurement error” in both components is not great. Although bimodal, the estimated density for \( X_i \) does not deviate considerably from normality; coefficients of skewness based on the density estimate are mild with \((-0.13, -0.87)\) for \((X_{1i}, X_{2i})^T\). Another possible reason may be that the number of longitudinal PDG measurements for each woman in the SWAN was in the range of \(8 \sim 25\), which is large enough to provide sufficient information on the longitudinal process so that even the naive approach can yield valid inference.

The ability of the proposed approach to estimate random effects density may provide insights which could aid subject-matter understanding. Figure 5.2a shows the estimated bivariate density of the random effects \( X_i \) for the preferred fit \((K = 2)\), which indicates the presence of an obvious second mode and suggests that there is evidence of deviation of random effects distribution from bivariate normality. Another perspective of the estimated density is that the subject-specific empirical Bayes estimates of \( X_i \) clump into two distinct subgroups which may be identified based on slope as shown in Figure 5.2b. Figures 5.2c and 5.2d depict the corresponding estimated marginal densities for \( X_{1i} \) and \( X_{2i} \), from which the intercepts seem to be normally distributed, while the shape of the estimated density of slopes shows evidence of bimodality. These results suggest the possibility of a subpopulation of individuals with lower rate of after-14-day symmetric rise and fall of log PDG after adjusting for the effects of BMI, age, and ethnicity. This pattern may be partially explained by the
observed covariates. Figure 5.3 shows the plots of empirical Bayes estimates of slopes versus observed covariates (box-plots for age and ethnicity groups), which suggests BMI and age are significantly different between the two groups of women. Since the primary logistic model incorporates BMI, age, and ethnicity effects (although BMI and age are associated with the bimodality), it is possible that the apparent non-normal pattern for slopes in the estimated density may reflect that an important covariate, which is not in the data set, has not been taken into account. The non-normal behavior in the estimated density also suggests that inference on individual effects under the usual normal random effects assumption could be misleading.

In the simulation we carried out based on the estimates and design for the SWAN data in Section 5.2.1, all methods yielded similar inferences. Evidently, the design and variation/correlation configuration for these data leads to only minimal differences among the procedures. Nonetheless, use of the proposed method offers the analyst assurance of credible estimation of the relationship. Moreover, the ability to estimate the random effects distribution provides data analyst considerable insight into the features of the subject-specific effects and may raise issues for further investigation.

### 5.4 Summary

A relevant framework to joint model a primary outcome and longitudinal profiles of a continuous response is to assume the longitudinal data follow a mixed model and the primary endpoint follows a generalized linear model depending on the underlying random effects for longitudinal measurements and other covariates. Interest may focus on the association between the primary endpoint and features of longitudinal
profiles; the longitudinal data process and the underlying random effects distribution may also be of interest. Normality of the random effects distribution is a routine assumption made by most existing methods for fitting the joint model, but such a restrictive distributional assumption may be unrealistic, raising concerns of sensitivity to violation of this assumption and of misleading inference on individual effects. We relax this assumption and propose a full likelihood-based approach which requires only the assumption that the random effects have a smooth density. Implementation using EM algorithm is described. The proposed method yields valid inference and shows potential efficiency gains over competing methods under various random effects distributions. Moreover, it provides reliable estimator for the underlying distribution of random effects, offering insight on the study population. We illustrate the approach via simulation and by application to the SWAN data.

The method has been implemented using the EM algorithm for logistic primary model; SAS code is available from the author.
Table 5.1: Simulation results for SWAN design. True parameter values are in parentheses below the parameters. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score; SNP, proposed semiparametric likelihood method, where $K = 0$ denotes normal random effects, and HQ denotes the estimates preferred by HQ. MEAN, Monte Carlo average; RB, relative bias (%); MSE, mean square error; CP, Monte Carlo coverage probability of 95% Wald confidence interval.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>RR</th>
<th>CS</th>
<th>SNP ($K = 0$)</th>
<th>SNP (HQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MEAN</td>
<td>MSE</td>
<td>MEAN</td>
<td>MSE</td>
<td>MEAN</td>
</tr>
<tr>
<td>(-7.22)</td>
<td>1.16</td>
<td>0.96</td>
<td>2.06</td>
<td>0.96</td>
<td>2.24</td>
</tr>
<tr>
<td>$\beta_{0,BMI}$</td>
<td>0.30</td>
<td>0.001</td>
<td>0.30</td>
<td>0.001</td>
<td>0.30</td>
</tr>
<tr>
<td>(0.29)</td>
<td>2.34</td>
<td>0.95</td>
<td>2.66</td>
<td>0.95</td>
<td>2.73</td>
</tr>
<tr>
<td>$\beta_{0,age}$</td>
<td>0.01</td>
<td>0.002</td>
<td>0.01</td>
<td>0.002</td>
<td>0.01</td>
</tr>
<tr>
<td>(0.01)</td>
<td>8.06</td>
<td>0.96</td>
<td>7.87</td>
<td>0.95</td>
<td>4.81</td>
</tr>
<tr>
<td>$\beta_{0,B}$</td>
<td>0.52</td>
<td>0.148</td>
<td>0.52</td>
<td>0.149</td>
<td>0.52</td>
</tr>
<tr>
<td>(0.54)</td>
<td>-3.60</td>
<td>0.97</td>
<td>-3.28</td>
<td>0.97</td>
<td>-3.57</td>
</tr>
<tr>
<td>$\beta_{0,W}$</td>
<td>-0.23</td>
<td>0.071</td>
<td>-0.23</td>
<td>0.072</td>
<td>-0.23</td>
</tr>
<tr>
<td>(-0.22)</td>
<td>5.36</td>
<td>0.97</td>
<td>4.77</td>
<td>0.96</td>
<td>2.89</td>
</tr>
<tr>
<td>$\beta_{0,C}$</td>
<td>-0.68</td>
<td>0.073</td>
<td>-0.68</td>
<td>0.073</td>
<td>-0.68</td>
</tr>
<tr>
<td>(-0.68)</td>
<td>-0.72</td>
<td>0.94</td>
<td>-0.43</td>
<td>0.94</td>
<td>-0.34</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.17</td>
<td>0.032</td>
<td>0.15</td>
<td>0.060</td>
<td>0.15</td>
</tr>
<tr>
<td>(0.15)</td>
<td>12.94</td>
<td>0.93</td>
<td>3.06</td>
<td>0.94</td>
<td>2.51</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.10</td>
<td>0.004</td>
<td>0.13</td>
<td>0.006</td>
<td>0.13</td>
</tr>
<tr>
<td>(0.13)</td>
<td>-24.21</td>
<td>0.93</td>
<td>-1.95</td>
<td>0.98</td>
<td>-1.99</td>
</tr>
</tbody>
</table>
Table 5.2: Simulation results for normal and bimodal underlying random effect distributions for \( P(Y_i = 1 | X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1} \), with true values \( \beta_0 = -2.5 \), \( \beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T \), \( \sigma^2_u = 0.5 \), \( n = 500 \). RB, relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval; RE, relative efficiency, i.e., Monte Carlo mean square errors for CS divided by that for the indicated estimator. Estimators: N, naive; RC, regression calibration; RR, refined regression calibration; SS, sufficiency score; CS, conditional score; SNP, proposed semiparametric method, where \( K = 0 \) denotes normal random effects, and AIC, HQ, and BIC are the estimates preferred by AIC, HQ, and BIC, respectively.

<table>
<thead>
<tr>
<th>Method</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>RE</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-25.9</td>
<td>0.22</td>
<td>0.22</td>
<td>0.18</td>
<td>0.30</td>
<td>-16.7</td>
<td>0.25</td>
<td>0.27</td>
<td>0.62</td>
<td>0.90</td>
</tr>
<tr>
<td>RC</td>
<td>-18.3</td>
<td>0.26</td>
<td>0.23</td>
<td>0.46</td>
<td>0.51</td>
<td>-5.2</td>
<td>0.30</td>
<td>0.30</td>
<td>0.87</td>
<td>2.01</td>
</tr>
<tr>
<td>RR</td>
<td>-0.7</td>
<td>0.38</td>
<td>0.36</td>
<td>0.95</td>
<td>0.99</td>
<td>32.2</td>
<td>0.91</td>
<td>0.70</td>
<td>0.99</td>
<td>0.15</td>
</tr>
<tr>
<td>SS</td>
<td>1.2</td>
<td>0.40</td>
<td>0.40</td>
<td>0.97</td>
<td>0.90</td>
<td>2.7</td>
<td>0.47</td>
<td>0.47</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>CS</td>
<td>1.0</td>
<td>0.38</td>
<td>0.40</td>
<td>0.96</td>
<td>1.00</td>
<td>2.2</td>
<td>0.46</td>
<td>0.47</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP (K = 0)</td>
<td>0.5</td>
<td>0.35</td>
<td>0.36</td>
<td>0.96</td>
<td>1.16</td>
<td>22.8</td>
<td>0.65</td>
<td>0.56</td>
<td>0.98</td>
<td>0.29</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>0.4</td>
<td>0.35</td>
<td>0.36</td>
<td>0.96</td>
<td>1.17</td>
<td>1.5</td>
<td>0.42</td>
<td>0.43</td>
<td>0.96</td>
<td>1.23</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>0.5</td>
<td>0.35</td>
<td>0.36</td>
<td>0.96</td>
<td>1.16</td>
<td>1.7</td>
<td>0.41</td>
<td>0.43</td>
<td>0.96</td>
<td>1.23</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>0.5</td>
<td>0.35</td>
<td>0.36</td>
<td>0.96</td>
<td>1.16</td>
<td>2.8</td>
<td>0.41</td>
<td>0.43</td>
<td>0.95</td>
<td>1.23</td>
</tr>
<tr>
<td>( \beta_{11} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-32.5</td>
<td>0.17</td>
<td>0.19</td>
<td>0.00</td>
<td>0.19</td>
<td>-19.5</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.38</td>
</tr>
<tr>
<td>RC</td>
<td>-18.1</td>
<td>0.25</td>
<td>0.23</td>
<td>0.39</td>
<td>0.53</td>
<td>-1.1</td>
<td>0.28</td>
<td>0.27</td>
<td>0.92</td>
<td>1.91</td>
</tr>
<tr>
<td>RR</td>
<td>-0.3</td>
<td>0.43</td>
<td>0.42</td>
<td>0.93</td>
<td>1.05</td>
<td>42.2</td>
<td>1.09</td>
<td>0.76</td>
<td>0.99</td>
<td>0.05</td>
</tr>
<tr>
<td>SS</td>
<td>2.1</td>
<td>0.46</td>
<td>0.49</td>
<td>0.95</td>
<td>0.87</td>
<td>3.1</td>
<td>0.40</td>
<td>0.40</td>
<td>0.96</td>
<td>0.91</td>
</tr>
<tr>
<td>CS</td>
<td>2.0</td>
<td>0.43</td>
<td>0.48</td>
<td>0.94</td>
<td>1.00</td>
<td>2.8</td>
<td>0.38</td>
<td>0.40</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP (K = 0)</td>
<td>0.9</td>
<td>0.39</td>
<td>0.42</td>
<td>0.94</td>
<td>1.25</td>
<td>32.3</td>
<td>0.74</td>
<td>0.58</td>
<td>0.73</td>
<td>0.10</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>0.6</td>
<td>0.38</td>
<td>0.42</td>
<td>0.94</td>
<td>1.28</td>
<td>3.4</td>
<td>0.37</td>
<td>0.36</td>
<td>0.94</td>
<td>1.02</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>0.9</td>
<td>0.39</td>
<td>0.42</td>
<td>0.94</td>
<td>1.25</td>
<td>3.8</td>
<td>0.37</td>
<td>0.36</td>
<td>0.95</td>
<td>1.02</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>0.9</td>
<td>0.39</td>
<td>0.42</td>
<td>0.94</td>
<td>1.25</td>
<td>5.4</td>
<td>0.37</td>
<td>0.37</td>
<td>0.94</td>
<td>0.92</td>
</tr>
<tr>
<td>( \beta_{12} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-15.0</td>
<td>0.20</td>
<td>0.20</td>
<td>0.62</td>
<td>0.72</td>
<td>-9.7</td>
<td>0.22</td>
<td>0.24</td>
<td>0.85</td>
<td>1.23</td>
</tr>
<tr>
<td>RC</td>
<td>-17.8</td>
<td>0.22</td>
<td>0.20</td>
<td>0.51</td>
<td>0.53</td>
<td>-13.0</td>
<td>0.23</td>
<td>0.24</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td>RR</td>
<td>-0.2</td>
<td>0.30</td>
<td>0.29</td>
<td>0.97</td>
<td>1.04</td>
<td>24.4</td>
<td>0.58</td>
<td>0.45</td>
<td>0.99</td>
<td>0.18</td>
</tr>
<tr>
<td>SS</td>
<td>1.9</td>
<td>0.32</td>
<td>0.32</td>
<td>0.97</td>
<td>0.88</td>
<td>1.7</td>
<td>0.32</td>
<td>0.34</td>
<td>0.96</td>
<td>1.01</td>
</tr>
<tr>
<td>CS</td>
<td>1.7</td>
<td>0.30</td>
<td>0.31</td>
<td>0.95</td>
<td>1.00</td>
<td>1.5</td>
<td>0.32</td>
<td>0.34</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP (K = 0)</td>
<td>1.0</td>
<td>0.28</td>
<td>0.29</td>
<td>0.97</td>
<td>1.22</td>
<td>17.1</td>
<td>0.42</td>
<td>0.39</td>
<td>0.96</td>
<td>0.35</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>1.0</td>
<td>0.28</td>
<td>0.29</td>
<td>0.97</td>
<td>1.22</td>
<td>1.4</td>
<td>0.31</td>
<td>0.33</td>
<td>0.96</td>
<td>1.09</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>1.0</td>
<td>0.28</td>
<td>0.29</td>
<td>0.97</td>
<td>1.22</td>
<td>1.8</td>
<td>0.31</td>
<td>0.33</td>
<td>0.96</td>
<td>1.04</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>1.0</td>
<td>0.28</td>
<td>0.29</td>
<td>0.97</td>
<td>1.22</td>
<td>2.7</td>
<td>0.31</td>
<td>0.33</td>
<td>0.96</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Table 5.3: Simulation results for skewed and heavy-tailed underlying random effect distributions for $P(Y_i = 1|X_i) = \{1 + \exp(\beta_0 + \beta_{11}X_{1i} + \beta_{12}X_{2i})\}^{-1}$, with true values $\beta_0 = -2.5$, $\beta_1 = (\beta_{11}, \beta_{12})^T = (3.0, 2.0)^T$, $\sigma_u^2 = 0.5$, $n = 500$; abbreviations for estimates and methods are as in Table 5.2.

<table>
<thead>
<tr>
<th>Method</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>RE</th>
<th>RB (%)</th>
<th>SD</th>
<th>SE</th>
<th>CP</th>
<th>RE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(c) Skewed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(d) Bivariate $t_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-25.1</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19</td>
<td>0.33</td>
<td>-28.9</td>
<td>0.20</td>
<td>0.21</td>
<td>0.09</td>
<td>0.28</td>
</tr>
<tr>
<td>RC</td>
<td>-17.2</td>
<td>0.25</td>
<td>0.23</td>
<td>0.53</td>
<td>0.59</td>
<td>-21.9</td>
<td>0.23</td>
<td>0.23</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>RR</td>
<td>1.9</td>
<td>0.40</td>
<td>0.39</td>
<td>0.97</td>
<td>0.89</td>
<td>-7.1</td>
<td>0.34</td>
<td>0.33</td>
<td>0.87</td>
<td>1.09</td>
</tr>
<tr>
<td>SS</td>
<td>4.2</td>
<td>0.40</td>
<td>0.43</td>
<td>0.98</td>
<td>0.86</td>
<td>2.1</td>
<td>0.42</td>
<td>0.42</td>
<td>0.98</td>
<td>0.89</td>
</tr>
<tr>
<td>CS</td>
<td>3.4</td>
<td>0.37</td>
<td>0.42</td>
<td>0.98</td>
<td>1.00</td>
<td>2.1</td>
<td>0.40</td>
<td>0.44</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP ($K = 0$)</td>
<td>2.4</td>
<td>0.35</td>
<td>0.37</td>
<td>0.98</td>
<td>1.19</td>
<td>-6.5</td>
<td>0.31</td>
<td>0.34</td>
<td>0.93</td>
<td>1.26</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>0.1</td>
<td>0.34</td>
<td>0.35</td>
<td>0.96</td>
<td>1.29</td>
<td>-3.5</td>
<td>0.35</td>
<td>0.35</td>
<td>0.94</td>
<td>1.19</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>0.2</td>
<td>0.34</td>
<td>0.35</td>
<td>0.96</td>
<td>1.27</td>
<td>-4.0</td>
<td>0.35</td>
<td>0.35</td>
<td>0.94</td>
<td>1.18</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>1.1</td>
<td>0.34</td>
<td>0.36</td>
<td>0.97</td>
<td>1.23</td>
<td>-5.0</td>
<td>0.33</td>
<td>0.34</td>
<td>0.94</td>
<td>1.24</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-31.6</td>
<td>0.19</td>
<td>0.19</td>
<td>0.00</td>
<td>0.21</td>
<td>-36.9</td>
<td>0.16</td>
<td>0.19</td>
<td>0.00</td>
<td>0.18</td>
</tr>
<tr>
<td>RC</td>
<td>-16.6</td>
<td>0.27</td>
<td>0.24</td>
<td>0.46</td>
<td>0.63</td>
<td>-23.4</td>
<td>0.24</td>
<td>0.23</td>
<td>0.15</td>
<td>0.41</td>
</tr>
<tr>
<td>RR</td>
<td>2.8</td>
<td>0.49</td>
<td>0.45</td>
<td>0.95</td>
<td>0.81</td>
<td>-8.8</td>
<td>0.38</td>
<td>0.37</td>
<td>0.81</td>
<td>1.05</td>
</tr>
<tr>
<td>SS</td>
<td>5.3</td>
<td>0.47</td>
<td>0.52</td>
<td>0.95</td>
<td>0.81</td>
<td>3.5</td>
<td>0.50</td>
<td>0.51</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>CS</td>
<td>4.4</td>
<td>0.43</td>
<td>0.50</td>
<td>0.96</td>
<td>1.00</td>
<td>3.4</td>
<td>0.47</td>
<td>0.54</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP ($K = 0$)</td>
<td>3.3</td>
<td>0.40</td>
<td>0.43</td>
<td>0.94</td>
<td>1.15</td>
<td>-8.3</td>
<td>0.36</td>
<td>0.38</td>
<td>0.87</td>
<td>1.21</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>0.3</td>
<td>0.38</td>
<td>0.40</td>
<td>0.95</td>
<td>1.40</td>
<td>-4.1</td>
<td>0.41</td>
<td>0.41</td>
<td>0.94</td>
<td>1.22</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>0.5</td>
<td>0.38</td>
<td>0.40</td>
<td>0.95</td>
<td>1.38</td>
<td>-4.7</td>
<td>0.41</td>
<td>0.41</td>
<td>0.94</td>
<td>1.21</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>1.9</td>
<td>0.39</td>
<td>0.41</td>
<td>0.94</td>
<td>1.29</td>
<td>-6.1</td>
<td>0.39</td>
<td>0.40</td>
<td>0.94</td>
<td>1.23</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-14.9</td>
<td>0.22</td>
<td>0.20</td>
<td>0.66</td>
<td>0.75</td>
<td>-17.7</td>
<td>0.19</td>
<td>0.21</td>
<td>0.57</td>
<td>0.58</td>
</tr>
<tr>
<td>RC</td>
<td>-17.5</td>
<td>0.23</td>
<td>0.21</td>
<td>0.59</td>
<td>0.58</td>
<td>-20.3</td>
<td>0.20</td>
<td>0.21</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>RR</td>
<td>1.4</td>
<td>0.33</td>
<td>0.31</td>
<td>0.93</td>
<td>0.91</td>
<td>-5.3</td>
<td>0.27</td>
<td>0.28</td>
<td>0.93</td>
<td>1.10</td>
</tr>
<tr>
<td>SS</td>
<td>3.1</td>
<td>0.33</td>
<td>0.33</td>
<td>0.95</td>
<td>0.90</td>
<td>2.7</td>
<td>0.31</td>
<td>0.33</td>
<td>0.97</td>
<td>0.94</td>
</tr>
<tr>
<td>CS</td>
<td>2.4</td>
<td>0.32</td>
<td>0.32</td>
<td>0.95</td>
<td>1.00</td>
<td>2.8</td>
<td>0.30</td>
<td>0.34</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP ($K = 0$)</td>
<td>2.0</td>
<td>0.30</td>
<td>0.29</td>
<td>0.95</td>
<td>1.15</td>
<td>-4.5</td>
<td>0.25</td>
<td>0.28</td>
<td>0.94</td>
<td>1.32</td>
</tr>
<tr>
<td>SNP (AIC)</td>
<td>-0.1</td>
<td>0.29</td>
<td>0.28</td>
<td>0.94</td>
<td>1.21</td>
<td>-3.4</td>
<td>0.27</td>
<td>0.29</td>
<td>0.94</td>
<td>1.19</td>
</tr>
<tr>
<td>SNP (HQ)</td>
<td>0.0</td>
<td>0.29</td>
<td>0.28</td>
<td>0.94</td>
<td>1.19</td>
<td>-3.5</td>
<td>0.27</td>
<td>0.29</td>
<td>0.94</td>
<td>1.18</td>
</tr>
<tr>
<td>SNP (BIC)</td>
<td>0.6</td>
<td>0.29</td>
<td>0.29</td>
<td>0.94</td>
<td>1.19</td>
<td>-4.1</td>
<td>0.26</td>
<td>0.29</td>
<td>0.94</td>
<td>1.23</td>
</tr>
</tbody>
</table>
Table 5.4: Percentage of simulated data sets with $K = 0, 1, 2$ preferred by each information criterion.

<table>
<thead>
<tr>
<th></th>
<th>$K = 0$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 0$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>96.0</td>
<td>2.0</td>
<td>2.0</td>
<td>0.0</td>
<td>18.5</td>
<td>81.5</td>
</tr>
<tr>
<td>HQ</td>
<td>99.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>26.5</td>
<td>73.5</td>
</tr>
<tr>
<td>BIC</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>54.0</td>
<td>45.0</td>
</tr>
<tr>
<td>(b) Bimodal mixture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>18.5</td>
<td>81.5</td>
</tr>
<tr>
<td>HQ</td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>26.5</td>
<td>73.5</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td>54.0</td>
<td>45.0</td>
</tr>
<tr>
<td>(c) Skewed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>4.0</td>
<td>26.0</td>
<td>70.0</td>
<td>5.5</td>
<td>21.5</td>
<td>73.0</td>
</tr>
<tr>
<td>HQ</td>
<td>5.5</td>
<td>32.5</td>
<td>62.0</td>
<td>9.0</td>
<td>27.0</td>
<td>64.0</td>
</tr>
<tr>
<td>BIC</td>
<td>19.5</td>
<td>53.0</td>
<td>27.5</td>
<td>18.5</td>
<td>38.0</td>
<td>43.5</td>
</tr>
<tr>
<td>(d) Bivariate $t_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 5.5: Parameter estimates and p-values for the SWAN data analysis via proposed semiparametric likelihood (SNP) method. Estimated standard errors are in parentheses below each estimate. Standard errors for estimates of $\sigma_u^2$ are multiplied by 10.

<table>
<thead>
<tr>
<th></th>
<th>SNP ($K = 0$)</th>
<th>SNP ($K = 1$)</th>
<th>SNP ($K = 2$ preferred)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-7.28</td>
<td>0.00</td>
<td>-7.20</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td></td>
<td>(2.16)</td>
</tr>
<tr>
<td>$\beta_{0,\text{BMI}}$</td>
<td>0.29</td>
<td>0.00</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>$\beta_{0,\text{age}}$</td>
<td>0.01</td>
<td>0.73</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>$\beta_{0,B}$</td>
<td>0.55</td>
<td>0.17</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>$\beta_{0,W}$</td>
<td>-0.21</td>
<td>0.43</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>$\beta_{0,C}$</td>
<td>-0.68</td>
<td>0.01</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.13</td>
<td>0.57</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td></td>
<td>(0.22)</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>0.14</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
<td>(0.08)</td>
</tr>
<tr>
<td>$\sigma_u^2$</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td>(0.02)</td>
</tr>
</tbody>
</table>
Table 5.6: Estimates and standard errors of the parameters in the SNP representation for the SWAN data analysis via proposed semiparametric likelihood method.

<table>
<thead>
<tr>
<th>SNP (K = 0)</th>
<th>SNP (K = 1)</th>
<th>SNP (K = 2 preferred)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.48</td>
<td>-0.56</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>2.99</td>
<td>2.28</td>
</tr>
<tr>
<td>$r_{11}$</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>$r_{21}$</td>
<td>0.32</td>
<td>0.05</td>
</tr>
<tr>
<td>$r_{22}$</td>
<td>1.54</td>
<td>1.06</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-</td>
<td>0.38</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-</td>
<td>0.22</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 5.7: Model selection criteria for $K = 0, 1, 2$ for the SWAN data analysis via proposed semiparametric likelihood method. Smaller values are preferred.

<table>
<thead>
<tr>
<th></th>
<th>$K = 0$</th>
<th>$K = 1$</th>
<th>$K = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Log likelihood</td>
<td>9583.75</td>
<td>9476.67</td>
<td>9462.37</td>
</tr>
<tr>
<td>AIC</td>
<td>1.0071</td>
<td>0.9961</td>
<td>0.9949</td>
</tr>
<tr>
<td>HQ</td>
<td>1.0089</td>
<td>0.9981</td>
<td>0.9973</td>
</tr>
<tr>
<td>BIC</td>
<td>1.0124</td>
<td>1.0021</td>
<td>1.0020</td>
</tr>
</tbody>
</table>
Figure 5.1: Estimated densities for the bimodal random effects simulation. 

(a) True density of $X_i$.  

(b) Monte Carlo average of estimated densities of $X_i$ preferred by HQ.

(c) and (d) True marginal densities (solid line), Monte Carlo average of estimated marginal densities preferred by AIC (long-dashed line), HQ (dashed line) and BIC (dotted line).

(e) and (f) Estimated marginal densities preferred by HQ for 200 data sets.
Figure 5.2: Fit to the SWAN data using SNP approach.  

(a) Estimated density of $X_i$.  

(b) Contour plot of the density in (a) with empirical Bayes estimates superimposed (Contours are 10, 25, 50, 75, 90, and 95%).  

(c) and (d) Corresponding estimated marginal densities for $K = 0$ (dotted line), $K = 1$ (dashed line), and $K = 2$ (solid line) with the histogram of empirical Bayes estimates of the intercept and slope superimposed, respectively.
Figure 5.3: Empirical Bayes estimates of the slope plotted against observed covariates in the SWAN data. (a) Body mass index. (b) Age. (c) Ethnicity.
Chapter 6
Discussion

A joint model of a primary outcome and longitudinal data has been studied, in which a generalized linear model assumed for the primary outcome has covariates that are underlying subject-specific random effects in a linear mixed model for the longitudinal response. Normality of random effects distribution is a routine assumption for most methods in literature, but it may obscure important features of among-individual variation.

Adapting a strategy of conditional estimation by Stefanski and Carroll (1987), we have proposed inferential strategies (sufficiency score and conditional score) which need no assumptions on the distribution of the random effects and are straightforward and fast to implement, requiring only a few seconds for data on 500 subjects and 5 longitudinal measurements for each subject. In contrast to methods predicated on a parametric (normality) assumption for the random effects, the methods yield valid inferences under departures from this assumption and are competitive when the assumption holds. While the refined regression calibration estimator of Wang et al. (2000), which performed the best in our simulations among existing estimators, is only available for the logistic and probit models, our approach applies to any generalized
linear model formulation.

The proposed estimators are semiparametric in that the random effects distribution is completely unspecified. Results of Stefanski and Carroll (1987, sec. 3) and Lindsay and Lesperance (1995) suggest that the conditional score estimator for our model, with \( t(S_i) = E(X_i|S_i) \), achieves the semiparametric efficiency bound. A modification along the lines suggested by Lindsay (1985) may offer further improvement over our approach.

We have also proposed a semiparametric likelihood approach for the joint model. The normality assumption on random effects is relaxed by approximating the random effects distribution using the SNP representation of Gallant and Nychka (1987), which includes normality as a special case and provides flexibility in capturing a broad range of nonnormal behavior, controlled by a tuning parameter. The method needs only mild, plausible assumption (smoothness) on the distribution of random effects, which provides robustness to departures from normality and a flexible framework for addressing questions of interest. An additional main advantage of the proposed approach is the potential for subject-matter insight on study populations provided by the density estimate. For primary endpoints which are normally distributed given the random effects, the marginal observed-data likelihood can be expressed in a closed form and inference may be carried out using standard optimization techniques. For nonnormal primary endpoints, EM algorithm is used for implementation of the approach, which needs numerical integration and thus involves increased computational burden. The proposed estimator is semiparametric in that the random effects density is not assumed of a specific parametric form.
In contrast to methods predicated on a parametric (normality) assumption for the random effects, the method yields valid inferences under departures from this assumption and are competitive when the assumption holds. While the refined regression calibration estimator of Wang et al. (2000), which performed better than naive and regression calibration in our simulations, is only available for the logistic and probit models, our approach applies to any generalized linear model formulation. Although the conditional estimation methods are fast to implement and performed the best among existing approaches, they can not provide insight into the random effects density, while our approach shows potential efficiency gains when the smoothness assumption holds and provides reliable estimator for the underlying random effects distribution. The additional flexibility afforded by the SNP representation is sufficient to capture the underlying features of the random effects.

We demonstrate that standard information criteria may be used to choose the tuning parameter which controls the degree of flexibility of the SNP representation and detect departures from normality. The combination of visual inspection of estimated SNP density with information criteria has worked well to select a feasible model by focusing on dominant features in practice to provide reliable estimated density. For a data set with 500 subjects and 5 longitudinal observations for each subject, computing times using our implementation including estimation, calculation of standard errors, and evaluation of information criteria were about 15, 30, and 45 minutes for $K = 0, 1, \text{ and } 2$, respectively, and this demonstrates the feasibility of the approach for practical use.

In general, the conditional estimation approach requires no assumption on the
random effects, while the semiparametric likelihood approach requires a mild assumption (smoothness) on the random effects. The semiparametric approach can provide estimator for the random effects distribution, while the conditional estimation approach can not provide any insight into the features of the random effects. The conditional estimation approach is fast to implement, while the semiparametric likelihood approach generally involves numerical integration and is computationally intensive. Both approaches have demonstrated good performance via simulation and by application. Semiparametric likelihood approach shows potential for improved efficiency.

For all methods studied, normality of within-subject longitudinal data measurement errors, which is often reasonable, is assumed, so the analyst should take care to explore the relevance of this assumption.

Alternatively, a mixture of normals (e.g., Verbeke and Lesaffre, 1996) may be used to represent the random effects density, in which the number of normal mixtures acts like a tuning parameter. This imposes the density constraints on the mixing probabilities, while the density constraint of the SNP representation can be met straightforwardly via a parameterization which leads to stable computation.

Further investigation of robustness of theses methods would be of interest. For example, an interesting feature of Table 5.2 and Table 5.3 is that the SNP estimator with $K = 0$ (assuming normality) performs as well as the SNP estimators preferred by information criteria when the underlying random effects distribution is symmetric such as scenarios (a) and (d), while the SNP estimator with $K = 0$ exhibits bias, poor coverage probabilities, and loss of efficiency relative to the SNP estimators with $K$.
chosen adaptively when the random effect distribution is asymmetric such as scenarios (b) and (c). A theoretical basis for this robustness property may exist, which may need further research.

The model as presented here can be extended to more complicated situations, e.g., nonlinear trajectories in individual random effects or multivariate longitudinal measurement may be entertained. Further research on formal methods for assessing possible departures from normal random effects and evaluating the veracity of apparent multimodality as in Figure 1.1b (e.g., using Chaudhuri and Marron, 1999) would be useful.
Bibliography


