Abstract

HUANG, XIANZHENG. Robustness in Latent Variable Models. (Under the direction of Dr. Marie Davidian and Dr. Leonard A. Stefanski.)

Statistical models involving latent variables are widely used in many areas of applications, such as biomedical science and social science. When likelihood-based parametric inferential methods are used to make statistical inference, certain distributional assumptions on the latent variables are often invoked. As latent variables are not observable, parametric assumptions on the latent variables cannot be verified directly using observed data. Even though semiparametric and nonparametric approaches have been developed to avoid making strong assumptions on the latent variables, parametric inferential approaches are still more appealing in many situations in terms of consistency and efficiency in estimation, and computational burden. The goals of our study are to gain insight into the sensitivity of statistical inference to model assumptions on latent variables, and to develop methods for diagnosing latent-model misspecification to enable one to reveal whether the parametric inference is robust under certain latent-model assumptions. We refer to such robustness as *latent-model robustness*.

We start with a simple class of latent variable models, the structural measurement error models, to first tackle the problem. We define theoretical conditions under which a certain degree of latent-model robustness is achieved and study some special structural measurement error models analytically to gain insight into the sensitivity of inference to latent-model assumptions under these specific contexts. Then we borrow

the idea of simulation-extrapolation (SIMEX), or remeasurement method, introduced by Cook and Stefanski (1994) to develop an empirical diagnostic tool that is able to reveal graphically whether or not robustness is attained under the imposed latent-variable assumptions. Testing procedures are proposed as a numerical supplement to the graphical diagnostic tool. These methods are then generalized and refined to adapt to a more complex class of latent variable models called joint models. For this generalization we focus on joint models that link a primary response, which can be a simple response or a censored time-to-event, to an error-prone longitudinal process. The performances of the proposed methods are demonstrated through application to simulated data and data from medical studies.

Robustness in Latent Variable Models

by

XIANZHENG HUANG

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

STATISTICS

Raleigh

2006

APPROVED BY:

Dr. Marie Davidian Co-Chair of Advisory Comittee	Dr. Leonard A. Stefanski Co-Chair of Advisory Comittee
Dr. Anastasios A. Tsiatis	Dr. Hao Helen Zhang

To those who care about me

as a daughter, sister, student,

or just as a friend,

near or far

Biography

Huang, Xianzheng was born in Guangxi, China. She earned her B.S. in Mathematical Statistics from Nankai University in China in 1997 and her Master's degree in Statistics from Oklahoma State University in U.S. in 2002. She was a lecturer in the Department of Mathematics in Tianjin Industry University from July 1997 to July 2000. After graduation, she is joining the Department of Statistics at the University of South Carolina as an Assistant Professor starting Fall 2006.

Acknowledgements

I was once told that most of the things happen to me result from what I choose. If I choose to be happy, happiness will come to me in one way or the other. If I choose to stay depressed, even the funniest clown in the world will not make me laugh. I do not know if it is true. But there are a few things that I do not think I could ever choose. My dear parents are among them. Yet having such parents is what I feel most fortunate for. They equipped me with the belief that life is more worth living if one has dreams and a strong will to make them come true. Then they set me free and let me fly as far as I want, while believing in that I can make wise choices and I will not settle for less in what I pursue. Their support and faith in me are among those key forces without which I might not be able to go this far on my journey.

In addition to my parents, my advisors, Dr. Davidian and Dr. Stefanski, are the other two persons who have walked me though several first-time-in-life experiences that are very special to me. They are like my parents in the academic world, who motivate me to jump towards higher goals and help me to improve. They are among those few persons whom I trust enough to let my heart speak in front of them. What they have taught me and shared with me is the most invaluable treasure I got out of these years, which will benefit me for the rest of my life. They made me realize that, one may be born to be smart or less smart, but what really counts is if one can be persistent in pursuing one's dream and keep trying hard. While doing research under their guidance, I started to believe that there is no upper bound in what one can achieve. There may be an upper bound of how hard one is willing to try, the bar I need to raise, and can be raised, if I want to reach higher.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Models

1.1 Introduction

Models involving unobservable latent quantities are widely used in a host of applications. One example is the structural measurement error models where the latent variable is the true value of a mismeasured regression predictor (Carroll et al., 1995, sec. 1.2). Another example is the so-called "joint" models in which the longitudinal response and endpoint are linked through shared dependence on latent random effects (Henderson et al., 2000; Tsiatis and Davidian, 2004). These joint models can be viewed as generalizations of structural measurement error models. Following Carroll et al. (1995, sec. 1.2), we refer to models involving unobservable random latent variables as structural models.

Provided that the model for the latent variable is correctly specified, likelihood-based approaches are appealing because they lead to consistent and efficient inference. However, intuition suggests that misspecification of this model may compromise in-

ference, although some recent empirical studies have exhibited striking robustness to the assumption on the latent variable (Song et al., 2002). The data analyst faces the difficulty that the extent to which inference may be sensitive to the choice of model for unobservable latent variables is not known in a given problem. Techniques for studying and diagnosing robustness in these models would thus be invaluable. We present a framework for assessing model robustness in a class of structural latent variable models.

We first focus on the particular subclass of structural measurement error models, and propose practical strategies for diagnosing misspecification of the model for the true predictor, the latent variable for this subclass. We then adapt the methods to more general and complicated structural latent variable models, namely joint models. To illustrate the proposed methods in the context of different structural latent variable models, we demonstrate the methods using analytical examples, simulated data, and data sets from several medical studies where these models are entertained.

In Section 1.2, the structural measurement error models are reviewed. We argue that model robustness in this setting refers to lack of bias (asymptotically) in the estimator for parameters of interest. In Section 1.3, we give the generic setup of two types of joint models of our interest. Each model setup is followed by an example from a medical study to which the proposed methods are later applied.

1.2 Structural Measurement Error Models

We consider the so-called classical measurement error model. Let Y be the response; $\boldsymbol{X}_{q\times 1}$ and $\boldsymbol{W}_{q\times 1}$ be the true and observed predictor, respectively; and $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$, where $\boldsymbol{U}_{q\times 1}$ is nondifferential measurement error (Carroll et al. 1995, Section 1.6) that follows a multivariate normal $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\sigma}_U^2)$ with $\boldsymbol{\sigma}_U^2$ known. In practice when $\boldsymbol{\sigma}_U^2$ is unknown but there are replicated measures on $\boldsymbol{W}$, estimating $\boldsymbol{\sigma}_U^2$ adds little variation or complexity in implementing the proposed methods. Assuming the conditional density of Y given $\boldsymbol{X} = \boldsymbol{x}$ is $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x};\boldsymbol{\theta})$, the joint density of $(Y, \boldsymbol{W})$ given $\boldsymbol{X} = \boldsymbol{x}$ is

$$f_{Y,\boldsymbol{W}|\boldsymbol{X}}(y, \boldsymbol{w}|\boldsymbol{x};\boldsymbol{\theta}) = f_{Y|\boldsymbol{X}}(y|\boldsymbol{x};\boldsymbol{\theta}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x};\boldsymbol{\sigma}_U^2), \tag{1.1}$$

where the conditional density of $\boldsymbol{W}$given $\boldsymbol{X}$, $f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x};\boldsymbol{\sigma}_U^2)$, is the $\mathrm{N}(\boldsymbol{x},\ \boldsymbol{\sigma}_U^2)$ density. To focus on the choice of model for $\boldsymbol{X}$, we assume the densities in (1.1) are known up to $\boldsymbol{\theta}_{p\times 1}$. Inference on $\boldsymbol{\theta}$ is of central interest.

Denote the independent random pairs, $\{Y_i, \boldsymbol{W}_i\}_{i=1}^n$, as a realization of the measurement error model, with $\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{U}_i$, $i = 1, \ldots, n$. Two ways of viewing the $\boldsymbol{X}_i$, $i = 1, \ldots, n$, lead to two types of measurement error models, functional models and structural models (Carroll et al., 1995, sec. 1.2). In a functional model, the $\boldsymbol{X}_i$ are viewed as unknown parameters, and the likelihood of the observed data based on model (1.1) is

$$L(\boldsymbol{\theta}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \prod_{i=1}^{n} f_{Y,\boldsymbol{W}|\boldsymbol{X}}(Y_i, \boldsymbol{W}_i|\boldsymbol{X}_i;\boldsymbol{\theta}). \tag{1.2}$$

In a structural model, the $\boldsymbol{X}_i$ are regarded as random variables. Under the assumption

that the density of $\boldsymbol{X}$ is $f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)})$, depending on parameter $\boldsymbol{\tau}^{(a)}$,

$$f_{Y,\boldsymbol{W}}(y,\boldsymbol{w};\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) = \int f_{Y|\boldsymbol{X}}(y|\boldsymbol{x};\boldsymbol{\theta}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{w}|\boldsymbol{x};\boldsymbol{\sigma}_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)})\, d\boldsymbol{x}, \tag{1.3}$$

is the modeled marginal density of $(Y,\boldsymbol{W})$, and the corresponding likelihood is

$$L(\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) = \prod_{i=1}^{n} f_{Y,\boldsymbol{W}}(Y_i,\boldsymbol{W}_i;\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}). \tag{1.4}$$

An example of a situation where such a model is suitable is the analysis of cardiovascular disease outcomes in the Framingham study (Kannel et al., 1986), which followed subjects for development of coronary heart disease (CHD) over several exam periods. An objective is to characterize the relationship between a response, an indicator of evidence of CHD at the end of an eight-year follow-up period after the second exam visit, and long-term systolic blood pressure (SBP). One may postulate a structural measurement error model, where the true predictor, long-term SBP, which cannot be measured directly, is viewed as a latent variable with some distribution in the population of subjects, information on which is only available through contaminated measurements of SBP taken during clinic visits.

Functional modeling makes minimal assumptions on the set of unobserved predictors and thus is generally applicable. However, functional-model inference is usually problematic. Maximizing the functional likelihood (1.2) with respect to $\boldsymbol{\theta}$ and $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ is often difficult and seldom results in consistent estimator for $\boldsymbol{\theta}$. Consequently, many functional-model inference methods are moment methods or conditional likelihood methods. Fuller (1987) and Carroll et al. (1995) describe a number of functional model inference methods.

Whereas historically, functional models and functional-model methods have been studied more than structural models and methods, recent emphasis has been on structural models and methods. The appeal of structural modeling is mainly due to the fact that inference can be based directly on the likelihood (1.4), thereby simplifying estimation relative to that in functional modeling, apart from the numerical problem of evaluating the integral in (1.3). Maximum likelihood estimation also offers the attraction of asymptotic efficiency when the assumed parametric model $f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x}; \boldsymbol{\tau}^{(a)})$ is correct.

A reason often cited for avoiding parametric structural modeling is that misspecification of the distributional model for $\boldsymbol{X}$ can result in inconsistent estimators for $\boldsymbol{\theta}$. With regard to robustness of inference on $\boldsymbol{\theta}$ to misspecification of this model, semiparametric modeling methods (Roeder, Carroll, and Lindsay, 1996; Schafer, 2001) and flexible-parametric modeling methods (Carroll, Roeder, and Wasserman, 1999; Richardson, 2002) provide some solutions. However, with respect to simplicity of implementation and efficiency, they can be more like functional methods than parametric structural methods. Hence, parametric structural modeling is preferable in practice as long as the analyst may be assured that the inferences are insensitive to an incorrect specification of the distributional model.

To study robustness to model specification on $\boldsymbol{X}$, we assume that measurement error exists with known variance $\boldsymbol{\sigma}_U^2$ ($\neq \mathbf{0}$) and that the unknown density of $\boldsymbol{X}$ is possibly misspecified in (1.4). It should be clear that only the interaction of these two factors can result in inconsistency of the estimator for $\boldsymbol{\theta}$. Accordingly, we characterize model robustness as lack of large-sample bias in this estimator, regardless of the size

of measurement error variance.

1.3 Joint Models

It is often of interest to characterize the association of a primary endpoint and a longitudinal process, as well as the features of the longitudinal process. One popular approach to tackle this dual-task problem is to adopt a regression model for the primary endpoint and a mixed-effect model for the longitudinal process, which are linked through joint dependence on latent random effects. It has been demonstrated in the literature that, with appropriate parametric modeling of the distribution of random effects, joint modeling can gain efficiency and provide insight into underlying features of the longitudinal process. Similar to what is pointed out in Section 1.2, one concern in this approach is the sensitivity of inference on the primary regression parameters to the model assumptions on the random effects.

For definiteness, we study two types of joint models that are of great interest in medical and public health application. The first type is for an error-prone longitudinal response and a primary simple response (Li et al., 2004; Wang et al., 2000). The second type is for an error-prone longitudinal response and a censored time-to-event (Song et al. 2002; Tsiatis and Davidian, 2001, 2004). Hereafter we refer to the first type as a simple-response joint model and the second type as a censored-endpoint joint model. Some common notations used in both joint models are given first.

Let $i(= 1, \ldots, n)$ be the subject index, and $j(= 1, \ldots, m_i)$ be the time index. The vector of longitudinal measures for subject i is denoted by $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{im_i})^T$,

recorded at times $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{im_i})^T$. The intra-subject errors, U_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, are independent and identically distributed (i.i.d.) as normal with mean zero and variance σ_U^2. Define $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{im_i})^T$, then $\boldsymbol{U}_i \sim N_{m_i}(\boldsymbol{0}, \sigma_U^2 \boldsymbol{I}_{m_i})$, where $\boldsymbol{I}_{m_i}$ is the $m_i \times m_i$ identity matrix. Let $\boldsymbol{\Omega}$ be the vector of all unknown parameters in the joint model and $\boldsymbol{\theta}$ be the vector of primary regression parameters. Inference on $\boldsymbol{\theta}$ is of central interest.

In the simple-response joint models, the joint model consists of two component models: a primary regression model relating the response Y_i and the $p \times 1$ unobservable explanatory variables $\boldsymbol{X}_i$ with density denoted by $f_{Y_i|\boldsymbol{X}_i}(Y_i|\boldsymbol{X}_i; \boldsymbol{\theta})$; and a mixed-effect model relating the longitudinal measurements $\boldsymbol{W}_i$ and $\boldsymbol{X}_i$, the unobserved subject-specific random effects. We consider a linear mixed model of the form $\boldsymbol{W}_i = \boldsymbol{D}_i \boldsymbol{X}_i + \boldsymbol{U}_i$, where $\boldsymbol{D}_i$ is an $m_i \times p$ ($m_i > p$) design matrix of rank p. It is further assumed that given $\boldsymbol{X}_i$, Y_i and $\boldsymbol{W}_i$ are independent. Define by $f^{(a)}_{\boldsymbol{X}_i}(\boldsymbol{X}_i; \boldsymbol{\tau}^{(a)})$ the density of the assumed model for $\boldsymbol{X}_i$, where $\boldsymbol{\tau}^{(a)}$ is the vector of parameters in the assumed model. In this case, $\boldsymbol{\Omega} = (\boldsymbol{\theta}^T, \boldsymbol{\tau}^{(a)T}, \sigma_U^2)^T$; and the contribution to the observed-data likelihood from subject i is

$$f_{Y_i, \boldsymbol{W}_i}(Y_i, \boldsymbol{W}_i; \boldsymbol{\Omega}) = \int f_{Y_i|\boldsymbol{X}_i}(Y_i|\boldsymbol{x}_i; \boldsymbol{\theta}) f_{\boldsymbol{W}_i|\boldsymbol{X}_i}(\boldsymbol{W}_i|\boldsymbol{x}_i; \sigma_U^2) f^{(a)}_{\boldsymbol{X}_i}(\boldsymbol{x}_i; \boldsymbol{\tau}^{(a)}) d\boldsymbol{x}_i, \qquad (1.5)$$

where $f_{\boldsymbol{W}_i|\boldsymbol{X}_i}(\boldsymbol{w}|\boldsymbol{x}_i; \sigma_U^2)$ is the density of $N_{m_i}(\boldsymbol{D}_i \boldsymbol{x}_i, \sigma_U^2 \boldsymbol{I}_{m_i})$.

In the censored-endpoint joint models, the first component model is a model for the time-to-event, denoted by the random variable T. For example, one may specify

a proportional hazards model with hazard rate given by

$$\begin{aligned}\lambda_i(u) &= \lim_{h\to 0} h^{-1} P(u \le T_i < u+h | T_i \ge u, X_i, F_i, C_i, \boldsymbol{t}_i) \\ &= \lambda_0(u)\exp\{\beta X_i(u) + \iota F_i\}, \end{aligned} \tag{1.6}$$

where $\lambda_0(u)$ is an unspecified baseline hazard function, $X_i(u)$ is the inherent value of a longitudinal response at time u, F_i is the value of the time-independent covariate, β and ι are parameters. The observed survival data on subject i include $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$, where T_i and C_i are the (potential) time to event and censoring time, respectively, and $I(\cdot)$ is the indicator function. Censoring, covariate errors, and timing of measurements are assumed noninformative. The second component model relates the observed longitudinal measures W_{ij} to the inherent longitudinal measures $X_i(t_{ij})$, $W_{ij} = X_i(t_{ij}) + U_{ij}$. It is further assumed that $X_i(u)$ depends on a $p \times 1$ vector of subject-specific random effects, $\boldsymbol{\alpha}_i$, via the relationship $X_i(u) = \boldsymbol{D}_i(u)\boldsymbol{\alpha}_i$, for instance, where $\boldsymbol{D}_i(u)$ is a $1\times p$ design matrix. Define by $f^{(a)}_{\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i; \boldsymbol{\tau}^{(a)})$ the assumed model for $\boldsymbol{\alpha}_i$. In this case $\boldsymbol{\Omega} = (\boldsymbol{\theta}^T, \boldsymbol{\tau}^{(a)T}, \lambda_0, \sigma_U^2)^T$ and $\boldsymbol{\theta} = (\beta, \iota)^T$. The contribution to the observed-data likelihood from subject i is

$$\begin{aligned} f(V_i, \Delta_i, \boldsymbol{W}_i, \boldsymbol{t}_i, F_i; \boldsymbol{\Omega}) &= \int f(V_i, \Delta_i | \boldsymbol{\alpha}_i, F_i; \boldsymbol{\theta}, \lambda_0) f_{\boldsymbol{W}_i|\boldsymbol{\alpha}_i}(\boldsymbol{W}_i | \boldsymbol{\alpha}_i, \boldsymbol{t}_i; \sigma_U^2) \\ &\qquad f^{(a)}_{\boldsymbol{\alpha}_i}(\boldsymbol{\alpha}_i | F_i; \boldsymbol{\tau}^{(a)}) d\boldsymbol{\alpha}_i, \end{aligned} \tag{1.7}$$

where $f_{\boldsymbol{W}_i|\boldsymbol{\alpha}_i}(\boldsymbol{w}|\boldsymbol{\alpha}_i, \boldsymbol{t}_i; \sigma_U^2)$ is the density of $N_{m_i}(\boldsymbol{D}_i\boldsymbol{\alpha}_i, \sigma^2 \boldsymbol{I}_{m_i})$, and assuming the hazard rate in (1.6),

$$\begin{aligned} f(V_i, \Delta_i | \boldsymbol{\alpha}_i, F_i; \boldsymbol{\theta}, \lambda_0) &= [\lambda_0(V_i)\exp\{\beta \boldsymbol{D}_i(V_i)\boldsymbol{\alpha}_i + \iota F_i\}]^{\Delta_i} \\ &\qquad \exp\Big[- \int_0^{V_i} \lambda_0(u)\exp\{\beta \boldsymbol{D}_i(u)\boldsymbol{\alpha}_i + \iota F_i\} du \Big]. \end{aligned}$$

An example where the simple-response joint model is appropriate is the Study of Women's Health Across the Nation (SWAN) (Sowers et al., 2003). Two objectives of SWAN study are to characterize the association between the evidence of osteopenia, a binary endpoint, and the underlying hormone patterns over the menstrual cycle in peri-menopausal women, and to understand the underlying hormone patterns of this population, which can only be observed in this study through the longitudinal progesterone levels derived from urine (PDG).

The censored-endpoint joint model is appropriate for data from the AIDS Clinical Trials Group (ACTG) Protocol 175 (Hammer et al., 1996). In this study, more than 2000 HIV-1-infected subjects enrolled between December 1991 and October 1992 were followed for their CD4 counts from week 8, and every 12 weeks thereafter, until November 30, 1994. The "event" defined in this study is a composite of $\geq$50% decline in CD4, progression to AIDS, or death. It is of interest to study the prognostic value of CD4 counts and its inherent trajectory over time for such a population.

In both of the proceeding examples, the longitudinal measurements, PDG and CD4 counts, are subject to assay error and intra-subject variation so that the true longitudinal process is unobservable. The joint model characterizes the true longitudinal process as an "inherent" trajectory represented above as $\boldsymbol{D}_i(u)\boldsymbol{\alpha}_i$, depending on unobserved latent random effects, plus a deviation due to these sources. Interest focuses on the relationship between the "inherent" trajectory and the simple-response or censored endpoint. Thus the model relating the simple-response endpoint or the censored endpoint to the longitudinal process depends on these unobserved latent variables.

Among many others, Verbeke and Lesffre (1997) studied the effect of misspecifying the model for random effects in linear mixed models for longitudinal data, and Heagerty and Kurland (2001) explored the impact of such model misspecification in generalized linear mixed models. The models they considered can be viewed as one of the two component models involved in joint models. Relatively limited study has been done on the consequences of the violation of the model assumption on the random effects in the context of joint models. Wang et al. (2000) considered the simple-response joint models and proposed three methods to estimate the primary regression parameters. One key assumption in their methods is that the random effects in the longitudinal model follow a multivariate normal distribution. A natural concern is the effect on inference when this normality assumption is violated. Li et al. (2004) also studied the simple-response joint models and proposed conditional score estimators (CSE) for the primary regression parameters, which require no assumption on the random effects. The drawbacks with the CSE approach are that one cannot capture the nature of the random-effect distribution and that the information in longitudinal data is not used efficiently. The latter drawback is common for semiparametric methods, like the conditional score method, in which the distribution of the random effects is left unspecified. Tsiatis and Davidian (2001) also proposed CSE for the primary regression parameters in the censored-endpoint joint model. Song et al. (2002) considered the censored-endpoint joint model and modeled the random effects using the seminonparametric (SNP) representation of Gallant and Nychka (1987) to gain flexibility and efficiency. Moreover, they observed in simulation studies remarkable robustness of the maximum likelihood estimator (MLE) to

the normality assumption on the random effects. Hsieh et al. (2006) investigated this robustness aspect of joint models via simulation and provided a heuristic explanation for this phenomenon.

1.4 A Brief Tour

As noted in Section 1.1, we start with structural measurement error models to present the framework for assessing latent-model robustness. In Chapter 2, we define two theoretical conditions needed to achieve robustness. We use several specific measurement error models as examples to demonstrate how to check these conditions and by so doing understand the reasons for (non)robustness. Even though these theoretical tools can shed light on the sensitivity of inference to model assumptions on the true predictors in structural measurement error models, checking the conditions analytically is often involved. A graphical device better suited to practical use to assess latent-model robustness is given in Chapter 3. This graphical diagnostic tool leads to a way of examining the theoretical robustness conditions empirically. We also present in Chapter 3 several test statistics that can provide numerical evidence of (non)robustness.

The graphical device and test statistics are applied to structural measurement error models in Chapter 4. We use examples, some of which also appear in Chapter 2, to demonstrate the performance of these methods. In addition to simulated examples, we also implement the methods to Framingham study data. For these data we are interested in the robustness of estimates for the regression parameter in a primary

model relating evidence of coronary heart disease to long-term systolic blood pressure. In the analysis, the long-term systolic blood pressure is the latent variable that is linked to the observed systolic blood pressure via an additive measurement error model.

In Chapter 5 we adapt these methods to joint models. We show in Section 5.1 that inference on the primary regression parameters is expected to be robust to model assumptions on latent variables in the mixed effects model for the longitudinal data, as well as to intra-subject random errors when there is sufficient longitudinal information. Due to some extra complexity in joint models compared to measurement error models, it becomes more computationally expensive to implement the methods demonstrated in Chapter 4, especially for the censored-endpoint joint models. To reduce computational burden when examining robustness under this complicated setting, we propose a refined graphical diagnostic method. This, along with the use of the test statistics defined in Chapter 3, is illustrated via simulated examples in Sections 5.2 and 5.3. Section 5.4 presents the analysis of SWAN data and ACTG 175 data. Because model fitting and parameter estimation for these data sets has been carried out by other authors (for example Li et al., (2004) analyzed SWAN data and Song et al., (2002) analyzed ACTG 175 data), in our analyses we focus on demonstrating how to use the graphical diagnostic tool and test statistics to reveal the potential impact of different model assumptions on random effects in these joint models.

We conclude with a summary of our work in Chapter 6. Some unresolved issues in the current study and thoughts on related future research are also discussed.

Chapter 2

Theoretical Robustness

In this chapter we focus on structural measurement error models. The view of model robustness described in Section 1.2 is formulated theoretically and a strategy for checking robustness is developed. The theory and methods are illustrated using several specific structural measurement error models.

2.1 Full Latent-Model Robustness

To formalize this idea, without loss of generality we henceforth take the true predictor X and its error-prone version W to be scalars, which suffices for the purpose of motivating and illustrating the proposed methods. We first define exact, or full, latent-model robustness. Using the assumed model for X given by the density

$f_X^{(a)}(x;\boldsymbol{\tau}^{(a)})$, consider the structural-model likelihood

$$\begin{aligned} L(\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) &= \prod_{i=1}^{n} f_{Y,W}(Y_i,W_i;\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) \\ &= \prod_{i=1}^{n}\int_{-\infty}^{\infty} f_{Y|X}(Y_i|x;\boldsymbol{\theta}) f_{W|X}(W_i|x;\sigma_U^2) f_X^{(a)}(x;\boldsymbol{\tau}^{(a)})\ dx; \end{aligned} \tag{2.1}$$

the maximum likelihood estimators (MLEs) for ($\boldsymbol{\theta}$,$\boldsymbol{\tau}^{(a)}$) under this assumed structural model are the values maximizing (2.1). Denote by $\boldsymbol{\theta}^*$ the true value of $\boldsymbol{\theta}$ determining the conditional distribution of Y given X. Let

$$\boldsymbol{\psi}(y,w,\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) = \begin{pmatrix} (\partial/\partial\boldsymbol{\theta}) \log\left\{f_{Y,W}(y,w;\boldsymbol{\theta},\boldsymbol{\tau}^{(a)})\right\} \\ (\partial/\partial\boldsymbol{\tau}^{(a)}) \log\left\{f_{Y,W}(y,w;\boldsymbol{\theta},\boldsymbol{\tau}^{(a)})\right\} \end{pmatrix}, \tag{2.2}$$

and define $\boldsymbol{\theta}(\,\cdot\,)$ and $\boldsymbol{\tau}^{(a)}(\,\cdot\,)$ as functions of σ_U implicitly via

$$E[\boldsymbol{\psi}\{Y,W,\boldsymbol{\theta}(\sigma_U),\boldsymbol{\tau}^{(a)}(\sigma_U)\}] = \mathbf{0}. \tag{2.3}$$

The expectation in (2.3) is with respect to the distribution of ($Y,\ W$) with density

$$f_{Y,W}(Y,W;\boldsymbol{\theta}) = \int_{-\infty}^{\infty} f_{Y|X}(Y|x;\boldsymbol{\theta}) f_{W|X}(W|x;\sigma_U^2) f_X^*(x)\ dx,$$

where $f_X^*(x)$ is the true density of X.

We say that the structural model MLE for $\boldsymbol{\theta}$ is robust to choice of model for X provided

$$\boldsymbol{\theta}(\sigma_U) \equiv \boldsymbol{\theta}^*,\ \text{for } \sigma_U \geq 0. \tag{2.4}$$

It is worth pointing out that the model for X does not have to be correctly specified for robustness of the structural model MLE for $\boldsymbol{\theta}$. For example, if the assumed model for X is flexible enough such that the moments of the true model, on which $\boldsymbol{\theta}(\sigma_U)$ depends, are estimated consistently, then robustness can be achieved in some simple

models. Two examples given next illustrate the consequence of using models with different degree of flexibility for X.

Example 2.1: *Y given X follows a normal distribution with mean $\beta_0+\beta_1 X$.* Assume $Y|X = x \sim \mathrm{N}(\beta_0 + \beta_1 x, \sigma_\epsilon^2)$. If one assumes X to be normal, then normality of X is not necessary for consistency of the structural MLE for $\boldsymbol{\theta} = (\beta_0,\ \beta_1,\ \sigma_\epsilon)^T$ (Fuller, 1987, p 17). The explanation lies in the facts that the regression coefficients are functions of the first- and second-order moments and that the population moments are consistently estimated regardless of the true distribution of X. The key to this positive finding is that the assumed normal model for X is flexible enough to permit consistent estimation of all required moments.

We now consider a less flexible normal model. Specifically, suppose that the distribution of X is assumed to be normal with mean equal to the variance. That is, assume that

$$f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x}; \tau^{(a)}) = \frac{1}{\sqrt{2\pi\tau^{(a)}}} \exp\left\{-(x - \tau^{(a)})^2/2\tau^{(a)}\right\}$$

is the model used to construct the likelihood (2.1) and define estimators for $\boldsymbol{\theta}$ and $\tau^{(a)}$. The functions $\boldsymbol{\theta}(\,\cdot\,)$ and $\tau^{(a)}(\,\cdot\,)$ defined through (2.3) give the probability limits ($n \to \infty$) of the MLEs for $\boldsymbol{\theta}$ and $\tau^{(a)}$. If the true distribution of X is not normal with mean and variance equal, then the assumed model is incorrect and too restrictive to permit consistent estimation of the first two moments of the true distribution of X. This will lead to potential bias in the estimator for $\boldsymbol{\theta}$, with the magnitude of the bias generally increasing with increasing σ_U.

Figure 2.1 (a)–(b) displays plots of the limiting MLEs for $\boldsymbol{\theta}$ and τ as functions

of σ_U for three true distributions of X, N(1, 1), N(0.5, 1), and N(1.5, 1), when the true values of the parameters in the model for $Y|X$ are $\beta_0 = 0$, $\beta_1 = 1$, and $\sigma_\epsilon = 1$. In the latter two cases, the assumed model is incorrect and too restrictive compared to the true density. As shown in the plots, the misspecification and lack of flexibility in modeling X result in asymptotic biases in $\widehat{\boldsymbol{\theta}}$ that increase in magnitude as σ_U increases.

In the foregoing example, we consider three true distributions of X while fixing the assumed model for X at a very restrictive distribution. In the next example, we fix the true distribution of X and compare the structural MLEs under several assumed models for X.

Example 2.2: *Y given X follows a Bernoulli distribution with mean probit($\beta_0 + \beta_1 X$).* Assume Y is binary and $P(Y = 1|X = x) = \Phi(\beta_0 + \beta_1 x)$, where $\Phi(\,\cdot\,)$ is the standard normal cumulative distribution function, and the true distribution $f_X^*(x)$ of X is the mixture normal, $0.1N(2.35, 0.64^2) + 0.9N(-0.26, 0.62^2)$. In this case $\boldsymbol{\theta}^* =$ the true value of $(\beta_0,\ \beta_1)^T = (0,\ 1)^T$, and the true density $f_X^*(x)$ is right-skewed with a small secondary mode.

Three assumed models for X are chosen to construct the likelihood (2.1). First, assume $X \sim N(\mu_X,\ \sigma_X^2)$. Second, assume X follows a distribution with the density defined as the second-order seminonparametric (SNP) density given by

$$\frac{1}{\eta}\phi\Big(\frac{x-\xi}{\eta}\Big)\Big\{a_0 + a_1\Big(\frac{x-\xi}{\eta}\Big) + a_2\Big(\frac{x-\xi}{\eta}\Big)^2\Big\}^2, \tag{2.5}$$

where $\phi(.)$ is the standard normal density function, (ξ, η, a_0, a_1, a_2) are unknown

parameters of the density, and $(a_0,\ a_1,\ a_2)$ are constrained so that (2.5) integrates to one. The general SNP density provides a flexible family that is able to capture certain features related to high-order moments that deviate from those of a normal distribution, and includes the normal as a special case. Third, assume X follows a mixture normal distribution. Compared to the true model of X, the first assumed model is incorrect and probably too restrictive, the second assumed model is also incorrect but more flexible than the first one, and the third model class includes the true distribution $f_X^*(x)$.

Denote the densities of $(Y,\ W)$ in (2.1) corresponding to the three assumed models for X as $f^{(n)}_{Y,\ W}(Y, W; \boldsymbol{\theta}, \boldsymbol{\tau}^{(n)}, \sigma_U)$, $f^{(s)}_{Y,\ W}(Y, W; \boldsymbol{\theta}, \boldsymbol{\tau}^{(s)}, \sigma_U)$, and $f^{(m)}_{Y,\ W}(Y, W; \boldsymbol{\theta}, \boldsymbol{\tau}^{(m)}, \sigma_U)$, i.e., the joint density assuming X follows a normal distribution, a distribution with SNP density (2.5), and a mixture normal distribution, respectively, where $\boldsymbol{\theta}{=}(\beta_0,\ \beta_1)^T$, $\boldsymbol{\tau}^{(n)} = (\mu_x,\ \sigma_x)^T$, $\boldsymbol{\tau}^{(s)} = (\xi,\ \eta,\ a_0,\ a_1,\ a_2)^T$, and $\boldsymbol{\tau}^{(m)} = (\mu_1,\ \sigma_1,\ \mu_2,\ \sigma_2,\ \alpha)^T$. The integral in (2.1) can be solved analytically when X is assumed to be normal or mixture normal but not when modeling X with SNP. The resulting joint densities for (Y, W) are given in the Appendix A.

Due to the complexity of these joint densities, it is tedious, if possible at all, to derive (2.2) analytically and computationally inefficient to find the functions $\boldsymbol{\theta}(\,\cdot\,)$ and $\boldsymbol{\tau}^{(\,\cdot\,)}(\,\cdot\,)$ by solving (2.3). Accordingly we maximize the expectations $E\{\log(f^{(n)}_{Y,W})\}$, $E\{\log(f^{(s)}_{Y,W})\}$, and $E\{\log(f^{(m)}_{Y,W})\}$ to obtain these functions of σ_U numerically. All three expectations are with respect to the true joint distribution of $(Y,\ W)$, of which the density is given by (A.5) with $\boldsymbol{\tau}^{(m)}$ replaced by $\boldsymbol{\tau}^*$.

Denote by $\theta^{(n)}(\sigma_U)$, $\theta^{(s)}(\sigma_U)$ and $\theta^{(m)}(\sigma_U)$ the parameter values defined by (2.3) under these assumed models for X. Theoretically, $\theta^{(m)}(\sigma_U) \equiv \theta^*$, as it results from the correct modelling. Hence we use it as the gold standard to which $\theta^{(n)}$ and $\theta^{(s)}$ are compared. The differences, $\theta^{(n)} - \theta^{(m)}$ and $\theta^{(s)} - \theta^{(m)}$, are plotted against σ_U in Figure 2.2 (a)–(b). The component curves in $\theta^{(n)} - \theta^{(m)}$ show deviation away from the zero-reference line that becomes more pronounced as σ_U increases. In contrast, the component curves in $\theta^{(s)} - \theta^{(m)}$ are flatter and stay closer to the zero-reference line along the range of σ_U. The plots indicate that $\theta^{(s)}(\sigma_U)$ is much more robust than $\theta^{(n)}(\sigma_U)$ and closely matches $\theta^{(m)}(\sigma_U)$.

Figure 2.2 (c)–(d) are Monte-Carlo estimated finite-sample versions of Fig. 2.2 (a)–(b). In the simulation study, 100 datasets each of size 500 were generated from the true structural measurement error model with the same parameter values given above. For each dataset, $\widehat{\theta}$ was computed by maximizing (2.1), depending on the assumed model for X. The expectations, $E(\widehat{\theta}^{(\cdot)})$, are estimated by the corresponding Monte-Carlo averages. Clearly, no procedure can do better than the true-model estimator, $\widehat{\theta}^{(m)}$, and we again use it as the gold standard to which $\widehat{\theta}^{(n)}$ and $\widehat{\theta}^{(s)}$ are compared. The Monte-Carlo averages of the differences, $\widehat{\theta}^{(n)} - \widehat{\theta}^{(m)}$ and $\widehat{\theta}^{(s)} - \widehat{\theta}^{(m)}$, as functions of σ_U are depicted in Figure 2.2 (c)–(d). Similar to the observations from Figure 2.2 (a)–(b), the component curves in $\widehat{\theta}^{(n)} - \widehat{\theta}^{(m)}$ deviate from the zero-reference line more dramatically as σ_U increases, while the component curves in $\widehat{\theta}^{(s)} - \widehat{\theta}^{(m)}$ overlap with the flat zero-reference line closely. This indicates the robustness of $\widehat{\theta}^{(s)}$ and the nonrobustness of $\widehat{\theta}^{(n)}$.

2.2 First-order Latent-Model Robustness

The condition for robustness in (2.4) is not easily verified except in very simple models. Also, it is not obvious that it can be satisfied in general, without making some assumptions on the true distribution of X, except in simple models. Thus its utility is limited. However, note that if (2.4) is satisfied, then the derivatives of $\boldsymbol{\theta}(\sigma_U)$ with respect to σ_U of any order are identically $\mathbf{0}$. More generally, whether (2.4) is satisfied or not, $\boldsymbol{\theta}(\sigma_U)$ has the MacLaurin series expansion (for σ_U near 0)

$$\boldsymbol{\theta}(\sigma_U) = \boldsymbol{\theta}^* + \frac{\sigma_U^2}{2}\boldsymbol{\theta}''(0) + \boldsymbol{o}\left(\sigma_U^2\right).$$

Thus, a necessary, first-order condition for robustness is that $\boldsymbol{\theta}''(0) = \mathbf{0}$. This condition is somewhat easier to verify than (2.4). The required derivatives $\boldsymbol{\theta}''(0)$ can be obtained in principle by implicit differentiation as in Stefanski (1985). The following two examples illustrate the first-order condition.

Example 2.3: *First-order latent-model robustness of location-scale models in simple linear regression.* Consider the simple linear regression model in which Y given X is $\mathrm{N}(\beta_0 + \beta_1 X, \sigma_\epsilon^2)$. Suppose that the distribution of X is modeled with a location-scale family; that is,

$$f_{\boldsymbol{x}}(x; \boldsymbol{\tau}) = \tau_2 h(\tau_1 + \tau_2 x)$$

for some fixed, known, but otherwise arbitrary density $h(\cdot)$. For this model it can be

shown that $\boldsymbol{\theta}''(0)$ is a non-singular matrix multiple of the vector

$$\begin{pmatrix} \tau_2^*\beta_1^* E\left\{\dfrac{h'(\tau_1^*+\tau_2^*X)}{h(\tau_1^*+\tau_2^*X)}\right\} \\ \beta_1^* + \tau_2^*\beta_1^* E\left\{\dfrac{Xh'(\tau_1^*+\tau_2^*X)}{h(\tau_1^*+\tau_2^*X)}\right\} \\ 0 \end{pmatrix}, \tag{2.6}$$

where β_1^* is the true value of the slope parameter and τ_1^* and τ_2^* are the probability limits of the MLEs for τ_1 and τ_2 in the location-scale model in the case of no measurement error. Regardless of whether the true density of X is in the assumed location-scale family, τ_1^* and τ_2^* satisfy the location-scale (asymptotic) likelihood equations

$$E\left\{\frac{h'(\tau_1^*+\tau_2^*X)}{h(\tau_1^*+\tau_2^*X)}\right\} = 0, \tag{2.7}$$

$$E\left\{\frac{Xh'(\tau_1^*+\tau_2^*X)}{h(\tau_1^*+\tau_2^*X)}\right\} + \frac{1}{\tau_2^*} = 0. \tag{2.8}$$

Equations (2.7) and (2.8) imply that (2.6) is equal to $\mathbf{0}$. Thus, estimation of $\boldsymbol{\theta} = (\beta_0,\ \beta_1,\ \sigma_\epsilon)^T$ in the simple linear regression measurement error model is first-order robust for arbitrary location-scale models for the distribution of X. The robustness associated with the normal distribution assumption noted in Example 3.1 is a special case of the first-order robustness of location-scale families in general.

Example 2.4: *First-order latent-model robustness of the normal distribution model in quadratic regression.* Consider the simple quadratic regression model in which Y given X is $\mathrm{N}(\beta_0 + \beta_1 X + \beta_2 X^2, \sigma_\epsilon^2)$. Suppose that the distribution of X is modeled as $\mathrm{N}(\tau_1, \tau_2)$. For this model it can be shown that $\boldsymbol{\theta}''(0)$ is a non-singular matrix multiple

of the vector

$$\begin{pmatrix} -2\beta_2^*\tau_2^* + \beta_1^* E(X) + 2\beta_2^* E(X^2) - 2\beta_2^*\tau_1^* E(X) - \beta_1^*\tau_1^* \\ -\beta_1^*\tau_2^* - 4\beta_2^*\tau_2^* E(X) - \beta_1^*\tau_1^* E(X) - 2\beta_2^*\tau_1^* E(X^2) + \beta_1^* E(X^2) + 2\beta_2^* E(X^3) \\ -6\tau_2^*\beta_2^* E(X^2) - 2\tau_2^*\beta_1^* E(X) + \beta_1^* E(X^3) + 2\beta_2^* E(X^4) - \beta_1^*\tau_1^* E(X^2) - 2\beta_2^*\tau_1^* E(X^3) \\ 0 \end{pmatrix}, \tag{2.9}$$

where β_1^* and β_2^* are the true values of the regression parameters, and τ_1^* and τ_2^* are the probability limits of the MLEs for τ_1 and τ_2 in the $\mathrm{N}(\tau_1, \tau_2)$ model in the case of no measurement error. Thus $\tau_1^* = E(X)$ and $\tau_2^* = E(X^2) - \{E(X)\}^2 = \sigma_X^2$.

The fourth component of (2.9) is identically 0. Substituting $\tau_1^* = E(X)$ and $\tau_2^* = E(X^2) - \{E(X)\}^2 = \sigma_X^2$ in (2.9) and simplifying reveals that the first component of (2.9) is also identically 0. The second component of (2.9) reduces to $2\beta_2^*\sigma_X^3\kappa_{X,3}$ where $\kappa_{X,3}$ is the skewness of X. The third component of (2.9) simplifies to $\beta_1^*\sigma_X^3\kappa_{X,3} + 2\beta_2^*\left[\sigma_X^4\left\{\kappa_{X,4} - 3\right\} + 3\mu_X\sigma_X^3\kappa_{X,3}\right]$ where $\kappa_{X,4}$ is the kurtosis of X. Thus, estimation of the coefficients in the quadratic model with an assumed normal model for X is first-order robust in general only if the distribution of X satisfies $\kappa_{X,3} = 0$ and $\kappa_{X,4} = 3$. Of course, if the model for X is correctly specified, that is, if X is normally distributed, then X has skewness=0 and kurtosis = 3 and all components of (2.9) are 0.

Both full robustness and first-order robustness are valuable analytic constructs for understanding sensitivity of inference to model misspecification. First-order robustness is easier to assess than full robustness, but is still quite involved for many models. The more relevant problem for data analysis is assessing the robustness to a choice of model for $\boldsymbol{X}$ in a particular application. We describe a practical approach in the context of structural measurement error models in the next chapter.

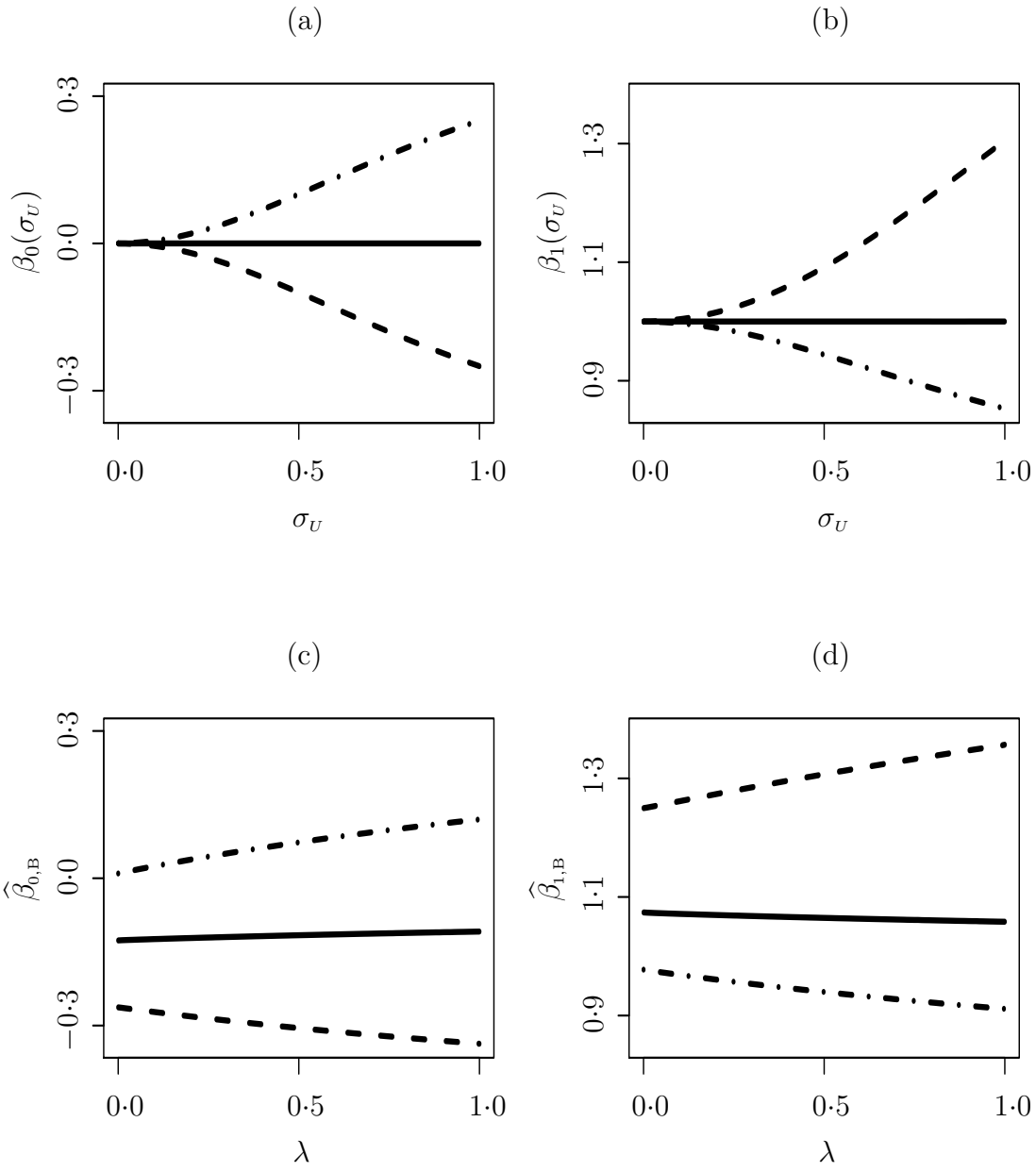


Figure 2.1: Figures (a)–(b) illustrate Example 2.1, plots of $\beta_0(\sigma_U)$ and $\beta_1(\sigma_U)$ for assumed model $N(\tau^{(a)}, \tau^{(a)})$ and three true X distributions, $N(1,1)$ (solid line), $N(0{\cdot}5,1)$ (dashed line) and $N(1{\cdot}5,1)$ (dashed-dotted line); (c)–(d) single-sample re-measurement versions of (a)–(b) as described in Example 4.1.

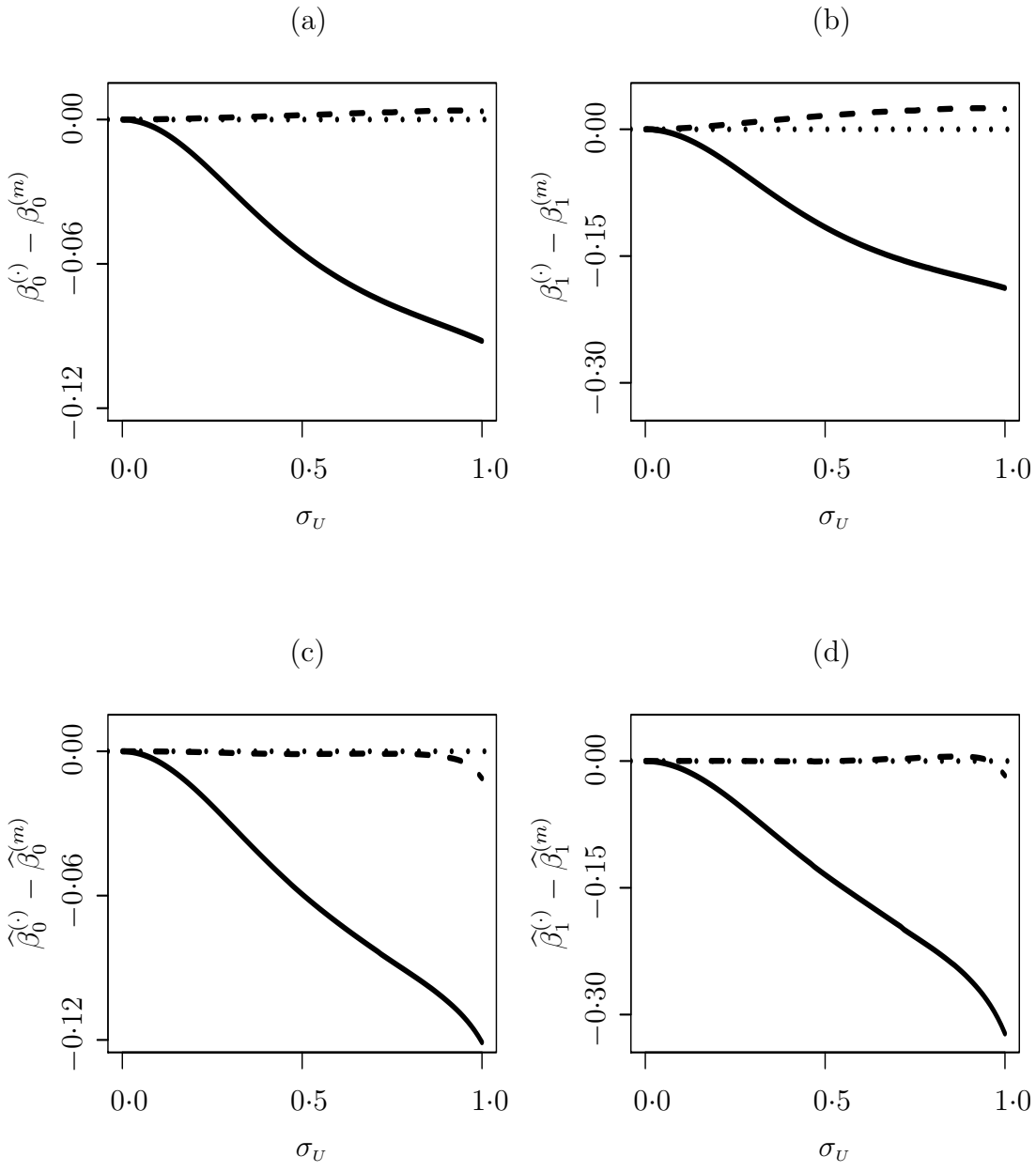


Figure 2.2: Plots for Example 2.2 with $Y|X$ linear-probit. (a)–(b), $\theta^{(n)}(\sigma_U) - \theta^{(m)}(\sigma_U)$ and $\theta^{(s)}(\sigma_U) - \theta^{(m)}(\sigma_U)$; (c)–(d) Monte-Carlo estimates based on 100 replicates, of finite-sample, $n = 500$, version of (a)–(b). The solid line and the dashed line correspond to the normal modelling and seminonparametric modelling, respectively. The dotted line is the reference line.

Chapter 3

Remeasurement Method and Test of Robustness

This chapter first reviews the remeasurement method, or simulation-extrapolation (SIMEX) method, as preparation for its use in diagnosing latent-model robustness in the following chapters. Then testing procedures for assessing latent-model robustness are developed based on the remeasurement method.

3.1 Remeasurement Method

The remeasurment method was developed originally for measurement error models. Thus it is natural, and notationally easier, to review it first in the context of such models. The generalization of the remeasurement method to more complex latent variable models is discussed in Chapters 4 and 5.

The remeasurement method (Cook & Stefanski, 1994; Stefanski & Cook, 1995;

Carroll et al., 1995, Ch. 4) is a simulation-based technique for determining the effects of measurement error, such as bias and variance, on a statistic. The idea is that the effects of measurement error from a particular dataset are determined by computing the statistic on simulated "remeasured" data sets, in which the variables measured with error are further contaminated with Monte-Carlo-generated, pseudo-measurement errors. Once the dependence of the statistic on the variance of the added pseudo-measurement errors is estimated using simple regression models, the biasing effects of measurement error can be lessened by extrapolating the fitted regression model to the case of no measurement error.

The discussion of theoretical robustness in Chapter 2 shows that, when an inadequate model for X is assumed, the bias in $\widehat{\theta}$ is manifested by a nonconstant plot of $\theta(\sigma_U)$. With a single data set there is only one true error variance σ_U^2 and only one calculated statistic, thus at first blush it appears that empirically mimicking the theory in Chapter 2 is not possible. However, data sets with different levels of measurement error can be created by simply adding noise to those variables measured with errors. Using the remeasurement method, one can construct empirical versions of the plots shown in Chapter 2 and thus check for lack of robustness. Specifically, this is done via the following four steps, in which we assume W and X are scalars.

Denote the observed data from the measurement error models defined in Section 1.2 as $\boldsymbol{Q} \triangleq \{\boldsymbol{Q}_i\}_{i=1}^n \triangleq \{Y_i, W_i\}_{i=1}^n$. Then for each of several chosen positive constant values of λ :

- Step 1. For $b = 1, \ldots, B$, generate the bth λ-remeasured data, denoted by

$\{\boldsymbol{Q}_{b,i}(\lambda)\}_{i=1}^n \triangleq \{Y_i, W_{b,i}(\lambda)\}_{i=1}^n$, in which W_i are replaced by

$$W_{b,i}(\lambda) = W_i + \sqrt{\lambda}\sigma_U Z_{b,i},\ i = 1, \ldots, n, \tag{3.1}$$

where $Z_{b,i}$ $(i = 1, \ldots, n)$ are i.i.d. standard normal random errors.

- Step 2. Estimate the parameters based on $\{\boldsymbol{Q}_{b,i}(\lambda)\}_{i=1}^n$. Denote the estimate for $\boldsymbol{\theta}$ as $\widehat{\boldsymbol{\theta}}_b(\lambda)$, $b = 1, \ldots, B$.

- Step 3. Compute $\widehat{\boldsymbol{\theta}}_B(\lambda) = \sum_{b=1}^B \widehat{\boldsymbol{\theta}}_b(\lambda)/B$.

- Step 4. Plot $\widehat{\boldsymbol{\theta}}_B(\lambda)$ versus $\lambda \geq 0$, where $\widehat{\boldsymbol{\theta}}_B(0)$ is the estimate based on the observed data $\boldsymbol{Q}$. We call such plots *SIMEX plots* in the sequel.

In practice σ_U^2 is usually unknown but replicate measurements of X are available. In this case σ_U in (3.1) is substituted by its estimate. For example, in the structural measurement error model described in Section 1.2, suppose there are m_i replicate measures of X_i, then an estimator for σ_U^2 is given by (Carroll et al. 1995),

$$\hat{\sigma}_U^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - \overline{W}_{i\cdot})^2}{\sum_{i=1}^n (m_i - 1)}, \tag{3.2}$$

where $\overline{W}_{i\cdot} = n^{-1} \sum_{j=1}^{m_i} W_{ij}$. In the joint models defined in Section 1.3, an estimate for σ_U^2 used in generating remeasured data is the estimated intra-subject error variance $\hat{\sigma}_U^2$ obtained along with the other parameter estimates in the joint models based on the observed data.

When remeasurment is used to assess and reduced bias, there is an extrapolation step that entails modeling $\widehat{\boldsymbol{\theta}}_B(\lambda)$ as a function of $\lambda \geq 0$ and extrapolating the fitted model to $\lambda = -1$.

For our purpose of diagnosing latent-model robustness we do not need the extrapolation step even though we still use the name SIMEX or remeasurement method to refer to our diagnostic method. Our diagnostic method exploits the fact that, if

$$f_{Y,W}(Y,W;\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}) = \int f_{Y|X}(Y|x;\boldsymbol{\theta}) f_{W|X}(W|x;\sigma_U^2) f_X^{(a)}(x;\boldsymbol{\tau}^{(a)}) dx$$

is a correct model for (Y,W), then

$$f_{Y,W(\lambda)}\{Y,W(\lambda);\boldsymbol{\theta},\boldsymbol{\tau}^{(a)}\} = \int f_{Y|X}(Y|x;\boldsymbol{\theta}) f_{W|X}\{W|x;(1+\lambda)\sigma_U^2\} f_X^{(a)}(x;\boldsymbol{\tau}^{(a)}) dx$$

is a correct model for $(Y, W+\lambda^{1/2}\sigma_U Z)$ for all $\lambda > 0$, where $Z \sim N(0,1)$ independently of (Y,W). Consequently, if the assumed model for X is correct, or robust in the sense defined in Chapter 2, an estimator for $\boldsymbol{\theta}$ derived from the latter model fitted to remeasured data $\{Y_i, W_i + \lambda^{1/2}\sigma_U Z_i\}_{i=1}^n$ should be consistent for $\boldsymbol{\theta}$ regardless of the size of λ, and therefore should exhibit no dependence on λ. Conversely, if the model is incorrect and nonrobust, then absolute bias will tend to increase with increasing measurement error, and this will be manifested by a dependence on λ. For simulation-extrapolation estimation, Carroll et al. (1995) recommend taking $\lambda \in [0, \lambda_{\max}]$ with $1 \leq \lambda_{\max} \leq 3$. For our diagnostic purposes, we take $\lambda_{\max} = 1$ or 3 in most of our examples. Note that the added variance is $\lambda\sigma_U^2$. Thus, if σ_U^2 is small, the amount of added noise will also be small provided λ is not extremely large.

Note that our method is not specific to parametric likelihood estimation. For example, if $\sum \psi(Y_i, W_i, \boldsymbol{\theta}, \boldsymbol{\tau}^{(a)}, \sigma_U^2)$ is a correct or robust estimating equation for $\boldsymbol{\theta}$, the same is true of $\sum \psi\{Y_i, W_{b,i}(\lambda), \boldsymbol{\theta}, \boldsymbol{\tau}^{(a)}, (1+\lambda)\sigma_U^2\}$, and robustness of the M-estimator obtained from $\psi(\cdot)$ can be checked with the remeasurement method. Hence, even though in Chapter 2 we consider MLEs and their asymptotic limits, one may

also diagnose other estimators for robustness. With this understanding, we next give an improved version of the remeasurement method and several test statistics for assessing robustness without restricting the discussion to measurement error models or to MLE.

The simulation step given above requires estimating $\boldsymbol{\theta}$ and other parameters in the full parameter vector $\boldsymbol{\Omega}$ repeatedly B times at each fixed $\lambda > 0$. This can be very computationally prohibitive, especially for estimators that are computationally intensive. We can reduce the computational burden by replacing Steps 2 and 3 above with Step 2* implemented in the following way. Suppose, based on the observed data $\boldsymbol{Q}$, an estimator for $\boldsymbol{\Omega}$ is obtained by solving the system of estimating equations,

$$\sum_{i=1}^{n} \boldsymbol{\psi}(\boldsymbol{Q}_i; \boldsymbol{\Omega}) = \mathbf{0}, \tag{3.3}$$

where the summand $\boldsymbol{\psi}(\boldsymbol{Q}_i; \boldsymbol{\Omega})$ satisfies

$$E\Big[\boldsymbol{\psi}\{\boldsymbol{Q}_i; \boldsymbol{\Omega}(0)\}\Big] = \mathbf{0}. \tag{3.4}$$

Denote the estimator by $\widehat{\boldsymbol{\Omega}}(0)$, where the "0" indicates that no additional noise has been added to the data. Now, instead of taking Steps 2 and 3, in Step 2*, for a fixed $\lambda > 0$, obtain the estimator $\widehat{\boldsymbol{\Omega}}(\lambda)$ by solving the averaged estimating equations

$$\sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}(\lambda); \boldsymbol{\Omega}\} = \mathbf{0}, \tag{3.5}$$

where

$$\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}(\lambda); \boldsymbol{\Omega}\} = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}(\lambda); \boldsymbol{\Omega}\}, \tag{3.6}$$

and $\boldsymbol{Q}_i^{(B)}(\lambda) = \{\boldsymbol{Q}_{b,i}(\lambda)\}_{b=1}^{B}$. In the original Steps 2 and 3, $\widehat{\boldsymbol{\theta}}_B(\lambda)$, along with the other parameters in $\boldsymbol{\Omega}$ that make up the full parameter estimate $\widehat{\boldsymbol{\Omega}}_B(\lambda)$, is obtained

by solving the estimating equations, with solutions denoted by $\widehat{\boldsymbol{\Omega}}_b$,

$$\sum_{i=1}^{n} \boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}(\lambda); \boldsymbol{\Omega}\} = \mathbf{0}, \tag{3.7}$$

for $b = 1, \ldots, B$. With Step 2*, $\widehat{\boldsymbol{\theta}}_B(\lambda)$ only requires solving the estimating equation (3.5) once. Furthermore, the estimating equations in (3.5) are usually "smoother" than those in (3.7), making it easier to solve (3.5) than (3.7). Hereafter we refer to the *improved remeasurement method* to distinguish from the traditional remeasurement method with Steps $1 \sim 4$.

The improved and traditional remeasurement method are interchangable as diagnostic devices because the solution to (3.5), $\widehat{\boldsymbol{\Omega}}(\lambda)$, is asymptotically equivalent to the average of the solutions to (3.7) for $b = 1, \ldots, B$, that is, $\widehat{\boldsymbol{\Omega}}_B(\lambda) = \sum_{i=1}^{B} \widehat{\boldsymbol{\Omega}}_b(\lambda)/B$. This asymptotic equivalence can be demonstrated by noting that, if the summand in (3.7), $\boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}(\lambda); \boldsymbol{\Omega}\}$, satisfies $E\Big[\boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}(\lambda); \boldsymbol{\Omega}(\lambda)\}\Big] = \mathbf{0}$, then $\widehat{\boldsymbol{\Omega}}_b(\lambda) \xrightarrow{p} \boldsymbol{\Omega}(\lambda)$, and obviously $\widehat{\boldsymbol{\Omega}}_B(\lambda) \xrightarrow{p} \boldsymbol{\Omega}(\lambda)$ follows. It also follows that $E\Big[\sum_{b=1}^{B} \boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}(\lambda); \boldsymbol{\Omega}(\lambda)\}/B\Big] = \mathbf{0}$, so that the summand in (3.5) satisfies

$$E\Big[\boldsymbol{\psi}\{\boldsymbol{Q}_i^{(B)}(\lambda); \boldsymbol{\Omega}(\lambda)\}\Big] = \mathbf{0}, \tag{3.8}$$

leading to $\widehat{\boldsymbol{\Omega}}(\lambda) \xrightarrow{p} \boldsymbol{\Omega}(\lambda)$.

3.2 Test of Robustness

Even though the remeasurement SIMEX plot provides a graphical indication of (non)robustness, objective assessment is also necessary. We next describe four statistics for testing the (non)robustness in $\widehat{\boldsymbol{\theta}}(\lambda)$. In this section we assume $\boldsymbol{\Omega}$ is finite di-

mensional of length q. Moreover we partition $\boldsymbol{\Omega}$ into three subsets, $\boldsymbol{\Omega} = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^2)^T$, where $\boldsymbol{\gamma}$ includes all the parameters in $\boldsymbol{\Omega}$ other than the primary regression parameter of central interest, $\boldsymbol{\theta}$, and the error variance σ_U^2. For example, in the measurement error model and simple-response joint models defined in Section 1.3, $\boldsymbol{\gamma} = \boldsymbol{\tau}^{(a)}$, the parameters in the presumed latent-variable model.

The hypotheses corresponding to the question of whether or not the estimator for $\boldsymbol{\Omega}$ is robust can be formulated as $H_0 : \boldsymbol{\Omega}(0) = \boldsymbol{\Omega}(\lambda)$ versus $H_a : \boldsymbol{\Omega}(0) \neq \boldsymbol{\Omega}(\lambda)$, where $\boldsymbol{\Omega}(0) = \{\boldsymbol{\theta}(0)^T, \boldsymbol{\gamma}(0)^T, \sigma_U^2(0)\}^T$ and $\boldsymbol{\Omega}(\lambda) = \{\boldsymbol{\theta}(\lambda)^T, \boldsymbol{\gamma}(\lambda)^T, \sigma_U^{2*}(\lambda)\}^T$ are determined by (3.4) and (3.8), respectively. It is worth pointing out that the relationship between $\sigma_U^2(0)$ and $\sigma_U^{2*}(\lambda)$ is different from the relationship between $\boldsymbol{\theta}(0)$ and $\boldsymbol{\theta}(\lambda)$, or that between $\boldsymbol{\gamma}(0)$ and $\boldsymbol{\gamma}(\lambda)$. Take MLE and simple-response joint model as an example, that is, one applies the remeasurement method to simple-response joint model and computes MLE for $\boldsymbol{\Omega}$. Under H_0, $\boldsymbol{\theta}(0) = \boldsymbol{\theta}(\lambda)$ are the limits of MLE as $n \to \infty$ for the same parameter corresponding to the primary regression model; similarly $\boldsymbol{\gamma}(0) = \boldsymbol{\gamma}(\lambda)$ are the limits of MLE for the same parameter corresponding to the presumed latent-variable model. But $\sigma_U^2(0)$ is the limit of MLE for the intra-subject error variance of the raw data before adding extra simulated noise, and $\sigma_U^{2*}(\lambda)$ is the limit of MLE for the intra-subject error variance of the λ-remeasured data. Therefore, under H_0, $\sigma_U^{2*}(\lambda) = (1 + \lambda)\sigma_U^2(0)$. We define $\sigma_U^2(\lambda) = \sigma_U^{2*}(\lambda)/(1 + \lambda)$ so that under H_0, $\sigma_U^2(\lambda) = \sigma_U^2(0)$. The corresponding estimators are similarly denoted with hats on the preceding notations. Specifically, $\widehat{\boldsymbol{\Omega}}(0) = \{\widehat{\boldsymbol{\theta}}(0)^T, \widehat{\boldsymbol{\gamma}}(0)^T, \widehat{\sigma}_U^2(0)\}^T$ satisfies

$$\sum_{i=1}^{n} \boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \widehat{\sigma}_U^2(0)\} = \mathbf{0}; \tag{3.9}$$

$\widehat{\boldsymbol{\Omega}}(\lambda) = \{\widehat{\boldsymbol{\theta}}(\lambda)^T, \widehat{\boldsymbol{\gamma}}(\lambda)^T, \widehat{\sigma}_U^{2*}(\lambda)\}^T$ satisfies

$$\sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \widehat{\sigma}_U^{2*}(\lambda)\} = \mathbf{0},$$

or equivalently,

$$\sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), (1+\lambda)\widehat{\sigma}_U^{2}(\lambda)\} = \mathbf{0}. \tag{3.10}$$

The four test statistics are based on the following four $q \times 1$ statistics, for a fixed $\lambda > 0$,

$$\begin{aligned}
\boldsymbol{T}_1 &= \sqrt{n}\{\widehat{\boldsymbol{\Omega}}(0) - \widehat{\boldsymbol{\Omega}}(\lambda)\}, && (3.11)\\
\boldsymbol{T}_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), (1+\lambda)\hat{\sigma}_U^{2}(0)\}, && (3.12)\\
\boldsymbol{T}_3 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2}(\lambda)\}, && (3.13)\\
\boldsymbol{T}_4 &= \frac{1}{2}(\boldsymbol{T}_2 - \boldsymbol{T}_3). && (3.14)
\end{aligned}$$

The first statistic, $\boldsymbol{T}_1$, is a direct assessment of the difference in the parameter estimates for different levels of measurement error variance, σ_U^2 versus $(1+\lambda)\sigma_U^2$, with large absolute value indicating lack of robustness. So high $\boldsymbol{T}_1$ in absolute value implies nonrobustness. The intuition leading to $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$ is that, under H_0, the estimates that solve the estimating equations evaluated at the raw data $\boldsymbol{Q}$ should also solve, at least approximately, the estimating equations evaluated at the λ-remeasured data $\boldsymbol{Q}^{(B)}$, and vice versa. Therefore under H_0, $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$ should be close to zero and significant deviation from zero implies nonrobust estimates. $\boldsymbol{T}_4$ is a symmetrized version of $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$.

The estimators for the variance-covariance matrix of $\boldsymbol{T}_k$, denoted by $\widehat{\boldsymbol{V}}_k^2$, are given

by

$$\widehat{\boldsymbol{V}}_k^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{R}_{ki} - \overline{\boldsymbol{R}}_{k\cdot})(\boldsymbol{R}_{ki} - \overline{\boldsymbol{R}}_{k\cdot})^T,$$

where $\overline{\boldsymbol{R}}_{k\cdot}$ is the average of $\boldsymbol{R}_{ki}$ over $i = 1, \ldots, n$ for k =1, 2, 3, 4;

$$\begin{aligned}\boldsymbol{R}_{1i} &= \widehat{\boldsymbol{A}}_1^{-1}\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\} - \\ &\quad \widehat{\boldsymbol{A}}_2^{-1}\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\};\end{aligned} \tag{3.15}$$

$$\begin{aligned}\boldsymbol{R}_{2i} &= \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), (1+\lambda)\hat{\sigma}_U^2(0)\} - \widehat{\boldsymbol{A}}_2\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), (1+\lambda)\hat{\sigma}_U^2(0)\} \\ &\quad \widehat{\boldsymbol{A}}_1^{-1}\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\};\end{aligned} \tag{3.16}$$

$$\begin{aligned}\boldsymbol{R}_{3i} &= \boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^2(\lambda)\} - \widehat{\boldsymbol{A}}_1\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^2(\lambda)\} \\ &\quad \widehat{\boldsymbol{A}}_2^{-1}\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\};\end{aligned} \tag{3.17}$$

$$\begin{aligned}\boldsymbol{R}_{4i} &= \frac{1}{2}\Big[\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), (1+\lambda)\hat{\sigma}_U^2(0)\} - \widehat{\boldsymbol{A}}_2\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), (1+\lambda)\hat{\sigma}_U^2(0)\} \\ &\quad \widehat{\boldsymbol{A}}_1^{-1}\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(0), \widehat{\boldsymbol{\gamma}}(0), \hat{\sigma}_U^2(0)\} - \\ &\quad \boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^2(\lambda)\} + \widehat{\boldsymbol{A}}_1\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^2(\lambda)\} \\ &\quad \widehat{\boldsymbol{A}}_2^{-1}\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \hat{\sigma}_U^{2*}(\lambda)\}\Big];\end{aligned} \tag{3.18}$$

$$\widehat{\boldsymbol{A}}_1\{\boldsymbol{Q}; \widehat{\boldsymbol{\theta}}(\cdot), \widehat{\boldsymbol{\gamma}}(\cdot), \hat{\sigma}_U^2(\cdot)\} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\boldsymbol{\psi}\{\boldsymbol{Q}_i; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^2\}}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^2)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\cdot), \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}(\cdot), \sigma_U^2=\hat{\sigma}_U^2(\cdot)} \tag{3.19}$$

is the empirical estimator for

$$\boldsymbol{A}_1\{\boldsymbol{\theta}(\cdot), \boldsymbol{\gamma}(\cdot), \sigma_U^2(\cdot)\} = E\Big[-\frac{\partial\boldsymbol{\psi}\{\boldsymbol{Q}_i; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^2\}}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^2)}\Big]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}(\cdot), \boldsymbol{\gamma}=\boldsymbol{\gamma}(\cdot), \sigma_U^2=\sigma_U^2(\cdot)};$$

and

$$\begin{aligned} & \widehat{\boldsymbol{A}}_2\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\theta}}(\cdot), \widehat{\boldsymbol{\gamma}}(\cdot), (1+\lambda)\hat{\sigma}_U^2(\cdot)\} \\ = & -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^{2*}\}}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^{2*})} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\cdot), \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}(\cdot), \sigma_U^{2*}=(1+\lambda)\sigma_U^2(\cdot)} \end{aligned} \tag{3.20}$$

is the empirical estimator for

$$\boldsymbol{A}_2\{\boldsymbol{\theta}(\cdot), \boldsymbol{\gamma}(\cdot), (1+\lambda)\sigma_U^2(\cdot)\} = E\Big[-\frac{\partial \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^{2*}\}}{\partial(\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^{2*})}\Big] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}(\cdot), \boldsymbol{\gamma}=\boldsymbol{\gamma}(\cdot), \sigma_U^{2*}=(1+\lambda)\sigma_U^2(\cdot)}.$$

Denote by $T_{k,d}$ the dth element in $\boldsymbol{T}_k$, and define $\widehat{\nu}_{k,d} = \sqrt{\widehat{\boldsymbol{V}}_k(d,d)}$, for $d = 1, \ldots, q$ and k =1, 2, 3, 4. The test statistics for testing the robustness of the dth parameter in $\boldsymbol{\Omega}$ are $T_{k,d}^* = T_{k,d}/\widehat{\nu}_{k,d}$, for $d = 1, \ldots, q$ and k =1, 2, 3, 4. Define $\boldsymbol{T}_k^* = (T_{k,1}^*, \ldots, T_{k,q}^*)^T$ for k=1, 2, 3, 4. The derivations of $\boldsymbol{R}_{ki}$ and the asymptotic distributions of $\boldsymbol{T}_k$ (k=1, 2, 3, 4) are given in Appendix B.

As indicated in the proof in Appendix B, the four test statistics are asymptotically equivalent and thus are expected to have similar operating characteristics in large sample. This equivalence was verified by some simulation studies that are not reported herein. The statistics $\boldsymbol{T}_1$, $\boldsymbol{T}_3$, and $\boldsymbol{T}_4$ depend on $\widehat{\boldsymbol{\Omega}}(\lambda)$ whereas $\boldsymbol{T}_2$ does not. Thus, $\boldsymbol{T}_2^*$ has the advantage of not requiring computation of $\widehat{\boldsymbol{\Omega}}(\lambda)$. However, taking the time to compute $\widehat{\boldsymbol{\Omega}}(\lambda)$ may be worthwhile when it is of interest to make the SIMEX plot to visualize how bias depends on error variance. In the following chapters, we use $\boldsymbol{T}_1^*$ for hypothesis testing of robustness when entertaining simpler latent variable models and focus on $\boldsymbol{T}_2^*$ when the models are more complicated and computing $\widehat{\boldsymbol{\Omega}}(\lambda)$ is time-consuming.

Chapter 4

Latent-Model Robustness in Measurement Error Models

In this chapter the traditional remeasurement method described in Section 3.1 is applied to structural measurement error models and the test statistic $\boldsymbol{T}_1^*$ is used to test robustness of $\widehat{\boldsymbol{\theta}}$. The utility of these methods are demonstrated by application to simulated data and data from the Framingham study descried in Section 1.2.

4.1 Simulated Examples

We present three examples based on simulated data. In each example, we construct the SIMEX plot or plot the deviation of parameter estimates from "gold standard" estimates as described in Example 2.2 in Section 2.1. Deviations from a horizontal plot indicate non-robustness to the chosen model for the true predictor X.

Example 4.1: *Y given X follows a normal distribution with mean $\beta_0 + \beta_1 X$.* A

sample of $n = 500$ was generated based on the simple linear regression model $Y|X \sim N(\beta_0 + \beta_1 X, \sigma_\epsilon^2)$ with $\beta_0 = 0$, $\beta_1 = 1$, and $\sigma_\epsilon = 1$, for each of the three cases $X \sim \mathrm{N}(1,\ 1)$, $X \sim \mathrm{N}(0.5,\ 1)$, and $X \sim \mathrm{N}(1.5,\ 1)$. Given X, Y and W were generated from the conditional models $Y|X \sim \mathrm{N}(X,\ 1)$ and $W|X \sim \mathrm{N}(X,\ 0.5)$. In all three cases the likelihood was constructed assuming the $\mathrm{N}(\tau^{(a)},\ \tau^{(a)})$ distribution for X. $B = 500$ λ-remeasured data sets were generated for each fixed λ varying from 0 to 1. Figure 2.1 (c)–(d) displays the empirical version of the plots in Figure 2.1 (a)–(b).

For the case $X \sim \mathrm{N}(1,\ 1)$, that is, when the assumed model is correct, the component curves of $\widehat{\boldsymbol{\theta}}_B(\lambda)$ and $\widehat{\tau}_B^{(a)}(\lambda)$ are expected to be horizontal lines. This case is easily recognized in each panel of Figure 2.1 (c)–(d). For the cases $X \sim \mathrm{N}(0.5,\ 1)$ and $X \sim \mathrm{N}(1.5,\ 1)$ the assumed model is incorrect, which in general should result in non-horizontal component curves of $\widehat{\boldsymbol{\theta}}_B(\lambda)$ and $\widehat{\tau}_B^{(a)}(\lambda)$. These cases are also readily identified for the three regression model parameters.

Example 4.2: *Y given X follows a Bernoulli distribution with mean probit($\beta_0 + \beta_1 X$).* Independent random pairs $\{y_j, w_j\}_{j=1}^{2000}$ ($n = 2000$) were generated from the measurement error model defined in Example 2.2 with measurement error variance $\sigma_U^2 = 0.16$. When generating remeasured data, we varied λ from 0 to 3 to obtain λ-remeasured data sets with reliability ratio (Carroll et al., 1995, p. 22) varying from 0.86 to 0.61. For each fixed λ, $B = 100$ remeasured data sets were generated. Maximum likelihood estimates were calculated for each of the three assumed models used in Example 2.2.

Using a plotting strategy similar to that in Example 2.2, treating the estimates from the mixture-normal modeling as the "gold standard," deviations of the normal and SNP estimates from the "gold standard" are plotted versus λ in Figure 4.1. The plot shows that normal modeling, which is over-restrictive relative to the true model, results in estimates that differ increasingly from the true values as the measurement error variance increases. Flexible modeling via the SNP family of models results in estimates that are very close to those from the correct modeling. In fact, the estimated moments of X (not shown) from SNP modeling are nearly equal to the moments of true mixture normal density up to very high orders, indicating the estimated SNP density approximates the true mixture normal density very well for this data set.

Example 4.3: *Y given X follows a Bernoulli distribution with mean logistic($\beta_0 + \beta_1 X$).* In this example, the setup of the measurement error model and the three assumed models for X are identical to those in Example 4.2, except that now $P(Y = 1|X = x) = \{1 + \exp(-\beta_0 - \beta_1 x)\}^{-1}$.

In addition to the three sets of $\widehat{\boldsymbol{\theta}}_B(\lambda)$'s under the three assumed models for X, the conditional score estimator of $\boldsymbol{\theta}$ (Stefanski and Carroll, 1987), denoted as $\widehat{\boldsymbol{\theta}}_*(\lambda)$, was also computed. The conditional score equation for this linear-logistic measurement error model is derived from the conditional density of $Y|\Delta$, where $\Delta = W + Y\sigma_{\mathrm{U}}^2\beta_1$. It can be shown that, if X is treated as an unknown parameter and both σ_U and β_1 are assumed known, then Δ is a sufficient statistic for X. Consequently, the conditional density of $Y|\Delta$, and thus the conditional score equation, is free of X. Hence, conditional score estimators that solve the conditional score equations are

expected to be robust to the model specification for X.

Based on a random sample of size $n = 2000$ generated from the logistic measurement error model, the conditional score estimates and three sets of MLEs were computed for the generated λ-remeasured data sets. Figure 4.2 displays these estimates as functions of λ. As expected, the conditional score estimates do not vary much as λ varies; moreover, correct modeling (mixture-normal modeling) and flexible modeling (SNP modeling) also yield robust estimates, that are very close to the conditional score estimates. In contrast, normal modeling leads to estimates with apparent bias increasing as λ increases.

4.2 Test of Robustness

Corresponding to the analysis of the simulated data set presented in Figure 4.1 for Example 4.2, Table 4.1 gives the values of $T^*_{1,1}$ and $T^*_{1,2}$ for testing significance of the differences $\widehat{\beta}_{0,\,B}(0) - \widehat{\beta}_{0,\,B}(3)$ and $\widehat{\beta}_{1,\,B}(0) - \widehat{\beta}_{1,\,B}(3)$, respectively, for each assumed model for X. Using 1.96 as the critical value, the associated p-values in Table 4.1 indicate that the changes in the estimates are much more significant for the normal-modeling than for the other two approaches.

Table 4.2 gives the values of $T^*_{1,1}$ and $T^*_{1,2}$ corresponding to Example 4.3 presented in Figure 4.2. The p-values associated with the values of $T^*_{1,1}$ and $T^*_{1,2}$ in Table 4.2 are nearly zero when normal modeling is assumed for X, and are high above 0.05 when X is modeled as SNP or mixture normal, or for CSE. This also justifies the claims made previously regarding robustness.

Under the scenario and parameter settings in Example 2.2, we examined operating characteristics of $T^*_{1,1}$ and $T^*_{1,2}$ via a Monte Carlo simulation study. In the simulation, 100 data sets of size $n = 1000$ were generated from the probit measurement error model. For each data set, the remeasurement method was applied with $B = 50$ and λ varying from 0 to 3, and $T^*_{1,1}$ and $T^*_{1,2}$ were computed for testing robustness of $\widehat{\beta}_{0,B}$ and $\widehat{\beta}_{1,B}$, respectively. Table 4.3 gives the rejection rates corresponding to $T^*_{1,1}$, defined as the proportion of MC replications with $|T^*_{1,1}| > 1.96$, and the rejection rates similarly defined corresponding to $T^*_{1,2}$, associated with MLEs for β_0 and β_1 under different assumed model for X. As $\widehat{\boldsymbol{\theta}}^{(m)}$ is expected to be robust with the assumed latent-variable model being the true model, the low rejection rates resulting from mixture-normal modeling suggest that the testing procedure with test statistics $T^*_{1,1}$ and $T^*_{1,2}$ has a reasonable size. The similarly low rejection rates for SNP modeling agree with the conclusion drawn in Examples 2.2 and 4.2 that flexible modeling on X leads to more robust estimates. With much higher rejection rates for normal modeling and the nonrobustness observations from Examples 2.2 and 4.2, the test appears to have high power to detect nonrobustness caused by an inadequate latent-model assumption.

4.3 Application to Framingham Study

We apply the remeasurement method and testing procedure to data on 1615 individuals from the Framingham study described in Section 1.2. We take W to be the average of two measurements of systolic blood pressure (SBP) for each individual at Exam 2. We regard W as a contaminated version of true long-term SBP, X. Y is the

binary indicator of evidence of CHD within the follow-up period. We use the SBP measures from Exams 2 and 3 to estimate σ_U^2 as in Carroll et al. (1995).

To construct the likelihood function, we assume $P(Y = 1|X)$ follows probit($\beta_0 + \beta_1 X$), with X modeled the same three ways as in Example 2.2. We want to investigate the robustness of the resulting parameter estimates under three assumed models for X with different degrees of flexibility. When generating λ-remeasured data sets, we varied λ from 0 to 3, with $B = 100$ remeasured data sets generated for each fixed λ. Figure 4.3 presents plots of the $\widehat{\beta}_{0,\,B}(\lambda)$ and $\widehat{\beta}_{1,\,B}(\lambda)$ versus λ for the three ways of modeling and shows that the SNP modeling leads to the most robust estimates while the normal modeling results in the least robust estimates. It again suggests gain in robustness from flexible modeling of X.

The values of $T_{1,1}^*$ and $T_{1,2}^*$ for assessing the robustness of MLEs for β_0 and β_1 resulting from different modelings and the corresponding p-values are given in Table 4.4. The fact that SNP modeling gives the smallest $T_{1,1}^*$ and $T_{1,2}^*$ in absolute value and normal modeling yields the largest $T_{1,1}^*$ and $T_{1,2}^*$ in absolute value is consistent with the visual impression of Figure 4.3.

Table 4.1: Values of $T^*_{1,1}$ and $T^*_{1,2}$ assessing robustness of the regression parameter estimates when $\lambda = 0$ and $\lambda = 3$ under three ways of modeling for the simulated data used in Example 4.2. Corresponding p-values are given in the parentheses.

	Normal	SNP	Mixture-Normal
Change in $\widehat{\beta}_{0,\,B}$	6.91 (0.00)	0.12 (0.90)	−0.83 (0.41)
Change in $\widehat{\beta}_{1,\,B}$	4.99 (0.00)	−0.20 (0.84)	−0.32 (0.75)

Table 4.2: Values of $T^*_{1,1}$ and $T^*_{1,2}$ assessing robustness of the conditional score estimates and MLEs for the regression parameter under three ways of modeling when $\lambda = 0$ and $\lambda = 3$ for the simulated data used in Example 4.3. Corresponding p-values are given in the parentheses.

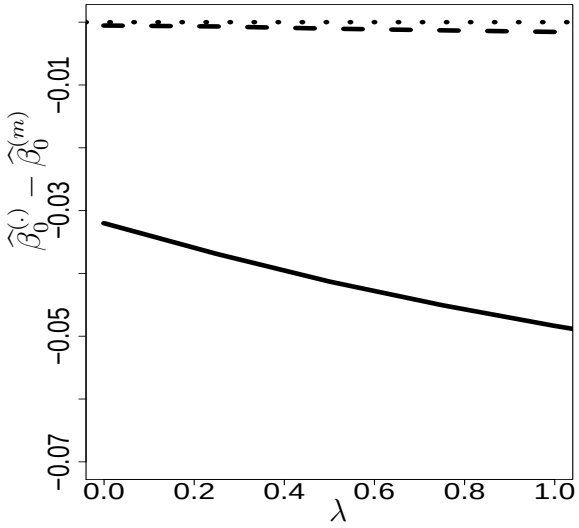
	Normal	SNP	Mixture-Normal	Conditional-score
Change in $\widehat{\beta}_{0,\,B}$	5.26 (0.00)	0.65 (0.52)	0.61 (0.54)	0.12 (0.90)
Change in $\widehat{\beta}_{1,\,B}$	3.45 (0.00)	0.72 (0.47)	0.73 (0.47)	−0.66 (0.51)

Table 4.3: Rejection rates (proportion of 100 data sets with $|T^*_{1,\cdot}| > 1.96$) in testing robustness of the estimates using $T^*_{1,1}$ and $T^*_{1,2}$ for β_0 and β_1, respectively, when λ varies from 0 to 3 under three ways of modeling for the simulated data. Numbers in the parentheses are estimated standard errors of the rejection rates.

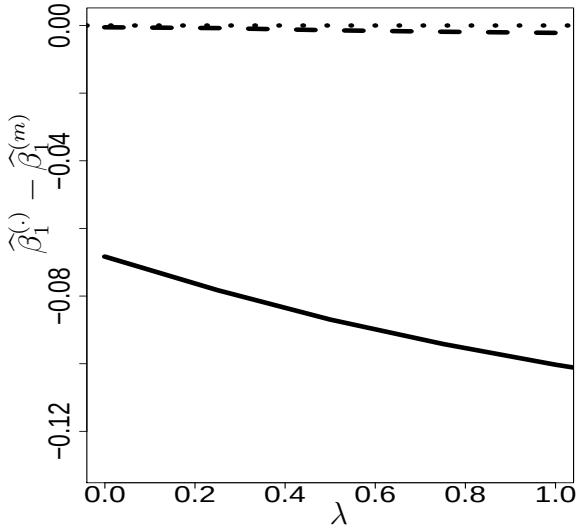
	Normal	SNP	Mixture-Normal
Test the change in $\widehat{\beta}_{0,\,B}$	1 (0)	0.08 (0.03)	0.07 (0.03)
Test the change in $\widehat{\beta}_{1,\,B}$	0.99 (0.01)	0.1 (0.03)	0.1 (0.03)

Table 4.4: Values of $T^*_{1,1}$ and $T^*_{1,2}$ assessing robustness of the regression parameter estimates when $\lambda = 0$ and $\lambda = 3$ under three ways of modeling for the Framingham data. Corresponding p-values are given in the parentheses.

	Normal	SNP	Mixture-Normal
Change in $\widehat{\beta}_{0,\,B}$	3.16 (0.00)	−0.18 (0.86)	−0.81 (0.42)
Change in $\widehat{\beta}_{1,\,B}$	−3.21 (0.00)	0.19 (0.85)	0.81 (0.42)

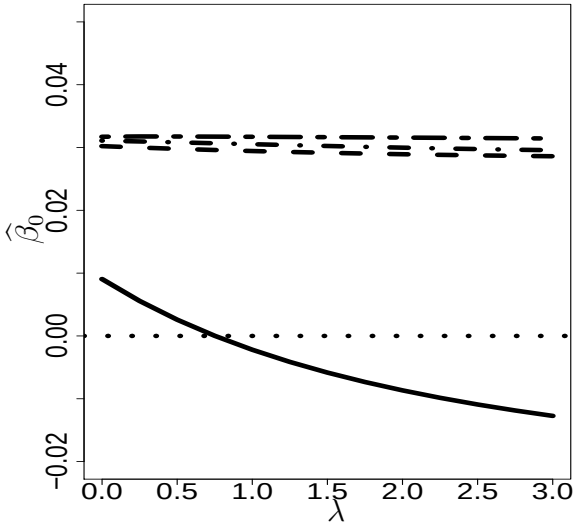


(a)

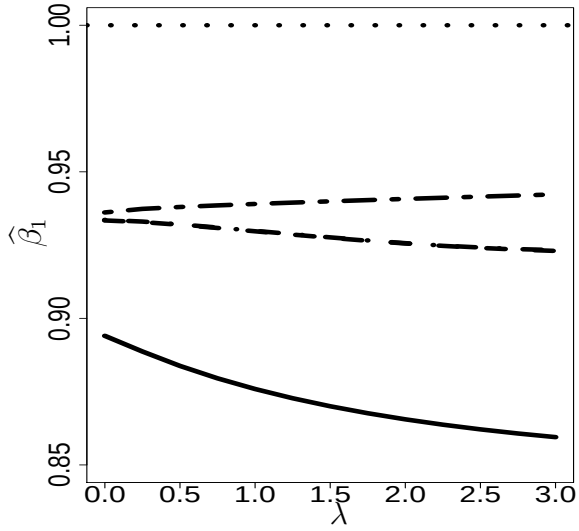


(b)

Figure 4.1: Deviations from the MLEs resulting from the mixture-normal modeling when modeling X as normal and SNP, Example 4.2; (a) corresponds to $\widehat{\beta}_{0,B}$; (b) corresponds to $\widehat{\beta}_{1,B}$. The correspondence of the line types and ways of modeling is the same as used in Figure 2.2.



(a)



(b)

Figure 4.2: MLEs under the three assumed models for X and conditional score estimates versus λ, Example 4.3; (a) corresponds to $\widehat{\beta}_{0,B}$ and $\widehat{\beta}_{0,*}$; (b) corresponds to $\widehat{\beta}_{1,B}$ and $\widehat{\beta}_{1,*}$. True values of β_0 and β_1 are marked by the dotted reference lines. Line types used for assumed models are identical to those used in Figure 2.2. The long-short-dash line corresponds to the conditional score estimates.

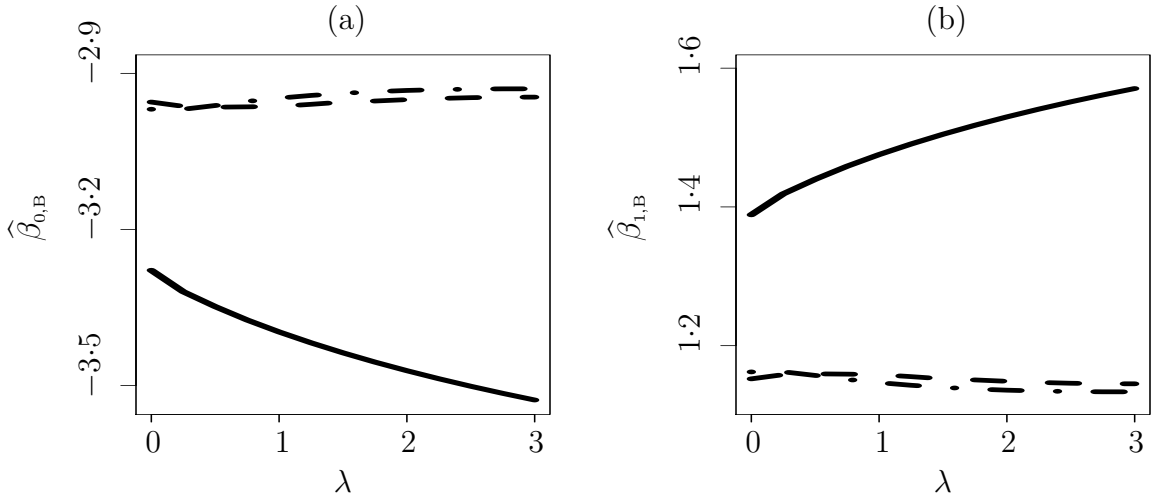


Figure 4.3: $\widehat{\theta}_{\mathrm{B}}^{(n)}$ (solid line), $\widehat{\theta}_{\mathrm{B}}^{(s)}$ (dashed line), and $\widehat{\theta}_{\mathrm{B}}^{(m)}$ (dashed-dotted line) resulting from applying the remeasurement method with $B = 100$ to Framingham study data.

Chapter 5

Latent-Model Robustness in Joint Models

5.1 Expected Robustness

We now consider joint models that have a linear mixed effect model for the longitudinal process, where the random effect is a latent variable. Insensitivity of inference on $\boldsymbol{\theta}$ to the assumed random-effect model has been reported in the literature. Hsieh et al. (2006) gave a heuristic explanation for this insensitivity "when reasonably large numbers of longitudinal measurements are available per subject" using a Laplace approximation technique. In this section, we prove more explicitly that the insensitivity to the assumed random-effect model is expected when the subject-specific information on the longitudinal process is sufficiently large.

Without loss of generality, we consider the simple-response joint model defined in Section 1.3 with observed-data density in (1.5). For brevity, in this section, we drop

the subscript, that is the subject index i, in (1.5) and the other notations defined for this model in Section 1.3. This gives the observed-data density as

$$f_{Y,\boldsymbol{W}}(Y,\boldsymbol{W};\boldsymbol{\Omega}) = \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x};\boldsymbol{\theta}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{W}|\boldsymbol{x};\sigma_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)}) d\boldsymbol{x}, \tag{5.1}$$

where $f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{W}|\boldsymbol{x};\sigma_U^2)$ is the density of $N_m(\boldsymbol{D}\boldsymbol{x},\sigma_U^2\boldsymbol{I}_m)$. Recall that m is the number of longitudinal measures for a subject, also the length of $\boldsymbol{W}$, and $\boldsymbol{D}$ is the design matrix in the linear mixed model for longitudinal process $\boldsymbol{W} = \boldsymbol{D}\boldsymbol{X} + \boldsymbol{U}$. Denote by $\widehat{\boldsymbol{X}}_m$ the ordinary least squares estimator (OLSE) for $\boldsymbol{X}$, i.e., $\widehat{\boldsymbol{X}}_m = (\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{W}_{m\times 1}$. Assuming σ_U^2 known and viewing $\boldsymbol{X}$ as an unknown parameter, $\widehat{\boldsymbol{X}}_m$ is a complete sufficient statistic for $\boldsymbol{X}$, and $\widehat{\boldsymbol{X}}_m|\boldsymbol{X} \sim N_p\{\boldsymbol{X},\sigma_U^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\}$. Therefore, by the Facterization Theorem,

$$f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{W}|\boldsymbol{x};\sigma_U^2) = f_{\boldsymbol{W}|\hat{\boldsymbol{X}}_m}(\boldsymbol{W}|\widehat{\boldsymbol{X}}_m;\sigma_U^2) f_{\hat{\boldsymbol{X}}_m|\boldsymbol{X}}(\widehat{\boldsymbol{X}}_m|\boldsymbol{x};\sigma_U^2),$$

where $f_{\boldsymbol{W}|\hat{\boldsymbol{X}}_m}(\boldsymbol{W}|\widehat{\boldsymbol{X}}_m;\sigma_U^2)$ is free of $\boldsymbol{X}$,

$$f_{\hat{\boldsymbol{X}}_m|\boldsymbol{X}}(\widehat{\boldsymbol{X}}_m|\boldsymbol{x};\sigma_U^2) = |\boldsymbol{G}_m|^{-1}\phi\{\boldsymbol{G}_m^{-1}(\widehat{\boldsymbol{X}}_m - \boldsymbol{x})\},$$

$\phi(\cdot)$ is the density of p-dimensional standard normal distribution, and $\boldsymbol{G}_m$ satisfies $\boldsymbol{G}_m\boldsymbol{G}_m^T = \sigma_U^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}$. It immediately follows that the observed-data density (5.1) is equal to

$$\begin{aligned} & \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x};\boldsymbol{\theta}) f_{\boldsymbol{W}|\boldsymbol{X}}(\boldsymbol{W}|\boldsymbol{x};\sigma_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)}) d\boldsymbol{x} \\ = \;& f_{\boldsymbol{W}|\hat{\boldsymbol{X}}_m}(\boldsymbol{W}|\widehat{\boldsymbol{X}}_m,\sigma_U^2) \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x},\boldsymbol{\theta}) f_{\hat{\boldsymbol{X}}_m|\boldsymbol{X}}(\widehat{\boldsymbol{X}}_m|\boldsymbol{x},\sigma_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)}) d\boldsymbol{x}. \end{aligned} \tag{5.2}$$

Next consider the integral in (5.2). Specifically, consider the difference

$$\begin{aligned}
& \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x},\boldsymbol{\theta}) f_{\hat{\boldsymbol{X}}_m|\boldsymbol{X}}(\widehat{\boldsymbol{X}}_m|\boldsymbol{x},\sigma_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x}|\boldsymbol{\tau}^{(a)}) d\boldsymbol{x} \\
& -f_{Y|\hat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m,\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m|\boldsymbol{\tau}^{(a)}) \qquad (5.3) \\
= & \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x},\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x}|\boldsymbol{\tau}^{(a)}) |\boldsymbol{G}_m|^{-1} \phi\{\boldsymbol{G}_m^{-1}(\widehat{\boldsymbol{X}}_m - \boldsymbol{x})\} d\boldsymbol{x} - \\
& \int f_{Y|\hat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m;\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m|\boldsymbol{\tau}^{(a)}) |\boldsymbol{G}_m|^{-1} \phi\{\boldsymbol{G}_m^{-1}(\boldsymbol{x} - \widehat{\boldsymbol{X}}_m)\} d\boldsymbol{x} \\
= & \int \Big\{ f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x};\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x}|\boldsymbol{\tau}^{(a)}) - f_{Y|\hat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m;\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m|\boldsymbol{\tau}^{(a)}) \Big\} \\
& \qquad |\boldsymbol{G}_m|^{-1} \phi\{\boldsymbol{G}_m^{-1}(\boldsymbol{x} - \widehat{\boldsymbol{X}}_m)\} d\boldsymbol{x} \\
= & \int \Big\{ f_{Y|\boldsymbol{X}}(Y|\widehat{\boldsymbol{X}}_m + \boldsymbol{G}_m\boldsymbol{z};\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m + \boldsymbol{G}_m\boldsymbol{z}|\boldsymbol{\tau}^{(a)}) - \\
& \qquad f_{Y|\hat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m,\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m|\boldsymbol{\tau}^{(a)}) \Big\} \phi(\boldsymbol{z}) d\boldsymbol{z}, \qquad (5.4)
\end{aligned}$$

where in (5.4) we have made the change of variables $\boldsymbol{x} = \widehat{\boldsymbol{X}}_m + \boldsymbol{G}_m\boldsymbol{z}$.

Assume $f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x}) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x})$ continuous and bounded in $\boldsymbol{x}$ for each fixed Y such that $|f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x}) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x})| \leq M$ for some constant $M > 0$ for each fixed Y. Then the absolute value of the integrand in (5.4) is bounded by $2M\phi(\boldsymbol{z})$, which is integrable. We now assume that as $m \to \infty$, the minimum eigenvalue of $\sigma_U^{-2}\boldsymbol{D}^T\boldsymbol{D}$ diverges to $+\infty$. In other words, we assume that the subject-specific information increases without bound as $m \to \infty$. It follows that $\sigma_U^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}$ and thus $\boldsymbol{G}_m$ converges to the $\boldsymbol{0}_{p\times p}$ matrix as $m \to \infty$, and consequently $\widehat{\boldsymbol{X}}_m + \boldsymbol{G}_m\boldsymbol{z} \to \widehat{\boldsymbol{X}}_m$ for each fixed $\boldsymbol{z}$. Therefore, by Lebesgue Dominated Convergence Theorem, (5.4), and thus (5.3), converges to zero in probability as $m \to \infty$. Hence for large m

$$\begin{aligned}
& f_{\boldsymbol{W}|\hat{\boldsymbol{X}}_m}(\boldsymbol{W}|\widehat{\boldsymbol{X}}_m,\sigma_U^2) \int f_{Y|\boldsymbol{X}}(Y|\boldsymbol{x},\boldsymbol{\theta}) f_{\hat{\boldsymbol{X}}_m|\boldsymbol{X}}(\widehat{\boldsymbol{X}}_m|\boldsymbol{x},\sigma_U^2) f_{\boldsymbol{X}}^{(a)}(\boldsymbol{x};\boldsymbol{\tau}^{(a)}) d\boldsymbol{x} \qquad (5.5) \\
\approx & \; f_{\boldsymbol{W}|\hat{\boldsymbol{X}}_m}(\boldsymbol{W}|\widehat{\boldsymbol{X}}_m,\sigma_U^2) f_{Y|\hat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m,\boldsymbol{\theta}) f_{\boldsymbol{X}}^{(a)}(\widehat{\boldsymbol{X}}_m|\boldsymbol{\tau}^{(a)}). \qquad (5.6)
\end{aligned}$$

Note that the parameter of interest, $\boldsymbol{\theta}$, appears only in the density of Y given $\widehat{\boldsymbol{X}}_m$, $f_{Y|\widehat{\boldsymbol{X}}_m}(Y|\widehat{\boldsymbol{X}}_m, \boldsymbol{\theta})$ in (5.6). Therefore, where m is sufficiently large, likelihoods constructed from (5.6) and (5.5) will be approximately equal. Clearly the MLE from the former does not depend on $f_X^{(a)}(\cdot)$ and thus neither will the MLE of the latter as $m \to \infty$.

When it is not clear in a specific application how much longitudinal information is sufficient to achieve a robust estimator, it is preferable and safer to model $\boldsymbol{X}$ in a way such that the resulting estimators for $\boldsymbol{\theta}$ are insensitive to intra-subject errors. We next apply the remeasurement method and the test statistics derived in Chapter 3 to test estimator robustness under various assumed random-effect models.

5.2 Simulated Examples

Two simulated examples are presented in this section to demonstrate the implementation and performance of traditional or improved remeasurement method. The steps of these methods are similar to those implemented in Chapter 4 for structural measurement error models. For completeness the methods are outlined next using the notations defined in joint-model setting.

Denote the observed data generated from the joint models as $\boldsymbol{Q} \triangleq \{\boldsymbol{Q}_i\}_{i=1}^n$, where $\boldsymbol{Q}_i = \{Y_i, \boldsymbol{W}_i\}_{i=1}^n$ in the simple-response joint models, and $\boldsymbol{Q}_i = \{V_i, \Delta_i, \boldsymbol{W}_i, \boldsymbol{t}_i, F_i\}$ in the censored-endpoint joint models. For each of several chosen positive constants values of λ :

- Step 1. For $b = 1, \ldots, B$, generate the bth λ-remeasured data, denoted by $\{\boldsymbol{Q}_{b,i}(\lambda)\}_{i=1}^{n}$, in which $\boldsymbol{W}_i$ in the observed data are replaced by

$$\boldsymbol{W}_{b,i}(\lambda) = \boldsymbol{W}_i + \sqrt{\lambda}\sigma_U \boldsymbol{Z}_{b,i}, \tag{5.7}$$

  for $i = 1, \ldots, n$, where $\boldsymbol{Z}_{b,i}$ $(i = 1, \ldots, n)$ are i.i.d. m_i-dimensional standard normal random errors. Denote by $\boldsymbol{Q}^{(B)}(\lambda)$ the entire B remeasured data for all subjects, and $\boldsymbol{Q}_i^{(B)}(\lambda)$ the B remeasured data for subject i, for $i = 1, \ldots, n$.

- Step 2. Estimate the parameters based on $\{\boldsymbol{Q}_{b,i}(\lambda)\}_{i=1}^{n}$. Denote the estimate for $\boldsymbol{\theta}$ as $\widehat{\boldsymbol{\theta}}_b(\lambda)$, $b = 1, \ldots, B$.

- Step 3. Compute $\widehat{\boldsymbol{\theta}}_B(\lambda) = \sum_{b=1}^{B} \widehat{\boldsymbol{\theta}}_b(\lambda)/B$.

- Step 4. Plot $\widehat{\boldsymbol{\theta}}_B(\lambda)$ versus $\lambda \geq 0$, where $\widehat{\boldsymbol{\theta}}_B(0)$ is the estimate based on the observed data $\boldsymbol{Q}$.

In practice σ_U in (5.7) is substituted by its estimate that is obtained based on $\boldsymbol{Q}$ the same time the full parameter estimate for $\boldsymbol{\Omega}$ is computed.

Partition the full parameter vector $\boldsymbol{\Omega}$ into three subsets, $\boldsymbol{\Omega} = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T, \sigma_U^2)^T$, where $\boldsymbol{\gamma}$ includes all the parameters in $\boldsymbol{\Omega}$ other than the primary regression parameter $\boldsymbol{\theta}$ and the intra-subject error variance σ_U^2. For simple-response joint models, $\boldsymbol{\gamma} = \boldsymbol{\tau}^{(a)}$, the parameters in the assumed random-effect model; for censored-endpoint joint models, $\boldsymbol{\gamma} = \{\boldsymbol{\tau}^{(a)T}, \lambda_0(u)\}^T$; that is, $\boldsymbol{\gamma}$ also includes the infinite dimensional baseline hazards.

Suppose the full parameter estimator obtained based on $\boldsymbol{Q}$, denoted by $\widehat{\boldsymbol{\Omega}}(0) =$

$\{\widehat{\boldsymbol{\theta}}^T(0), \widehat{\boldsymbol{\gamma}}^T(0), \hat{\sigma}_U^2(0)\}^T$, satisfies

$$\sum_{i=1}^{n} \boldsymbol{\psi}\{\boldsymbol{Q}_i; \widehat{\boldsymbol{\theta}}^T(0), \widehat{\boldsymbol{\gamma}}^T(0), \hat{\sigma}_U^2(0)\} = \mathbf{0}, \tag{5.8}$$

and the full parameter estimator obtained based on $\{\boldsymbol{Q}_{b,i}\}_{i=1}^n$ in Step 2 given above, denoted by $\widehat{\boldsymbol{\Omega}}_b(\lambda) = (\widehat{\boldsymbol{\theta}}_b^T(\lambda), \widehat{\boldsymbol{\gamma}}_b^T(\lambda), \hat{\sigma}_{U,b}^{2*}(\lambda))^T$, satisfies

$$\sum_{i=1}^{n} \boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}; \widehat{\boldsymbol{\theta}}_b^T(\lambda), \widehat{\boldsymbol{\gamma}}_b^T(\lambda), \hat{\sigma}_{U,b}^{2*}(\lambda)\} = \mathbf{0}. \tag{5.9}$$

The full parameter estimator obtained in Step 3 is given by $\widehat{\boldsymbol{\Omega}}_B(\lambda) = \sum_{b=1}^{B} \widehat{\boldsymbol{\Omega}}_b(\lambda)/B$. In the improved remeasurement method, Steps 2 and 3 are replaced by Step 2*, in which the full parameter estimator based on λ-remeasured data is obtained by solving the average of estimating equations in (5.9) for $b = 1, \ldots, B$. Denote the full parameter estimator from Step 2* as $\widehat{\boldsymbol{\Omega}}(\lambda) = \{\widehat{\boldsymbol{\theta}}^T(\lambda), \widehat{\boldsymbol{\gamma}}^T(\lambda), \widehat{\sigma}_U^{2*}(\lambda)\}^T$. Then $\widehat{\boldsymbol{\Omega}}(\lambda)$ satisfies

$$\sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \widehat{\boldsymbol{\theta}}(\lambda), \widehat{\boldsymbol{\gamma}}(\lambda), \widehat{\sigma}_U^{2*}\} = \mathbf{0}, \tag{5.10}$$

where $\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^{2*}\} = \sum_{b=1}^{B} \boldsymbol{\psi}\{\boldsymbol{Q}_{b,i}; \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_U^{2*}\}/B$.

Example 5.1: *Simulation for the simple-response joint model.* The simple-response joint model defined in Section 1.3 is a direct generalization of the structural measurement error model. The implementation of remeasurement method and test of robustness are parallel with the examples presented in Chapter 4. For instance, as in Example 4.3, we consider MLE and CSE in this example, where the MLEs are obtained by directly maximizing the observed-data likelihood with different assumed random-effect models. The only concept that needs some explanation is the multivariate version of the reliability ratio. This multivariate generalization of the reliability

ratio is described in Appendix C.

In this simulation, a data set of size $n = 500$ is generated from a simple-response joint model. Specifically, a binary response Y_i is generated from the logistic model $P(Y_i = 1|\boldsymbol{X}_i) = \{1 + \exp(-\beta_0 - \boldsymbol{\beta}_1\boldsymbol{X}_i)\}^{-1}$, where $\boldsymbol{X}_i = (X_{1i},\, X_{2i})^T$ and $\boldsymbol{\beta}_1 = (\beta_{11},\ \beta_{12})^T$, so that the primary parameter vector $\boldsymbol{\theta} = (\beta_0,\ \boldsymbol{\beta}_1^T)^T$, with the true values in the simulation $(-2,\ 1,\ 1)^T$. The longitudinal measurements $\boldsymbol{W}_i$, with $m_i = 5$ measurements on each subject i taken at times $t_{ij} = j$ $(j = 1, \ldots, 5)$, follow the linear mixed model $\boldsymbol{W}_i = \boldsymbol{D}_i\boldsymbol{X}_i + \boldsymbol{U}_i$, where the design matrix $\boldsymbol{D}_i$ is 5×2 with j^{th} row $(1,\ j)$, and $\boldsymbol{U}_i \sim N_5(\boldsymbol{0},\, 0.6\boldsymbol{I}_5)$. The $\boldsymbol{X}_i$'s are generated from a two-component location mixture bivariate normal (BVN), $(1-p)N_2(\boldsymbol{\delta},\ \boldsymbol{I}_2)+pN_2(\boldsymbol{0},\ \boldsymbol{I}_2)$, with $p = 0.4$ and $\boldsymbol{\delta} = (5,\ 0)^T$. In the remeasurement method, we take $B = 50$, and λ ranges from 0 to 2. This range of λ, along with the true value of $\sigma_U^2 (= 0.6)$, yield remeasured data with multivariate reliability ratio varying from 0.93 to 0.84 (see Appendix C).

We consider the following four estimators for $\boldsymbol{\theta}$. One is the CSE studied by Li et al. (2004) for the same joint model under consideration. Denote the CSEs computed based on the bth λ-remeasured data as $\widehat{\boldsymbol{\theta}}_b^{(c)}(\lambda)$ and the average of $\widehat{\boldsymbol{\theta}}_b^{(c)}(\lambda)$ for $b = 1, \ldots, B$ as $\widehat{\boldsymbol{\theta}}_{_B}^{(c)}(\lambda)$. The other three estimators are MLEs when the assumed models for $\boldsymbol{X}$ are a two-component location mixture BVN (MixBVN), BVN, and a flexible distribution defined by the bivariate second-order, seminonparametric (SNP) density $f_{\boldsymbol{X}}^{(s)}(\boldsymbol{x}; \boldsymbol{\tau}^{(s)}) = P_2^2\{\boldsymbol{G}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\}\phi\{\boldsymbol{G}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\}|\boldsymbol{G}|^{-1}$, where the second-order polynomial $P_2(\boldsymbol{z}) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{20}z_1^2 + a_{11}z_1z_2 + a_{02}z_2^2$ for $\boldsymbol{z} = (z_1, z_2)^T$, and the polynomial coefficients are constrained so that $f_{\boldsymbol{X}}^{(s)}(\boldsymbol{x}; \boldsymbol{\tau}^{(s)})$ integrates to one. Denote the three MLEs computed in Step 2 as $\widehat{\boldsymbol{\theta}}_b^{(m)}(\lambda)$, $\widehat{\boldsymbol{\theta}}_b^{(n)}(\lambda)$, and $\widehat{\boldsymbol{\theta}}_b^{(s)}(\lambda)$, respectively; and

define the final estimates obtained in Step 3 similarly with b replaced by B in the subscript.

Figure 5.1 (a) and (b) present the SIMEX plots of four estimates for the simulated data set. In order to see the typical trend in these four estimators as functions of λ in the long run, the above experiment is repeated on $N = 30$ sets of "observed" data (in contrast to remeasured data) of size $n = 500$ generated independently from the same joint model. This gives 30 sets of estimates, and each set gives SIMEX plots similar to Figure 5.1 (a) and (b). Figure 5.1 (c) and (d) show the average of these 30 MC replications. In Figure 5.1, the range of the vertical axis is set to be two standard deviations of $\widehat{\boldsymbol{\theta}}_B^{(n)}(0)$ below and above the average of $\widehat{\boldsymbol{\theta}}_B^{(c)}(0)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(0)$, $\widehat{\boldsymbol{\theta}}_B^{(n)}(0)$, and $\widehat{\boldsymbol{\theta}}_B^{(s)}(0)$.

The nonrobustness of $\widehat{\boldsymbol{\theta}}_B^{(n)}(\lambda)$ stands out in Figure 5.1. This matches the intuition that when the true model for $\boldsymbol{X}$ is a MixBVN while a relatively restrictive model such as the BVN is assumed for $\boldsymbol{X}$, bias due to measurement error is expected. Thus in this example the longitudinal information with five equally spaced measures per subject is not great enough to ensure that approximating (5.5) with (5.6) is reasonable. The fairly flat trend in $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$ is also expected, as it results from the correct modeling for $\boldsymbol{X}$. It is established elsewhere (Li et al., 2004) that $\widehat{\boldsymbol{\theta}}_B^{(c)}$ should also be consistent and robust. This is reflected in the nearly constant plots of $\widehat{\boldsymbol{\theta}}_B^{(c)}(\lambda)$, which are close to $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$. The nearly constant plots of $\widehat{\boldsymbol{\theta}}_B^{(s)}$ suggest a gain in robustness by using flexible modeling for the distribution of $\boldsymbol{X}$.

Example 5.2: *Simulation for the censored-endpoint joint model.* The censored-

endpoint joint model defined in Section 1.3 is more complex than all the previous models we have considered so far. The complexity comes from the survival model, for which we focus on proportional hazards model. The joint model is semiparametric due to the unspecified baseline hazard function $\lambda_0(u)$, and therefore $\mathbf{\Omega}$ has infinite dimension. It is no longer feasible to compute MLE by directly maximizing the observed-data likelihood as before. In this example we adopt the expectation-maximization (EM) algorithm as in Wulfsohn & Tsiatis (1997), Song et al. (2002), and Hsieh et al. (2006) to obtain the MLEs of $\mathbf{\Omega}$. This algorithm exploits the fact that the MLE of $\lambda_0(u)$ only has mass at each failure time and is equal to zero else where. Because the EM algorithm is very computationally intensive, we use the improved remeasurement method, in which the full parameter estimator solve (5.10).

In this simulation we use the simulation setting in Song et al. (2002) and generate a data set with $n = 500$ subjects. The longitudinal measures $W_{ij} = X_i(t_{ij}) + U_{ij}$, where $U_{ij} \sim N(0, \sigma_U^2)$, are generated at times t_{ij} =0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80 weeks, with a 10% missing rate at time $u \geq 16$. The intra-subject error variance is $\sigma_U^2 = 0.6$, and $X_i(u) = \alpha_{0i} + \alpha_{1i}u$, where the subject-specific random effects $\boldsymbol{\alpha}_i = (\alpha_{0i}, \alpha_{1i})^T$ are generated from a two-component location mixture BVN constructed in the manner described in Davidian and Gallant (1993) with mixing proportion $p = 0.5$, sep = 4, $\mathrm{E}(\boldsymbol{\alpha}_i) = (4.173, -0.0103)^T$, and $\{\mathrm{var}(\alpha_{0i}), \mathrm{cov}(\alpha_{0i}, \alpha_{1i}), \mathrm{var}(\alpha_{0i})\} = (4.96, -0.0456, 0.012)$, where "sep" is a mixture separation measure in Davidian and Gallant (1993). The hazard rate defined in (1.6) is assumed to be $\lambda_i(u) = \lambda_0(u)\exp\{\theta X_i(u)\}$ with $\theta = -1$ and $\lambda_0(u) = 1$ for $u \geq 16$ and zero otherwise.

The same four estimators as those in Example 5.1 are considered in this experiment but the first-order SNP is used instead of the second-order SNP to simplify the computation. In the improved remeasurement method we set $B = 30$ and λ =0, 1. The CSE is computed according to Tsiatis and Davidian (2001).

Figure 5.2 (a) shows the SIMEX plot of four estimates from one simulated data set. Figure 5.2 (b) depicts the plot from $N = 50$ MC replications of this experiment, i.e., (b) can be viewed as the average of 50 plots identically distributed as that in (a). The range of the vertical axis in Figure 5.2 is set to be one standard deviation of $\widehat{\theta}_B^{(n)}(0)$ below and above the average of $\hat{\theta}_B^{(c)}(0)$, $\hat{\theta}_B^{(m)}(0)$, $\hat{\theta}_B^{(n)}(0)$, and $\hat{\theta}_B^{(s)}(0)$.

Even though $\hat{\theta}_B^{(n)}$ appears to be slightly less robust than the other three estimates in plot (a), after averaging the results from 50 MC replications, the four estimators perform similarly in terms of robustness. As CSE is expected to be robust (Tsiatis and Davidian, 2001) as is $\hat{\theta}_B^{(m)}$, plot (b) suggests that misspecifying the random-effect model as BVN does not affect the consistency of MLE, which agrees with the observations in Song et al. (2002) and Hsieh et al. (2006). This is an example where the MLE is insensitive to model assumptions on random effects when the longitudinal information is great enough. Under this simulation setting, there are approximately seven measures per subject on average and the distribution of the time points differs from subject to subject.

5.3 Test of Robustness

The values of $T^*_{1,d}$ (d=1, 2, 3) assessing the robustness of $\widehat{\beta}_0$, $\widehat{\beta}_{11}$, and $\widehat{\beta}_{12}$, respectively, and the associated p-values corresponding to the simulation in Example 5.1 and Figure 5.1 (a) and (b) are given in Table 5.1. It is clear from Table 5.1 that the change in $\widehat{\boldsymbol{\theta}}^{(n)}_B(\lambda)$ as λ increases from 0 to 2 is much more significant than the changes in the other three estimates. For the simulation in Example 5.2, where the primary regression parameter is a scalar θ, with plot given in Figure 5.2 (a), we obtained values of $T^*_{2,1}$ corresponding to $\widehat{\theta}^{(c)}_B$, $\widehat{\theta}^{(m)}_B$, $\widehat{\theta}^{(n)}_B$, and $\widehat{\theta}^{(s)}_B$ as 0.84 (0.40), -0.09 (0.93) -1.68 (0.09), and -0.64 (0.52), respectively, with the associated p-values in the parentheses. These values of $T^*_{2,1}$ seem to agree with the observations from Figure 5.2. Because of the semiparametric nature of the joint model in Example 5.2, we use a simplifying approximation in constructing $T^*_{2,1}$ to test the robustness in the MLEs. This approximation is not needed for a fully parametric model or when the dimension of $\boldsymbol{\Omega}$ is not overly large. But the validity of the approximation calls for further investigation.

In Example 5.2, in order to estimate the parameters of interest in the joint model via EM algorithm, the original infinite dimensional full parameter vector $\boldsymbol{\Omega} = (\theta, \boldsymbol{\tau}^{(a)T}, \sigma^2_U, \lambda_0)^T$ is regarded as finite dimensional with $(2 + t + L)$ parameters, $\{\theta, \boldsymbol{\tau}^{(a)T}, \sigma^2_U, \lambda_0(u_1), \ldots, \lambda_0(u_L)\}^T$, where t is the dimension of $\boldsymbol{\tau}^{(a)}$ and $(u_1, \ldots, u_L)$ are the observed failure times. Define $\boldsymbol{\lambda}^*_0 = \{\lambda_0(u_1), \ldots, \lambda_0(u_L)\}^T$. This view of $\boldsymbol{\Omega}$ converts the problem from that of a semiparametric to parametric with the number of unknown parameters the same order of the sample size n. The resulting high dimensional parameter space makes it infeasible to construct the test statistics defined

in Section 3.2. We use the following approximation to compute $T^*_{2,1}$.

Denote the contribution of subject i to the log likelihood as, assuming a scalar primary regression parameter θ as in Example 5.2,

$$l_i(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau},\sigma_U^2,\boldsymbol{\lambda}_0^*)=\log\{f(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*)\}.$$

where the observed-data density $f(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*)$ is given by (1.7). Suppose one is only interested in θ, define

$$\chi(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*)=\frac{\partial}{\partial\theta}l_i(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*),$$

$$\chi^{(B)}(\boldsymbol{Q}_i^{(B)};\theta,\boldsymbol{\tau}^{(a)},\sigma_U^{2*},\boldsymbol{\lambda}_0^*)=\frac{1}{B}\sum_{b=1}^{B}\chi(\boldsymbol{Q}_{b,i};\theta,\boldsymbol{\tau}^{(a)},\sigma_U^{2*},\boldsymbol{\lambda}_0^*),$$

and the $(t+1+L)\times 1$ vector

$$\boldsymbol{\zeta}(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*)=\frac{\partial}{\partial(\boldsymbol{\tau}^{(a)T},\sigma_U^2,\boldsymbol{\lambda}_0^{*T})^T}l_i(\boldsymbol{Q}_i;\theta,\boldsymbol{\tau}^{(a)},\sigma_U^2,\boldsymbol{\lambda}_0^*).$$

Then the test statistic for testing the robustness of $\hat{\theta}$ is $T^*_{2,1}=T_{2,1}/\widehat{\nu}_{2,1}$, where,

$$T_{2,1}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\chi^{(B)}\{\boldsymbol{Q}_i^{(B)};\hat{\theta}(0),\widehat{\boldsymbol{\tau}}^{(a)}(0),(1+\lambda)\hat{\sigma}_U^2(0),\widehat{\boldsymbol{\lambda}}_0^*(0)\},$$

and

$$\widehat{\nu}_{2,1}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_{2i,1}-\overline{R}_{2,1})^2},$$

with $\overline{R}_{2,1} = n^{-1}\sum_{i=1}^{n} R_{2i,1}$, and $R_{2i,1}$ as the first element of $\boldsymbol{R}_{2i}$ in (3.16) given by

$$
\begin{aligned}
R_{2i,1} \;=\; & \chi^{(B)}\{\boldsymbol{Q}_i^{(B)}; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), (1+\lambda)\hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} - \\
& \Big[\frac{1}{n}\sum_{j=1}^{n} \frac{\partial \chi^{(B)}\{\boldsymbol{Q}_j^{(B)}; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^{2*}, \boldsymbol{\lambda}_0^*\}}{\partial(\theta, \boldsymbol{\tau}^{(a)T}, \sigma_U^{2*}, \boldsymbol{\lambda}_0^{*T})} \Big|_{\hat{\theta}(0), \hat{\boldsymbol{\tau}}^{(a)}(0), (1+\lambda)\hat{\sigma}_U^2(0), \hat{\boldsymbol{\lambda}}_0^*(0)}\Big] \qquad (5.11) \\
& \Big[\frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial(\theta, \boldsymbol{\tau}^{(a)T}, \sigma_U^2, \boldsymbol{\lambda}_0^{*T})^T} \begin{pmatrix} \chi\{\boldsymbol{Q}_j; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^2, \boldsymbol{\lambda}_0^*\} \\ \boldsymbol{\zeta}\{\boldsymbol{Q}_j; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^2, \boldsymbol{\lambda}_0^*\} \end{pmatrix} \Big|_{\hat{\theta}(0), \hat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0), \hat{\boldsymbol{\lambda}}_0^*(0)}\Big]^{-1} \\
& \begin{pmatrix} \chi\{\boldsymbol{Q}_i; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} \\ \boldsymbol{\zeta}\{\boldsymbol{Q}_i; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} \end{pmatrix}. \qquad (5.12)
\end{aligned}
$$

Because L is usually large, it is cumbersome or practically impossible to compute (5.12), we replace $\boldsymbol{\zeta}(\cdot)$ in (5.12) with the first $(t+1)$ subvector consisting of its first $(t+1)$ elements corresponding to $(\boldsymbol{\tau}^{(a)T}, \sigma_U^2)^T$. Denote this subvector as $\boldsymbol{\eta}(\cdot)$. Then we compute $R_{2i,1}$ as

$$
\begin{aligned}
R_{2i,1} \;=\; & \chi^{(B)}\{\boldsymbol{Q}_i^{(B)}; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), (1+\lambda)\hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} - \\
& \Big[\frac{1}{n}\sum_{j=1}^{n} \frac{\partial \chi^{(B)}\{\boldsymbol{Q}_j^{(B)}; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^{2*}, \widehat{\boldsymbol{\lambda}_0}^*(0)\}}{\partial(\theta, \boldsymbol{\tau}^{(a)T}, \sigma_U^{2*})} \Big|_{\hat{\theta}(0), \hat{\boldsymbol{\tau}}^{(a)}(0), (1+\lambda)\hat{\sigma}_U^2(0)}\Big] \\
& \Big[\frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial(\theta, \boldsymbol{\tau}^{(a)T}, \sigma_U^2)^T} \begin{pmatrix} \chi\{\boldsymbol{Q}_j; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^2, \widehat{\boldsymbol{\lambda}}_0^*(0)\} \\ \boldsymbol{\eta}\{\boldsymbol{Q}_j; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^2, \widehat{\boldsymbol{\lambda}}_0^*(0)\} \end{pmatrix} \Big|_{\hat{\theta}(0), \hat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0)}\Big]^{-1} \\
& \begin{pmatrix} \chi\{\boldsymbol{Q}_i; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} \\ \boldsymbol{\eta}\{\boldsymbol{Q}_i; \hat{\theta}(0), \widehat{\boldsymbol{\tau}}^{(a)}(0), \hat{\sigma}_U^2(0), \widehat{\boldsymbol{\lambda}}_0^*(0)\} \end{pmatrix}. \qquad (5.13)
\end{aligned}
$$

If the parameter space of θ is orthogonal to that of $\boldsymbol{\lambda}_0^*$ in the sense that

$$
\frac{\partial^2 l_i(\boldsymbol{Q}_i; \theta, \boldsymbol{\tau}^{(a)}, \sigma_U^2, \boldsymbol{\lambda}_0^*)}{\partial\theta\partial\boldsymbol{\lambda}_0^{*T}} = \mathbf{0}, \qquad (5.14)
$$

then the last L elements in expression (5.11) are zeros, and (5.12) is equal to (5.13). As this orthogonality does not hold in general, the asymptotic difference between

(5.12) and (5.13) and the adjustment to obtain a reasonable variance estimate $\widehat{\nu}_{2,1}$ needs further investigation.

Using the simulation setting in Example 5.2 except that $n = 200$ and only three MLEs are considered, we carried out a simulation with $N = 200$ MC replications to study the operating characteristics of $T^*_{2,1}$ replacing (5.12) with (5.13). The averages of the absolute values of $T^*_{2,1}$ from $N = 200$ replications are 0.85 (0.04), 0.85 (0.05), and 0.76 (0.04) for BVN-, SNP-, and mixture BVN-modeling, respectively, with the standard errors of these averages given in the parentheses. The proportions of $T^*_{2,1}$ that satisfy $|T^*_{2,1}| > 1.96$ are 0.05 (0.01), 0.07 (0.02), and 0.04 (0.01) for BVN-, SNP-, and mixture BVN-modeling, respectively, with the standard errors of these averages given in the parentheses. These results agree with the observation from SIMEX plots and also reconcile the finding in Song et al. (2002), that this is a situation where the MLE is robust to the assumption on the random-effect model.

5.4 Application to SWAN and ACTG 175

5.4.1 Application to SWAN

We now analyze a data set with similar structure as the data generated from Example 5.1. We apply the diagnostic methods to a data set from the SWAN study introduced in Section 1.3. The data set contains information on $n = 632$ subjects. The response is the indicator of absence of osteopenia (bone mineral density above the 33rd percentile). Specifically, for subject i $(= 1, \ldots, 632)$, $Y_i = 1$ indicates absence of osteopenia, and $Y_i = 0$ indicates presence. The observed longitudinal measurements,

$\boldsymbol{W}_i$, are the natural log of progesterone levels from urine (PDG) collected over one menstrual cycle of subject i, with the length of cycle being standardized to a reference of 28 days. For $i = 1, \ldots, 632$, the number of longitudinal measurements, m_i, varies from 6 to 14 among $i = 1, \ldots, 632$ subjects. Adopting the joint model studied in Li et al. (2004) but excluding the observable explanatory variables considered in their paper, we consider the logistic primary model as in Example 5.1, $P(Y_i = 1|\boldsymbol{X}_i) = \{1 + \exp(-\beta_0 - \boldsymbol{\beta}_1 \boldsymbol{X}_i)\}^{-1}$, and the piecewise linear mixed model for the longitudinal measurements, $W_{ij} = X_{1i} + X_{2i}(t_{ij} - 1.4)_+ - 2X_{2j}(t_{ij} - 2.1)_+ + U_{ij}$, for $i = 1, \ldots, 632$ and $j = 1, \ldots, m_i$, where $u_+ = u$ if $u > 0$ and 0 otherwise, time t_{ij} is in units of 10 days, U_{ij} are i.i.d. N(0, σ^2) random errors, X_{1i} is the subject-specific underlying log PDG up to day 14, and X_{2i} is the subject-specific "slope" of the symmetric rise (days 14–21) and fall (days 21–28).

Before carrying out the remeasurement method, to gain some idea of how noisy the observed longitudinal measures are, we estimated the multivariate reliability ratio for this data set based on the definition of reliability ratio in this context given in (C.2), i.e., $RR = \text{trace}\{(\boldsymbol{V}_{\boldsymbol{X}} + \boldsymbol{V}_U)^{-1}\boldsymbol{V}_X\}/2$, as $\boldsymbol{X}$ is a bivariate random variable in this case. We first estimated the intra-subject error variance σ_U^2 by

$$\hat{\sigma}_U^2 = n^{-1} \sum_{i=1}^{n} \Big\{ \sum_{j=1}^{m_i} (W_{ij} - \widehat{W}_{ij})^2/(m_i - 2) \Big\},$$

where $\widehat{W}_{ij}$ is the OLS estimate for W_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. This gives an estimate of $\boldsymbol{V}_U$ as $\widehat{\boldsymbol{V}}_U = n^{-1} \sum_{i=1}^{n} \hat{\sigma}_U^2 (\boldsymbol{D}_i^T \boldsymbol{D}_i)^{-1}$ for $i = 1, \ldots, n$. Next $\boldsymbol{V}_X$ in (C.2) is estimated by $\widehat{\boldsymbol{V}}_X = \sum_{i=1}^{n} (\widehat{\boldsymbol{X}}_i - \overline{\boldsymbol{X}})(\widehat{\boldsymbol{X}}_i - \overline{\boldsymbol{X}})^T/(n-1) - \widehat{\boldsymbol{V}}_U$, where $\widehat{\boldsymbol{X}}_i$ is the OLS estimate for $\boldsymbol{X}_i$, and $\overline{\boldsymbol{X}}$ is the average of $\widehat{\boldsymbol{X}}_i$ for $i = 1, \ldots, n$. Finally the reliability ratio for this data set is estimated by $\text{trace}\{(\widehat{\boldsymbol{V}}_{\boldsymbol{X}} + \widehat{\boldsymbol{V}}_U)^{-1}\widehat{\boldsymbol{V}}_X\}/2$. We found

that the estimated multivariate reliability ratio of the observed longitudinal measures is 0.93. For the remeasurement method we computed $\widehat{\boldsymbol{\theta}}_B^{(c)}(\lambda)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$ and $\widehat{\boldsymbol{\theta}}_B^{(n)}(\lambda)$, with $B = 50$ and λ ranging from 0 to 2, corresponding to the estimated multivariate reliability ratio ranging from 0.93 to 0.78.

Figure 5.3 shows the SIMEX plots for β_0 and β_{12}. Table 5.2 gives the values of $T^*_{1,d}$ (d=1, 2, 3) for different types of estimators. In Figure 5.3, the range of the vertical axis is set to be one standard deviation of $\widehat{\boldsymbol{\theta}}_B^{(c)}(0)$ below and above the average of $\widehat{\boldsymbol{\theta}}_B^{(c)}(0)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(0)$, and $\widehat{\boldsymbol{\theta}}_B^{(n)}(0)$. Even though Figure 5.3 seems to indicate slight nonrobustness of $\widehat{\boldsymbol{\theta}}_B^{(n)}(\lambda)$ compared to $\widehat{\boldsymbol{\theta}}_B^{(c)}(\lambda)$ and $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$, the values of $T^*_{1,d}$ (d=1, 2, 3) in Table 5.2 suggest that none of $\widehat{\boldsymbol{\theta}}_B^{(c)}(2)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(2)$, and $\widehat{\boldsymbol{\theta}}_B^{(n)}(2)$ changes significantly from their counterpart values at $\lambda = 0$. The observed robustness can be explained by the finding in Li et al. (2004) that the estimated density for $\boldsymbol{X}$ "does not deviate considerably from multivariate normality."

5.4.2 Application to ACTG 175

We now analyze the data from the ACTG 175 described in Section 1.3, which has a similar structure as the data generated for Example 5.2. For this application we model the true $\log_{10}$ CD4 count by $X_i(u) = \alpha_{i0} + \alpha_{i1}u$ as in Example 5.2 and Song et al. (2002) to describe the longitudinal trajectory after week 12. Different from the study in Song et al. (2002), we assume a proportional hazards model for the event time T, a composite of $\geq 50\%$ decline in CD4, progression to AIDS, or death, with $X_i(u)$ as the only covariate, and hazard rate as $\lambda_0(u) = \lambda_0(u)\exp\{\theta X_i(u)\}$, to simplify the problem somewhat. The data set includes information on 2279 patients,

among which there were 350 events. There were approximately 8 CD4 measures per subject on average.

Due to the large sample size, it becomes fairly time-consuming to implement the traditional remeasurement method. Therefore we did not make the SIMEX plot for this example and only compute $T^*_{2,1}$ corresponding to the MLE for θ with presumed random-effect model as BVN, first-order SNP, and mixture BVN. With $B = 30$, we found $T^*_{2,1}$ to be 13.84, 13.82, and 19.21, respectively, for these three MLEs. As for the parameter estimates based on the raw data, the estimated first two moments of the random effects are similar to those obtained by Song et al. (2002), but the estimated θ is not comparable with the estimates presented in Song et al. (2002). For instance, in Table 3 of Song et al. (2002), $\hat{\theta} = -2.487$ with estimated standard error 0.091 when the presumed random-effect model is BVN, while we got $\hat{\theta}^{(n)} = -3.405$ with estimated standard error 0.066. However because we assume the proportional hazards model without a second treatment indicator convariate but only $X_i(u)$ while Song et al. (2002) include this covariate in their model, it may not be meaningful to compare to their estimates.

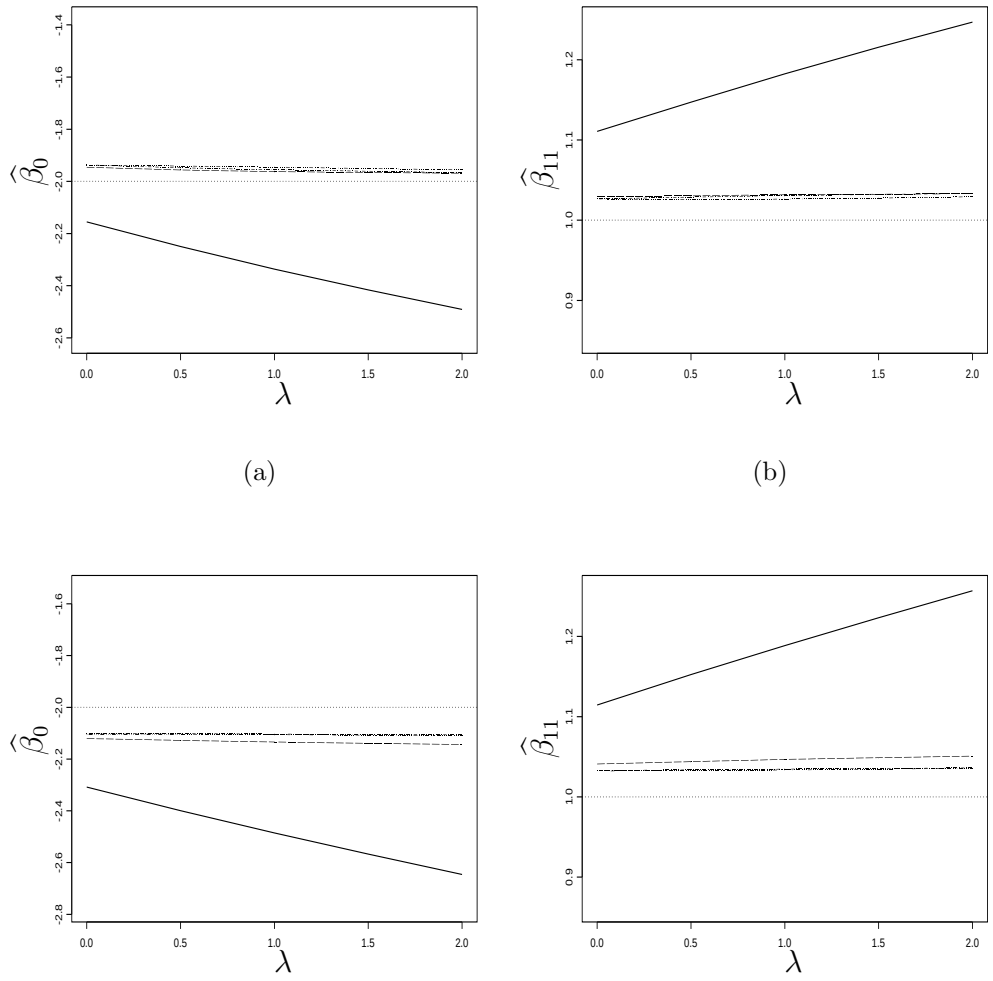
The values of $T^*_{2,1}$ for these data suggest that none of three presumed models lead to a robust MLE even though Song et al. (2002) indicate that this is an example where MLE is robust to model assumptions on random effect. The discrepancy in our finding and that of Song et al. may be due to the use of a different assumed proportional hazards model. It may also be a result of overoptimistic $\widehat{\nu}_{2,1}$ bearing in mind the concern of approximating (5.12) with (5.13) brought up in Section 5.3 when the joint model is actually semiparametric.

Table 5.1: Values of $T^*_{1,d}$ (d=1, 2, 3) used to assess the changes in $\widehat{\boldsymbol{\theta}}^{(c)}_{B}(\lambda)$, $\widehat{\boldsymbol{\theta}}^{(m)}_{B}(\lambda)$, $\widehat{\boldsymbol{\theta}}^{(n)}_{B}(\lambda)$, and $\widehat{\boldsymbol{\theta}}^{(s)}_{B}(\lambda)$, as λ increases from 0 to 2 corresponding to the simulation in Example 5.1 and Figure 5.1 (a). Corresponding p-values are in the parentheses.

	CSE	Mixture BVN-MLE	BVN-MLE	SNP-MLE
$\widehat{\beta}_{0,B}$	0.47 (0.64)	0.74 (0.46)	3.60 (0.00)	1.52 (0.13)
$\widehat{\beta}_{11,B}$	−0.58 (0.56)	−0.66 (0.51)	−3.61 (0.00)	−1.43 (0.15)
$\widehat{\beta}_{12,B}$	−0.33 (0.74)	−0.52 (0.63)	−1.99 (0.05)	−0.04 (0.97)

Table 5.2: Values of $T^*_{1,d}$ (d=1, 2, 3) used to assess changes in $\widehat{\boldsymbol{\theta}}^{(c)}_{B}(\lambda)$, $\widehat{\boldsymbol{\theta}}^{(m)}_{B}(\lambda)$, and $\widehat{\boldsymbol{\theta}}^{(n)}_{B}(\lambda)$ as λ increases from 0 to 2 for the SWAN data. Corresponding p-values are in the parentheses.

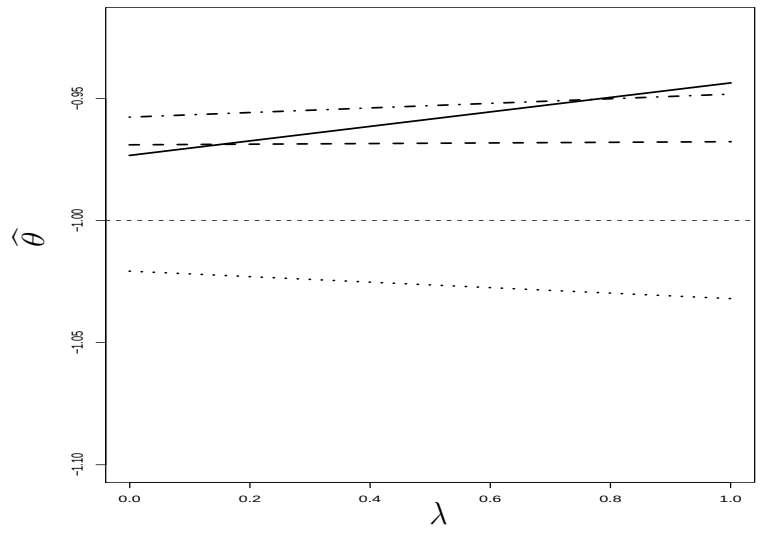
	CSE	Mixture BVN-MLE	BVN-MLE
$\widehat{\beta}_{0,B}$	0.47 (0.64)	−0.46 (0.65)	−1.70 (0.09)
$\widehat{\beta}_{11,B}$	0.99 (0.32)	−0.01 (0.99)	−0.68 (0.50)
$\widehat{\beta}_{12,B}$	−0.48 (0.63)	0.43 (0.67)	1.70 (0.09)



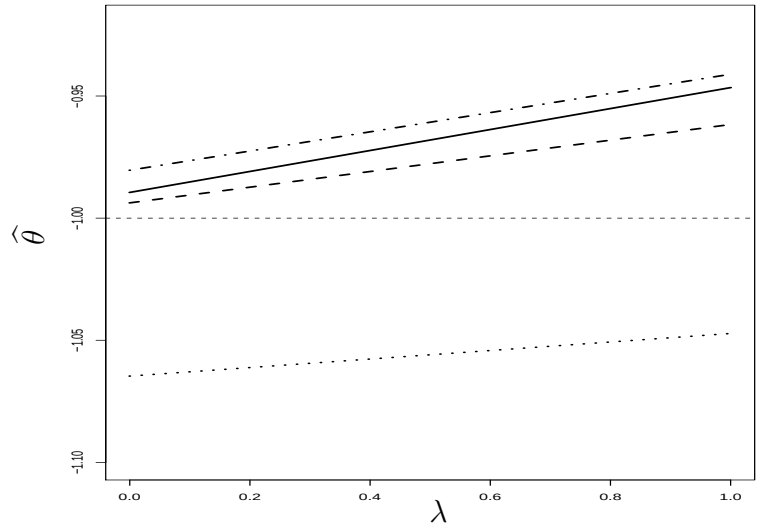
(a) (b)

(c) (d)

Figure 5.1: Plots (a) and (b) show MLE's assuming mixture BVN, BVN, and SNP random effects, and CSE's as function of λ obtained at Step 3 of remeasurement method with $B = 50$ to one "observed" data set. Plots (c) and (d) show averages of $N = 30$ sets of estimates plotted in (a) and (b) for $N = 30$ MC replications. Only the plots of the first two regression parameters, β_0 and β_{11}, are shown. The line types for $\widehat{\boldsymbol{\theta}}_B^{(c)}(\lambda)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$, $\widehat{\boldsymbol{\theta}}_B^{(n)}(\lambda)$, and $\widehat{\boldsymbol{\theta}}_B^{(s)}(\lambda)$ are dash-multiple-dotted line, dash-dotted line, solid line, and dashed line, respectively. Horizontal lines are reference lines at the true values, $\beta_0 = -2$ and $\beta_{11} = 1$.

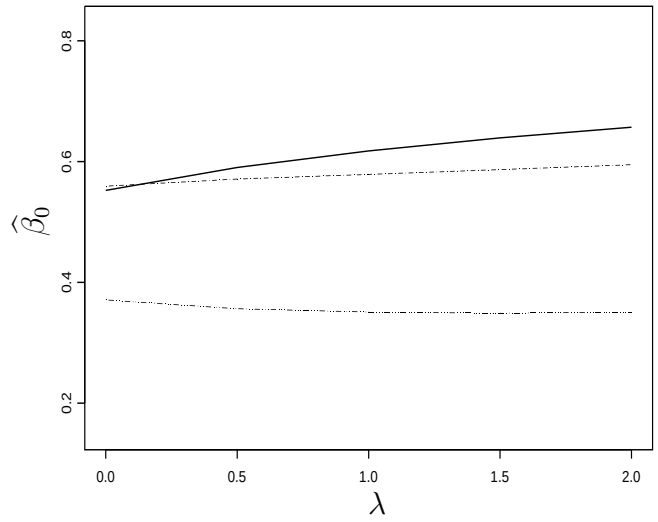


(a)

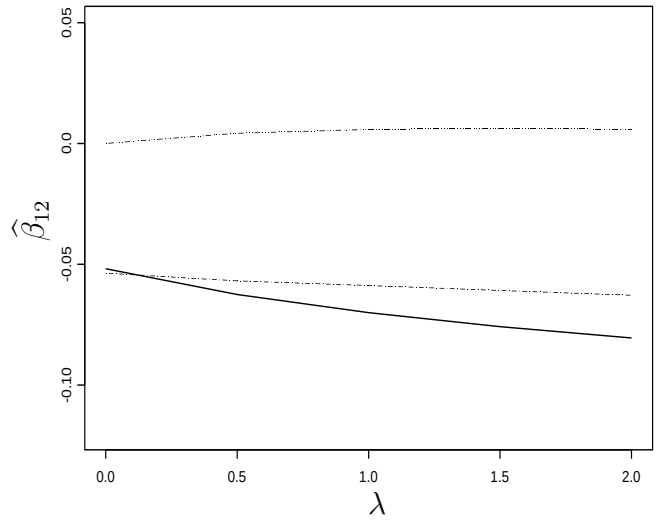


(b)

Figure 5.2: Plot (a) depicts CSEs and MLEs obtained at Step 2* of remeasurement method with $B = 30$ when $\boldsymbol{\alpha}_i$ is modeled as MixBVN, BVN, and SNP, for θ at λ =0, 1. Plot (b) shows the average of $N = 50$ sets of estimates ploted in (a) resulting from $N = 50$ MC replications. The dotted line is for CSE, the dashed line, solid line, and dash-dotted line are for MLE assuming mixture BVN, BVN, and first-order SNP random effects, respectively. The horizontal line is the reference line at the true θ value $\theta = -1$.



(a)



(b)

Figure 5.3: Plots of MLE's corresponding to mixture BVN and BVN modeling, $\widehat{\boldsymbol{\theta}}_B^{(m)}$, $\widehat{\boldsymbol{\theta}}_B^{(n)}$, and CSE, $\widehat{\boldsymbol{\theta}}_B^{(c)}$, obtained at Step 3 with $B = 50$ as functions of λ, using SWAN data. Line types for $\widehat{\boldsymbol{\theta}}_B^{(c)}(\lambda)$, $\widehat{\boldsymbol{\theta}}_B^{(m)}(\lambda)$, and $\widehat{\boldsymbol{\theta}}_B^{(n)}(\lambda)$ are the same as those in Figure 5.1.

Chapter 6

Discussion

We proposed methods for diagnosing estimator robustness to distributional specification of latent variable models in the structural measurement error models and joint models. We defined analytic robustness conditions in Chapter 2. In Chapters 4 and 5 we demonstrated the implementation and performance of remeasurement method and its improved version along with several test statistics. The models we studied cover a wide range of latent-variable models that are suitable in many applications. In principle, these methods can be applied to any latent variable models with some sort of error model whose effects on the observed data can be simulated so that remeasured data can be generated. Therefore, these methods provide data analyst with a systematic approach useful for practical use.

The traditional remeasurement method is usually time-consuming, and the improved version saves a considerable amount of computation time, as does the test statistic $\boldsymbol{T}_2^*$ in (3.12) compared to the other proposed test statistics in (3.11), (3.13), and (3.14) given in Section 3.2. However if the latent variable model involves non-

or semiparametric component, or when the dimension of the entire parameter space $\boldsymbol{\Omega}$ is large, as in the censored-endpoint joint models based on a proportional hazards model, it is problematic to obtain a variance estimator that leads to a valid $\boldsymbol{T}_k^*$ (k=1, 2, 3, 4), namely the variance estimator for $\boldsymbol{T}_k$, because it becomes formidable to compute (3.15)–(3.18) directly. We took an "ad hoc" approach of dropping the nuisance infinite dimensional part out of $\boldsymbol{\Omega}$, computing the variance estimate as if this part of $\boldsymbol{\Omega}$ were fixed constants with values equal to their estimates. This approach is exact only when the estimates for the ignored infinite dimensional part are orthogonal to the estimates for the rest of $\boldsymbol{\Omega}$ in the sense of (5.14) described in Section 5.3. When the orthogonality does not hold, it may lead to an overoptimistic variance estimator, and thus adjustment is needed to obtain valid test statistics. One may use bootstrap to obtain a variance estimate for $\boldsymbol{T}_k$. Yet the combination of two computationally expensive methods, remeasurement method plus bootstrap, makes this approach unattractive in practice. Further study is needed to find a more efficient and reliable solution to this issue.

One more involved question closely related to our current study is the diagnosis of model misspecification on random effects in generalized linear mixed effect models where there is no such error model as $\boldsymbol{W} = \boldsymbol{D}\boldsymbol{X} + \boldsymbol{U}$, that is, the observed data are only realizationa of Y with no $\boldsymbol{W}$ as in the models we have studied. The methods we have proposed all rely on manipulating some sort of error model. Without an error model, new approaches are required to reveal model misspecification in generalized linear mixed effect models. This is an area for future research.

Bibliography

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

Cook, J. and Stefanski, L. A. (1994). Simulation extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89,** 1314–1328.

Davidian, M. and Gallant, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80,** 475–488.

Gallant, A. R. and Nychka, D. W, (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica* **55,** 363–390.

Gleser, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *Journal of the American Statistical Association* **87,** 696–707.

Hammer, S. M., Katezstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch,

M. S., and Merigan, T. C., for the AIDS Clinical Group Study 175 Study Team. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* **335,** 1081–1089.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88,** 973–985.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **4,** 465–480.

Hsieh, F., Tseng, Y. K., and Wang, J. L. (2006) Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* in press.

Li, E., Zhang, D., and Davidian, M. (2004) Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal parameters. *Biometrics* **60,** 1–7.

Song, X., Davidian, M., and Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58,** 742–753.

Sowers, M. R., Finkelstein, J. , Ettinger, B., Bondarenko, I. Neer, R., Cauley, J., Sherman, S., and Greendale, G. (2003). The association of endogenous hormone concentrations and bone mineral density measures in pre- and peri-menopausal women of four ethnic groups: SWAN. *Osteoporosis International* **14,** 44–52.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74,** 703–716.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88,** 447–458.

Tsiatis, A. A. and Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* **14,** 793–818.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis* **23,** 541–556.

Wang, C. Y., Wang, N., and Wang, S. (2000). Regression analysis when covariates are regression parameters of a random effects models for observed longitudinal measurements. *Biometrics* **56,** 487–495.

Zhang, D. and Davidian, M. (2001). Linear mixed model with flexible distribution of random effects for longitudinal data. *Biometrics* **57,** 795–802.

Appendix

Appendix A

Joint densities of (Y, W) in Example 2.2

The closed-form expression of $f^{(n)}_{Y,W}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(n)}, \sigma_U)$ is

$$f^{(n)}_{Y,W}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(n)}, \sigma_U) = c(w, \mu_x, \sigma_x, \sigma_U)[\Phi\{h(w, \boldsymbol{\theta}, \sigma_U)\}]^y[1 - \Phi\{h(w, \boldsymbol{\theta}, \sigma_U)\}]^{1-y},$$

for y=0, 1, and $-\infty < w < +\infty$, where $\boldsymbol{\tau}^{(n)} = (\mu_x,\ \sigma_x)^T$,

$$\begin{aligned}
c(w, \mu_x, \sigma_x, \sigma_U) &= \{2\pi(\sigma_U^2 + \sigma_x^2)\}^{-1/2} \exp\Big\{ - \frac{(w - \mu_x)^2}{2(\sigma_U^2 + \sigma_x^2)}\Big\}, && \text{(A.1)}\\
h(w, \boldsymbol{\theta}, \boldsymbol{\tau}^{(n)}, \sigma_U) &= \frac{\beta_1 + \beta_x B(w, \mu_x, \sigma_x, \sigma_U)}{\{1 + A(\beta_x, \sigma_x, \sigma_U)\}^{1/2}}, && \text{(A.2)}\\
A(\beta_x, \sigma_x, \sigma_U) &= \frac{\beta_x^2 \sigma_U^2 \sigma_x^2}{\sigma_U^2 + \sigma_x^2}, && \text{(A.3)}\\
B(w, \mu_x, \sigma_x, \sigma_U) &= \frac{w\sigma_x^2 + \mu_x \sigma_U^2}{\sigma_U^2 + \sigma_x^2}. && \text{(A.4)}
\end{aligned}$$

The closed-form expression of $f^{(m)}_{Y,W}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(m)}, \sigma_U)$ is

$$\begin{aligned}
f^{(m)}_{\mathrm{Y,W}}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(m)}, \sigma_U) &= \alpha c(w, \mu_1, \sigma_1, \sigma_U) p_1^y (1 - p_1)^{1-y} + \\
&\quad (1 - \alpha) c(w, \mu_2, \sigma_2, \sigma_U) p_2^y (1 - p_2)^{1-y}, && \text{(A.5)}
\end{aligned}$$

for y =0, 1, $-\infty < w < +\infty$, where $p_1 = \Phi\{h_1(w, \boldsymbol{\gamma}, \sigma_U)\}$, $p_2 = \Phi\{h_2(w, \boldsymbol{\gamma}, \sigma_U)\}$, $\boldsymbol{\tau}^{(m)} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha)^T$; $c(.)$ is defined from (A.1) by replacing μ_x and σ_x with either μ_1 and σ_1 or μ_2 and σ_2; $h_1(.)$ is defined from (A.2)–(A.4) by replacing μ_x and σ_x with μ_1 and σ_1; $h_2(.)$ is defined from (A.2)–(A.4) by replacing μ_x and σ_x with μ_2 and σ_2.

Finally, $f^{(s)}_{\mathrm{Y,W}}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(s)}, \sigma_U)$ is

$$f^{(s)}_{\mathrm{Y,W}}(y, w; \boldsymbol{\theta}, \boldsymbol{\tau}^{(s)}, \sigma_U) = \int_{-\infty}^{\infty} \{\Phi(\beta_1 + \beta_x x)\}^y \{1 - \Phi(\beta_1 + \beta_x x)\}^{1-y} \times \frac{1}{\sigma_U}\phi\Big(\frac{w - x}{\sigma_U}\Big) \times \frac{1}{\eta}\phi\Big(\frac{x - \xi}{\eta}\Big)\Big\{a_0 + a_1\Big(\frac{x - \xi}{\eta}\Big) + a_2\Big(\frac{x - \xi}{\eta}\Big)^2\Big\}^2 dx,$$

for y =0, 1, $-\infty < w < +\infty$, where $\boldsymbol{\tau}^{(s)} = (\xi, \eta, a_0, a_1, a_2)^T$.

Appendix B

Estimation of var(T_k) in Section 3.2

Recall that $\boldsymbol{Q}_i$ is the observed data for subject i with measurement error variance σ_U^2, and $\boldsymbol{Q}_i^{(B)}$ is the B λ-remeasured data for subject i, where the measurement error variance is $(1+\lambda)\sigma_U^2$. Suppose that $E\Big[\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}\Big] = \mathbf{0}$ uniquely determines $\boldsymbol{\Omega}(0) = \{\boldsymbol{\theta}(0)^T, \boldsymbol{\tau}^{(a)T}(0), \sigma_U^2(0)\}^T$, and $E\Big[\boldsymbol{\psi}(\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big] = \mathbf{0}$ uniquely determines $\boldsymbol{\Omega}(\lambda) = \{\boldsymbol{\theta}^T(\lambda), \boldsymbol{\tau}^{(a)T}(\lambda), \sigma_U^{2*}(\lambda)\}^T$, where the expectations are with respect to the true density of $\boldsymbol{Q}_i$ and $\boldsymbol{Q}_i^{(B)}$, respectively. Denote estimators for $\boldsymbol{\Omega}(0)$ and $\boldsymbol{\Omega}(\lambda)$ by $\widehat{\boldsymbol{\Omega}}(0)$ and $\widehat{\boldsymbol{\Omega}}(\lambda)$, obtained by solving (3.3) and (3.5), respectively.

The test statistics $\boldsymbol{T}_k^*$ (k=1, 2, 3, 4) are for testing hypothesis $H_0 : \boldsymbol{\Omega}(0) = \boldsymbol{\Omega}(\lambda)$ versus $H_a : \boldsymbol{\Omega}(\lambda) \neq \boldsymbol{\Omega}(0)$. The variance-covariance estimators $\widehat{\boldsymbol{V}}_k^2$ in Section 3.2 are derived next.

We first derive a variance-covariance estimator for $\boldsymbol{T}_1$. By the theory for M-estimating equation, one has the following approximation using the influence functions (Casella

and Berger, 2002, p.517),

$$\sqrt{n}\{\widehat{\boldsymbol{\Omega}}(0)-\boldsymbol{\Omega}(0)\}\approx\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\},\tag{B.1}$$

$$\sqrt{n}\{\widehat{\boldsymbol{\Omega}}(\lambda)-\boldsymbol{\Omega}(\lambda)\}\approx\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{A}_2^{-1}\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}.\tag{B.2}$$

Subtracting (B.2) from (B.1) yields

$$\begin{aligned}\boldsymbol{T}_1 &\approx \sqrt{n}\{\boldsymbol{\Omega}(0)-\boldsymbol{\Omega}(\lambda)\}+\\ &\quad\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}-\boldsymbol{A}_2^{-1}\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big]\\ &\triangleq \sqrt{n}\{\boldsymbol{\Omega}(0)-\boldsymbol{\Omega}(\lambda)\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{R}_{1i}^{*}.\end{aligned}\tag{B.3}$$

Because the summands in (B.3) $\boldsymbol{R}_{1i}^{*}$ for $i=1,\ldots,n$, are independent with mean zero, by Central Limit Theorem (CLT), $\boldsymbol{T}_1$ has asymptotic distribution as normal with variance-covariance matrix $\boldsymbol{V}_1$ =var($\boldsymbol{R}_{1i}^{*}$). Define

$$\boldsymbol{B}_1\{\boldsymbol{\Omega}(0)\}=\mathrm{var}\Big[\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}\Big],$$

$$\boldsymbol{B}_2\{\boldsymbol{\Omega}(\lambda)\}=\mathrm{var}\Big[\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big],$$

$$\boldsymbol{C}\{\boldsymbol{\Omega}(0),\boldsymbol{\Omega}(\lambda)\}=\mathrm{cov}\Big[\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\},\ \boldsymbol{A}_2^{-1}\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big],$$

then the asymptotic variance-covariance matrix of $\boldsymbol{T}_1$ is given by, dropping the arguments in $\boldsymbol{A}_1(\cdot)$, $\boldsymbol{A}_2(\cdot)$, $\boldsymbol{B}_1(\cdot)$, $\boldsymbol{B}_2(\cdot)$, $\boldsymbol{C}(\cdot)$ when it causes no confusion given the context,

$$\begin{aligned}\boldsymbol{V}_1 &= \mathrm{var}\Big[\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}-\boldsymbol{A}_2^{-1}\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big]\\ &= \boldsymbol{A}_1^{-1}\boldsymbol{B}_1(\boldsymbol{A}_1^{-1})^T+\boldsymbol{A}_2^{-1}\boldsymbol{B}_2(\boldsymbol{A}_2^{-1})^T-2\boldsymbol{C}.\end{aligned}$$

Substituting $\widehat{\boldsymbol{\Omega}}(0)$, $\widehat{\boldsymbol{\Omega}}(\lambda)$, and $\widehat{\boldsymbol{A}}_1(\cdot)$, $\widehat{\boldsymbol{A}}_2(\cdot)$ gives $\boldsymbol{R}_{1i}$ in (3.15) as an estimator for $\boldsymbol{V}_1 = \boldsymbol{R}^*_{1i}$, so that an estimator for $\text{var}(\boldsymbol{R}^*_{1i})$, $\widehat{\boldsymbol{V}}_1$, follows.

Moreover if H_a is in the form,

$$H_a : \boldsymbol{\Omega}(\lambda) = \boldsymbol{\Omega}(0) + \boldsymbol{\Delta}/\sqrt{n}, \tag{B.4}$$

the asymptotic noncentrality parameter (NCP) of $\boldsymbol{T}_1$ is equal to

$$\text{NCP}_1 = \boldsymbol{\Delta}^T \boldsymbol{V}_1^{-1} \boldsymbol{\Delta} = \boldsymbol{\Delta}^T (\boldsymbol{A}_1^{-1} \boldsymbol{B}_1 \boldsymbol{A}_1^{-1T} + \boldsymbol{A}_2^{-1} \boldsymbol{B}_2 \boldsymbol{A}_2^{-1T} - 2\boldsymbol{C})^{-1} \boldsymbol{\Delta}. \tag{B.5}$$

Next we derive a variance estimator for $\boldsymbol{T}_2$. Under H_0, applying a first-order Taylor expansion of $\boldsymbol{T}_2$ around $\boldsymbol{\Omega}(0)$ leads to, dropping the remainder terms of order $o_p(1)$ for MLE,

$$\begin{aligned}
\boldsymbol{T}_2 &\approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\Omega}(0)\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\boldsymbol{\psi}}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\Omega}(0)\}\{\widehat{\boldsymbol{\Omega}}(0) - \boldsymbol{\Omega}(0)\} \\
&\approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\Omega}(0)\} - \boldsymbol{A}_2\{\boldsymbol{\Omega}(0)\} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\} \boldsymbol{\psi}\{\boldsymbol{Q}_i; \boldsymbol{\Omega}(0)\}, \qquad \text{(B.6)} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)}; \boldsymbol{\Omega}(0)\} - \boldsymbol{A}_2\{\boldsymbol{\Omega}(0)\} \boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\} \boldsymbol{\psi}\{\boldsymbol{Q}_i; \boldsymbol{\Omega}(0)\} \right] \\
&\triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{R}^*_{2i}. \qquad \text{(B.7)}
\end{aligned}$$

where (B.6) follows from (B.1) and that $\widehat{\boldsymbol{A}}_2\{\boldsymbol{Q}^{(B)}; \widehat{\boldsymbol{\Omega}}(0)\}$ converges to $\boldsymbol{A}_2\{\boldsymbol{\Omega}(0)\}$ in probability. Because $\boldsymbol{R}^*_{2i}$, for $i = 1, \ldots, n$, are mean-zero independent random quantities, by CLT, under H_0, $\boldsymbol{T}_2$ has asymptotic distribution as normal with mean zero

and some variance-covariance matrix denoted by $\boldsymbol{V}_2$ =var($\boldsymbol{R}^*_{2i}$), that is,

$$\begin{aligned}
\boldsymbol{V}_2 &= \text{var}\Big[\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\} - \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}\Big] \\
&= \boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{B}_1(\boldsymbol{A}_1^{-1})^T\boldsymbol{A}_2^T - 2\text{cov}(\boldsymbol{\psi}^{(B)},\, \boldsymbol{\psi})(\boldsymbol{A}_1^{-1})^T\boldsymbol{A}_2^T \\
&= \boldsymbol{A}_2[\boldsymbol{A}_1^{-1}\boldsymbol{B}_1(\boldsymbol{A}_1^{-1})^T + \boldsymbol{A}_2^{-1}\boldsymbol{B}_2(\boldsymbol{A}_2^{-1})^T - 2\text{cov}\{(\boldsymbol{A}_2^{-1})^T\boldsymbol{\psi}^{(B)},\, \boldsymbol{A}_1^{-1}\boldsymbol{\psi})]\boldsymbol{A}_2^T \\
&= \boldsymbol{A}_2\boldsymbol{V}_1\boldsymbol{A}_2^T. \qquad \text{(B.8)}
\end{aligned}$$

$\boldsymbol{R}_{2i}$ given in (3.16) is obtained by substituting the parameter estimates in $\boldsymbol{R}^*_{2i}$, and $\widehat{\boldsymbol{V}}_2$, the sample variance-covariance of $\boldsymbol{R}_{2i}$, is an estimator for $\boldsymbol{V}_2 = \text{var}(\boldsymbol{R}^*_{2i})$.

Under H_a given by (B.4), applying the first-order Taylor expansion of $\boldsymbol{T}_2$ around $\boldsymbol{\Omega}(\lambda)$ leads to,

$$\begin{aligned}
\boldsymbol{T}_2 &\approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\boldsymbol{\psi}}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\{\widehat{\boldsymbol{\Omega}}(0) - \boldsymbol{\Omega}(0) - \boldsymbol{\Delta}/\sqrt{n}\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\} + \frac{1}{n}\sum_{i=1}^{n}\dot{\boldsymbol{\psi}}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\}\Big[\sqrt{n}\{\widehat{\boldsymbol{\Omega}}(0) - \boldsymbol{\Omega}(0)\} - \boldsymbol{\Delta}\Big] \\
&\approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\} - \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\} \\
&\quad + \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\Delta} \\
&= \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\Delta} + \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\boldsymbol{\psi}^{(B)}\{\boldsymbol{Q}_i^{(B)};\boldsymbol{\Omega}(\lambda)\} - \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{A}_1^{-1}\{\boldsymbol{\Omega}(0)\}\boldsymbol{\psi}\{\boldsymbol{Q}_i;\boldsymbol{\Omega}(0)\}\Big] \\
&= \boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\boldsymbol{\Delta} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{R}^*_{2i} \qquad \text{(B.9)}
\end{aligned}$$

Therefore the mean of the asymptotic distribution of $\boldsymbol{T}_2$ is $\boldsymbol{A}_2\{\boldsymbol{\Omega}(\lambda)\}\Delta$. It follows that the NCP of $\boldsymbol{T}_2$ is equal to

$$\text{NCP}_2 = (\boldsymbol{A}_2\boldsymbol{\Delta})^T(\boldsymbol{A}_2\boldsymbol{V}_1\boldsymbol{A}_2^T)^{-1}(\boldsymbol{A}_2\boldsymbol{\Delta}) = \text{NCP}_1. \qquad \text{(B.10)}$$

Following the same line of reasoning, the variance-covariance estimator for $\boldsymbol{T}_3$, $\widehat{\boldsymbol{V}}_3$, can be derived and its asymptotic distribution with mean zero under H_0 and variance equal to $\boldsymbol{V}_3 = \boldsymbol{A}_1\boldsymbol{V}_1\boldsymbol{A}_1^T$. Under H_a, and specifically (B.4), the asymptotic mean of $\boldsymbol{T}_3$ becomes $-\boldsymbol{A}_1\{\boldsymbol{\Omega}(0)\}\boldsymbol{\Delta}$, leading to NCP same as NCP_1. With the derivations to reach the results for $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$, the results for $\boldsymbol{T}_4$ follows immediately. It should be obvious at this point how $\boldsymbol{R}_{4i}$ is derived. The asymptotic variance of $\boldsymbol{T}_4$ is $\boldsymbol{V}_4 = (\boldsymbol{A}_1 + \boldsymbol{A}_2)\boldsymbol{V}_1(\boldsymbol{A}_1 + \boldsymbol{A}_2)^T/4$. Moreover, under H_a in (B.4), $\boldsymbol{T}_4$ has asymptotic distribution as normal with mean $(\boldsymbol{A}_1 + \boldsymbol{A}_2)\boldsymbol{\Delta}/2$ and NCP of $\boldsymbol{T}_4$ is the same as NCP_1.

Having the same NCP, $\boldsymbol{T}_k$ (k=1, 2, 3, 4) are equivalent in terms of hypothesis testing when H_a is in the form of (B.4).

Appendix C

Definition of Reliability Ratio

An important concept often used in the measurement-error field is the reliability ratio (Carroll et al. 1995, p. 22). For a simple scalar measurement error model, $W = X + U$, where W is a measurement of the unobservable variable X contaminated by an additive measurement error U, which is a mean-zero normal random variable with variance σ_U^2, denoting the variance of X as σ_X^2, the reliability ratio is defined as

$$RR = \sigma_X^2/(\sigma_X^2 + \sigma_U^2). \tag{C.1}$$

The reliabilty ratio is a useful general measure of the effects of measurement error. When $RR \approx 1$, it means that σ_U^2/σ_X^2 is small and thus bias due to measurement error will be small. In fact, in simple linear regression, the relative bias in the estimator for slope due to measurement error is just RR.

When it comes to the simple-response joint model defined in Section 1.3, viewing $\boldsymbol{U}$ as the measurement-error vector when measuring $\boldsymbol{DX}$, denoting the covariance matrix of $\boldsymbol{X}_{p\times 1}$ as $\boldsymbol{V_X}$, an analogy to (C.1) is the reliability matrix $(\boldsymbol{V_X} + \boldsymbol{V}_U)^{-1}\boldsymbol{V}_X$,

where $\boldsymbol{V}_U$ is the covariance matrix of the "best" estimator, i.e., OLSE, for $\boldsymbol{X}$ given in Section 5.1 just as σ_U^2 is the variance of the "best" estimator for X, namely W, in the simple scalar measurement error model. Gleser (1992) has shown the importance of the reliability matrix $(\boldsymbol{V}_{\boldsymbol{X}} + \boldsymbol{V}_U)^{-1}\boldsymbol{V}_X$ in linear models. We define a scalar function of the reliability matrix as

$$RR = \frac{1}{p}\text{trace}\{(\boldsymbol{V}_{\boldsymbol{X}} + \boldsymbol{V}_U)^{-1}\boldsymbol{V}_X\}, \tag{C.2}$$

and use it as a general measure of the amount of measurement error in the longitudinal data. Note that in the absence of intra-subject errors, that is, when $\sigma_U = 0$, $\boldsymbol{V}_U = \boldsymbol{0}$, and RR defined in (C.2) is equal to one. It is obvious that $0 < RR < 1$ in the presence of intra-subject errors. Therefore, the reliability ratio defined in (C.2) has the same range and implication as the reliability ratio in the measurement-error field.