ABSTRACT

ELLIOTT, LAINE E. Adjusting for Measurement Error. (Under the direction of Dr. Len Stefanski and Dr. Marie Davidian.)

A variety of complications arise when imperfect measurements, W, are observed in place of a true variable of interest, X. In the context of linear and non-linear regression models where X is a covariate, regression parameter estimators obtained when W is substituted for X may be substantially biased. Many strategies for correcting for measurement error depend on the specific modeling or regression context and can be intractable in highly non-linear models. In addition, previous methods often assume that the measurement error is normally distributed. In our work, we focus on re-creating the distribution of X from the observed W, either as the primary quantity of interest or as a means to improving parameter estimation. We obtain estimators of X for which the first M sample moments are unbiased for the corresponding moments of X. We investigate the benefit of substituting these estimates in density estimation, logistic regression and survival models. We compare this moment adjusted imputation (MAI) approach to existing alternatives in applications with normally distributed measurement error. We identify an important case of chi-square measurement error and propose a variety of methods to adjust for it, including a version of MAI. We find that MAI is often superior and has the advantage that once the estimates of X are obtained, they can be substituted in any model, including complicated non-linear models.

Adjusting for Measurement Error

by
Laine E. Elliott

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

Statistics

Raleigh, North Carolina

2009

APPROVED BY:

Dr. Butch Tsiatis	Dr. Daowen Zhang
Dr. Len Stefanski Co-chair of Advisory Committee	Dr. Marie Davidian Co-chair of Advisory Committee

DEDICATION

To my husband, Steven Thomas, and my parents, Reed and Chris Elliott.

BIOGRAPHY

Laine Elliott was born on April 24, 1979, in Camarillo, CA. She graduated from Rio Mesa High School in 1997. In 2004, Laine earned a B.S. in statistics, at California Polytechnic University, San Luis Obispo. She began graduate studies at North Carolina State University in the fall of 2004.

ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. Marie Davidian and Dr. Len Stefanski. My thesis research experience has been challenging and enjoyable because of their support. Our collaborator, Steve Brunelli, has contributed advice on applications for our methodology and conducted data analysis to augment this research.

I am grateful to the Department of Statistics at NCSU for financial support and excellent opportunities. I would also like to thank Terry Byron for his technical support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Moment Adjusted Imputation for Measurement Error Models

1.1 Introduction

In clinical studies, biological covariates are often measured only at baseline, and this measurement includes noise due to natural fluctuations or other sources. The relevant quantity of interest may be the average over fluctuations. For example, Gheorgiade et al. (2006) studied systolic blood pressure at hospital admission in patients hospitalized with acute heart failure using data from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF) registry. The relationship between blood pressure and mortality was studied in-hospital by fitting logistic regression models and post-discharge by fitting Cox proportional hazard models. The value of blood pressure was determined by a single in-hospital measurement. However, many studies have demonstrated large fluctuations in systolic blood pressure, and the average of many longitudinal blood pressure measurements is more strongly correlated with outcomes (Brueren et al., 1997; Pickering et al., 2005; Marshall, 2008). In other words, outcomes are more directly related to an underlying blood pressure level, averaged over fluctuations. Similarly, in descriptive analysis an unobserved "true" value may be more relevant than a noisy baseline measure.

Complications arise in the analysis of data when an imperfect measurement, W, is observed in place of a latent variable, X, and desired inferences involve X. When interest

focuses on estimation of the density of X, this presents an obvious problem because the mis-measured data, W, do not have the same distribution as the unobserved X. When X is a covariate in a linear or non-linear regression model for an outcome Y, estimators obtained when W is substituted for X may be substantially biased (Liu et al., 1978) and statistical power may be compromised (Freudenheim and Marshall, 1988; Carroll et al., 2006).

The effects of measurement error have been extensively studied. Many strategies to adjust for measurement error depend on the specific modeling or regression context. Correction for measurement error in linear model covariates is commonly achieved by regression calibration (RC) (Carroll and Stefanski, 1990; Gleser, 1990), which substitutes an estimate of the conditional mean $E(X|W)$ for the unknown X. The linear regression based on this substitution estimates the underlying parameters of interest. Regression calibration is also implemented in non-linear models because of its simplicity, but is typically most effective for general linear models when the measurement error is not large (Rosner, Spiegelman and Willett, 1989; Carroll et al., 2006). Alternatives for non-linear models are described by Carroll et al. (2006). These include structural models, which regard X as a random variable, and functional models in which X is treated as a fixed parameter. Structural methods, like maximum likelihood, yield efficient estimation but require that the density of X, $f_X(x)$, be known or well approximated. Alternatively, conditional score methods use estimating equations derived from the distribution of the observed data conditional on sufficient statistics for the unobserved X and include estimators that are efficient among functional methods. These have been developed for generalized linear models (Stefanski and Carroll, 1985, 1987), survival analysis (Tsiatis and Davidian, 2001), and joint models for longitudinal data and a primary endpoint (Li, Zhang and Davidian, 2004).

The preceding methods target estimation of parameters in a specific regression context; a different method must be implemented for every type of regression model in which X is used. This would be very time consuming in the OPTIMIZE-HF study where the mis-measured variable is used in multiple analyses. An alternative approach is to focus on re-creating the distribution of X from the observed W, either as the primary quantity of interest or as a means to improving parameter estimation. This has been explored from a Bayesian perspective (Louis, 1984; Shen and Louis, 1998; Freedman et al., 2004). Freedman et al. (2004) aim to replace the mis-measured data W with estimators that have the same joint distribution with Y as X, asymptotically. They implement a more

practical approximation to this idea by focusing on only the first two moments of the joint distribution. Their method, called moment reconstruction (MR), is based on empirical Bayes estimators, $\widehat{X}_{MR}$, which are derived so that $E(\widehat{X}_{MR})$, $\mathrm{Var}(\widehat{X}_{MR})$ and $\mathrm{Cov}(\widehat{X}_{MR}, Y)$ are equal to the corresponding moments of X. They impute the estimates $\widehat{X}_{MR}$ in a variety of applications and observe good results for normally distributed X.

When X has a normal distribution, it suffices to match two moments; bias in regression parameter estimators will be eliminated regardless of the particular model. Except in linear regression, this will not be true for other latent variable distributions. Moreover, this is not adequate when the density of X is of specific interest. This suggests an extension of moment reconstruction to higher-order moments and cross-products. To the best of our knowledge, the idea of computing higher-order, moment-adjusted estimates of the true X originated with the unpublished dissertation research of Bay (1997). His work focused on linear regression and applications to estimating distribution functionals such as the interquartile range. We expand on this work, calling the method Moment Adjusted Imputation (MAI). Our method retains the convenience of other imputation methods like regression calibration and moment reconstruction, in that, once the adjusted values are obtained, they can be used across a variety of analyses on the same data set using standard software.

In this chapter, we demonstrate the benefit of MAI, particularly for density estimation and logistic regression, where X is non-normal. In Section 1.2 we define the MAI algorithm and relate it to other imputation methods. In Sections 1.3 and 1.4 we compare adjustment procedures by simulation studies of kernel density estimation and regression models, respectively. We adjust the previous OPTIMIZE-HF analysis to account for measurement error and obtain estimates that describe the features of "true" blood pressure in Section 1.5.

1.2 The Method

In this section we describe the method developed by Bay (1997). Consider a collection of mis-measured observations W_i, $i = 1, \ldots, n$. These may be mis-measured versions of a scalar covariate X_i in a regression model of interest, or subject-specific estimates of scalar random effects X_i in a mixed effects model, subject to estimation uncertainty represented by U_i. Assume the W_i are independent across $i = 1, \ldots, n$, $W_i = X_i + U_i$ for $U_i \sim N(0, \sigma_{ui}^2)$,

where $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2, $\boldsymbol{U} = (U_1, \ldots, U_n)^T$ independent of $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, and U_i are mutually independent. No assumptions are made about the unobservable latent variables $X_1, \ldots, X_n$; they could be an independent, identically distributed (iid) random sample from some unknown distribution, as in a structural model, or fixed constants. We focus on the iid case where notation is simpler. Assume that σ_{ui} are known, as is common in measurement error models.

The objective is to construct adjusted versions of the W_i, say $\widehat{X}_i$, where the first M sample moments of $\widehat{X}_i$ unbiasedly estimate the corresponding moments of X_i; that is, $E(n^{-1}\sum_{i=1}^n \widehat{X}_i^r) = E(X^r)$, $r = 1, \ldots, M$. The distribution of $\widehat{X}_i$ resembles that of X_i up to M moments. If the latent variable X_i is a predictor in a regression model, $E(Y_i|X_i)$, we also match cross-product moments. This has intuitive appeal but can be motivated considering the simple case of linear regression. In linear regression, where $E(Y_i|X_i) = \beta_0 + \beta_X X_i$, the naive estimator of $\boldsymbol{\beta} = (\beta_0, \beta_X)^T$ based on W_i is

$$\widehat{\boldsymbol{\beta}}^N = \begin{bmatrix} n & \sum_i W_i \\ \sum_i W_i & \sum_i W_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i Y_i \\ \sum_i W_i Y_i \end{bmatrix}.$$

Inspection of $\widehat{\boldsymbol{\beta}}^N$ suggests matching the first two moments of X_i as well as the cross-product with Y_i so that $M = 2$ and $E(n^{-1}\sum_{i=1}^n \widehat{X}_i Y_i) = E(XY)$. Freedman et al. (2004) match these same moments. It is straightforward to show that

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} n & \sum_i \widehat{X}_i \\ \sum_i \widehat{X}_i & \sum_i \widehat{X}_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_i Y_i \\ \sum_i \widehat{X}_i Y_i \end{bmatrix}$$

is consistent for $\boldsymbol{\beta}$. Similarly, in the linear regression model including an error-free covariate Z_i, i.e. $E(Y_i|X_i) = \beta_0 + \beta_X X_i + \beta_Z Z_i$, both β_X and β_Z are consistently estimated if $\widehat{X}_i$ also satisfy $E(n^{-1}\sum_{i=1}^n \widehat{X}_i Z_i) = E(XZ)$. In non-linear models, where parameter estimates depend on higher-order moments and cross-products, we propose to match these as well.

In general, we want to find $\widehat{X}_i$ with $E(n^{-1}\sum_{i=1}^n \widehat{X}_i^r V_{ik}) = E(X^r V_k)$ for $r = 1, 2, \ldots, M$, where V_{ik} is the (i, k) element of $\boldsymbol{V} = \boldsymbol{(1, Y, Z)}$, 1 is an $n \times 1$ vector of ones, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, and $\boldsymbol{Z}$ is an $n \times (K-2)$ matrix whose columns are the values of $K-2$ error-free covariates for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Because $\boldsymbol{V}$ includes a vector of ones, matching cross-products with the columns of $\boldsymbol{V}$ includes matching moments. We

make the common *surrogacy* assumption that W_i is conditionally independent of V_{ik} given X_i, and use this in the following implementation (Carroll et al. 1995, Section 2.5).

1.2.1 Implementation

The first step is to find estimators $\widehat{m}_{rk}$ so that $E(\widehat{m}_{rk}) = E(X^r V_k)$, $k = 1, \ldots, K$. Based on the normality of U_i, unbiased estimators for the moments of X_i can be found as follows. Define Hermite polynomials by the recursion formula $H_0(z) = 1$, $H_1(z) = z$, $H_r(z) = zH_{r-1}(z) - (r-1)H_{r-2}(z)$ for $r = 2, 3, \ldots$ (Cramer, 1957). Stulajter (1978) proved that, if $W \sim N(\mu, \sigma^2)$, then $E\{\sigma^r H_r(W/\sigma)\} = \mu^r$ (Stefanski, 1989; Cheng and Van Ness, 1999). Letting $P_{ri}(w) = \sigma_{ui}^r H_r(w/\sigma_{ui})$, we have $E\{P_{ri}(W_i)|X_i\} = X_i^r$. The estimators $\widehat{m}_{rk} = n^{-1}\sum_{i=1}^n P_{ri}(W_i)V_{ik}$ are unbiased under the surrogacy assumption that W_i is conditionally independent of V_{ik} given X_i because $E\{E(\widehat{m}_{rk}|\boldsymbol{X}, \boldsymbol{V})\} = E[n^{-1}\sum_{i=1}^n E\{P_{ri}(W_i)|X_i\}V_{ik}]$ $= n^{-1}\sum_{i=1}^n E(X_i^r V_{ik}) = E(X^r V_k)$.

The adjusted $\widehat{X}_i$ are obtained by minimizing $\sum_{i=1}^n (W_i - X_i)^2$ subject to constraints on the moments and cross-products. For each of the K columns of $\boldsymbol{V}$, M_k constraints are imposed. For a particular matrix $\boldsymbol{V}$, the vector $\boldsymbol{M} = (M_1, \ldots, M_K)$ describes the number of cross-products matched, with each of its columns. Using Lagrange multipliers $(\lambda_{11}, ..., \lambda_{M_K K}) = \Lambda$, the objective function is

$$Q_{MK}(X_1, ..., X_n, \Lambda) = n^{-1}\sum_{i=1}^n \frac{1}{2}(W_i - X_i)^2 + \sum_{k=1}^K \sum_{r=1}^{M_k} \frac{\lambda_{rk}}{r}(n^{-1}\sum_{i=1}^n X_i^r V_{ik} - \widehat{m}_{rk}). \quad (1.1)$$

We take the derivative of Q_{MK} with respect to $(X_1, ..., X_n, \Lambda)$, equate this to 0, and solve for $(\widehat{X}_1, ..., \widehat{X}_n, \widehat{\Lambda})$ by Newton-Raphson (Appendix A.1). The resulting adjusted data are defined implicitly as $\widehat{X}_i = h(W_i, \boldsymbol{V}_i, \widehat{\Lambda})$. The solution $\widehat{X}_i$ is then substituted for X_i in the standard methods of estimation that would be performed if X_i were observed.

1.2.2 Implications

In a simple case it is possible to obtain an analytical solution that minimizes objective function (1.1). When we are interested in a single X in the absence of additional covariates, the estimator that matches two moments is $\widehat{X}_i = W_i\widehat{a} + \overline{W}(1 - \widehat{a})$, where $\widehat{a} = \left(\widehat{\sigma}_x^2/\widehat{\sigma}_w^2\right)^{1/2}$, $\widehat{\sigma}_w^2 = n^{-1}\sum_{i=1}^n (W_i - \overline{W})^2$, $\widehat{\sigma}_x^2 = \widehat{\sigma}_w^2 - \overline{\sigma}_u^2$, and $\overline{\sigma}_u^2 = n^{-1}\sum_{i=1}^n \sigma_{ui}^2$. A very similar estimator was developed via a multi-stage loss function by Louis (1984), with the

intention of matching only two moments. Except for the exponent in $\widehat{a} = \left(\widehat{\sigma}_x^2/\widehat{\sigma}_w^2\right)^{1/2}$, this estimator resembles the empirical Bayes estimator, which is known to have variance smaller than the posterior expected variance (Louis, 1984). The proposed approach puts more weight on W_i than does empirical Bayes. Thus, MAI provides an alternative to empirical Bayes when one is concerned about the problem of "overshrinkage."

MAI maintains the desirable properties of RC and MR with additional benefits. In fact, in simple linear regression we can replicate the RC and MR parameter estimates by matching two moments and a cross-product with the response, so $\boldsymbol{M} = (M_1, M_2) = (2, 1)$ for $\boldsymbol{V} = \boldsymbol{(1, Y)}$. The adjusted data are $\widehat{X}_i = W_i\widehat{a} + \overline{W}(1-\widehat{a}) + (Y_i - \overline{Y})\left(\widehat{\sigma}_{wy}/\widehat{\sigma}_y^2\right)(1-\widehat{a})$, where $\overline{W} = \sum_{i=1}^n W_i$, $\overline{Y} = \sum_{i=1}^n Y_i$, $\widehat{\sigma}_y^2 = n^{-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$, $\widehat{a} = \{(\widehat{\sigma}_x^2\widehat{\sigma}_y^2 - \widehat{\sigma}_{wy}^2)/(\widehat{\sigma}_w^2\widehat{\sigma}_y^2 - \widehat{\sigma}_{wy}^2)\}^{1/2}$, and $\widehat{\sigma}_{wy} = n^{-1}\sum_{i=1}^n (W_i - \overline{W})(Y_i - \overline{Y})$. The parameter estimator $\widehat{\beta}_X = \sum_{i=1}^n (\widehat{X}_i - \overline{\widehat{X}}_i)(Y_i - \overline{Y}_i)/\sum_{i=1}^n (\widehat{X}_i - \overline{\widehat{X}}_i)^2$ is equivalent to the RC estimator, $(\widehat{\sigma}_w^2/\widehat{\sigma}_x^2)\widehat{\beta}_X^N$, which is also identical to MR (Bay, 1997; Freedman et al., 2004). The application of our method with $\boldsymbol{M} = (2, 1)$ does not result in the same $\widehat{X}_i$ as Freedman et al. (2004), but our $\widehat{X}_i$ have similar properties. The $\widehat{X}_i$ converge to $X_i^* = W_i a + E(X)W(1-a) + \{Y_i - E(Y)\}\left(\sigma_{wy}/\sigma_y^2\right)(1-a)$, where $a = \{(\sigma_x^2\sigma_y^2 - \sigma_{wy}^2)/(\sigma_w^2\sigma_y^2 - \sigma_{wy}^2)\}^{1/2}$. (X^*, Y) has the same distribution as (X, Y) when this is multivariate normal. This allows for consistent estimation of the regression error variance (Freedman et al., 2004).

The proposed method replicates the adjusted estimator of Cheng and Schneeweiss (1998) for polynomial regression, which is obtained without adjusting data. In polynomial regression, there is a closed form solution for the coefficients that depends only on sample moments and cross-products, so the unbiased estimators $\widehat{m}_{rk}$ can be substituted directly into this solution. These authors use the same Hermite polynomials as a method to obtain $\widehat{m}_{rk}$ for normally distributed measurement error. Our method creates adjusted data that have these unbiased moments and therefore replicates that of Cheng and Schneeweiss (1998). For consistent estimation of a quadratic polynomial regression in X it is necessary to substitute unbiased estimators for four moments and second order cross-products, i.e. $\boldsymbol{M} = (4, 2)$ for $\boldsymbol{V} = \boldsymbol{(1, Y)}$. Although closed form solutions are not available for many non-linear models, the non-linearity is often well approximated by a lower-order polynomial. This suggests that these estimators are largely determined by lower-order moments and cross-products, so MAI could result in negligible bias.

1.2.3 Practical considerations

In practice, the $\widehat{m}_{rk}$ may not be a valid moment sequence. If our aim is only to match moments, but not cross-products ($k = 1$), it is well known (Shohat and Tamarkin, 1943) that a sequence of $2q + 1$ moments is valid if

$$\begin{bmatrix} 1 & \widehat{m}_{11} \\ \widehat{m}_{11} & \widehat{m}_{21} \end{bmatrix} \geq 0, \ldots, \begin{bmatrix} 1 & \widehat{m}_{11} & \cdots & \widehat{m}_{q1} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{m}_{q1} & \widehat{m}_{(q+1)1} & \cdots & \widehat{m}_{(2q)1} \end{bmatrix} \geq 0.$$

Checking these inequalities will identify the number of valid moments for a given data set. We address how many of these valid moments should be used, depending on the context.

For the purpose of matching an arbitrary number of moments and cross-products it is less clear how to identify a valid collection. For this discussion, let there be a single error-free covariate Z. We consider a limited set of $\boldsymbol{M} = (M_1, M_2 = M_1/2, ...M_K = M_1/2)$, where the number of moments, M_1, is even and the order of cross-products is $(M_1/2 + 1)$. This corresponds to matching the variance-covariance matrix of $(1, X, \ldots, X^{M_1/2}, Y, Z)$, and therefore has a nice interpretation. It is also the set of moments that must be matched to achieve consistent parameter estimation in polynomial regression of order $M_1/2$. In addition, simulations indicate that letting M_1 be odd can lead to "outlying" $\widehat{X}_i$ if the true distribution of X_i is extremely skewed (Bay, 1997). When $M_1 = 4$ we obtain $\widehat{m}_{11}$, $\widehat{m}_{21}$, $\widehat{m}_{31}$, $\widehat{m}_{41}$, $\widehat{m}_{12}$, $\widehat{m}_{22}$, $\widehat{m}_{13}$, and $\widehat{m}_{23}$, which have expectation $E(X)$, $E(X^2)$, $E(X^3)$, $E(X^4)$, $E(XY)$, $E(X^2Y)$, $E(XZ)$, and $E(X^2Z)$, respectively. Guided by this structure, we check the following to guarantee that we are targeting a valid variance covariance matrix:

$$\begin{bmatrix} 1 & \widehat{m}_{11} & \widehat{m}_{21} & n^{-1}\sum Y_i & n^{-1}\sum Z_i \\ \widehat{m}_{11} & \widehat{m}_{21} & \widehat{m}_{31} & \widehat{m}_{12} & \widehat{m}_{13} \\ \widehat{m}_{21} & \widehat{m}_{31} & \widehat{m}_{41} & \widehat{m}_{22} & \widehat{m}_{23} \\ n^{-1}\sum Y_i & \widehat{m}_{12} & \widehat{m}_{22} & n^{-1}\sum Y_i^2 & n^{-1}\sum Y_i Z_i \\ n^{-1}\sum Z_i & \widehat{m}_{13} & \widehat{m}_{23} & n^{-1}\sum Y_i Z_i & n^{-1}\sum Z_i^2 \end{bmatrix} \geq 0.$$

Occasionally, we encounter numerical problems when these moments can not be matched with a data set of size n. Even if the moments form a valid sequence, there may not be an empirical distribution function that takes jumps of size $1/n$ at each of n points having these moments. When this occurs, the Newton-Raphson algorithm will not converge to a

solution. It is rare to encounter this problem when only moments, but not cross-products, are involved. When this does occur, we recommend matching fewer moments.

1.3 Histogram and Kernel Density Estimation

In this and the next section, we demonstrate the utility of MAI in several representative analysis contexts. A simple and useful application of moment matching is to adjust mis-measured data to resemble the underlying error-free variable when interest focuses on insight about the distribution of the latent variable. A histogram or kernel density estimate (KDE) based on mis-measured data is too flat and dispersed. We illustrate this for three distributions of X: N(0,1), chi-square with 4 degrees of freedom, and a bimodal mixture of normals, which is 30% N(0,1) and 70% N(5,1), where the latter two are standardized to have mean 0 and variance 1. Measurement error with variance $\sigma_u^2 = 1$ is added to the data, corresponding to large measurement error, where the reliability ratio (RR), defined as $\text{var}(X)/\text{var}(W)$, is 0.5. Kernel density estimates for a single simulated data set for each distribution are displayed in Figure 1.3. Density estimates based on $\widehat{X}_{M=4}$ have features more like the density estimate that would be obtained from the true X. Because the normal distribution is completely defined by its first two moments, there is no benefit in matching additional moments. However, there is great improvement from matching four moments when X is chi-square or bimodal, as the KDE based on $\widehat{X}_{M=4}$ is substantially closer to that based on X.

These examples illustrate of the potential improvement in density estimation. We conducted a Monte Carlo simulation to investigate the extent of improvement and to identify the best number of moments to match under a variety of conditions. Several situations were considered, including the three distributions for X described above, two levels of measurement error, with reliability ratios of 0.75 and 0.5, and three sample sizes. The sample sizes $n = 300$, 1000, and 2000 were typical for measurement error applications (Stefanski and Carroll, 1985; Freedman et al., 2004).

We compare KDE based on our $\widehat{X}_i$ to that based on alternative methods of obtaining adjusted data. The first is regression calibration, where $E(X_i|W_i)$ is estimated by the best linear unbiased estimator $\widehat{X}_{RC,i} = \widehat{\mu}_x + (\widehat{\sigma}_{wx}/\widehat{\sigma}_w^2)(W_i - \widehat{\mu}_w)$, where $\widehat{\mu}_x$, $\widehat{\sigma}_{wx}$, and $\widehat{\sigma}_w^2$ are estimates obtained by method of moments. This is the form of the maximum likeli-

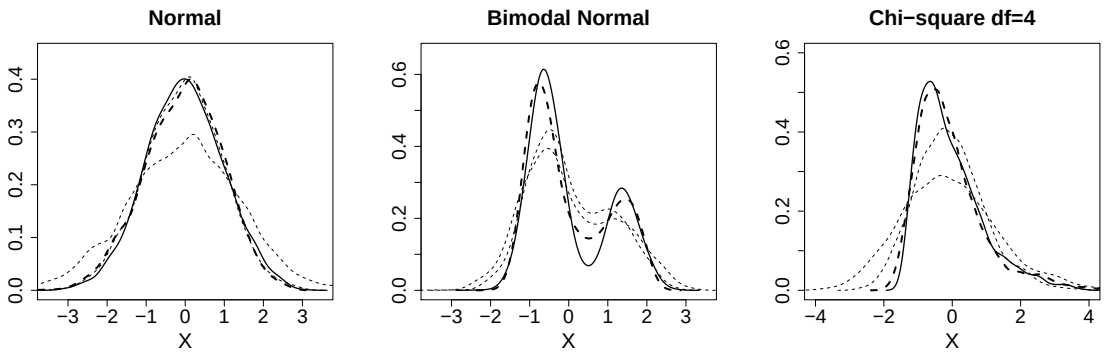

Figure 1.1: *Kernel Density Estimation; solid line: KDE of X, dark-dashed line: KDE of $\widehat{X}_{M=4}$, light-dotted lines: KDE of W and $\widehat{X}_{M=2}$; $n = 2000$ and $RR = 0.50$*

hood estimator for X_i when X is normally distributed, however, the same linear estimator is used regardless of the actual distribution of X. To account for non-normality, we also consider a different estimate of $E(X_i|W_i)$ called $\widehat{X}_{SNP,i}$, obtained by assuming X_i has density represented by the flexible family of semi-nonparametric (SNP) densities $f_X(x|\mu, \sigma, \boldsymbol{\alpha})$ (Gallant and Nychka, 1987; Carroll et al., 2006), which involve parameters μ, σ, and $\boldsymbol{\alpha}$ and can approximate many potential latent variable distributions. The family has a convenient form so that $f_W(w|\mu, \sigma, \boldsymbol{\alpha})$ can be obtained by integration over x, and the parameters $(\mu, \sigma, \boldsymbol{\alpha})$ can be estimated from the observed data by maximum likelihood. This approach naturally provides density estimation in the form of $f_X(x|\widehat{\mu}, \widehat{\sigma}, \widehat{\boldsymbol{\alpha}})$. We took the extra step of estimating $E(X_i|W_i)$ as $\widehat{X}_{SNP,i} = \int x f_{X,W}(x, w_i|\widehat{\mu}, \widehat{\sigma}, \widehat{\boldsymbol{\alpha}})dx / f_W(w_i|\widehat{\mu}, \widehat{\sigma}, \widehat{\boldsymbol{\alpha}})$.

The versions of $\widehat{X}_i$ are evaluated according to their closeness to the underlying X_i as measured by $MSE(\widehat{X}) = B^{-1}\sum_{b=1}^{B} n^{-1}\sum_{i=1}^{n}(\widehat{X}_{i,b} - X_{i,b})^2$ and B simulated data sets, and by computing the integrated squared error between the empirical distributions functions, given by $ISE(G_{\widehat{X}}) = B^{-1}\sum_{b=1}^{B}\int\{G_{\widehat{X},b}(t) - G_{X,b}(t)\}^2 dt$, where $G_X(t) = n^{-1}\sum_{i=1}^{n} I_{(X_i \le t)}$, $-\infty < t < \infty$. When X is assumed to have an SNP density we can also estimate the cumulative distribution function, cdf_{SNP}, directly from $f_X(x|\widehat{\mu}, \widehat{\sigma}, \widehat{\boldsymbol{\alpha}})$. Density estimation using the SNP family is well established, so we also calculate $ISE(cdf_{SNP}) = B^{-1}\sum_{b=1}^{B}\int\{cdf_{SNP,b}(t) - G_{X,b}(t)\}^2 dt$ as a gold standard. We report an MSE ratio, $MSE(W)/MSE(\widehat{X})$, and an ISE ratio, $ISE(G_W)/ISE(G_{\widehat{X}})$ so that larger ratios indicate a greater reduction in error. Standard errors for these ratios are obtained by the delta

method and are reported as a "coefficient of variation" which is the ratio standard error divided by the ratio itself.

Table 1.1: Simulation results for three latent variable distributions, $f_X(x)$; two reliability ratios (RR), $B = 500$ simulated data sets, and $n = 1000$. Statistics reported: (a) $MSE(W)/MSE(\widehat{X})$, where $MSE(\widehat{X}) = B^{-1}\sum_{b=1}^{B} n^{-1}\sum_{i=1}^{n}(\widehat{X}_{i,b} - X_{i,b})^2$ (coefficient of variation ≈ 0.001), and (b) $ISE(G_W)/ISE(G_{\widehat{X}})$, where $ISE(G_{\widehat{X}}) = B^{-1}\sum_{b=1}^{B}\int\{G_{\widehat{X},b}(t) - G_{X,b}(t)\}^2 dt$, for $G_X(t) = n^{-1}\sum_{i=1}^{n} I_{(X_i \le t)}$, $-\infty < t < \infty$ (coefficient of variation ≈ 0.02). Adjusted data $\widehat{X}$: RC, regression calibration; $M = 2$, $M = 4$, $M = 6$, MAI matching 2, 4 or 6 respectively; SNP, semi-nonparametric.

Distribution	RR	$\widehat{X}_{RC}$	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	$\widehat{X}_{M=6}$	$\widehat{X}_{SNP}$	cdf_{SNP}
			(a) $\frac{MSE(W)}{MSE(\widehat{X})}$				
Normal	0.75	1.33	1.24	1.24	1.23	1.32	-
	0.50	1.99	1.71	1.70	1.55	1.99	-
Chi Sq df=4	0.75	1.33	1.24	1.38	1.37	1.43	-
	0.50	1.99	1.71	1.88	1.79	2.15	-
Bimodal	0.75	1.33	1.24	1.50	1.43	1.64	-
	0.50	2.00	1.71	1.79	1.64	2.15	-
			(b) $\frac{ISE(G_W)}{ISE(G_{\widehat{X}})}$				
Normal	0.75	1.09	7.90	7.03	4.20	1.08	5.99
	0.50	1.34	23.72	11.65	0.82	1.33	9.64
Chi Sq df=4	0.75	1.39	2.41	6.96	5.15	0.61	1.08
	0.50	1.74	4.39	10.99	3.15	1.19	1.91
Bimodal	0.75	0.81	1.32	5.05	1.94	2.35	2.71
	0.50	0.86	1.90	4.13	1.29	1.48	4.99

Simulation results for sample size $n = 1000$ are displayed in Table 1.1. Results for $n = 300$ and $n = 2000$ were similar and are available in Appendix A.2. For normally distributed X, there is very little difference in the MSE ratio based on matching two or four moments. However, there is no additional benefit from matching six moments. When it is not important to get each $\widehat{X}_i$ close to the original X_i, but instead we want an ensemble of $\widehat{X}_1, ..., \widehat{X}_n$ that have a similar distribution to $X_1, ..., X_n$, then ISE ratios indicate that it is better to match only two moments. However, when $f_X(x)$ is chi-square or bimodal

normal mixture, both ratios indicate that it is better to match four moments rather than two or six. Moments greater than four may not be as essential in describing distributions, and their estimators are likely to be highly variable. A general recommendation is to match four moments.

A comparison of MAI $\widehat{X}_{M=4}$ to $\widehat{X}_{RC}$ and $\widehat{X}_{SNP}$ indicates that the latter two methods do a better job at getting each $\widehat{X}_i$ close to the original X_i. These have slightly larger MSE ratios that those for $\widehat{X}_{M=4}$. This is true regardless of the underlying distribution of X. It is not surprising that conditional expectations give good estimation of the individual $\widehat{X}_i$. However expectations are known to be less variable than the original data, so the distribution of $\widehat{X}_{RC}$ and $\widehat{X}_{SNP}$ may not resemble that of X. In fact, the ISE ratios for these methods are close to 1, or even smaller than 1, indicating that the empirical distribution based on these $\widehat{X}$ is no better than that of W. Thus, when interest focuses on density estimation, RC and SNP are not a good approaches. The ISE ratios for $\widehat{X}_{M=4}$ reflect large improvement in density estimation relative to W. This confirms the impression suggested by Figure 1.3. The ISE ratios for $\widehat{X}_{M=4}$ are at least as large as those of our gold standard, cdf_{SNP}, particularly for chi-square X. The SNP density estimator is based on the normal distribution and is not ideal for estimating skewed densities. Here, MAI provides a good alternative.

1.4 Simulations in Regression Models

In this section, we evaluate various methods that adjust for covariate measurement error in common non-linear models over a range of conditions, including (1) Distribution of X : N(0,1), standardized chi-square df=4, standardized bimodal mixture of normals (see Section 3) (2) Sample size: $n = 300$, $n = 1000$, $n = 2000$ (3) Measurement error variance: moderate $\sigma_u^2 = 0.33$ (reliability ratio 0.75), large $\sigma_u^2 = 1.0$ (reliability ratio 0.50) (4) Underlying model parameters: levels depend on the specific model. An additional error-free covariate Z was generated from the same distribution as X, and the correlation between Z and X was 0.4. The underlying model parameters control the extent to which the model deviates from linearity. To get a general understanding of how these methods perform in different circumstances, we vary the model parameters and strength of the covariate effects.

We compare parameter estimation based on three imputation methods as well as

the conditional score method (Stefanski and Carroll, 1987; Tsiatis and Davidian, 2001). For RC, $\widehat{X}_{RC}$ is the best estimated linear unbiased estimator of $E(X|W,Z)$. We use the modification of MR proposed by Freedman et al. (2004) that involves conditioning on the error-free covariate so that $\widehat{X}_{MR,i} = \widehat{E}(W|Y_i,Z_i)(1-\widehat{G}) + W_i\widehat{G}$, where $\widehat{G} = \widehat{\text{var}}(X|Y_i,Z_i)^{1/2}/\widehat{\text{var}}(W|Y_i,Z_i)^{1/2}$. For the MAI method we use $\widehat{X}_{4,2,2}$ for which $E(X)$, $E(X^2)$, $E(X^3)$, $E(X^4)$, $E(XY)$, $E(X^2Y)$, $E(XZ)$, $E(X^2Z)$ are matched. Some alternatives are discussed in the context of specific models.

When $\widehat{X}_i$ are used in a regression model, the usual standard errors for regression parameter estimates will not be correct. When the regression parameter estimators are M-estimators, standard errors may be obtained by the empirical sandwich approach. The equations that determine $(\widehat{X}_1,...,\widehat{X}_n,\widehat{\Lambda})$ can be stacked with the usual equations in which the unknown X_i are replaced with $h(W_i,\boldsymbol{V}_i,\widehat{\Lambda})$. Details are included in the Appendix A.3. Alternatively, standard errors can be obtained by bootstrapping. We estimate standard errors for the Cox model parameters by bootstrapping.

1.4.1 Logistic Regression

A common model to describe the relationship between a binary outcome Y_i and covariates is logistic regression. The model for the response is $P(Y=1|X,Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$ where $F(v) = 1+\exp(-v)^{-1}$, and Z is a scalar error free covariate. We simulate data from two parameter settings, $(\beta_0,\beta_X,\beta_Z) = (-1.5,1,1)$ and $(\beta_0,\beta_X,\beta_Z) = (-0.6,0.3,0.3)$. The first is similar to Freedman et al. (2004) and corresponds to substantial non-linearity, strong covariate effects, and event rate $P(Y=1) \approx 0.30$. For the second set of parameter values, $P(Y=1|X,Z)$ is nearly linear in the range of X, the effect of X is moderate, and the event rate $P(Y=1) \approx 0.36$. The observed data are Y_i, W_i, and Z_i, for $i = 1,\ldots,n$.

Boxplots of the estimated coefficients $\widehat{\beta}_X$ from 500 simulations are displayed in Figure 1.2 for the case where $(\beta_0,\beta_X,\beta_Z) = (-1.5,1,1)$, $n = 2000$, and $\sigma_u^2 = 1.0$. When X is normally distributed, the RC estimator for β_X shows slight bias, but has the least variability. The other methods are unbiased and have similar variability. RC and MR are expected to perform well when X is normally distributed, and there is nothing to be gained from information about higher order moments. However, it is reassuring to see that the increase in variability from matching additional moments is not substantial. When the

latent variable distribution is either chi-square or bimodal, the RC and MR estimators are biased. Only MAI and CS are unbiased, and these have similar variability.

Other results are presented in Appendix A.4. When the underlying coefficients are $(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$, results are similar to those in Figure 1.2. As expected, when the underlying coefficients are $(\beta_0, \beta_X, \beta_Z) = (-0.6, 0.3, 0.3)$, all adjustment methods are nearly identical in terms of estimator bias and variability. The measurement error in X also effects the estimation of the error-free covariate effect, β_Z. The naive estimator for β_Z is biased. This can be corrected by adjusting for measurement error in X. However, in our simulations the regression calibration estimator for β_Z is biased, even when the latent variable is normally distributed. The other adjustment procedures perform similarly and demonstrate negligible bias in β_Z (Appendix A.4).

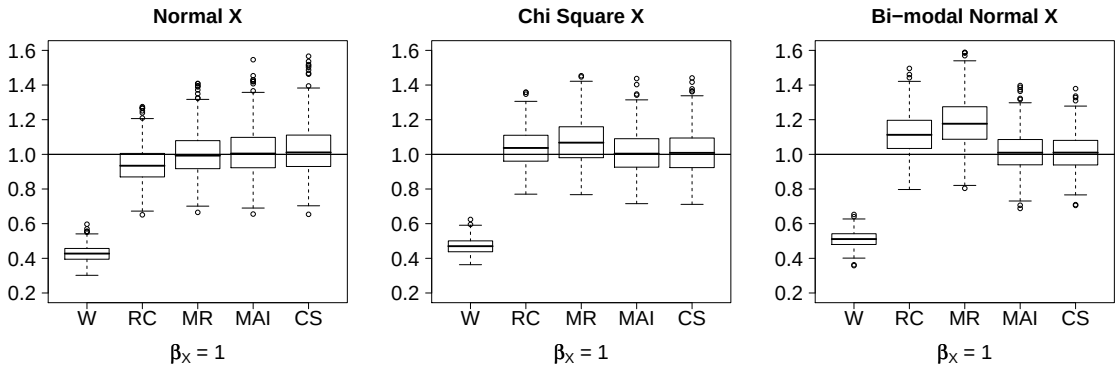

Figure 1.2: *Boxplots of* $\widehat{\beta}_X$ *from* $B = 500$ *simulated data sets where* $P(Y = 1|X, Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$ *with true values* $(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$, $\sigma_u^2 = 1.0$, *and* $n = 2000$.

Moment matching is a good alternative in logistic regression when the underlying latent variable distribution is unknown. In logistic regression, we recommend that four, moments and two cross-products with every model variable be matched. This level of matching is necessary to render negligible bias in our simulations (Appendix A.4). In Appendix A.4, Table A.10, we compare the sandwich variance to the Monte Carlo variance of parameter estimates and observe that the sandwich formula estimates are reasonable.

1.4.2 Cox Proportional Hazard Model

Another common non-linear model is the Cox proportional hazard model for a time to event. For subject $i = 1, \ldots, n$, let T_i denote the failure time and C_i denote the censoring

time. The failure time T_i is not available for all subjects, but instead $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ are observed. As before, X_i denotes a covariate of interest for which $W_i = X_i + U_i$ is observed in its place, and Z_i is an error free covariate. The hazard of failure $\lambda(t|X, Z)$ is related to the covariates by the proportional hazard model

$$\begin{aligned}\lambda(t|X,Z) &= \lim_{h\to 0^+}\left\{h^{-1}P(t \leq T \leq t+h|T \geq t, X, Z)\right\} \\ &= \lambda_0(t)\exp(\beta_X X + \beta_Z Z),\end{aligned}$$

where $\lambda_0(t)$ is an underlying baseline hazard function.

We consider two scenarios. One is similar to that of Wang (2006), where failure times occur according to the hazard function $\lambda(t|X, Z)$ with $\lambda_0 = 0.2$, $\beta_X = 0.7$, and $\beta_Z = 0.7$, and 50% of subjects are censored uniformly. This implies a very strong covariate effect with a hazard ratio of 2 for each unit change in X and a hazard ratio of 66 for the largest value of X compared to the smallest [$\exp(0.7) \approx 2$ and $\exp\{0.7\text{range}(X)\} \approx 66$]. As a moderate alternative, we generate failure times from hazard function $\lambda(t|X, Z)$ with $\lambda_0 = 1$, $\beta_X = 0.3$, and $\beta_Z = 0.3$, and 40% of subjects are censored. This corresponds to a hazard ratio of 1.4 for each unit change in X and a hazard ratio of 6 overall [$\exp(0.3) \approx 1.4$ and $\exp\{0.3\text{range}(X)\} \approx 6$]. This is strong enough to be statistically significant in a sample of $n = 300$, but moderately so.

In logistic regression, the available data consist of (Y, W, Z) whereas for time to event data we have (Y, δ, W, Z). In this case we extend the moment matching to include δ and target the joint distribution of (Y, δ, X, Z). We considered alternative matching approaches. For example, the simplest approach would match the variance-covariance matrix of $(1, X, \ldots, X^{M_1/2}, Y, \delta, Z)$. We could also match the variance-covariance matrix of $(1, X, \ldots, X^{M_1/2}, Y, Z)$ within each level of δ. Furthermore, we could match on risk sets, which is to re-match $(1, X, \ldots, X^{M_1/2}, Y, Z)$ at different points in time for those subjects who are still at risk. We tried all of these approaches and saw little difference in the results. We therefore recommend the first and simplest method. The adjusted data are $\widehat{X}_{MAI}$ for which $E(X)$, $E(X^2)$, $E(X^3)$, $E(X^4)$, $E(XY)$, $E(X^2Y)$, $E(X\delta)$, $E(X^2\delta)$, $E(XZ)$, and $E(X^2Z)$ are matched. We compared this to a lesser adjustment for which only $E(X)$, $E(X^2)$, $E(XY)$, $E(X\delta)$, $E(XZ)$ are matched and observed more bias and similar variability in the resulting estimator. Among the MAI methods we prefer $\widehat{X}_{MAI}$ and results are presented for this version only.

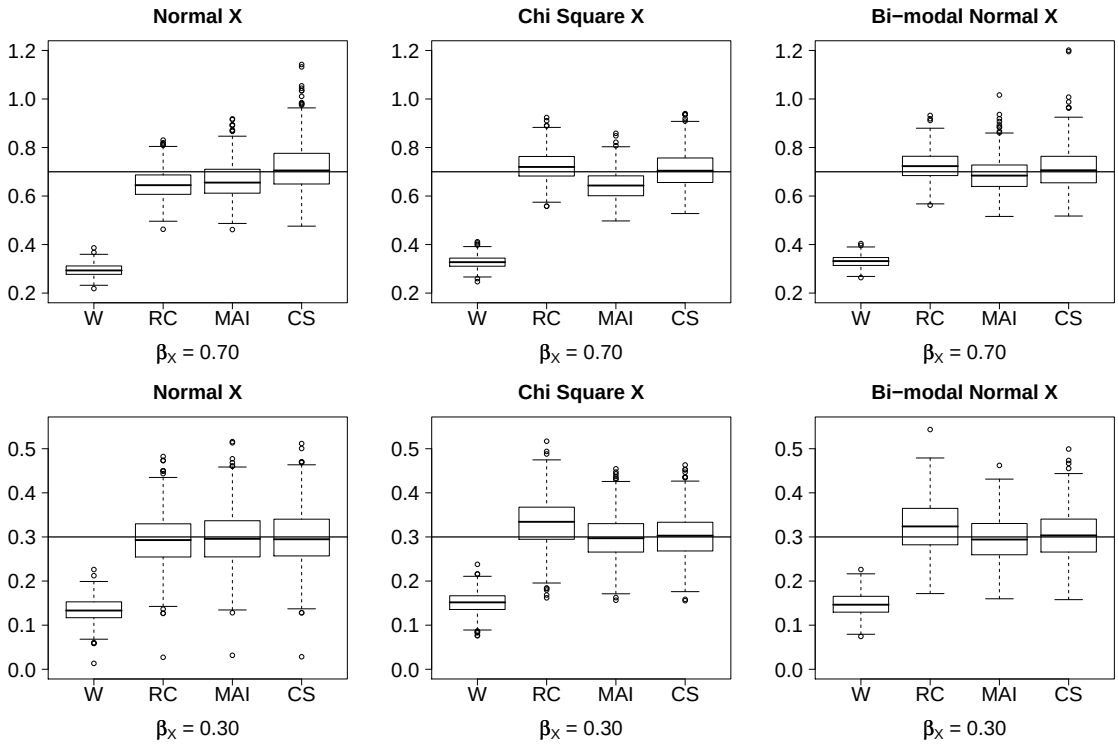

Figure 1.3: *Boxplots of* $\widehat{\beta}_X$ *from* $B = 500$ *data sets where* $\lambda(t|X, Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$ *with true values* $\{\lambda_0(t), \beta_X, \beta_Z\} = (0.2, 0.7, 0.7)$ *or* $\{\lambda_0(t), \beta_X, \beta_Z\} = (1, 0.3, 0.3)$*,* $\sigma_u^2 = 1.0$*, and* $n = 2000$*.*

Boxplots of the estimated coefficients $\widehat{\beta}_X$ from 500 simulations, where $n = 2000$ and $\sigma_u^2 = 1.0$, are displayed in Figure 1.3. When the true parameter values are $\lambda_0 = 0.2$, $\beta_X = 0.7$, and $\beta_Z = 0.7$, only the CS estimator is unbiased, regardless of the distribution of X. The bias in RC and MAI estimators is evident, though relatively small. The variability in these estimators is similar, though the RC estimator is somewhat less variable. When the true parameter values are $\lambda_0 = 1$, $\beta_X = 0.3$ and $\beta_Z = 0.3$, and X has a normal distribution, all of the methods give virtually unbiased estimation with similar variability. However, for a chi-square or bimodal latent variable, the RC estimator is biased. Both the MAI and CS estimators are unbiased and have similar variability. The results for other sample sizes and levels of measurement error are similar (Appendix A.5).

As in logistic regression, the measurement error in X impacts estimation of β_Z, and the naive estimator is biased. In Appendix A.5, we see that the RC estimator for β_Z is over-corrected, particularly for the larger underlying parameter values. MAI and CS

estimators are nearly unbiased and have similar variability to RC.

In our simulations, the conditional score approach is preferable. However, the conditional score approach may be excessively time consuming or infeasible for complicated Cox models. Imputation approaches, although imperfect, offer a practical solution. Both RC and MAI are easy to implement and provide great improvement over the naive method. For the estimation of β_X neither can be recommended over the other based on our simulations. However, when β_Z is also of interest MAI is preferable to RC.

1.5 Application to OPTIMIZE-HF data

We redo the OPTIMIZE-HF analyses performed by Gheorgiade et al. (2006) accounting for measurement error. The data set consists of $n = 48,612$ subjects, aged 18 or older, with heart failure. The distribution of systolic blood pressure across subjects is described by the KDE (Figure 1.4). There are two outcomes of interest, in-hospital mortality and post-discharge mortality. We use the models reported by Gheorgiade et al. (2006), which include baseline systolic blood pressure, many baseline covariates that are regarded as error-free, and linear splines and truncation that account for non-linearity in continuous covariates. Their model for in-hospital mortality can be written as

$$P(Y = 1|\boldsymbol{Z}, X) = \{1 - \exp(\beta_0 + \boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_1 S)\}^{-1},$$

where $\boldsymbol{Z}$ includes error-free covariates listed in Appendix A.6, Table A.17; S is a truncated version of blood pressure, i.e. $S{=}-\{XI(X < 160) + 160I(X \geq 160)\}$; and X represents true systolic blood pressure. A pre-specified subset of patients ($n = 5,791$) were followed for 60 to 90 days after discharge. In this group, post-discharge mortality is described by the Cox proportional hazard model

$$\lambda(t|X, \boldsymbol{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_1 S_1 + \beta_2 S_2)$$

where $\boldsymbol{Z}$ includes error free covariates which are listed in Appendix A.6, Table A.18; S_1 and S_2 fit a linear spline to blood pressure, i.e. $S_1{=}-\{XI(X < 140) + 140I(X \geq 140)\}$ and $S_2 = -\{0I(X < 140) + XI(X \geq 140)\}$; and X is true systolic blood pressure. Gheorgiade et al. (2006) fit these models using baseline systolic blood pressure, W, in place of X. We adjust the mis-measured W and impute $\widehat{X}$ in place of X. We report the odds ratios per

10-mm Hg change in S and hazard ratios for S_1, and S_2 per 10-mm Hg change. Ninety-five percent confidence intervals for the odds ratios and hazard ratios are based on standard errors obtained from 1000 bootstrap samples.

The adjustment procedures assume that the measurement error variance, σ_u^2, is known. In practice, this is usually replaced by a good estimate. It would be best to estimate the measurement error variance from replicate measures of systolic blood pressure, taken over a period of time. Replicate measures were not available in the OPTIMIZE-HF data set; however, variability in blood pressure has been extensively studied. One source is the Framingham data set (Carroll et al., 2006), which includes four measurements of blood pressure, two taken at the first exam and two taken at a second exam. The average standard deviation in four measurements is 9 mm Hg, which corresponds to a reliability ratio of about 0.75. Based on the information from other external studies, the measurement error may actually be larger (Marshall, 2008). For the purpose of illustration, we use a reliability ratio of 0.75 for adjustment. In practice it is very important to obtain replicate data.

Let W denote baseline systolic blood pressure in 10-mm Hg units and $\widehat{X}_{M=4}$ denote the MAI adjusted data based on matching four moments. The estimated distribution of baseline systolic blood pressure is altered by adjustment (Figure 1.4). The adjusted version indicates a distribution with higher peak and smaller tails and conveys the impression that patients' blood pressures are more similar to each other.

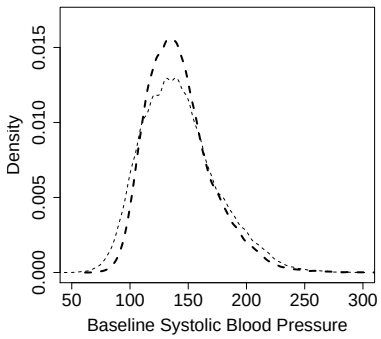

Figure 1.4: *Kernel Density Estimate; light-dotted lines: KDE of W, dark-dashed line: KDE of $\widehat{X}_{M=4}$*

In Table 1.2 we compare the adjusted odds ratios and hazard ratios to those obtained by Gheorgiade et al. (2006). The adjusted estimates indicate a stronger effect of

systolic blood pressure. The RC estimates move in the same direction as MAI, but are closer to the naive estimates. The impact of adjustment is not substantial in this case; however, we are assuming a relatively moderate amount of measurement error. Many studies have reported higher variability in replicate blood pressure measurements and adjustment could be more important in estimating effect size.

Table 1.2: Parameter Estimates for the OPTIMIZE-HF data analysis. Confidence intervals are included in parenthesis, following the estimates

	Unadjusted	*RC*	*MAI*
	(a) Odds ratios for logistic regression (1.5)		
β_1	1.21 (1.17, 1.25)	1.28 (1.22, 1.33)	1.36 (1.44, 1.52)
	(b) Hazard ratios for Cox model (1.5)		
β_1	1.18 (1.10, 1.26)	1.24 (1.13, 1.36)	1.38 (1.22, 1.54)
β_2	1.08 (1.01, 1.15)	1.10 (1.00, 1.21)	1.06 (0.94, 1.17)

1.6 Discussion

We have introduced MAI as a means for adjusting mis-measured data to reflect the latent variable distribution and improve parameter estimation in non-linear regression models. The method does not require any assumptions on the latent variable distribution. Moreover, it performs well for a variety of distributions. For the purpose of density estimation, MAI is typically superior to the alternatives we considered. In simulations of logistic regression, the method is similar to MR when the latent variable is normally distributed, but is a superior imputation method when the latent variable is non-normal. In the Cox proportional hazards model, RC and MAI provide substantial improvement over the naive approach, but do not eliminate bias. Of the functional approaches that we considered, the conditional score is the only method that eliminates bias in the Cox model parameter estimators.

The OPTIMIZE-HF study of systolic blood pressure performed by Gheorgiade et al. (2006) is illustrative of a typical data analysis. The mis-measured variable is included in descriptive analyses and in multiple models. The models are complex, involving splines

to account for non-linearity. In practice, models could include splines, squared terms or interactions with the latent variable of interest. These are easily accommodated by imputation methods, and other approaches such as conditional score may be difficult or impossible to implement. In these circumstances, an imputation approach may be desirable.

In this paper we develop MAI for the case of normally distributed measurement error. The method depends on correct specification of the measurement error distribution. Analysts should take care to verify that normality of measurement error is a reasonable assumption. The MAI method can be extended to other types of measurement error, as long as the distribution is known. Work on such extensions is reported elsewhere (Chapter 3).

Chapter 2

Extension of Moment Adjusted Imputation to Multiple Mis-measured Covariates

2.1 Introduction

In Chapter 1, we focus on measurement error in a single latent variable of interest X. In a regression setting, we assume that other model variables are measured without error or that the error is negligible. In some situations, this may not be reasonable. When multiple covariates are measured with error, we could apply a univariate adjustment separately. However, this would not account for correlation between the latent variables, which could be exploited to improve the adjustment. For example, systolic and diastolic blood pressure are often measured simultaneously and with error. Because these are likely to be correlated, we should consider their joint distribution in making adjustments. In addition, the measurement error in observed systolic and diastolic blood pressures will be correlated. In this chapter, we propose an extension of Moment Adjusted Imputation (MAI) that accounts for these features of multivariate mis-measured data.

2.2 The Method

The notation used in Chapter 1 is not convenient for multivariate mis-measured data. Here we introduce notation for the current problem that is similar to Chapter 1 but not identical. Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iG})^T$ be a $(G \times 1)$ vector of latent variables for $i = 1, \ldots, n$. The observed data are $\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{U}_i$ where $\boldsymbol{U}_i \sim MVN(\boldsymbol{0}, \Sigma_{ui})$, $MVN(\boldsymbol{\mu}, \Sigma)$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, $\boldsymbol{0}$ is a $G \times 1$ vector of zeros, $\boldsymbol{U}_i$ is independent of $\boldsymbol{X}_i$, and $\boldsymbol{U}_i$ are mutually independent. We assume that Σ_{ui} is known. The latent variables $\boldsymbol{X}_i$ may be of particular interest, as in density estimation or as predictors in a regression model. In the latter case, we also have a response Y_i and potentially a vector of $(K-1)$ error-free covariates $\boldsymbol{Z}_i$. These additional variables are collected to create $\boldsymbol{V}_i$=$(Y_i, \boldsymbol{Z}_i^T)$, with components V_{ik} for $k = 1, \ldots, K$. Note that the definition of $\boldsymbol{V}$ differs from Chapter 1. We assume that Y_i and $\boldsymbol{Z}_i$ are not related to the measurement error in $\boldsymbol{X}_i$, so that $\boldsymbol{V}_i$ is conditionally independent of $\boldsymbol{W}_i$ given $\boldsymbol{X}_i$ (Carroll et al. 1995, Section 2.5). The goal is to obtain adjusted versions of the $\boldsymbol{W}_i$, $\widehat{\boldsymbol{X}}_i$, that reflect the joint distribution of $\boldsymbol{X}_i$ and possibly the joint distribution of $\boldsymbol{X}_i$ and other variables. In terms of moments, we require that $E(n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i)$=$E(\boldsymbol{X}_i)$, $E(n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i\widehat{\boldsymbol{X}}_i^T)$=$E(\boldsymbol{X}_i\boldsymbol{X}_i^T)$, $E(n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i V_{ik})$=$E(\boldsymbol{X}_i V_{ik})$ for $k = 1, \ldots, K$, and $E(n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i^r)$=$E(\boldsymbol{X}_i^r)$ for $r = 3, \ldots, M$, and the r^{th} power is applied component-wise.

2.2.1 Implementation

The first step is to define unbiased estimators for the unknown quantities $\boldsymbol{m}_r = E(\boldsymbol{X}_i^r)$, $(G \times 1)$ and $r = 1, \ldots, M$, $\boldsymbol{m}^* = E(\boldsymbol{X}_i\boldsymbol{X}_i^T)$, $(G \times G)$, $\boldsymbol{m}_k^V = E(\boldsymbol{X}_i V_{ik})$, $(G \times 1)$. Because $\boldsymbol{W}_i|\boldsymbol{X}_i \sim MVN(\boldsymbol{X}_i, \Sigma_{ui})$, we know that $E(\boldsymbol{W}_i) = E\{E(\boldsymbol{W}_i|\boldsymbol{X}_i)\} = E(\boldsymbol{X}_i)$ so $\widehat{\boldsymbol{m}}_1$=$n^{-1}\sum_{i=1}^n \boldsymbol{W}_i$. Unbiased estimators for the higher order moments are defined using the recursion formula $H_0(z) = 1$, $H_1(z) = z$, $H_r(z) = zH_{r-1}(z) - (r-1)H_{r-2}(z)$ for $r = 2, 3, \ldots$ (Cramer, 1957). Stulajter (1978) proved that if $W \sim N(\mu, \sigma^2)$, then $E\{\sigma^r H_r(W/\sigma)\} = \mu^r$ (Stefanski, 1989; Cheng and Van Ness, 1999). Let W_{ig} and X_{ig} denote the g^{th} component of $\boldsymbol{W}_i$ and $\boldsymbol{X}_i$, respectively, and let $\Sigma_{ui,gg'}$ denote the element of Σ_{ui} in the g^{th} row and g'^{th} column. Marginally, $W_{ig}|X_{ig} \sim N(X_{ig}, \Sigma_{ui,gg})$. Letting $P_r(w, \sigma) = \sigma^r H_r(w/\sigma)$, we have $E\{P_r(W_{ig}, \Sigma_{ui,gg})\}$= $E[E\{P_r(W_{ig}, \Sigma_{ui,gg})|X_{ig}\}]$=$E(X_{ig}^r)$. The g^{th} component of $\widehat{\boldsymbol{m}}_r$ is $\widehat{m}_{rg} = n^{-1}\sum_{i=1}^n P_r(W_{ig}, \Sigma_{ui,gg})$ for $r = 1, \ldots, M$. In addition, $E(\boldsymbol{W}_i\boldsymbol{W}_i^T|\boldsymbol{X}_i)$=

$\boldsymbol{X}_i\boldsymbol{X}_i^T + \Sigma_{ui}$, so $\widehat{\boldsymbol{m}}^*$=$n^{-1}\sum_{i=1}^n(\boldsymbol{W}_i\boldsymbol{W}_i^T - \Sigma_{ui})$ is unbiased for $\boldsymbol{m}^*$. In some cases the estimate $\widehat{\boldsymbol{m}}^*$ may be non-positive definite, corresponding to an invalid moment sequence. We only perform adjustment for a valid sequence of moments estimates, as defined in Section 2.3. Under the *surrogacy* assumption that $\boldsymbol{V}_i$ are conditionally independent of $\boldsymbol{W}_i$ given $\boldsymbol{X}_i$, $E(\boldsymbol{W}_iV_{ik}) = E\{E(\boldsymbol{W}_iV_{ik}|\boldsymbol{X}_i, V_{ik})\} = E\{V_{ik}E(\boldsymbol{W}_i|\boldsymbol{X}_i)\} = E(\boldsymbol{X}_iV_{ik})$. Therefore, the estimators $\widehat{\boldsymbol{m}}_k^V$=$n^{-1}\sum_{i=1}^n \boldsymbol{W}_iV_{ik}$ are unbiased for $\boldsymbol{m}_k^V$.

The adjusted $\widehat{\boldsymbol{X}}_i$ are obtained by minimizing

$$\sum_{i=1}^n(\boldsymbol{W}_i - \boldsymbol{X}_i)^T\Sigma_{ui}^{-1}(\boldsymbol{W}_i - \boldsymbol{X}_i) \tag{2.1}$$

subject to the following constraints on sample moments and cross products: $n^{-1}\sum_{i=1}^n \boldsymbol{X}_i^r$ $= \widehat{\boldsymbol{m}}_r$, $n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T = \widehat{\boldsymbol{m}}^*$, and $n^{-1}\sum_{i=1}^n \boldsymbol{X}_iV_{ik} = \widehat{\boldsymbol{m}}_k^V$. This is minimized by taking the derivative with respect to $\boldsymbol{X}_i$, and constraints are imposed by Lagrange multipliers. In an implementation, we set the derivatives equal to 0 and solve numerically using the R function multiroot().

2.2.2 Special Case: Matching two moments and cross product

In the simple case where $\boldsymbol{V}_i = \boldsymbol{Y}_i$ and we are only interested in matching two moments of $\boldsymbol{X}_i$, the $\widehat{\boldsymbol{X}}_i$ can be obtained analytically. In order to simplify notation, we minimize

$$\sum_{i=1}^n(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i) \tag{2.2}$$

rather than (2.1). The weighted minimization problem can be simplified to (2.2) by making a change of variables. We demonstrate this later. Based on (2.2), the objective function is

$$\begin{aligned} &n^{-1}\textstyle\sum_{i=1}^n(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i) + \boldsymbol{\lambda}_1^T(n^{-1}\sum_{i=1}^n \boldsymbol{X}_i - \boldsymbol{W}_i) \\ &+\boldsymbol{\lambda}_2^T\mathrm{vech}(n^{-1}\textstyle\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T - \boldsymbol{W}_i\boldsymbol{W}_i^T + \Sigma_{ui}) + \boldsymbol{\lambda}_3^T(n^{-1}\sum_{i=1}^n \boldsymbol{X}_iY_i - \boldsymbol{W}_iY_i) \end{aligned} \tag{2.3}$$

where $\boldsymbol{\lambda}_1^T$ and $\boldsymbol{\lambda}_3^T$ are vectors of Lagrange multipliers and $\boldsymbol{\lambda}_2^T$ is a vector. Let I_G denote the identity matrix of dimension G. Taking the derivative with respect to $\boldsymbol{X}_i$ gives $(\boldsymbol{X}_i - \boldsymbol{W}_i) +$ $\boldsymbol{\lambda}_1 + (\{\mathrm{vech}^{-1}(\boldsymbol{\lambda}_2) + \mathrm{I}_G\}\boldsymbol{X}_i + \boldsymbol{\lambda}_3Y_i = 0$, and the solution for $\boldsymbol{X}_i$ is $\boldsymbol{X}_i = \boldsymbol{A}(\boldsymbol{W}_i - \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_3Y_i)$ where $\boldsymbol{A} = \{2\mathrm{I}_G\mathrm{vech}^{-1}(\boldsymbol{\lambda}_2) + \mathrm{I}_G\}^{-1}$, where vech^{-1} re-creates a symmetric matrix from its vech half so that if $\boldsymbol{A} = \mathrm{vech}(\boldsymbol{B})$ for symmetric matrix $\boldsymbol{B}$, then $\mathrm{vech}^{-1}(\boldsymbol{A}) = \boldsymbol{B}$. The solution for $\boldsymbol{X}_i$ depends on the unknown $\Lambda = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_3^T, \boldsymbol{\lambda}_2^T)^T$, which must be estimated to

obtain $\widehat{\boldsymbol{X}}_i$. Taking the derivative with respect to Λ provides additional equations that we can solve to obtain $\widehat{\Lambda}$.

Rather than solving for Λ directly, it is easier to solve for "coefficients" in the equation for $\boldsymbol{X}_i$. Note that the $\boldsymbol{X}_i$ have the form $\boldsymbol{A}\boldsymbol{W}_i + \boldsymbol{B} + \boldsymbol{C}Y_i$, and we solve for the "coefficients" $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ as follows:

Constraint 1: $\overline{\widehat{\boldsymbol{X}}}=\overline{\boldsymbol{W}}$ for $\overline{\widehat{\boldsymbol{X}}}=n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i$, and $\overline{\boldsymbol{W}}=n^{-1}\sum_{i=1}^n \boldsymbol{W}_i$
So $\widehat{\boldsymbol{A}}\overline{\boldsymbol{W}} + \widehat{\boldsymbol{B}} + \widehat{\boldsymbol{C}}\overline{Y}=\overline{\boldsymbol{W}}$ and $\widehat{\boldsymbol{B}} = (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} - \widehat{\boldsymbol{C}}\overline{Y}$.
Substituting this for $\widehat{\boldsymbol{B}}$ we have $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$.

Constraint 3: $n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i Y_i = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i Y_i$
Equivalently $\boldsymbol{S}_{\widehat{X}Y} = \boldsymbol{S}_{WY}$ for

$$\boldsymbol{S}_{\widehat{X}Y} = n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{X}}_i(Y_i - \overline{Y}) \text{ and } \boldsymbol{S}_{WY} = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i(Y_i - \overline{Y}).$$

Substituting $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$ implies that

$$\widehat{\boldsymbol{A}}\boldsymbol{S}_{WY} + \widehat{\boldsymbol{C}}s_Y^2 = \boldsymbol{S}_{WY}, \text{ where } s_Y^2 = n^{-1}\sum_{i=1}^n (Y_i - \overline{Y})^2.$$

This gives $\widehat{\boldsymbol{C}} = (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{S}_{WY}/s_Y^2$.

Constraint 2: $n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T = n^{-1}\sum_{i=1}^n \boldsymbol{W}_i\boldsymbol{W}_i^T - \Sigma_{ui}$
Equivalently $\boldsymbol{S}_{\widehat{X}\widehat{X}} = \boldsymbol{S}_{WW} - \Sigma_{ui}$.

$$\boldsymbol{S}_{\widehat{X}\widehat{X}} = n^{-1}\sum_{i=1}^n (\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i)(\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i)^T$$
$$\boldsymbol{S}_{WW} = n^{-1}\sum_{i=1}^n (\boldsymbol{W}_i - \overline{\boldsymbol{W}})(\boldsymbol{W}_i - \overline{\boldsymbol{W}})^T$$

Substituting $(\widehat{\boldsymbol{X}}_i - \overline{\widehat{\boldsymbol{X}}}_i) = \widehat{\boldsymbol{A}}(\boldsymbol{W}_i - \overline{\boldsymbol{W}}) + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{S}_{WY}(Y_i - \overline{Y})/s_Y^2$ we have

$$\boldsymbol{S}_{\widehat{X}\widehat{X}} = \widehat{\boldsymbol{A}}\boldsymbol{S}_{WW}\widehat{\boldsymbol{A}}^T + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{M} + \widehat{\boldsymbol{A}}\boldsymbol{M}(\mathrm{I}_G - \widehat{\boldsymbol{A}})^T + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\boldsymbol{M}\widehat{\boldsymbol{A}}^T = \boldsymbol{S}_{WW} - \Sigma_{ui},$$
where $\boldsymbol{M} = \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T/s_Y^2$.

Algebraic simplification leads to the equation

$$\widehat{\boldsymbol{A}}(\boldsymbol{S}_{WW} - \boldsymbol{M})\widehat{\boldsymbol{A}}^T = \boldsymbol{S}_{WW} - \Sigma_{ui} - \boldsymbol{M}, \text{ or}$$
$$\widehat{\boldsymbol{A}}(\boldsymbol{S}_{WW}s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T)\widehat{\boldsymbol{A}}^T = (\boldsymbol{S}_{WW} - \Sigma_{ui})s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T.$$

The matrix $\boldsymbol{V}_1 = (\boldsymbol{S}_{WW}s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T)$ is positive definite by the Cauchy Schwartz inequality. If $\boldsymbol{V}_2 = (\boldsymbol{S}_{WW} - \Sigma_{ui})s_Y^2 - \boldsymbol{S}_{WY}\boldsymbol{S}_{WY}^T$ is non-negative definite, then the equation can be solved to yield $\widehat{\boldsymbol{A}} = \boldsymbol{V}_1^{-1/2}(\boldsymbol{V}_1^{1/2}\boldsymbol{V}_2\boldsymbol{V}_1^{1/2})^{1/2}\boldsymbol{V}_1^{-1/2}$. The second matrix $\boldsymbol{V}_2$ is usually non-negative definite, unless the measurement error is very large and/or the correlation between $\boldsymbol{X}_i$ and Y_i is extremely high.

This defines our adjusted data $\widehat{\boldsymbol{X}}_i = \widehat{\boldsymbol{A}}\boldsymbol{W}_i + (\mathrm{I}_G - \widehat{\boldsymbol{A}})\overline{\boldsymbol{W}} + \widehat{\boldsymbol{C}}(Y_i - \overline{Y})$.

Now we show that the weighted minimization (2.1) can be framed in terms of (2.2). Let $\boldsymbol{X}_i^* = \Sigma_{ui}^{-1/2}\boldsymbol{X}_i$ and $\boldsymbol{W}_i^* = \Sigma_{ui}^{-1/2}\boldsymbol{W}_i$ Then $\boldsymbol{W}_i^*|\boldsymbol{X}_i^* \sim MVN(\boldsymbol{X}_i^*, \mathrm{I}_G)$, and $\sum_{i=1}^n(\boldsymbol{W}_i - \boldsymbol{X}_i)^T\Sigma_{ui}^{-1}(\boldsymbol{W}_i - \boldsymbol{X}_i) = \sum_{i=1}^n(\boldsymbol{W}_i^* - \boldsymbol{X}_i^*)^T(\boldsymbol{W}_i^* - \boldsymbol{X}_i^*)$. We can define the moment constraints based on the distribution of $\boldsymbol{W}_i^*|\boldsymbol{X}_i^*$ and perform an unweighted minimization to obtain $\widehat{\boldsymbol{X}}_i^*$. On the original scale we have $\widehat{\boldsymbol{X}}_i = \Sigma_{ui}^{1/2}\widehat{\boldsymbol{X}}_i^*$.

2.2.3 Practical Considerations

The estimators $\widehat{\boldsymbol{m}}_r$, $\widehat{\boldsymbol{m}}^*$, and $\widehat{\boldsymbol{m}}_k^V$ are not always a valid set of moments in finite samples. This is primarily a problem if the sample size n is small or the measurement error in $\boldsymbol{W}_i$ is very large. Before adjusting data, we first check that our estimators represent a valid sample variance-covariance matrix. When $M = 2$, our adjustment corresponds to matching the variance-covariance matrix of $(1, \boldsymbol{X}_i^T, \boldsymbol{V}_i^T)^T$. For simplicity, let $\boldsymbol{X}_i$ be (2×1) $(G = 2)$, and let $\boldsymbol{V}_i$ include only the response Y_i and a single error free covariate Z_i. Then we check the following condition to guarantee that we are targeting a valid set of moments:

$$\begin{bmatrix} 1 & \widehat{\boldsymbol{m}}_1^T & n^{-1}\sum_{i=1}^n Y_i & n^{-1}\sum_{i=1}^n Z_i \\ \widehat{\boldsymbol{m}}_1 & \widehat{\boldsymbol{m}}^* & (\widehat{\boldsymbol{m}}_k^V)^T & \\ n^{-1}\sum_{i=1}^n Y_i & \widehat{\boldsymbol{m}}_k^V & n^{-1}\sum_{i=1}^n Y_i^2 & n^{-1}\sum_{i=1}^n Y_iZ_i \\ n^{-1}\sum_{i=1}^n Z_i & & n^{-1}\sum_{i=1}^n Y_iZ_i & n^{-1}\sum_{i=1}^n Z_i^2 \end{bmatrix} \geq 0.$$

For $M = 4$, we match higher order moments for the G components of $\boldsymbol{X}_i$. In this case, we also verify the inequality

$$\begin{bmatrix} 1 & \widehat{\boldsymbol{m}}_{1g} & \widehat{\boldsymbol{m}}_{2g} \\ \widehat{\boldsymbol{m}}_{1g} & \widehat{\boldsymbol{m}}_{2g} & \widehat{\boldsymbol{m}}_{3g} \\ \widehat{\boldsymbol{m}}_{2g} & \widehat{\boldsymbol{m}}_{3g} & \widehat{\boldsymbol{m}}_{4g} \end{bmatrix} \geq 0$$

for all $g = 1, \ldots, G$.

In general, we do not have an analytical solution for the $\widehat{\boldsymbol{X}}_i$, and we solve for these numerically, as described in Section 2.1. Even for a valid sequence of moments, numerical problems occur for some data sets, and a solution is not obtained. In our simulations (see Section 2.4), this happens for about 10% of data sets, even when $\boldsymbol{X}_i$ is only 2 dimensional. In the following section, we suggest an alternative method of obtaining adjusted data that generally avoids numerical problems.

2.3 Alternative Implementation of Adjustment

Our goal is to obtain adjusted data $\widehat{\boldsymbol{X}}_i$ that have unbiased moments for the corresponding moments of $\boldsymbol{X}_i$. In Section 2.1, we do this by imposing constraints on the moments and minimizing the "distance" between our observed data $\boldsymbol{W}_i$ and the adjusted data $\widehat{\boldsymbol{X}}_i$, as measured by (2.1). This measure incorporates all of the G components of $\widehat{\boldsymbol{X}}_i$ and their cross products and weights the components inversely according to their measurement error. A simpler alternative minimizes the unweighted norm (2.2). This is more crude and does not account for differential measurement error nor correlated measurement errors.

Even (2.2) can be difficult to minimize numerically, since nG adjusted data points are obtained, and many constraints may be involved. This approach can be approximated by performing a univariate adjustment sequentially; adjusting $(W_{1g}, \ldots, W_{ng})$ for each $g = 1, \ldots, G$. At each step $(\widehat{X}_{1g}, \ldots, \widehat{X}_{ng})$, are obtained such that $E(n^{-1}\sum_{i=1}^{n}\widehat{X}_{ig}^r) = E(X_{ig}^r) = m_{rg}$, for $r = 1, \ldots, M$, and $E(n^{-1}\sum_{i=1}^{n}\widehat{X}_{ig}V_{ik}) = E(X_{ig}V_{ik}) = m_{kg}^V$, and we account for the correlation between X_{ig} and $X_{ig'}$ by requiring that $E(n^{-1}\sum_{i=1}^{n}\widehat{X}_{ig}\widehat{X}_{ig'}) = E(X_{ig}X_{ig'}) = m_{gg'}^*$ for all $g \neq g'$.

The estimators $\widehat{\boldsymbol{m}}_r$, $\widehat{\boldsymbol{m}}^*$, and $\widehat{\boldsymbol{m}}_k^V$ are obtained exactly as in Section 2.1, based on the distribution $\boldsymbol{W}_i|\boldsymbol{X}_i \sim MVN(\boldsymbol{X}_i, \Sigma_{ui})$. The adjustment is performed sequentially beginning with $g = 1$. For each g, we obtain $(\widehat{X}_{1g}, \ldots, \widehat{X}_{ng})$ by minimizing $\sum_{i=1}^{n}(W_{ig} - X_{ig})^2$ with the following constraints imposed by Lagrange multipliers: $n^{-1}\sum_{i=1}^{n}X_{ig}^r = \widehat{m}_{rg}$, $n^{-1}\sum_{i=1}^{n}X_{ig}V_{ik} = \widehat{m}_{kg}^V$, and $n^{-1}\sum_{i=1}^{n}X_{ig}X_{ig'} = \widehat{m}_{gg'}^*$ for $g' < g$.

At each step, a separate objective function is defined, and Newton Raphson is used to solve for $(\widehat{X}_{1g}, \ldots, \widehat{X}_{ng})$. At the first step, adjusted data are not available, so we can not

impose constraints to match $\widehat{m}^*_{1g}$. The objective function is

$$n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_{i1}-X_{i1})^2+\sum_{r=1}^{M}\frac{\lambda_r}{r}(n^{-1}\sum_{i=1}^{n}X_{i1}^r-\widehat{m}_{r1})+\sum_{k=1}^{K}\lambda_k^V(n^{-1}\sum_{i=1}^{n}X_{i1}V_{ik}-\widehat{m}_{k1}^V),$$

where λ_r and λ_k^V are Lagrange multipliers. The adjusted data, $(\widehat{X}_{11},\ldots,\widehat{X}_{n1})$, can be used to adjust W_{i2} so that $(\widehat{X}_{12},\ldots,\widehat{X}_{n2})$ have $E(n^{-1}\sum_{i=1}^{n}\widehat{X}_{i1}\widehat{X}_{i2})=E(X_{i1}X_{i2})$. The second objective function, which incorporates this additional constraint, is

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_{i2}-X_{i2})^2+\sum_{r=1}^{M}\frac{\lambda_r}{r}(n^{-1}\sum_{i=1}^{n}X_{i2}^r-\widehat{m}_{r2})+\sum_{k=1}^{K}\lambda_k^V(n^{-1}\sum_{i=1}^{n}X_{i2}V_{ik}-\widehat{m}_{k2}^V)\\+\lambda_1^*(n^{-1}\sum_{i=1}^{n}\widehat{X}_{i1}X_{i2}-\widehat{m}_{12}^*),\end{aligned}$$

where λ_r, λ_k^V and λ_1^* are Lagrange multipliers. This process is continued at each step using the previously adjusted data. So $(\widehat{X}_{1G},\ldots,\widehat{X}_{nG})$ have $E(n^{-1}\sum_{i=1}^{n}\widehat{X}_{ig}\widehat{X}_{iG})=E(X_{ig}X_{iG})$, for $g=1,\ldots,(G-1)$. The final objective function is

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_{ig}-X_{ig})^2+\sum_{r=1}^{M}\frac{\lambda_r}{r}(n^{-1}\sum_{i=1}^{n}X_{ig}^r-\widehat{m}_{rg})+\sum_{k=1}^{K}\lambda_k^V(n^{-1}\sum_{i=1}^{n}X_{ig}V_{ik}-\widehat{m}_{kg}^V)\\+\sum_{g=1}^{G-1}\lambda_g^*(n^{-1}\sum_{i=1}^{n}\widehat{X}_{ig}X_{iG}-\widehat{m}_{gG}^*),\end{aligned}$$

where λ_r, λ_k^V and λ_g^* are Lagrange multipliers.

The sequential approach imposes the same moment constraints defined in Section 2.2. However, the full set of constraints is not implemented at every stage of adjustment. The sequential nature requires that additional constraints be added at each step. This suggests that the order of adjustment should be carefully considered. We recommend that W_{ig} be adjusted in the order of least measurement error to greatest measurement error. This implies that the variables with greatest measurement error are subject to the most constraints and adjusted most. In the following section, we compare the sequential approach to joint adjustment based on minimizing (2.1) or (2.2).

2.4 Simulations in Logistic Regression

We performed simulations to compare parameter estimation in logistic regression based on each these methods of implementing MAI. We let $\boldsymbol{X}_i=(X_{i1},X_{i2})$, and a re-

sponse Y_i is observed with probability $P(Y_i = 1|X_{i1}, X_{i2}) = F(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})$, where $F(v) = 1 + \exp(-v)^{-1}$. Both X_{i1} and X_{i2} have a chi-square distribution with four degrees of freedom, standardized to have mean zero and variance one, with $\text{Corr}(X_{i1}, X_{i2}) \approx 0.5$. The coefficients $\boldsymbol{\beta}$=$(\beta_0, \beta_1, \beta_2)$ are $(1.5, 0.5, 0.5)$. This corresponds to substantial non-linearity over the range of $\boldsymbol{X}_i$.

We generated $B = 500$ simulated data sets, each having sample size of $n = 1000$. In place of $\boldsymbol{X}_i$, we observe $\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{U}_i$ for $\boldsymbol{U}_i \sim N(\boldsymbol{0}, \Sigma_u)$ and $\Sigma_u = \text{vech}\{1, 0.5(0.2)^{1/2}, 0.2\}$. Because the off-diagonal of Σ_u is non-zero, the measurement error in these two variables is correlated. The measurement error in X_{i1} is large and corresponds to a reliability ratio of 0.5. The measurement error in X_{i2} is smaller with a reliability ratio of 0.83. Joint minimization of (2.1) weights components inversely by their measurement error variance and may provide better adjustment by accounting for these features of the measurement error. We perform sequential adjustment, first for X_{i2} then X_{i1}, adjusting the variable with least measurement error first. For each of the three methods of data adjustment, the estimated $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are compared in Table 2.1.

Table 2.1: *Estimation of $\boldsymbol{\beta}$ for $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$. True value of $\boldsymbol{\beta} = (1.5, 0.5, 0.5)$; B, bias; SD, standard deviation; MSE-R, $MSE_W/MSE_{\widehat{\boldsymbol{X}}}$. Adjusted data $\widehat{\boldsymbol{X}}$: Seq, sequential adjustment; J, joint adjustment (2.2); JW, joint adjustment (2.1).*

		Matching 2 moments			Matching 4 moments		
Stat.	W	$\widehat{\boldsymbol{X}}_{Seq}$	$\widehat{\boldsymbol{X}}_{J}$	$\widehat{\boldsymbol{X}}_{JW}$	$\widehat{\boldsymbol{X}}_{Seq}$	$\widehat{\boldsymbol{X}}_{J}$	$\widehat{\boldsymbol{X}}_{JW}$
		Estimation of β_1; true value is 0.5					
B	-0.35	-0.09	-0.09	-0.09	0.04	0.04	0.06
SD	0.06	0.14	0.14	0.14	0.23	0.22	0.22
MSE-R	1.00	4.53	4.54	4.65	2.47	2.49	2.39
		Estimation of β_2; true value is 0.5					
B	-0.05	-0.04	-0.04	-0.03	-0.01	-0.01	0.02
SD	0.09	0.12	0.12	0.12	0.14	0.14	0.13
MSE-R	1.00	0.72	0.72	0.74	0.57	0.26	0.62

We see no practical difference in the coefficient estimation provided by these three methods of adjusting data (Table 2.1). The bias and standard deviation of estimators are virtually identical across the three methods. There is no apparent loss from performing sequential adjustment. There is a noticeable difference in estimator bias and standard devi-

ation depending on whether two or four moments are matched. It appears that the impact of adjustment on logistic model parameter estimation comes primarily from the moment constraints rather than from differences in the minimization. We investigate alternative scenarios in Appendix B and observe similar results. However, the various implementations of MAI may differ in situations that we did not consider. The weighted joint minimization (2.1) has the benefit of accounting for differential measurement error and correlated measurement error, which could be important in many other applications. There is no benefit numerically, nor theoretically, from performing unweighted joint minimization (2.2), so we do not use this approach. The sequential approach runs more quickly and encounters fewer numerical problems, so we favor this approach when computing time is limited.

In the previous comparisons, we implement sequential adjustment in order of least to greatest measurement error. In Appendix B, we address the impact of adjustment order on logistic regression parameter estimation. In the cases that we considered, the order of adjustment had a negligible impact of the parameter estimates. We have not observed a case where the order of adjustment was important in logistic regression.

Chapter 3

Coefficient Estimation in Logistic Regression Where Covariates Include Variance Components From a Model for Longitudinal or Repeated Measurements

3.1 Introduction

Dichotomous health outcomes are frequently modeled by logistic regression, where the covariates may include subject-specific random effects and/or residual variances that characterize the longitudinal features of a continuous response. In this chapter we investigate the relationship between variability in longitudinal blood pressure and short-term mortality in hemodialysis patients. In healthy patients, blood pressure variability is often measured using 24-hour ambulatory monitoring, where patients wear a device to measure blood pressure at regular intervals throughout a typical 24 hour period. Describing variability in hemodialysis patients is more complex because their treatment may induce fluctuations over longer periods than 24 hours. For this reason, in a retrospective analysis of the Accelerated Mortality on Renal Replacement (AMoRR) cohort, Brunelli et al.

(2008) measured blood pressure variability over a three month period of dialysis treatment (91 to 180 days). Blood pressure was not measured until 91 days in order to allow subjects to "acclimate" to dialysis therapy. Patients underwent hemodialysis three times per week (typically), and blood pressure was measured and recorded immediately prior to each treatment. Afterwards, patient follow-up was conducted to assess mortality (181 to 365 days). Thus, the focus was on long-term variability and mortality among patients who had survived at least 180 days after the introduction of hemodialysis treatment. The AMoRR study includes adults (>18 years) who were incident to hemodialysis between June 2004 and August 2005. Using these same data[1], we wish to evaluate the relationship between short-term mortality, occurring before 180 days, and blood pressure variability. For the analysis presented here, blood pressure measurements were obtained from 91 to 120 days, and patient follow-up began at 121 days.

Individual subject longitudinal blood pressure trajectories are often well described by a simple linear random coefficient model with subject-specific intercept and slope. Variability in blood pressure may be correlated with the slope and intercept that describe a subject's longitudinal trajectory (Brunelli et al., 2008). In this context, it is also plausible that slopes and intercepts may be associated with mortality. To assess the independent effect of variability, we model short-term mortality by logistic regression with longitudinal blood pressure slope, intercept, and intra-subject variance parameters as covariates. These longitudinal parameters must be estimated from the blood pressure measurements taken over 91 to 120 days. The average number of replicates observed during this period is 12 with standard deviation of 1.5. Consequently, the subject-specific estimators are subject to substantial variability about the corresponding unknown parameters. This uncertainty of estimation introduces a problem of measurement error that must be addressed in fitting the primary logistic regression model.

This problem has been thoroughly studied for primary outcome models that include subject-specific random effects like longitudinal slope and intercept as covariates. Wang, Wang, and Wang (2000) characterize these "joint models" and show that a naive approach that substitutes ordinary least squares estimators, based on each subject's longitudinal data, for the random effects, leads to biased inference on parameters in the primary

[1]Our collaborator Steve Brunelli had sole access to the data. R code for this analysis was developed on simulated data.

model. To adjust for this measurement error bias, they propose a variety of structural methods including regression calibration (Carroll et al., 2006) and a psuedo-expected estimating equation approach. These methods rely on normality of the underlying random effects. In order to avoid parametric assumptions on the random effects parameters, Li, Zhang and Davidian (2004) introduce a conditional score approach for generalized linear primary models, which includes logistic regression as a special case. They follow the strategy of Stefanski and Carroll (1987) and derive unbiased estimating equations by conditioning on sufficient statistics for the unknown random effects parameters. The common feature in these joint models is that the estimated random effect parameters have a normal distribution, conditional on the true random effects. In other words, the measurement error is normally distributed. This is suitable for joint modeling applications where interest focuses on longitudinal slope and intercept parameters.

Many recent studies define joint models where the primary outcome model depends on subject-specific variance components. Yang et al. (2007) identified an association between hemoglobin variability and mortality in patients with end stage renal disease, even after controlling for the absolute level and temporal trend in hemoglobin. Havlik et al. (2002) reported increased risk of late-life white matter brain lesions for patients with higher variability in systolic blood pressure during mid-life. Similar work is abundant (Iribarren et al., 1995; Grove et al., 1997; Yang et al., 2007; Brunelli et al., 2008). These outcome models often include both random effects and variance components to describe the longitudinal data. Coefficient estimation should account for both sources of measurement error, which are correlated. Moreover, the measurement error in the variance estimators has a chi-square, rather than normal, distribution.

It seems that the problem of measurement error is not well appreciated in this context, as the previous studies all take a naive approach to model fitting. This is not surprising as, to the best of our knowledge, there is very little methodology to address this problem. One exception is a maximum likelihood method proposed by Lyles et al. (1999). These authors note that variance estimators have high variability relative to estimators of a mean from a similar sample size, so the problem of measurement error may be even more pronounced in this scenario. They study a continuous outcome, decline in CD4 counts after seroconversion, modeled by linear regression that depends on both the subject-specific mean and variance of longitudinal CD4 counts measured before seroconversion. They obtain a full

likelihood for the observed data by assuming that the subject-specific means are normally distributed and independent of log-normally distributed subject-specific variances. Based on these assumptions, they fit the model by maximum likelihood as well as a regression calibration type approach. In many applications, the investigator may not have confidence in parametric assumptions about the true longitudinal parameters and variances. This suggests that a functional approach that does not require these assumptions is desirable.

In this chapter, we introduce and compare three functional approaches when the outcome of interest is modeled by logistic regression. In Section 3.2 we define the joint model of interest, and methods for fitting the model are derived in Section 3.3. The first is a conditional score method, which is limited to a specific formulation of the logistic regression model. The second is a corrected score method. The exact implementation of this method is complicated and computationally intensive. A slight modification gives an approximately unbiased score function that avoids excessive computation. This approach also requires a specific formulation of the logistic regression model and is not easily adapted to alternative versions. Finally, we describe a modification of MAI (Chapter 1), which accommodates measurement error due to estimation of variance parameters. MAI is flexible and easy to implement for many complicated logistic regression models. We compare these approaches by simulation in Section 3.4. In Section 3.5, we present the analysis of short-term blood pressure variability and mortality in hemodialysis patients from the AMoRR study. Some conclusions and extensions are presented in Section 3.6.

3.2 Joint Model

Data are observed for subjects $i = 1, \ldots, n$. The binary outcome of interest is denoted Y_i, and baseline error-free covariates are included in an $(p \times 1)$ vector $\boldsymbol{Z}_i$. In addition, repeated measurements of a continuous response $\boldsymbol{S}_i = (S_{i1}, \ldots, S_{ir_i})^T$ are collected. These may be longitudinal measurements collected at times $t_{i1}, \ldots, t_{ir_i}$. We formulate the model for longitudinal data, although replicate measurements can be accommodated as a special case. The $\boldsymbol{S}_i$ can be described by a linear mixed model given by $\boldsymbol{S}_i = \boldsymbol{D}_i\boldsymbol{\gamma}_i + \boldsymbol{U}_i$, where $\boldsymbol{D}_i$ is the $(r_i \times q)$ design matrix, and $\boldsymbol{U}_i = (U_{it}, \ldots, U_{ir_i})^T$ are the within-subject errors such that $\boldsymbol{U}_i \sim N(\boldsymbol{0}, \sigma_i^2 \mathrm{I}_{r_i})$. In particular, we focus on the linear random coefficient model where $\boldsymbol{D}_i$ has rows $(1, t_{ij})$, $j = 1, \ldots, r_i$, and $\boldsymbol{\gamma}_i = (\gamma_{0i}, \gamma_{1i})^T$. The primary endpoint

Y_i is related to the baseline covariates $\boldsymbol{Z}_i$ and longitudinal parameters $(\boldsymbol{\gamma}_i, \sigma_i^2)$ through the logistic regression model

$$P(Y_i = 1|\boldsymbol{\gamma}_i, \sigma_i^2, \boldsymbol{Z}_i; \boldsymbol{\beta}) = F\{\beta_0 + \boldsymbol{\gamma}_i^T \boldsymbol{\beta}_\gamma + T(\sigma_i^2)\beta_\sigma + \boldsymbol{Z}_i^T \boldsymbol{\beta}_Z\}, \tag{3.1}$$

where $F(t) = 1/(1 + e^{-t})$; $\boldsymbol{\beta}_\gamma = (\beta_{\gamma_0}, \beta_{\gamma_1})^T$; $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_\gamma^T, \beta_\sigma, \boldsymbol{\beta}_Z^T)^T$; and the function $T(\cdot)$ represents a transformation of the variance, typically the square root or logarithm.

In this chapter, our interest focuses on estimation of β_σ, although all of the components of $\boldsymbol{\beta}$ are important in many applications. The naive approach to estimating $\boldsymbol{\beta}$ is to replace the longitudinal model parameters $(\boldsymbol{\gamma}_i, \sigma_i^2)$ with estimates $(\widehat{\boldsymbol{\gamma}}_i, \widehat{\sigma}_i^2)$, obtained, for example, by individual specific ordinary least squares, and fit the logistic regression model for Y_i as if these parameters were known. In the following section, we develop estimators for $\boldsymbol{\beta}$ that account for the estimation of $(\boldsymbol{\gamma}_i, \sigma_i^2)$. These methods require that we know the sampling distributions of $\widehat{\boldsymbol{\gamma}}_i|\boldsymbol{\gamma}_i$, and $\widehat{\sigma}_i^2|\sigma_i^2$. We obtain $\widehat{\boldsymbol{\gamma}}_i$ by ordinary least squares on the longitudinal data for the i^{th} subject so that $\widehat{\boldsymbol{\gamma}}_i = (\boldsymbol{D}_i^T \boldsymbol{D}_i)^{-1} \boldsymbol{D}_i^T \boldsymbol{S}_i$, and $\widehat{\sigma}_i^2 = \nu_i^{-1} \sum_{j=1}^{r_i} (S_{ij} - \widehat{S}_{ij})^2$ for $\nu_i = r_i - q$ with $q = 2$. For these estimators we have $\widehat{\boldsymbol{\gamma}}_i|\boldsymbol{\gamma}_i \sim N\{\boldsymbol{\gamma}_i, \sigma_i^2(\boldsymbol{D}_i^T \boldsymbol{D}_i)^{-1}\}$ and $\widehat{\sigma}_i^2|\sigma_i^2 \sim (\sigma_i^2/\nu_i)\chi_{\nu_i}^2$. The measurement error in $\widehat{\boldsymbol{\gamma}}_i$ is normally distributed and in $\widehat{\sigma}_i^2$ is proportional to a chi-square distribution. The vector of estimators $\widehat{\boldsymbol{\gamma}}_i$ is conditionally independent of $\widehat{\sigma}_i^2$ given $(\boldsymbol{\gamma}_i, \sigma_i^2)$. In addition, we make the common *surrogacy* assumption that Y_i and $\boldsymbol{Z}_i$ are independent of $(\widehat{\boldsymbol{\gamma}}_i, \widehat{\sigma}_i^2)$ conditional on $(\boldsymbol{\gamma}_i, \sigma_i^2)$ (Carroll et al. 2006, Section 2.5). The following estimators require no assumptions on the distributions of the underlying parameters $\boldsymbol{\gamma}_i$ and σ_i^2, as these are typically unknown in practice.

3.3 Proposed Methods

3.3.1 Conditional Score

The estimating equations for logistic regression model parameters, based on the maximum likelihood (ML) score function, are biased estimating equations when estimators are naively substituted for true covariates. When the measurement error in covariates is normally distributed, an unbiased score function can be obtained by conditioning on sufficient statistics for the unknown random effects parameters (Stefanski and Carroll, 1987; Wang and Huang, 2001; Li, Zhang and Davidian, 2004). This method is widely used to obtain

consistent estimators for the logistic regression model parameters in the presence of measurement error. The sufficient statistics are identified by factorization of the observed data likelihood. This involves the measurement error distribution, which, in the present case, is not normally distributed but chi-square. We account for this, and follow the approach of Stefanski and Carroll (1987).

We start with the special case of Model (3.1) defined by $P(Y=1|\sigma^2;\boldsymbol{\beta}) = F\{\beta_0 + \beta_\sigma(1/\sigma^2)\}$ for $\boldsymbol{\beta} = (\beta_0, \beta_\sigma)^T$. This model describes the relationship between the probability of an event and the inverse variance, which is often referred to as precision in the Bayesian statistics literature. In the exponential family form, the corresponding density of Y for a given true predictor σ^2 is

$$f_Y(y;\sigma^2,\boldsymbol{\beta}) = \exp\left[y\left(\beta_0 + \beta_\sigma\frac{1}{\sigma^2}\right) + \log\left\{1 - F\left(\beta_0 + \beta_\sigma\frac{1}{\sigma^2}\right)\right\}\right]. \tag{3.2}$$

The observed variance estimates $\widehat{\sigma}^2$ have conditional density

$$f_{\widehat{\sigma}^2}(\widehat{\sigma}^2;\sigma^2) = \frac{(\nu/\sigma^2)}{2^{\nu/2}\Gamma(\nu/2)}\left(\frac{\nu\widehat{\sigma}^2}{\sigma^2}\right)^{\nu/2-1}\exp\left(\frac{-\nu\widehat{\sigma}^2}{2\sigma^2}\right)I(\widehat{\sigma}^2 > 0). \tag{3.3}$$

Using the surrogacy assumption that Y and $\widehat{\sigma}^2$ are independent given σ^2, the joint density of the observed pair $(Y, \widehat{\sigma}^2)$ given σ^2 and $\boldsymbol{\beta}$ is the product of (C.1) and (C.2),

$$\begin{aligned} f_{Y,\widehat{\sigma}^2}(y,\widehat{\sigma}^2;\sigma^2,\boldsymbol{\beta}) = {} & \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\exp\left\{y\beta_0 + (\nu/2)\log(\nu) + (\nu/2-1)\log(\widehat{\sigma}^2)\right\} \times \\ & \exp\left[\left(y\beta_\sigma - \frac{\nu}{2}\widehat{\sigma}^2\right)\left(\frac{1}{\sigma^2}\right) + \log\left\{1 - F\left(\beta_0 + \beta_\sigma\frac{1}{\sigma^2}\right)\right\} + (\nu/2)\log\left(\frac{1}{\sigma^2}\right)\right]. \end{aligned} \tag{3.4}$$

If σ^2 is viewed as a parameter in density (3.4), and β_σ is treated as fixed and known, then the statistic $\Delta = \Delta(Y, \widehat{\sigma}^2, \beta_\sigma) = \widehat{\sigma}^2 - Y(2/\nu)\beta_\sigma$ is complete and sufficient for σ^2. Thus the conditional distribution of $Y|\Delta$ does not depend on the unknown σ^2. Let $f_{Y|\Delta}(y|\delta;\boldsymbol{\beta})$ denote the condititional distribution of Y given $\Delta = \delta$. Treat δ as a fixed conditioning argument in the following calculations. The Jacobian of the transformation from $(Y, \widehat{\sigma}^2)$ to $(Y, \widehat{\sigma}^2 - Y(2/\nu)\beta_\sigma)$ has determinant one. Thus, $P(Y=1, \Delta=\delta|\sigma^2) = P(Y=1, \widehat{\sigma}^2 = \delta + (2/\nu)\beta_\sigma|\sigma^2)$. Using the surrogacy assumption we have $P\{Y=1, \widehat{\sigma}^2 = \delta + (2/\nu)\beta_\sigma|\sigma^2\}$ $= P(Y=1|\sigma^2)P\{\widehat{\sigma}^2 = \delta + (2/\nu)\beta_\sigma|\sigma^2\} = F\{\beta_0 + \beta_\sigma(1/\sigma^2)\}f_{\widehat{\sigma}^2}\{\delta + (2/\nu)\beta_\sigma;\sigma^2\}$. Similarly, $P(Y=0, \Delta=\delta|\sigma^2) = P(Y=0, \widehat{\sigma}^2 = \delta|\sigma^2) = \left[1 - F\{\beta_0 + \beta_\sigma(1/\sigma^2)\}\right]f_{\widehat{\sigma}^2}(\delta;\sigma^2)$. Using

$P(\Delta=\delta|\sigma^2)$ =$P(Y=1,\Delta=\delta|\sigma^2)$+ $P(Y=0,\Delta=\delta|\sigma^2)$ we have

$$\begin{aligned}
P(Y=1|\Delta=\delta) &= \frac{1}{1+\dfrac{1-F\{\beta_0+\beta_\sigma(1/\sigma^2)\}}{F\{\beta_0+\beta_\sigma(1/\sigma^2)\}}\dfrac{f_{\widehat{\sigma}^2}(\delta;\sigma^2)}{f_{\widehat{\sigma}^2}\{\delta+(2/\nu)\beta_\sigma;\sigma^2\}}} \\
&= \frac{1}{1+\exp\left\{-\left(\beta_0+\beta_\sigma\frac{1}{\sigma^2}\right)\right\}\left[\dfrac{f_{\widehat{\sigma}^2}(\delta;\sigma^2)}{f_{\widehat{\sigma}^2}\{\delta+(2/\nu)\beta_\sigma;\sigma^2\}}\right]} \\
&= \begin{cases} 0, & \delta>0, \delta+2\beta_\sigma/\nu<0 \\ Q(\delta)=\dfrac{1}{1+\exp^{-\beta_0}\left(1-2\beta_\sigma/\nu\delta\right)^{1-\nu/2}}, & \delta>0, \delta+2\beta_\sigma/\nu>0 \\ 1, & \delta<0. \end{cases}
\end{aligned}$$

This does not depend on σ^2, and the conditional density is

$$f_{Y|\Delta}(y|\delta;\boldsymbol{\beta}) = \begin{cases} I(y=0), & \delta>0, \delta+2\beta_\sigma/\nu<0 \\ Q(\delta)^y\{1-Q(\delta)\}^{1-y}, & \delta>0, \delta+2\beta_\sigma/\nu>0 \\ I(y=1), & \delta<0. \end{cases} \tag{3.5}$$

Based on this density, and with $\boldsymbol{\beta}=(\beta_0,\beta_\sigma)^T$, we obtain a score function as $\Psi(y,\widehat{\sigma}^2,\boldsymbol{\beta})$=$\partial/\partial\boldsymbol{\beta}\{\log f_{Y|\Delta}(y|\delta;\boldsymbol{\beta})\}$ evaluated at $\delta=\widehat{\sigma}^2-y(2/\nu)\beta_\sigma$, so that

$$\Psi(y,\widehat{\sigma}^2,\boldsymbol{\beta}) = \left[\{y-Q(\delta)\}\begin{pmatrix}1\\ \beta^*\end{pmatrix} I\left(\delta>0,\delta+2\beta_\sigma/\nu>0\right)\right]_{\delta=\widehat{\sigma}^2-y(2/\nu)\beta_\sigma}, \tag{3.6}$$

where β^*=$(\nu/2-1)/(\beta_\sigma+\nu\delta)/2$. This is unbiased for $\boldsymbol{\beta}$ because $E\left\{\Psi(Y,\widehat{\sigma}^2,\boldsymbol{\beta})\right\}=$ $E\left[E\left\{\Psi(Y,\widehat{\sigma}^2,\boldsymbol{\beta})|\Delta\right\}\right]=0$.

In applications, the distribution of $1/\sigma_i^2$ can be extremely skewed (Figure 3.1), and therefore inclusion of $1/\sigma_i^2$ as a linear term in the logistic model might be questioned. In addition, the problem becomes quite complicated when the model involves other mismeasured covariates. In Appendix C.1, we demonstrate another specific case where the conditional score approach may be developed, namely in the model $P(Y=1|\sigma^2;\boldsymbol{\beta})=$ $F\{\beta_0+\beta_{\gamma_0}\left(\gamma_0/\sigma^2\right)+\beta_\sigma(1/\sigma^2)\}$ for $\boldsymbol{\beta}=(\beta_0,\beta_{\gamma_0},\beta_\sigma)$. In practice, the parameters in this model may be difficult to interpret, even if the model provides a reasonable fit to the data. We do not pursue this method any further, though additional discussion is included in Appendix C.1.

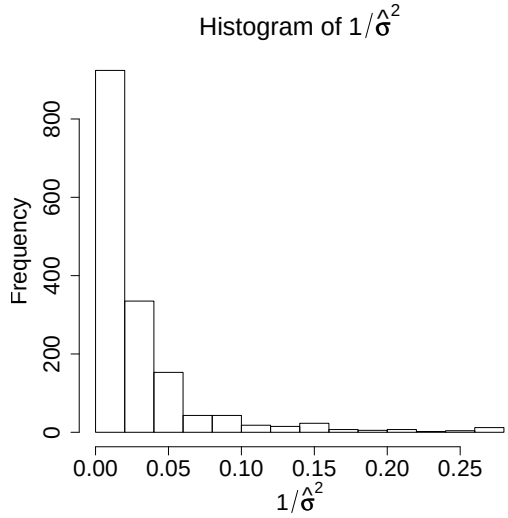

Figure 3.1: *Subject-specific variances in systolic blood pressure, estimated from four replicates (Exam 1-4) on* $n = 1,615$ *subjects in the Framingham Heart Study data set (Carroll et al., 2006).*

3.3.2 Corrected Score

The conditional score method is not very general, as it requires a logistic regression model of questionable practical utility. In order to accommodate an alternative specification of the regression model, we now consider a corrected score approach. Nakamura (1990) defined a corrected score function as "one whose expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true independent variables." Nakamura (1990) derived corrected score functions for certain generalized linear models involving covariates measured with error, where the "usual" score was the derivative of the log likelihood. In general, a correction of the likelihood score does not exist for logistic regression (Stefanski, 1989; Huang and Wang, 2001). Huang and Wang (2001) define an alternative to the likelihood score function that is unbiased under the logistic regression model in the absence of measurement error, and they show how this can be corrected for measurement error. In addition to applications with standard mismeasured data, the corrected score approach has been adopted for joint modeling (Wang, 2006). Thus, it is a natural strategy to consider in the present problem.

Let $T(\sigma^2) = \log(\sigma^2)$ define the transformation of the subject-specific variance.

This transformation is often applied in the literature (Lyles et al., 1999) because it reduces skewness in the distribution of the subject-specific variances. It is also convenient for the derivation of a score function. We begin with the score equation that is obtained by maximum likelihood when all of the covariates are known and not estimated, $\Psi_{ML}(Y, \boldsymbol{\gamma}^T, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta}) = \left\{Y - 1/(1+e^{-t})\right\}\left\{1, \boldsymbol{\gamma}, \log(\sigma^2), \boldsymbol{Z}^T\right\}^T$, for $t = \beta_0 + \boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \log(\sigma^2)\beta_\sigma + \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}}$. This is unbiased and has expectation $\mathbf{0}$ at the true value of $\boldsymbol{\beta}$ under model (3.1). If we naively substitute estimators, $\Psi_{ML}(Y, \widehat{\boldsymbol{\gamma}}^T, \widehat{\sigma}^2, \boldsymbol{Z}, \boldsymbol{\beta})$ is not unbiased. We want to assess the bias and correct for it. This is problematic for $\Psi_{ML}(Y, \boldsymbol{\gamma}^T, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta})$, but much simpler for a modified score equation. Following the idea of Huang and Wang (2001), we multiply Ψ_{ML} by $(1+e^{-t})$ to obtain the estimating function $\Psi(Y, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta}) = (1+e^{-t})\Psi_{ML}(Y, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta}) = \left\{Y(1+e^{-t}) - 1\right\}\left\{1, \boldsymbol{\gamma}^T, \log(\sigma^2), \boldsymbol{Z}^T\right\}^T$. Like the ML score, $\Psi(Y, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta})$ is unbiased when $(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z})$ are known because
$E\left\{Y(1+e^{-t}) - 1 | \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta}\right\} = (1+e^{-t})/(1+e^{-t}) - 1 = 0.$

When $\boldsymbol{\gamma}^T$ and σ^2 are estimated, $\Psi(Y, \widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{Z}, \boldsymbol{\beta})$ is not unbiased. To account for this, we now define a corrected score that has expectation $\Psi(Y, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta})$, with respect to the measurement error distribution. Begin with the expanded version of $\Psi(Y, \boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}, \boldsymbol{\beta})$,

$$\begin{bmatrix} (Y-1) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma} \\ (Y-1)\boldsymbol{\gamma} + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma} \\ (Y-1)\log(\sigma^2) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)\log(\sigma^2)(\sigma^2)^{-\beta_\sigma} \\ (Y-1)\boldsymbol{Z} + Y\boldsymbol{Z}\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma} \end{bmatrix} =$$

$$\begin{bmatrix} (Y-1) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})E_1(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) \\ (Y-1)\boldsymbol{\gamma} + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})E_2(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) \\ (Y-1)\log(\sigma^2) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})E_3(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) \\ (Y-1)\boldsymbol{Z} + Y\boldsymbol{Z}\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})E_1(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) \end{bmatrix},$$

where $E_1(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) = \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma}$, $E_2(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) = \boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma}$, and $E_3(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) = \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. We can substitute unbiased estimators $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\log(\sigma^2)}$ for $\boldsymbol{\gamma}$ and $\log(\sigma^2)$, respectively. In addition, let $\widehat{E}_t(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})$ be unbiased estimators such that $E\{\widehat{E}_t(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta}) | \boldsymbol{\gamma}, \sigma^2\} = E_t(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta})$ for $t = 1, 2, 3$. Substituting the estimators

for unknown quantities we obtain,

$$\Psi_*(Y,\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{Z},\boldsymbol{\beta}) = \begin{bmatrix} (Y-1) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\widehat{E}_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) \\ (Y-1)\widehat{\boldsymbol{\gamma}} + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\widehat{E}_2(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) \\ (Y-1)\widehat{\log}(\sigma^2) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\widehat{E}_3(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) \\ (Y-1)\boldsymbol{Z} + Y\boldsymbol{Z}\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})\widehat{E}_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) \end{bmatrix}.$$

$\Psi_*(Y,\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{Z},\boldsymbol{\beta})$ is a corrected score, because $E\{\Psi_*(Y,\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{Z},\boldsymbol{\beta})|\boldsymbol{\gamma},\sigma^2,Y\} = \Psi(Y,\boldsymbol{\gamma},\sigma^2,\boldsymbol{Z},\boldsymbol{\beta})$, where the expectation is with respect to the measurement error distribution and the surrogacy assumption holds, i.e. Y is independent of $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\sigma}^2$ given $\boldsymbol{\gamma}$ and σ^2.

In Appendix C.3, we suggest how such estimators $\widehat{E}_1$, $\widehat{E}_2$, and $\widehat{E}_3$, can be defined. The implementation of this correction is quite complicated, and preliminary simulations show that a much simpler approximation works well and is practically unbiased. Here, we derive an Approximately Corrected Score (ACS), and we use the approximation for the remainder of this chapter.

An approximately corrected score function is based on estimators for the quantities $\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$, $\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$, $(\sigma^2)^{-\beta_\sigma}$, and $\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. The estimators, derived in Appendix C.2, are

$$\begin{aligned}
&T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) = \exp\left\{-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma - \boldsymbol{\beta}_\gamma^T\widehat{\sigma}^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{\beta}_\gamma/2\right\}, && E(T_1|\boldsymbol{\gamma},\sigma^2) \cong \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)\\
&T_2(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) = \left\{\widehat{\boldsymbol{\gamma}} + \widehat{\sigma}^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{\beta}_\gamma\right\}T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}), && E(T_2|\boldsymbol{\gamma},\sigma^2) \cong \boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)\\
&g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) = (\nu\widehat{\sigma}^2/2)^{-\beta_\sigma}\Gamma(\nu/2)/\Gamma(\nu/2-\beta_\sigma), && E(g_1|\sigma^2) = (\sigma^2)^{-\beta_\sigma}\\
&g_2(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) = -\left\{\log(2/\nu\widehat{\sigma}^2) + D(\nu/2-\beta_\sigma)\right\}g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}), && E(g_2|\sigma^2) = \log(\sigma^2)(\sigma^2)^{-\beta_\sigma}.
\end{aligned}$$

The function $D(t)$ denotes the di-gamma function. The estimators $g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta})$ and $g_2(\widehat{\sigma}^2,\nu,\boldsymbol{\beta})$ are defined only if $\nu/2 > \beta_\sigma$; violation of this condition is most likely to arise if the number of longitudinal replicates, and hence ν, is small. For a moderate number of longitudinal replicates, like $r = 10$, this would only occur if β_σ were greater than 4, which corresponds to an odds ratio greater than 50, per unit change in $\log(\sigma^2)$. This is a limitation of the method, though it may be rare in practice.

In $\Psi(Y,\boldsymbol{\gamma},\sigma^2,\boldsymbol{Z},\boldsymbol{\beta})$, the unknown quantities are replaced with their estimators to

define an approximately unbiased score

$$\Psi_A(Y,\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{Z},\nu,\boldsymbol{\beta}) = \begin{bmatrix} (Y-1)+Y\exp(-\beta_0-\boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) \\ (Y-1)\widehat{\boldsymbol{\gamma}}+Y\exp(-\beta_0-\boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})T_2(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) \\ (Y-1)\widehat{\log}(\sigma^2)+Y\exp(-\beta_0-\boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})g_2(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) \\ (Y-1)\boldsymbol{Z}+Y\boldsymbol{Z}\exp(-\beta_0-\boldsymbol{Z}^T\boldsymbol{\beta}_{\mathbf{Z}})T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta}) \end{bmatrix}. \tag{3.7}$$

An estimator $\widehat{\boldsymbol{\beta}}_A$ satisfying $\sum_{i=1}^n \Psi_A(Y_i,\widehat{\boldsymbol{\gamma}}_i,\widehat{\sigma}_i^2,\nu_i,\boldsymbol{Z}_i,\widehat{\boldsymbol{\beta}}_A)=\mathbf{0}$ is called an ACS estimator. This score is not exactly unbiased, firstly because T_1 and T_2 involve the $\widehat{\sigma}^2$ rather than σ^2, and additionally because g_1 and g_2 also depend on $\widehat{\sigma}^2$ and are therefore correlated with T_1 and T_2. Consequently,
$E\{T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta})|\boldsymbol{\gamma},\sigma^2\} \neq E_1(\boldsymbol{\gamma},\sigma^2,\boldsymbol{\beta})$ even though
$E\{T_1(\widehat{\boldsymbol{\gamma}},\sigma^2,\boldsymbol{\beta})g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta})|\boldsymbol{\gamma},\sigma^2\} = E\{T_1(\widehat{\boldsymbol{\gamma}},\sigma^2,\boldsymbol{\beta})|\boldsymbol{\gamma},\sigma^2\}\ E\{g_1(\widehat{\sigma}^2,\nu,\boldsymbol{\beta})|\boldsymbol{\gamma},\sigma^2\} =$
$E_1(\boldsymbol{\gamma},\sigma^2,\boldsymbol{\beta})$. However, T_1 and T_2 depend on $\widehat{\sigma}^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}$, which shrinks in magnitude as the number of replicates increases and $\boldsymbol{D}^T\boldsymbol{D}$ increases. The correlation between the estimators T and g becomes quite small with adequate replication, and the estimation of $\widehat{\sigma}^2$ becomes very good so that $T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})$ is nearly $T_1(\widehat{\boldsymbol{\gamma}},\sigma^2,\boldsymbol{\beta})$. In this sense, $\Psi_A(Y,\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{Z},\nu,\boldsymbol{\beta})$ can be considered approximately unbiased.

Solutions to the score equations can be obtained by a Newton-Raphson algorithm. The naive estimate for $\boldsymbol{\beta}$ can be can be used a starting value. In practice, the ACS approach is less susceptible to numerical problems than the corrected score defined in Appendix C.3. In addition, our simulations indicate no discernable bias in the approximate score estimating equations, nor resulting estimators, with as few as 10 replicates per subject (Section 3.4). The estimator $\widehat{\boldsymbol{\beta}}_A$ is an M-estimators, so that under regularity conditions it is asymptotically normal, and standard errors may be obtained by the empirical sandwich estimator (Carroll et al., 2006). In Appendix C.5, we show that sandwich standard errors provide reliable inference.

3.3.3 Moment Adjusted Imputation

The ACS is a functional approach that yields bias reduction asymptotically, while making no assumptions on the underlying distributions of subject-specific random effects and variances. There are some disadvantages to its implementation. Primarily, the score equation is difficult to solve numerically and frequently has convergence problems, even for

the approximate version. In addition, our current derivation requires that the logistic regression model depend on the log transformation of subject-specific variances, $T(\sigma^2) = \log(\sigma^2)$. In practice, the analyst may want to modify the model to include other transformations of the covariates, quadratic terms, splines or categorization of continuous variables. The corrected score approach can be adapted for some alternative specifications of the logistic regression model, but even when this is possible, new derivations are required. Alternatively, we suggest that Moment Adjusted Imputation (MAI) (Chapters 1 and 2) can be implemented in this situation. In Chapters 1 and 2, we account for normally-distributed measurement error by obtaining new, estimated values for the covariate data, which have been constrained to have unbiased moments for the distribution of the unknown population covariate. The adjusted data are imputed in place of the underlying, unobserved covariate and the logistic regression model can be fit by standard methods. This approach offers greater flexibility than the corrected score.

We briefly review the MAI method and describe the modifications that are necessary to accommodate the present problem. Let $\boldsymbol{X}_i$ represent the (3×1) vector of unobserved subject-specific random effects and transformed variances, $\left\{\gamma_{1i}, \gamma_{0i}, T(\sigma_i^2)\right\}^T$, for subjects $i = 1, \ldots, n$. The components of $\boldsymbol{X}_i$ should be ordered from least to greatest relative measurement error, or largest to smallest reliability ratio. In some scenarios, this order may be different than the order presented here. The $\boldsymbol{X}_i$ need not be identically distributed, but we make this assumption to simplify notation. The estimates of these quantities, $\boldsymbol{W}_i = \left\{\widehat{\gamma}_{1i}, \widehat{\gamma}_{0i}, T(\widehat{\sigma}_i^2)\right\}^T$, are observed in place of $\boldsymbol{X}_i$. We also have a response Y_i and other error-free covariates collected in a $(K-1)\times 1$ vector $\boldsymbol{Z}_i$. These additional variables are collected to create $\boldsymbol{V}_i$=$(Y_i, \boldsymbol{Z}_i^T)$, with components V_{ik} for $k = 1, \ldots, K$. We make the usual surrogacy assumption that $\boldsymbol{V}_i$ and $\boldsymbol{W}_i$ are independent, given $\boldsymbol{X}_i$.

The goal is to obtain adjusted versions of the $\boldsymbol{W}_i$, $\widehat{\boldsymbol{X}}_i$, that reflect the joint distribution of $\boldsymbol{X}_i$ and other model variables by imposing constraints on the moments of $\widehat{\boldsymbol{X}}_i$. Various approaches to obtaining adjusted data are compared in Chapter 2. Here, we perform adjustment sequentially for the three components of $\boldsymbol{W}_i$, denoted W_{ig} for $g = 1, 2, 3$ as in Section 2.2. For each of these, we minimize $\sum_{i=1}^n (W_{ig} - X_{ig})^2$ constrained so that $E(n^{-1}\sum_{i=1}^n \widehat{X}_{ig}^r) = E(X_{ig}^r) = m_{rg}$, for $r = 1, \ldots, M$, and $E(n^{-1}\sum_{i=1}^n \widehat{X}_{ig}V_{ik}) = E(X_{ig}V_{ik})$ $= m_{kg}^V$, for $k = 1, \ldots, K$. Moreover, we require that $E(n^{-1}\sum_{i=1}^n \widehat{X}_{ig}\widehat{X}_{ig'}) = E(X_{ig}X_{ig'}) = m_{gg'}^*$ for all $g \neq g'$.

In order to implement the constrained minimization, we must unbiasedly estimate m_{rg}, m^V_{kg}, and $m^*_{gg'}$. In Chapter 2, we show how to estimate these quantities when $\boldsymbol{W}_i|\boldsymbol{X}_i \sim MVN(\boldsymbol{X}_i, \Sigma_{ui})$, and Σ_{ui} is known. In the present problem, $(W_{i2}, W_{i1})^T = \widehat{\boldsymbol{\gamma}}_i|\boldsymbol{\gamma}_i \sim N\{\boldsymbol{\gamma}_i, \Sigma_{ui} = \sigma_i^2(\boldsymbol{D}_i^T\boldsymbol{D}_i)^{-1}\}$. In the following derivations of moment estimators, Σ_{ui} is regarded as known, though $\widehat{\sigma}_i^2$ is substituted for σ_i^2 in the implementation of MAI. The moment estimators defined in Chapter 2 are $\widehat{m}_{r1} = n^{-1}\sum_{i=1}^n P_r(\widehat{\gamma}_{1i}, \Sigma_{ui,22})$ and $\widehat{m}_{r2} = n^{-1}\sum_{i=1}^n P_r(\widehat{\gamma}_{0i}, \Sigma_{ui,11})$, for $r = 1, \ldots, M$, where $P_r(w, \sigma) = \sigma^r H_r(w/\sigma)$, and H_r are the Hermite polynomials defined as $H_0(z) = 1$, $H_1(z) = z$, $H_r(z) = zH_{r-1}(z) - (r-1)H_{r-2}(z)$ for $r = 2, 3, \ldots$ (Cramer, 1957). The cross products are estimated by $\widehat{m}^V_{k1} = n^{-1}\sum_{i=1}^n \widehat{\gamma}_{1i}V_{ik}$ and $\widehat{m}^V_{k2} = n^{-1}\sum_{i=1}^n \widehat{\gamma}_{0i}V_{ik}$, for $k = 1, \ldots, K$. Finally, $\widehat{m}^*_{21}$=$\widehat{m}^*_{12}$=$n^{-1}\sum_{i=1}^n(\widehat{\gamma}_{0i}\widehat{\gamma}_{1i} - \Sigma_{ui,12})$. These cross product estimators are unbiased as long as the surrogacy assumption is not violated (Chapter 2), and assuming Σ_{ui} is known.

The measurement error in $W_{i3} = T(\sigma_i^2)$ is not normally distributed and hence previous definitions of moment estimators are not applicable. Moments and cross products involving W_{i3} are estimated based on the chi-square variation in $\widehat{\sigma}_i^2$. We develop moment estimators for two variance transformations, $T(\sigma_i^2) = \sigma_i$ and $T(\sigma_i^2)$= $\log(\sigma_i^2)$. First, consider the square root transformation. Like the log transformation, σ_i tends to be less skewed than σ_i^2 and is often preferred for the interpretation as subject-specific standard deviation. Given that $\widehat{\sigma}_i^2 \sim (\sigma_i^2/\nu_i)\chi^2_{\nu_i}$ we can integrate $\widehat{\sigma}_i^r$ over it's distribution to find that $E(\widehat{\sigma}_i^r|\sigma_i) = \sigma_i^r/H(r, \nu_i)$ where $H(r, \nu) = (\nu/2)^{r/2}\Gamma(\nu/2)/\Gamma(\nu/2 + r/2)$. So $E\{\widehat{\sigma}_i^r H(r, \nu_i)|\sigma_i\} = \sigma_i^r$. The estimators $\widehat{m}_{r3} = n^{-1}\sum_{i=1}^n \widehat{\sigma}_i^r H(r, \nu_i)$ are unbiased for $E(\sigma_i^r)$=m_{r3} and $r = 1, \ldots, M$. Assuming that V_{ik} are independent of $\widehat{\sigma}_i^2$ given σ_i^2, the cross products with V_{ik} can be estimated by $\widehat{m}^V_{k3}$= $n^{-1}\sum_{i=1}^n \widehat{\sigma}_i^r H(r, \nu_i)V_{ik}$, for $k = 1, \ldots, K$. Based on the surrogacy assumption, these are unbiased for m^V_{k3} because $E[E\{\widehat{\sigma}_i^r H(r, \nu_i)V_{ik}|\sigma_i, V_{ik}\}] = E[V_{ik}E\{\widehat{\sigma}_i^r H(r, \nu_i)|\sigma_i\}]$= $E(\sigma_i^r V_{ik})$. Under the joint model described in Section 3.2, $\widehat{\boldsymbol{\gamma}}_i$ are conditionally independent of $\widehat{\sigma}_i^2$ given $(\boldsymbol{\gamma}_i, \sigma_i^2)$, so the estimators $\widehat{m}^*_{32}$=$\widehat{m}^*_{23}$= $n^{-1}\sum_{i=1}^n \widehat{\sigma}_i H(1, \nu_i)\widehat{\gamma}_{0i}$ and $\widehat{m}^*_{31}$ = $\widehat{m}^*_{13}$ = $n^{-1}\sum_{i=1}^n \widehat{\sigma}_i H(1, \nu_i)\widehat{\gamma}_{1i}$ are also unbiased because $E[E\{\widehat{\sigma}_i^r H(r, \nu_i)\widehat{\gamma}_{0i}|\sigma_i, \gamma_{0i}\}]$ = $E[E\{\widehat{\gamma}_{0i}|\gamma_{0i}\}E\{\widehat{\sigma}_i^r H(r, \nu_i)|\sigma_i\}] = E(\sigma_i^r\gamma_{0i}) = m^*_{32}$ and $E[E\{\widehat{\sigma}_i^r H(r, \nu_i)\widehat{\gamma}_{1i}|\sigma_i, \gamma_{1i}\}]$ = $E[E\{\widehat{\gamma}_{1i}|\gamma_{1i}\}E\{\widehat{\sigma}_i^r H(r, \nu_i)|\sigma_i\}] = E(\sigma_i^r\gamma_{1i}) = m^*_{31}$. Estimators for the moments of $\log(\sigma_i^2)$ are more complicated, but can still be defined in closed form. These are provided in Appendix C.4.

3.4 Simulations

3.4.1 Model (3.1) with slope, intercept and variance covariates

We conducted simulations to compare the alternative approaches to fitting model (3.1) with adjustment for measurement error. The model without error-free covariates is

$$P(Y_i = 1|\boldsymbol{\gamma}_i, \sigma_i^2; \boldsymbol{\beta}_\gamma, \beta_\sigma) = F\{\beta_0 + \boldsymbol{\gamma}_i^T \boldsymbol{\beta}_\gamma + T(\sigma_i^2)\beta_\sigma\} \tag{3.8}$$

where $F(t) = 1/(1 + e^{-t})$. The underlying data were generated to resemble the AMoRR blood pressure data. In this data set, the event rate was around 7 percent. It is common to see such a low event rate when the outcome of interest is mortality. We generated the underlying subject-specific variances from a chi-square distribution with 6 degrees of freedom scaled by a multiple of 60, and the subject-specific intercepts and slopes according to a multivariate normal distribution, $(\gamma_0, \gamma_1)^T \sim MVN\{\mathbf{0}, \Sigma\}$ where $\Sigma = \text{vech}\{16^2, 0.4(16)(4), 4^2\}$. Because the off-diagonal is non-zero, these coefficients are correlated. This distribution produces data that resemble the subject-specific blood pressure intercepts, slopes and variances from the AMoRR trial.

We studied the two cases where $T(\sigma^2)$=σ and where $T(\sigma^2)$=log(σ^2), two sample sizes ($n = 1000, 5000$), and two levels of replication ($r = 10, 40$) on each subject. The true covariate effects were either moderately significant at $n = 1000$, or weakly significant. When $T(\sigma^2) = \sigma$, these are $(\beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma) = (0.02, 0.08, 0.08)$ and $(0.01, 0.04, 0.04)$, respectively. When $T(\sigma^2) = \log(\sigma^2)$, these are $(\beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma) = (0.02, 0.08, 0.80)$ and $(0.01, 0.04, 0.40)$, respectively. In this chapter, we focus on the results where $n = 5000$ and $r = 10$ because these values reflect our application of interest.

For the model where $T(\sigma^2)$=log(σ^2), boxplots of the estimates from $B = 500$ simulated data sets are displayed in Figure 3.2. The naive approach results in substantial bias in the estimator for β_σ, although the bias in estimators for β_{γ_0} and β_{γ_1} is much less. Implementation of MAI essentially eliminates the bias, regardless of whether two or four moments are matched. There is a tradeoff between bias and variance, which is particularly pronounced in the estimation of β_{γ_0} and β_{γ_1}. However, we are primarily interested in β_σ, for which the reduction in bias is large relative to the increase in variance. This is quantified by a comparison of mean squared errors in Table 3.1. The ACS estimator is nearly unbiased, suggesting that the approximation is good, however the variability is much larger than

for MAI (Figure 3.2). In addition, the Newton-Raphson algorithm for obtaining the ACS estimates failed to converge in 74 data sets, and these are not displayed.

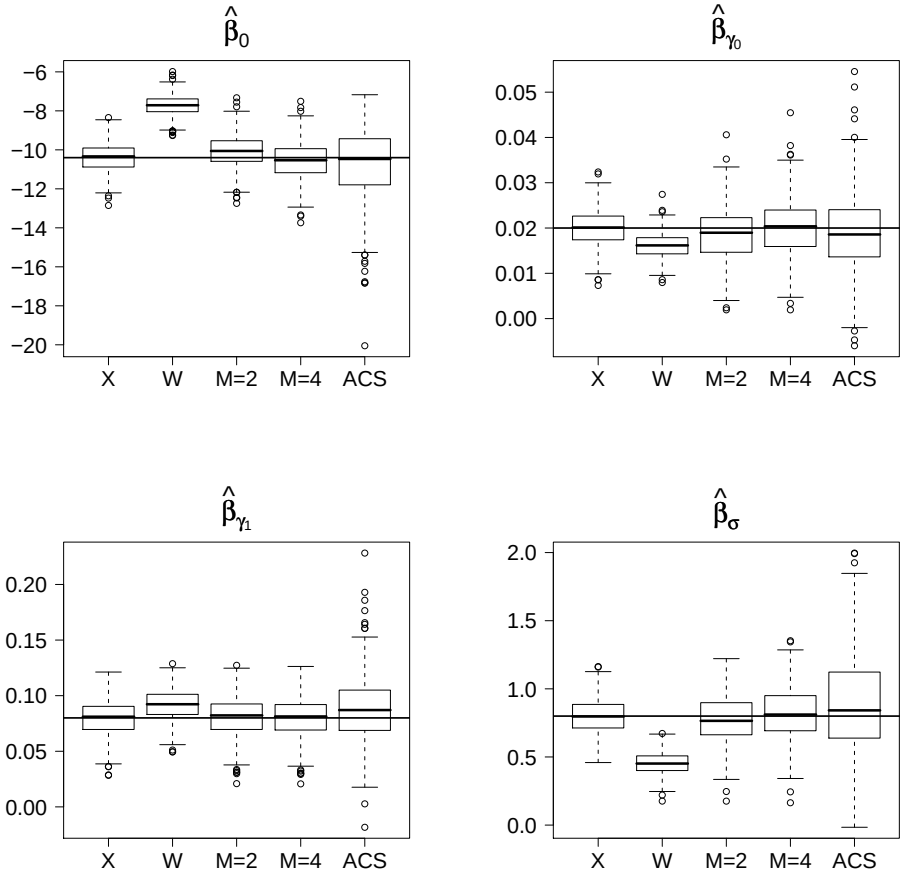

Figure 3.2: *Boxplots of* $\widehat{\boldsymbol{\beta}}$ *where* $P(Y=1|\boldsymbol{\gamma},\sigma^2;\boldsymbol{\beta}) = F(\beta_0 + \boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \log(\sigma^2)\beta_\sigma)$ *with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma)$=$(-10.4, 0.02, 0.08, 0.8)$, $r = 10$, *and* $n = 5000$. *Methods: X, true covariates; W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments; ACS, Approximately Corrected Score (based on 426 data sets for which the algorithm converged).*

In addition to plots, summary statistics for the case where $T(\sigma^2)$=$\log(\sigma^2)$ are displayed in Tables 3.1 and 3.2. We do not include the ACS method in these tables because it takes an exceedingly long time to run and demonstrates excessive variability relative to MAI. Results for the case where $T(\sigma^2)$=σ are similar and are included in Appendix C.5.1. Parameter estimators based on the naive method are compared to the adjusted versions according to relative bias, Monte Carlo standard deviation of relative bias, mean squared

error (MSE) and coverage probability. The scale of MSE's for these parameters is very small, so we report the ratio of MSE based on naive estimates divided by the MSE for adjusted estimates. Thus, large values of the MSE ratio indicate greater reduction in MSE relative to the naive method. The coverage probability is calculated by the proportion of data sets for which a 95% Wald confidence interval contains the true parameter value. Standard errors used in these confidence intervals were obtained by the sandwich method. On average, they are quite similar to the Monte Carlo standard deviation.

The primary parameter of interest is β_σ. In all scenarios, the relative bias in $\widehat{\beta}_\sigma$ is greatly reduced by adjustment. There is a tradeoff in terms of variability. In some cases the MSE ratio is less than one, indicating that the adjusted parameter estimator has larger MSE than the naive estimator. However, when the sample size is large, the MSE ratio is typically greater than one, and the adjusted parameter estimator has smaller MSE than the naive estimator. The coverage probability of the naive estimator is quite poor in most scenarios, and it is restored to 95% by adjustment.

The bias in naive estimators for β_{γ_0} and β_{γ_1} is less than for β_σ. Although MAI reduces bias, there is a substantial tradeoff in terms of variability in the estimation of β_{γ_0} and β_{γ_1}. The MSE ratios for these parameters are often less than one, indicating that the naive method is preferable, in terms of MSE. In the present problem, covariates γ_0 and γ_1 are included primarily for the purpose of adjusting the model so the additional variability in their estimates is not of direct concern. The coverage probability of the naive estimators is close to 95% in many cases, though it is improved by adjustment.

In general, the MAI estimators have minimal bias across a range of situations. In terms of MSE there is little difference between MAI which matches two or four moments, though the relative bias is often smaller for the version which matches four. As expected, the adjustment is most beneficial for data sets with less replication, larger sample size and stronger covariate effects.

Table 3.1: *Estimation of* $\boldsymbol{\beta}$ *where* $P(Y = 1|\gamma_0, \gamma_1, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \gamma_0\beta_{\gamma_0} + \gamma_1\beta_{\gamma_1} + log(\sigma^2)\beta_\sigma)$ *with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma)$=$(-10.4, 0.02, 0.08, 0.8)$*, two levels of replication (r), two sample sizes (n). RB, relative bias (standard error approximately 0.02 for n=1000, 0.01 for n=5000); SD, Monte Carlo standard deviation of* $\widehat{\beta}$ *multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability; MSE-R,* $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ *(standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.08). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments; ACS, Approximately Corrected Score.*

			β_{γ_0}			β_{γ_1}			β_σ		
r	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	−0.21	−0.10	−0.06	0.17	0.08	0.08	−0.42	−0.05	0.05
		SD	0.61	1.38	1.50	3.09	3.93	3.97	18.01	40.68	47.22
		SE/SD	1.00	1.05	1.05	1.00	0.97	0.97	1.00	1.02	1.03
		CP	0.87	0.95	0.95	0.95	0.95	0.95	0.51	0.94	0.94
		MSE-R	1.00	0.28	0.24	1.00	0.72	0.70	1.00	0.89	0.66
	5000	RB	−0.18	−0.04	−0.01	0.14	0.03	0.04	−0.43	−0.08	0.01
		SD	0.26	0.59	0.63	1.38	1.72	1.74	8.01	18.00	20.55
		SE/SD	1.00	1.02	1.02	1.00	0.97	0.97	1.00	1.03	1.03
		CP	0.68	0.94	0.94	0.88	0.95	0.95	0.01	0.92	0.95
		MSE-R	1.00	0.57	0.50	1.00	1.07	1.05	1.00	3.46	2.95
40	1000	RB	−0.01	0.02	0.04	0.04	0.01	0.01	−0.14	−0.04	0.00
		SD	0.76	0.96	0.99	3.45	3.68	3.69	22.77	28.29	30.25
		SE/SD	1.00	1.00	1.00	1.00	1.03	1.03	1.00	0.99	0.99
		CP	0.96	0.96	0.96	0.94	0.94	0.94	0.93	0.95	0.95
		MSE-R	1.00	0.63	0.58	1.00	0.89	0.89	1.00	0.80	0.71
	5000	RB	−0.04	−0.01	0.00	0.02	−0.01	−0.01	−0.14	−0.04	0.00
		SD	0.34	0.43	0.45	1.55	1.65	1.66	10.68	13.26	14.10
		SE/SD	1.00	1.02	1.02	1.00	1.04	1.04	1.00	1.03	1.03
		CP	0.94	0.96	0.96	0.95	0.94	0.94	0.79	0.95	0.95
		MSE-R	1.00	0.66	0.62	1.00	0.88	0.88	1.00	1.29	1.19

Table 3.2: *Estimation of* $\boldsymbol{\beta}$ *where* $P(Y = 1|\gamma_0, \gamma_1, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \gamma_0\beta_{\gamma_0} + \gamma_1\beta_{\gamma_1} + log(\sigma^2)\beta_\sigma)$ *with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma)=(-6.4, 0.01, 0.04, 0.4)$*, two levels of replication (r), two sample sizes (n). RB, relative bias (standard error approximately 0.05 for n=1000, 0.02 for n=5000); SD, Monte Carlo standard deviation of* $\widehat{\beta}$ *multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability; MSE-R,* $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ *(standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.06). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

			β_{γ_0}			β_{γ_1}			β_σ		
r	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	−0.19	−0.06	−0.04	0.10	−0.01	−0.01	−0.42	−0.06	0.01
		SD	0.60	1.31	1.35	3.16	4.16	4.17	16.30	35.48	38.20
		SE/SD	1.00	1.03	1.03	1.00	1.05	1.05	1.00	0.98	0.99
		CP	0.92	0.93	0.94	0.94	0.94	0.94	0.80	0.96	0.96
		MSE-R	1.00	0.23	0.22	1.00	0.59	0.58	1.00	0.44	0.38
	5000	RB	−0.15	−0.01	0.01	0.15	0.02	0.02	−0.41	−0.04	0.01
		SD	0.26	0.56	0.57	1.40	1.76	1.77	6.92	15.25	16.29
		SE/SD	1.00	1.00	1.00	1.00	1.03	1.03	1.00	0.95	0.95
		CP	0.90	0.94	0.94	0.91	0.93	0.94	0.37	0.96	0.97
		MSE-R	1.00	0.28	0.27	1.00	0.74	0.73	1.00	1.35	1.19
40	1000	RB	−0.07	−0.05	−0.04	0.06	0.03	0.03	−0.09	0.03	0.05
		SD	0.74	0.93	0.95	3.29	3.51	3.51	20.97	26.00	26.90
		SE/SD	1.00	1.00	1.00	1.00	1.01	1.01	1.00	0.97	0.97
		CP	0.94	0.94	0.94	0.94	0.94	0.94	0.95	0.95	0.95
		MSE-R	1.00	0.64	0.62	1.00	0.88	0.88	1.00	0.67	0.62
	5000	RB	−0.05	−0.04	−0.03	0.07	0.04	0.04	−0.11	0.00	0.03
		SD	0.34	0.42	0.43	1.45	1.56	1.56	9.59	11.87	12.25
		SE/SD	1.00	1.02	1.02	1.00	1.01	1.01	1.00	1.00	1.00
		CP	0.95	0.95	0.95	0.94	0.95	0.95	0.92	0.95	0.95
		MSE-R	1.00	0.65	0.63	1.00	0.89	0.89	1.00	0.79	0.73

3.4.2 Model (3.1) with variance covariate only

We also consider the special case of model (3.1), where the association between longitudinal variability and a binary outcome is evaluated without adjustment for other covariates. The model is

$$P(Y_i = 1|\sigma_i^2; \beta_\sigma) = F\{\beta_0 + T(\sigma_i^2)\beta_\sigma\} \tag{3.9}$$

where $F(t) = 1/(1 + e^{-t})$. As previously, the true subject-specific variances were generated from a scaled chi-square distribution. The value of β_σ was either 0.08 or 0.04 for $T(\sigma^2) = \sigma$ and 0.80 or 0.40 for $T(\sigma^2) = \log(\sigma^2)$. The value of the intercept β_0 was chosen to maintain an event rate near 7%. In this chapter, results are presented for $T(\sigma^2)$=log(σ^2), and similar results for $T(\sigma^2) = \sigma^2$ are included in Appendix C.5.3. In this reduced model, (3.9), the ACS method is implemented more quickly, and it is compared to MAI in our simulation results.

In Tables 3.3 and 3.4, we see substantial benefit from adjustment under most conditions. All of the adjustment methods reduce bias, however, the version of MAI that matches four moments over-corrects when the sample size is small. As expected, there is additional variability in the adjusted estimators, particularly the ACS estimator. For the MAI method of adjustment, the MSE ratios are typically greater than one. In addition, the coverage probability of the naive estimator is quite poor, and it is restored to 95% by either method of adjustment, MAI or ACS. When the sample size is large ($n = 5000$), and the replicates are few ($r = 10$), the performance of MAI is particularly good, and its MSE ratios are between 2.5 and 7.0. As in model (3.8), adjustment is most beneficial for data sets with less replication, larger sample size, and stronger covariate effects.

Table 3.3: *Estimation of $\boldsymbol{\beta}$ where $P(Y = 1|\sigma^2; \boldsymbol{\beta}) = F(\beta_0 + log(\sigma^2)\beta_\sigma)$ with true values (β_0, β_σ)=$(-7.2, 0.80)$, two levels of replication (r), two sample sizes (n). RB, relative bias (standard error approximately 0.02 for n=1000, 0.01 for n=5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.04 and 0.07). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

			β_σ			
r	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	ACS
10	1000	RB	−0.40	0.01	−0.13	−0.02
		SD	16.13	27.91	20.20	134.98
		SE/SD	1.00	1.01	1.41	0.51
		CP	0.50	0.96	0.96	0.94
		MSE-R	1.00	1.64	1.54	0.07
	5000	RB	−0.43	−0.05	−0.04	0.05
		SD	7.34	12.60	11.94	23.10
		SE/SD	1.00	0.98	1.08	0.97
		CP	0.00	0.92	0.95	0.96
		MSE-R	1.00	7.13	6.96	2.25
40	1000	RB	−0.14	−0.01	−0.10	0.15
		SD	19.87	22.90	17.80	41.56
		SE/SD	1.00	1.00	1.31	0.77
		CP	0.91	0.96	0.97	0.94
		MSE-R	1.00	0.98	0.92	0.28
	5000	RB	−0.15	−0.03	−0.02	0.01
		SD	8.56	9.86	9.44	13.52
		SE/SD	1.00	1.05	1.13	0.99
		CP	0.76	0.95	0.97	0.95
		MSE-R	1.00	2.10	2.05	1.16

Table 3.4: *Estimation of $\boldsymbol{\beta}$ where $P(Y = 1|\sigma^2; \boldsymbol{\beta}) = F(\beta_0 + log(\sigma^2)\beta_\sigma)$ with true values (β_0, β_σ)=$(-4.8, 0.40)$, two levels of replication (r), two sample sizes (n). RB, relative bias (standard error approximately 0.03 for n=1000, 0.01 for n=5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.02 to 0.07). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments; ACS, Approximately Corrected Score.*

			β_σ			
r	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	ACS
10	1000	RB	−0.35	0.08	0.04	0.19
		SD	16.07	27.02	25.55	48.75
		SE/SD	1.00	0.95	1.03	0.78
		CP	0.80	0.94	0.95	0.98
		MSE-R	1.00	0.62	0.58	0.19
	5000	RB	−0.41	−0.03	−0.02	0.02
		SD	6.66	11.11	11.42	13.38
		SE/SD	1.00	1.03	1.03	1.01
		CP	0.33	0.95	0.95	0.97
		MSE-R	1.00	2.53	2.42	1.76
40	1000	RB	−0.10	0.02	0.03	0.15
		SD	18.79	21.48	21.50	29.39
		SE/SD	1.00	0.99	1.01	0.84
		CP	0.94	0.95	0.95	0.94
		MSE-R	1.00	0.80	0.77	0.41
	5000	RB	−0.13	−0.02	−0.01	0.01
		SD	8.15	9.30	9.46	10.42
		SE/SD	1.00	1.02	1.02	1.00
		CP	0.90	0.95	0.96	0.96
		MSE-R	1.00	1.10	1.06	0.88

3.5 Application to AMoRR data

We use both the ACS and MAI approaches to investigate the relationship between variability in longitudinal diastolic blood pressure (DBP) and short-term mortality in the AMoRR data set. The analysis includes $n = 7382$ subjects who had $r_i \geq 5$ pressure measurements, observed between 91 and 120 days and who survived until 120 days. Follow-up for short-term mortality began at 121 days and concluded at 180 days. During this time, 215 deaths were observed, and approximately 4% of patients were censored.

We measure the variability in DBP by the mean squared error from a linear regression, fit for each individual subjects' longitudinal DBP data. This quantity, $\widehat{\sigma}_i^2$, is defined as in Section (3.2) and is mis-measured for an unknown true diastolic blood pressure (DBP) variability, σ_i^2. In addition, the data include subject-specific slope and intercept estimates, $\widehat{\gamma}_i$, which describe the trajectory of longitudinal DBP, and error-free variables, age, diabetes, and body mass index. In order to assess the independent association between DBP variability and short-term mortality, we fit the fully adjusted logistic regression model (3.1),

$$P(Y_i = 1|\boldsymbol{\gamma}_i, \sigma_i^2, \boldsymbol{Z}_i; \boldsymbol{\beta}) = F\{\beta_0 + \boldsymbol{\gamma}_i^T \boldsymbol{\beta}_\gamma + T(\sigma_i^2)\beta_\sigma + \boldsymbol{Z}_i^T \boldsymbol{\beta}_Z\},$$

where $\boldsymbol{Z}_i$ includes age, diabetes, and body mass index. Body mass index is fit as a categorical variable with four levels to account for non-linearity. Additional error-free covariates were available in the AMoRR data set, but are not included in this analysis.

Tables 3.5 and 3.6 display the logistic regression parameter estimates that are obtained by unadjusted and adjusted analysis. These include results for two transformations of DBP variability, the log transformation, $T(\sigma^2)$=log(σ^2) (Table 3.5), and the square root transformation, $T(\sigma^2)$=σ (Table 3.6). The ACS method is applicable only for the log variance transformation, and these results are included only in Table 3.5. We are primarily interested in the estimate of β_σ, controlling for other covariates. However, the estimates for $\boldsymbol{\beta}_\gamma$ are also included. We saw little difference in $\widehat{\boldsymbol{\beta}}_Z$, with and without adjustment, and do not report these quantities.

Regardless of the variance transformation, there is no statistically significant relationship between DBP variability and short-term mortality. Even so, the parameter estimate for β_σ is modified by adjustment. The results are very similar for both variance transformations, so we limit our discussion to the log transformation. The naive approach, which ignores measurement error, gives an odds ratio estimate of 1.10 (95% CI 0.93-1.26),

per standard deviation increase in DBP variability. The MAI odds ratio estimate is 1.26 (95% CI 0.93-1.62), and the odds ratio based on ACS is 1.59 (95% CI 0.49-2.69). As expected, the naive odds ratio is closer to one than the adjusted odds ratios. The ACS odds ratio is larger than the MAI and has a substantially larger confidence interval. All of the confidence intervals contain one, and there is no significant association between DBP variability and short term mortality, regardless of adjustment.

In this analysis, adjustment for measurement error had only a modest impact on the estimation of β_σ, and the corresponding odds ratio. The adjustment could be more substantial for stronger covariate effects, such as those considered in Section 3.4.

Table 3.5: Parameter estimates for the model $P(Y = 1|\boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}; \boldsymbol{\beta}) = F(\beta_0 + \boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \log(\sigma^2)\beta_\sigma + \boldsymbol{Z}^T\boldsymbol{\beta}_Z)$. Methods: Naive, model fit with $\widehat{\gamma}$ and $\log(\widehat{\sigma}^2)$ unadjusted for measurement error; MAI, Moment Adjusted Imputation; ACS Approximately Corrected Score. Quantities are reported per standard deviation increase in the covariate, and are followed by standard errors in parenthesis.

	Method	Estimate	P-value	Odds ratio
β_σ	Naive	0.09 (0.07)	0.22	1.10 (0.08)
	MAI	0.24 (0.14)	0.08	1.27 (0.18)
	ACS	0.46 (0.35)	0.19	1.59 (0.56)
β_{γ_0}	Naive	-0.42 (0.10)	0.00	0.66 (0.07)
	MAI	-0.50 (0.13)	0.00	0.61 (0.08)
	ACS	-0.30 (0.17)	0.08	0.74 (0.13)
β_{γ_1}	Naive	-0.24 (0.08)	0.00	0.79 (0.07)
	MAI	-0.32 (0.19)	0.10	0.73 (0.14)
	ACS	-0.09 (0.14)	0.52	0.91 (0.13)

Table 3.6: Parameter Estimates for the model $P(Y = 1|\boldsymbol{\gamma}, \sigma^2, \boldsymbol{Z}; \boldsymbol{\beta}) = F(\beta_0 + \boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \sigma\beta_\sigma + \boldsymbol{Z}^T\boldsymbol{\beta}_Z)$. Methods: Naive, model fit with $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\sigma}$ unadjusted for measurement error; MAI, Moment Adjusted Imputation. Quantities are reported per standard deviation increase in the covariate, and are followed by standard errors in parenthesis.

	Method	Estimate	P-value	Odds ratio
β_σ	Naive	0.13 (0.07)	0.08	1.14 (0.08)
	MAI	0.23 (0.12)	0.06	1.25 (0.15)
β_{γ_0}	Naive	-0.43 (0.10)	0.00	0.65 (0.06)
	MAI	-0.44 (0.13)	0.00	0.64 (0.08)
β_{γ_1}	Naive	-0.24 (0.08)	0.00	0.79 (0.06)
	MAI	-0.23 (0.22)	0.28	0.79 (0.17)

3.6 Discussion

We have compared three approaches to adjusting for measurement error in logistic regression models where covariates include variance components. The conditional score and corrected score approaches are limited to certain to primary outcome models that depend on $1/\sigma^2$ and $\log(\sigma^2)$, respectively. The MAI method is flexible and can be used with many transformations of σ^2. These methods require no assumptions on the distribution of subject-specific parameters, but make use of the sampling distributions to adjust for measurement error. In our simulations, the MAI estimator was preferable to the ACS estimator because the ACS had high variability and frequent outliers. The MAI estimator was generally preferable to the naive estimator for estimating the association between subject-specific variance and outcomes, particularly when the number of replicates was few and the sample size large.

In this chapter, our simulations involve data sets with at least ten longitudinal replicates. Many studies collect fewer longitudinal replicates (Yang et al., 2007). This leads us to wonder whether adjustment would be beneficial in these cases. In preliminary simulations we considered fewer replicates, such as $r = 5$. In this case, the measurement error in a variance estimates is very large, with reliability ratio around 0.30. The numerical algorithms necessary for MAI and ACS methods tend to fail. MAI can be implemented, but problems occur more frequently with fewer replicates. Thus, replication is not only important for reducing measurement error directly, but to facilitate adjustment. We stress the need for replication when subject-specific variances are included as covariates in a primary outcome model.

All of the methods of adjustment rely on the surrogacy assumption, that Y_i and $\boldsymbol{Z}_i$ are independent of $(\widehat{\boldsymbol{\gamma}}_i, \widehat{\sigma}_i^2)$ conditional on $(\boldsymbol{\gamma}_i, \sigma_i^2)$. This is a common assumption in measurement error literature and is justifiable in many cases. In the present problem, this means that measurement error in the subject-specific estimates is not related to the probability of observing an event, nor to the value of error-free covariates. This could be violated if the data collection is not homogeneous across subjects. For example, if data are collected whenever a patient visits the doctor, sicker patients will have more medical appointments and consequently have greater replication and less measurement error. Because these patients also have a higher probability of adverse outcomes, the surrogacy assumption may be

violated. Alternatively, if data collection is designed to achieve a common level of replication for all patients, with minimal missing data, the surrogacy assumption may be reasonable. Further research is needed to account for violations of the surrogacy assumption, and/or assess the consequences in models involving variance parameters as covariates.

In this chapter, we focus on dichotomous data that can be modeled by logistic regression. Biological outcomes are often quantified by time-to-event data. Future research could extend the methods in this chapter to survival models such as the Cox proportional hazard model. In addition, we study a relatively simple specification of the logistic regression model involving linear predictors. More complicated models involving quadratic terms or splines are often required in practice. The MAI method can easily be used to adjust continuous covariates that are later transformed to fit quadratic terms or splines. Future work could to assess the performance of MAI in these models.

Bibliography

Bay, J. (1997). Unpublished dissertation thesis.

Brueren, M. M., van Limpt, P., Schouten, H. J. A., de Leeuw, P. W., and van Ree, J. W. (2008). Is a series of blood pressure measurements by the general practitioner or the patient a reliable alternative to ambulatory blood pressure measurement? A study in general practice with reference to short-term and long-term between-visit variability. *American Journal of Hypertension* **101,** 879-885.

Brunelli, S.M., Thadhani, R.I., Lynch, K.E., Ankers E.D., Joffe, M.M., Boston, R., Chang, Y., and Feldman, H.I. (2008). Association between long-term blood pressure variability and mortality among incident hemodialysis patients. *American Journal of Kidney Diseases* **52,** 716-726.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. Boca Raton, Florida: Chapman and Hall.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association* **85,** 652-663.

Cheng, C. L. and Schneeweiss, H. (1998). Polynomial regression with errors in the variables. *Journal of the Royal Statistical Socieity, Series B* **60,** 189-199.

Cheng, C. L. and Van Ness, J. (1999). *Statistical Regression with Measurement Error*. London: Arnold Publishers.

Cramer, H. (1957). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Freedman, L. S., Fainberg, V., Kipnis, V., Midthune, D., and Carroll, R. J. (2004). A new method for dealing with measurement error in explanatory variables of regression models. *Biometrics* **60,** 172-181.

Freedman, L. S., Midthune, D., Carroll, R. J., and Kipnis, V. (2008). A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Statistics in Medicine* **27,** 5195-5216.

Freudenheim, J. L. and Marshall, J. R. (1988). The problem of profound mismeasurement and the power of epidemiologic studies of diet and cancer. *Nutrition and Cancer* **11,** 243250.

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

Garcia-Vera, M. and Sanz, J. (1999). How many self-measured blood pressure readings are needed to estimate hypertensive patients' true blood pressure. *Journal of Behavioral Medicine* **22,** 93-113.

Gheorghiade, M., Abraham, W. T., Albert, N. M., Greenberg, B. H., O'Connor, C. M., She, L., Stough, W. G., Yancy, C. W., Young, J. B., and Fonarow, G. C. (2006). Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. *Journal of the American Medical Association* **296,** 2217-2226.

Gleser, L. J. (1990). In *Statistical Analysis of Measurement Error Models and Application*, P.J. Brown and W.A. Fuller (eds). Providence: American Mathematical Society.

Grove, J.S., Reed, D.M., Tano, K., and Hwang, L.J. (1997). Variability in systolic blood pressure - a risk factor for coronary heart disease? *American Journal of Epidemiology* **145,** 771-776.

Havlik, R.J., Foley, D.J., Sayer, B., Masaki, K., White, L., and Launer, L.J. (2002). Variability in midlife systolic blood pressure is related to late-life brain white matter lesions. *Stroke* **33,** 26-30.

Huang, X., Stefanski, L. A., and Davidian, M. (2006). Latent-model robustness in structural measurement error models. *Biometrika* **93,** 53-64.

Huang, Y., and Wang C.Y. (2001). Consistent functions methods for logistic regression with errors in covariates. *Journal of the American Statistical Association* **96,** 1469-1482.

Iribarren, C.I., Sharp, D.S., Burchfiel, C.M., and Petrovitch, H.P. (1995). Association of weight loss and weight fluctuation with mortality among Japanese American men. *The New England Journal of Medicine* **333,** 686-692.

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

Liu, K., Stammler, J., Dyer, A., Mc Keever, J., and McKeever, P. (1978). Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *Journal of Chronic Diseases* **31,** 399418.

Louis, T. A. (1984). Estimating a population of parameter values using bayes and empirical bayes methods. *Journal of the Americal Statistical Association* **79,** 393-398.

Lyles, R.H., Munoz, A., Xu, J., Taylor, J.M.G., and Chmiel, J.S. (1999). Adjusting for measurement error to asses health effects of variability in biomarkers. *Statistics in Medicine* **18,** 1069-1086.

Marshal, T. P. (2008). Blood pressure variability: the challenge of variability. *American Journal of Hypertension* **21,** 3-4.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **77,** 127-37.

Pickering, T. G., Hall, J. E., Appel, L. J., Falkner, B. E., Graves, J., Hill, M. N., Jones, D. W., Kurtz, T., Sheps, S. G., and Roccella, E. J. (2005). Recommendations for blood pressure measurement in humans and experimental animals: Part 1: Blood pressure measurement in humans: A statement for professionals for the subcommittee of professional and public education of the American Heart Association Council on High Blood Pressure Research *Hypertension* **45,** 142-161.

Rosner, B., Spiegelman, D., and Willett, W. C. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* **8,** 1051-1070.

Shen, W. and Louis, T. A. (1998). Triple-goal estimates in two stage hierarchical models. *Journal of the Royal Statistical Society, Series B* **60,** 455-471.

Shohat, J. A. and Tamarkin, J. D. (1943). *The problem of moments*. Providence, R.I.: American Mathematical Society.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics, Series A* **18,** 4335-4358.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics* **13,** 1335-1351.

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

Stefanski, L. A., Novick, S.J., and Devanarayan, V. (2005). Estimating a nonlinear function of a normal mean. *Biometrika* **92,** 732-736.

Stulajter, F. (1978). Nonlinear estimators of polynomials in mean values of a Gaussian stochastic process. *Kybernetika* **14,** 206-220.

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.

Wang, C. Y. (2006). Corrected score estimator for joint modeling of longitudinal and faiulre time data. *Statistica Sinica* **16,** 235-253.

Wang, C. Y. and Huang, Y. (2001). Functional methods for logistic regression on random-effect-coefficients for longitudinal measurements. *Statistics and Probability Letters* **53,** 347-356.

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

Yang, W., Israni, R.K., Brunelli, S.M., Joffe, M.M., Fishbance, S., and Feldman, H.I. (2007). Hemoglobin variability and mortality in ESRD. *Journal of the American Society of Nephrology* **18,** 3164-3170.

Appendices

Appendix A

Supplement to Chapter 1

A.1 Implementation of Newton Raphson

We wish to minimize $n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_i-X_i)^2$ subject to moment constraints imposed by Lagrange multipliers. The objective function is

$$Q_{MK}(X_1,...,X_n,\Lambda)=n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_i-X_i)^2+\sum_{k=1}^{K}\sum_{r=1}^{M_k}\frac{\lambda_{rk}}{r}(n^{-1}\sum_{i=1}^{n}X_i^rV_{i,k}-m_{rk}). \quad \text{(A.1)}$$

Differentiation with respect to X_i and $\Lambda=(\lambda_{11},...,\lambda_{M_kK})$ gives the equations $n^{-1}(X_i - W_i+\sum_{k=1}^{K}\sum_{r=1}^{M_K}\lambda_{rk}X_i^{r-1}V_{i,k})=0$ for $i=1,\ldots,n$, and $(1/r)(n^{-1}\sum_{i=1}^{n}X_i^rV_{i,k}-m_{rk})=0$ for $k=1,\ldots,K$ and $r=1,\ldots,M_k$. These can be solved jointly for $\widehat{\Theta}=(\widehat{X}_1,...,\widehat{X}_n,\widehat{\Lambda})$ using software packages, such as R multiroot(). However, this may be slow or encounter convergence problems. Alternatively, we implemented our own Newton-Raphson iteration to solve these equations. The Newton-Raphson update at iteration t is

$$\begin{aligned}\Delta_{\lambda,t+1} &= \Lambda_{t+1}-\Lambda_t=-(\boldsymbol{B}_t^T\boldsymbol{A}_t^{-1}\boldsymbol{B}_t)^{-1}\{\boldsymbol{B}_t^T\boldsymbol{A}_t^{-1}\ell_x(\Theta_t)-\ell_\lambda(\Theta_t)\},\\ \Delta_{x,t+1} &= \boldsymbol{X}_{t+1}-\boldsymbol{X}_t=-\boldsymbol{A}_t^{-1}\{\ell_x(\Theta_t)+\boldsymbol{B}_t\Delta_{\lambda,t+1}\},\end{aligned}$$

where $\ell_x(\Theta)$ is the $(n\times 1)$ vector with components $n^{-1}(X_i-W_i+\sum_{k=1}^{K}\sum_{r=1}^{M_K}\lambda_{rk}X_i^{r-1}V_{i,k})$, $\ell_\lambda(\Theta)$ is the $(\sum_{k=1}^{K}M_k\times 1)$ vector with components $(1/r)(n^{-1}\sum_{i=1}^{n}X_i^rV_{i,k}-m_{rk})$, $\boldsymbol{A}$ is an $(n\times n)$ diagonal matrix with elements $n^{-1}(1+\sum_{k=1}^{K}\sum_{r=2}^{M_K}\lambda_{rk}(r-1)X_i^{r-2}V_{i,k})$, and $\boldsymbol{B}$ is the $(n\times\sum_{k=1}^{K}M_k)$ matrix with i^{th} row $n^{-1}(V_{1i},X_iV_{1i},...X_i^{M_1-1}V_{1i},...V_{ki},X_iV_{ki},...X_i^{M_k-1}V_{ki})$.

This algorithm requires starting values for $(\widehat{X}_1, ..., \widehat{X}_n, \widehat{\Lambda})$. We let $\widehat{X}_{i,1} = \gamma_0 + \gamma_1 W_i + \cdots + \gamma_P W_i^P$ for $P = \sum_{k=1}^{K} M_k - 1$ denote the initial adjusted value for $\widehat{X}_i$. The parameters $(\gamma_0, \gamma_1, \ldots, \gamma_{M-1})$ are chosen to minimize $\sum_{k=1}^{K} \sum_{r=1}^{M_k} (m_{rk} - n^{-1} \sum_{i=1}^{n} \widehat{X}_{i,1}^r V_k)^2$. We perform this minimization using standard software like the R nlm() function. The $\widehat{X}_{i,1}$, along with $\widehat{\Lambda}$=$\mathbf{0}$, provide good starting values for the Newton-Raphson algorithm. Sometimes convergence is not achieved, and a simpler alternative works well. This is to let $\widehat{X}_{i,1} = W_i + U_i$ where $U_i \sim N(0, \hat{\sigma}_w^2/10)$ and $\hat{\sigma}_w^2 = n^{-1} \sum_{i=1}^{n} (W_i - \bar{W})^2$. In all of the cases we considered, these starting values were adequate to reach convergence.

A.2 Simulations in Kernel Density Estimation

Table A.1: Simulation results for three latent variable distributions, $f_X(x)$; two reliability ratios (RR), $B = 500$ simulated data sets, and $n = 300$. Statistics reported: (a) $MSE(W)/MSE(\widehat{X})$, where $MSE(\widehat{X}) = B^{-1}\sum_{b=1}^{B} n^{-1}\sum_{i=1}^{n}(\widehat{X}_{i,b} - X_{i,b})^2$ (coefficient of variation $\approx$ 0.001), and (b) $ISE(G_W)/ISE(G_{\widehat{X}})$, where $ISE(G_{\widehat{X}}) = B^{-1}\sum_{b=1}^{B}\int\{G_{\widehat{X},b}(t) - G_{X,b}(t)\}^2 dt$, for $G_X(t) = n^{-1}\sum_{i=1}^{n} I_{(X_i \le t)}$, $-\infty < t < \infty$ (coefficient of variation ≈ 0.02). Adjusted data $\widehat{X}$: RC, regression calibration; $M = 2$, $M = 4$, $M = 6$, MAI matching 2, 4 or 6 respectively; SNP, semi-nonparametric.

Distribution	RR	$\widehat{X}_{RC}$	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	$\widehat{X}_{M=6}$	$\widehat{X}_{SNP}$	cdf_{SNP}
			(a) $\frac{MSE(W)}{MSE(\widehat{X})}$				
Normal	0.75	1.33	1.24	1.24	1.18	1.33	-
	0.50	1.98	1.70	1.65	1.54	1.98	-
Chi Sq df=4	0.75	1.33	1.24	1.37	1.31	1.40	-
	0.50	1.99	1.70	1.82	1.67	2.11	-
Bimodal	0.75	1.32	1.24	1.47	1.31	1.62	-
	0.50	2.00	1.70	1.74	1.66	2.05	-
			(b) $\frac{ISE(W)}{ISE(\widehat{X})}$				
Normal	0.75	1.09	3.10	2.62	0.93	1.06	2.16
	0.50	1.32	7.84	2.50	1.07	1.26	7.63
Chi Sq df=4	0.75	1.27	2.00	3.26	1.43	0.65	1.09
	0.50	1.67	3.58	3.33	1.62	1.28	3.31
Bi-modal	0.75	0.81	1.26	2.86	1.16	2.07	4.59
	0.50	0.86	1.75	2.66	1.55	1.01	2.00

Table A.2: Simulation results for estimation of $f_X(x)$, $B = 500$, $n = 1000$ (Table entries are as in Table A.1)

Distribution	RR	$\widehat{X}_{RC}$	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	$\widehat{X}_{M=6}$	$\widehat{X}_{SNP}$	cdf_{SNP}
			(a) $\frac{MSE(W)}{MSE(\widehat{X})}$				
Normal	0.75	1.33	1.24	1.24	1.23	1.32	-
	0.50	1.99	1.71	1.70	1.55	1.99	-
Chi Sq df=4	0.75	1.33	1.24	1.38	1.37	1.43	-
	0.50	1.99	1.71	1.88	1.79	2.15	-
Bimodal	0.75	1.33	1.24	1.50	1.43	1.64	-
	0.50	2.00	1.71	1.79	1.64	2.15	-
			(b) $\frac{ISE(W)}{ISE(\widehat{X})}$				
Normal	0.75	1.09	7.90	7.03	4.20	1.08	5.99
	0.50	1.34	23.72	11.65	0.82	1.33	9.64
Chi Sq df=4	0.75	1.39	2.41	6.96	5.15	0.61	1.03
	0.50	1.74	4.39	10.99	3.15	1.19	4.00
Bi-modal	0.75	0.81	1.32	5.05	1.94	2.35	4.45
	0.50	0.86	1.90	4.13	1.29	1.48	8.92

Table A.3: Simulation results for estimation of $f_X(x)$, $B = 500$, $n = 2000$ (Table entries are as in Table A.1)

Distribution	RR	$\widehat{X}_{RC}$	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	$\widehat{X}_{M=6}$	$\widehat{X}_{SNP}$	cdf_{SNP}
			(a) $\frac{MSE(W)}{MSE(\widehat{X})}$				
Normal	0.75	1.33	1.24	1.24	1.24	1.33	-
	0.50	2.00	1.71	1.70	1.59	2.00	-
Chi Sq df=4	0.75	1.33	1.24	1.38	1.38	1.43	-
	0.50	2.00	1.71	1.89	1.82	2.18	-
Bimodal	0.75	1.33	1.24	1.51	1.48	1.65	-
	0.50	2.00	1.71	1.82	1.68	2.15	-
			(b) $\frac{ISE(W)}{ISE(\widehat{X})}$				
Normal	0.75	1.13	15.00	13.20	9.71	1.13	2.16
	0.50	1.40	47.49	24.90	1.14	1.39	7.89
Chi Sq df=4	0.75	1.39	2.51	9.08	10.05	0.59	1.13
	0.50	1.81	4.60	18.75	4.76	1.21	3.00
Bi-modal	0.75	0.81	1.33	6.18	3.75	2.40	4.01
	0.50	0.85	1.92	5.91	1.41	1.54	15.22

A.3 Approximating variance for regression model parameters

When $\widehat{X}_i$ are used in a regression model, the usual standard errors for regression parameter estimates will not be correct. When the regression parameter estimators are M-estimators, standard errors may be obtained by the empirical sandwich approach. This is possible because the adjusted data are functions of observed data $(W_i, \boldsymbol{V}_i)$ and $\widehat{\Lambda}$, which is an M-estimator solving an equation of the form $n^{-1}\sum_{i=1}^{n}\Psi_\Lambda(W_i, \boldsymbol{V}_i, \widehat{\Lambda})=0$. The equations that define $\widehat{\Lambda}$ can be stacked with the usual regression parameter estimating equations to obtain approximate variances. We illustrate this in a variety of cases, beginning with simple linear regression for simplicity. We then show the extension to logistic regression models that are considered in Chapter 1.

A.3.1 Linear regression, matching only two moments and a cross product with response

Suppose we are interested in the linear regression model $Y = \beta_0 + \beta_X X + \epsilon$, for $\epsilon \sim N(0, \sigma_\epsilon^2)$. In place of X we observe W, and we obtain adjusted values by MAI. For simplicity, we match only two moments of X and a cross product with Y. The objective function, with Lagrange multipliers $\Lambda = (\lambda_{11}, \lambda_{21}, \lambda_{12})$, is

$$n^{-1}\sum_{i=1}^{n}(W_i - X_i)^2 + \lambda_{11}(n^{-1}\sum_{i=1}^{n}X_i - m_{11}) + \lambda_{21}(n^{-1}\sum_{i=1}^{n}X_i^2 - m_{21}) + \lambda_{12}(n^{-1}\sum_{i=1}^{n}X_iY_i - m_{12}). \tag{A.2}$$

As described in Section 1.2.1, we take the derivative with respect to $(X_1, ..., X_n, \Lambda)$, equate this to 0, and solve for $(\widehat{X}_1, ..., \widehat{X}_n, \widehat{\Lambda})$. In this simple case, the equations have a closed form solution. The derivative of this objective function with respect to X_i is $n^{-1}(X_i - W_i + \lambda_{11} + \lambda_{21}X_i + \lambda_{12}Y_i)$. Setting this equal to 0 and solving gives $X_i = h(W_i, Y_i, \Lambda) = (W_i - \lambda_{11} - \lambda_{12}Y_i)/(1+\lambda_{21})$. The derivative of the objective function with respect to $\Lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21})^T$ gives the equations $n^{-1}\sum_{i=1}^{n}\Psi_\Lambda(W_i, Y_i, \sigma_{ui}^2, \Lambda)=0$, where

$$\Psi_\Lambda(w, y, \sigma_u^2, \Lambda) = \begin{bmatrix} \{h(w, y, \Lambda) - w\}/n \\ \{h^2(w, y, \Lambda) - w^2 + \sigma_u^2\}/2n \\ \{h(w, y, \Lambda) - w\}y/n \end{bmatrix}.$$

Our estimator for Λ is $\widehat{\Lambda}$, which solves $n^{-1}\sum_{i=1}^{n}\Psi_{\Lambda}(W_i, Y_i, \sigma_{ui}^2, \widehat{\Lambda})$=0. We see that the first n equations define the form of $h(W_i, Y_i, \Lambda)$, and $\widehat{X}_i = h(W_i, Y_i, \widehat{\Lambda})$ depends on the observed data and M-estimators, $\widehat{\Lambda}$.

Using the adjusted data, we fit the regression model Y=$\beta_0 + \beta_X\widehat{X} + \epsilon = \beta_0 + \beta_X h(W, Y, \widehat{\Lambda}) + \epsilon$. The least squares estimator $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_X)$ solves
$n^{-1}\sum_{i=1}^{n}\Psi_{\beta,i}(W_i, Y_i, \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}) = 0$ for

$$\Psi_{\beta}(w, y, \boldsymbol{\beta}, \Lambda) = \begin{bmatrix} y - \beta_0 - \beta_X h(w, y, \Lambda) \\ \{y - \beta_0 - \beta_X h(w, y, \Lambda)\}h(w, y, \Lambda) \end{bmatrix}.$$

Because this equation involves $\widehat{\Lambda}$ we can account for the adjustment by stacking equations, i.e.,

$$\Psi(w, y, \boldsymbol{\beta}, \Lambda) = \begin{bmatrix} \Psi_{\Lambda}(w, y, \sigma_u^2, \Lambda) \\ \Psi_{\beta}(w, y, \boldsymbol{\beta}, \Lambda) \end{bmatrix}.$$

Letting $\theta = (\boldsymbol{\beta}, \Lambda)$ we see that $\widehat{\theta} = (\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})$ is an M-estimator solving
$n^{-1}\sum_{i=1}^{n}\Psi(W_i, Y_i, \widehat{\boldsymbol{\beta}}, \widehat{\Lambda})$=0. The solution $\widehat{\boldsymbol{\beta}}$ estimates the regression parameters for $Y = \beta_0 + \beta_X\widehat{X} + \epsilon$. These are the same estimators that are obtained from ordinary least squares of Y on $\widehat{X}$. For the purpose of variance estimation, it is necessary to consider the stacked equations $\Psi(w, y, \boldsymbol{\beta}, \Lambda)$.

The sandwich variance approximation follows from the normal approximation $n^{1/2}(\widehat{\theta} - \theta_0) \dot{\sim} MVN(\mathbf{0}, A^{-1}B(A^{-1})^T)$, where θ_0 is defined by $E\{\Psi(W, Y, \theta_0)\} = 0$, $B = E\{\Psi(W, Y, \theta_0)\Psi^T(W, Y, \theta_0)\}$, $A = E\{\dot{\Psi}(W, Y, \theta_0)\}$, and $\dot{\Psi}(w, y, \theta) = \partial\Psi(w, y, \theta)/\partial\theta^T$. A and B can be estimated by $\widehat{A}_n = n^{-1}\sum_{i=1}^{n}\dot{\Psi}(W_i, Y_i, \widehat{\theta})$ and
$\widehat{B}_n = n^{-1}\sum_{i=1}^{n}\Psi(W_i, Y_i, \widehat{\theta})\Psi(W_i, Y_i, \widehat{\theta})^T$. Then the empirical sandwich variance of our estimators is $\widehat{\boldsymbol{V}} = \widehat{A}_n^{-1}\ \widehat{B}_n\ \{\widehat{A}_n^{-1}\}^T/n$. In this case, the variance estimator for $\widehat{\beta}_0$ is $\widehat{V}_{44}$ and for $\widehat{\beta}_X$ is $\widehat{V}_{55}$.

The variance estimator for $\widehat{\boldsymbol{\beta}}$ can alternatively be calculated by partitioning $\widehat{A}_n$ and $\widehat{B}_n$ as follows. For $\Psi_{\Lambda,i}$=$\Psi_{\Lambda}(W_i, Y_i, \widehat{\theta})$ and $\Psi_{\beta,i}$=$\Psi_{\beta}(W_i, Y_i, \widehat{\theta})$ let

$$\widehat{A}_n = \begin{bmatrix} \widehat{A}_{11} = n^{-1}\sum_{i=1}^{n}\partial\Psi_{\Lambda,i}/\partial\Lambda^T & \widehat{A}_{12} = n^{-1}\sum_{i=1}^{n}\partial\Psi_{\Lambda,i}/\partial\beta^T \\ \widehat{A}_{21} = n^{-1}\sum_{i=1}^{n}\partial\Psi_{\beta,i}/\partial\Lambda^T & \widehat{A}_{22} = n^{-1}\sum_{i=1}^{n}\partial\Psi_{\beta,i}/\partial\beta^T \end{bmatrix}$$

$$\widehat{B}_n = \begin{bmatrix} \widehat{B}_{11} = n^{-1}\sum_{i=1}^{n}\Psi_{\Lambda,i}\Psi_{\Lambda,i}^T & \widehat{B}_{12} = n^{-1}\sum_{i=1}^{n}\Psi_{\Lambda,i}\Psi_{\beta,i}^T \\ \widehat{B}_{21} = n^{-1}\sum_{i=1}^{n}\Psi_{\beta,i}\Psi_{\Lambda,i}^T & \widehat{B}_{22} = n^{-1}\sum_{i=1}^{n}\Psi_{\beta,i}\Psi_{\beta,i}^T \end{bmatrix}.$$

Then $\widehat{V}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\widehat{A}_{11}^{-1}\{\widehat{B}_{12} - \widehat{A}_{21}\widehat{A}_{11}^{-1}\widehat{B}_{12} - \widehat{B}_{12}^T(\widehat{A}_{11}^{-1})^T\widehat{A}_{21}^T + \widehat{A}_{21}\widehat{A}_{11}^{-1}\widehat{B}_{11}(\widehat{A}_{11}^{-1})^T\widehat{A}_{21}^T\}(\widehat{A}_{11}^{-1})^T$.

All of the operations required to obtain this variance estimator can be performed numerically. The estimating equations can by solved by the R multiroot() function or methods described in Section A.1, and derivatives can be computed by the R jacobian() function. This can be a slow process, and we improve the speed by calculating $\dot{\Psi}(w, y, \theta) = \partial\Psi(w, y, \theta)/\partial\theta^T$ analytically. For the case of linear regression, matching two moments and a cross product,

$$\dot{\Psi}(w, y, \theta) = \begin{bmatrix} \partial h/\partial\Lambda^T & 0 & 0 \\ h(w, y, \Lambda)\partial h/\partial\Lambda^T & 0 & 0 \\ y\partial h/\partial\Lambda^T & 0 & 0 \\ \beta_X\partial h/\partial\Lambda^T & 1 & h(w, y, \Lambda) \\ -(y - \beta_0 - 2\beta_X h(w, y, \Lambda))\partial h/\partial\Lambda^T & h(w, y, \Lambda) & h^2(w, y, \Lambda) \end{bmatrix},$$

where $\partial h/\partial\Lambda^T = \{-1/(1 + \lambda_{21}), -(w - \lambda_{11} - y\lambda_{12})/(1 + \lambda_{21})^2, -y/(1 + \lambda_{21})\}$.

A.3.2 Linear regression, matching four moments and two cross product with response

Allowing a little more complexity, we can fit the same model using $\widehat{X}$ for which four moments of X and two cross products with Y have been matched. The objective function, with Lagrange multipliers $\Lambda = (\lambda_{11}, \ldots, \lambda_{41}, \lambda_{12}, \lambda_{22})$, is

$$n^{-1}\sum_{i=1}^{n}\frac{1}{2}(W_i - X_i)^2 + \sum_{r=1}^{4}\frac{\lambda_{r1}}{r}(n^{-1}\sum_{i=1}^{n}X_i^r - m_{r1}) + \sum_{r=1}^{2}\frac{\lambda_{r2}}{r}(n^{-1}\sum_{i=1}^{n}X_i^r Y_i - m_{r2}). \quad \text{(A.3)}$$

The derivative with respect to X_i is equated to 0 and we get

$$X_i - W_i + \sum_{r=1}^{4}\lambda_{r1}X_i^{r-1} + \sum_{r=1}^{2}\lambda_{r2}X_i^{r-1}Y_i = 0. \quad \text{(A.4)}$$

It is no longer straightforward to solve for X_i. Instead, we define $h(W_i, Y_i, \Lambda)$ implicitly as the solution to this equation. The estimating equations for Λ are

$n^{-1}\sum_{i=1}^{n}\Psi_\Lambda(W_i, Y_i, \sigma_{ui}^2, \Lambda)=0$ where

$$\Psi_\Lambda(w, y, \sigma_u^2, \Lambda) = \begin{bmatrix} \{h(w,y,\Lambda) - w\}/n \\ \{h^2(w,y,\Lambda) - w^2 + \sigma_u^2\}/2n \\ \{h^3(w,y,\Lambda) - w^3 + 3w\sigma_u^2\}/3n \\ \{h^4(w,y,\Lambda) - w^4 + 6w^2\sigma_u^2 + 3\sigma_u^4\}/4n \\ \{h(w,y,\Lambda) - w\}y/n \\ \{h^2(w,y,\Lambda) - w^2 + \sigma_u^2\}y/2n \end{bmatrix}.$$

These are the only changes to Section A.3.2 needed to account for additional matching and $\Psi(w, y, \boldsymbol{\beta}, \Lambda) = \{\Psi_\Lambda^T(w, y, \sigma_u^2, \Lambda), \Psi_\beta^T(w, y, \boldsymbol{\beta}, \Lambda)\}^T$ with $\Psi_\beta(w, y, \boldsymbol{\beta}, \Lambda)$ defined previously. Numerical calculation of the empirical sandwich variance is straightforward based on these equations. For the analytical derivative there are additional considerations. Firstly, $\dot{\Psi}(w, y, \theta)$ involves $\partial h/\partial\Lambda^T$, but we no longer have an explicit form of $h(w, y, \Lambda)$. In the present problem, $\partial h/\partial\Lambda^T$ can be found by noting that $w = h(w, y, \Lambda) + \sum_{r=1}^{4}(\lambda_{r1}/n)h(w, y, \Lambda)^{r-1} + \sum_{r=1}^{2}(\lambda_{r2}/n)h(w, y, \Lambda)^{r-1}y = g\{h(w, y, \Lambda), y, \Lambda\}$. Differentiating g with respect to Λ gives

$$0 = \frac{\partial g}{\partial \Lambda^T} = \frac{\partial g}{\partial h}\frac{\partial h}{\partial \Lambda^T} + \frac{\partial g}{\partial \Lambda^T}\frac{\partial \Lambda}{\partial \Lambda^T}.$$

Solving for $\partial h/\partial\Lambda^T$ gives $\partial h/\partial\Lambda^T = -\frac{\partial g/\partial\Lambda^T}{\partial g/\partial h}$. For the case of linear regression, matching four moments and two cross products with the response,

$$\partial h/\partial\Lambda^T = \frac{-(1, h, h^2, h^3, y, yh)}{1 + \sum_{r=2}^{4}(r-1)\lambda_{r1}h^{r-2} + 2\lambda_{22}hy},$$

and

$$\dot{\Psi}(w, y, \theta) = \begin{bmatrix} \partial h/\partial\Lambda^T & 0 & 0 \\ h(w,y,\Lambda)\partial h/\partial\Lambda^T & 0 & 0 \\ h^2(w,y,\Lambda)\partial h/\partial\Lambda^T & 0 & 0 \\ h^3(w,y,\Lambda)\partial h/\partial\Lambda^T & 0 & 0 \\ y\partial h/\partial\Lambda^T & 0 & 0 \\ h(w,y,\Lambda)y\partial h/\partial\Lambda^T & 0 & 0 \\ \beta_X\partial h/\partial\Lambda^T & 1 & h(w,y,\Lambda) \\ -(y - \beta_0 - 2\beta_X h(w,y,\Lambda))\partial/\partial\Lambda^T & h(w,y,\Lambda) & h^2(w,y,\Lambda) \end{bmatrix}.$$

A.3.3 Linear regression, matching moments and cross products

In addition to the latent variable of interest X, we may have error-free covariates $\boldsymbol{Z}$. The linear regression model is $Y = \beta_0 + \boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_X X + \epsilon$, for $\epsilon \sim N(0, \sigma_\epsilon^2)$. Some of the error-free covariates, $\boldsymbol{Z}$, may be highly correlated with X, and we wish to match cross-products with these components as well as with the response Y. Let $\boldsymbol{Z}^*$ be the subset of $\boldsymbol{Z}$ which are correlated with X. Then we wish to match cross-products with the columns of the $(n \times K)$ matrix $\boldsymbol{V}$=$\boldsymbol{(1, Y, Z^*)}$. We match four moments of X and two cross products with the components of $\boldsymbol{V}$, i.e. $M = (4, 2, \ldots, 2)$.

The objective function, with Lagrange multipliers $\Lambda = (\lambda_{11}, \ldots, \lambda_{M_k K})$, is

$$n^{-1} \sum_{i=1}^{n} \frac{1}{2}(W_i - X_i)^2 + \sum_{k=1}^{K} \sum_{r=1}^{M_k} \frac{\lambda_{rk}}{r}(n^{-1} \sum_{i=1}^{n} X_i^r V_{i,k} - m_{rk}). \tag{A.5}$$

Differentiation with respect to X_i gives the equation $n^{-1}(X_i - W_i + \sum_{k=1}^{K} \sum_{r=1}^{M_K} \lambda_{rk} X_i^{r-1} V_{i,k})$ $= 0$ and $h(W_i, \boldsymbol{V}_i, \Lambda)$ is the implicitly defined solution to this equation.

The estimating equations for Λ are $n^{-1} \sum_{i=1}^{n} \Psi_\Lambda(W_i, \boldsymbol{V}_i, \sigma_{ui}^2, \Lambda)$=0 where

$$\Psi_\Lambda(w, \boldsymbol{v}, \sigma_u^2, \Lambda) = \begin{bmatrix} \{h(w, \boldsymbol{v}, \Lambda) - w\}/n \\ \{h^2(w, \boldsymbol{v}, \Lambda) - w^2 + \sigma_u^2\}/2n \\ \{h^3(w, \boldsymbol{v}, \Lambda) - w^3 + 3w\sigma_u^2\}/3n \\ \{h^4(w, \boldsymbol{v}, \Lambda) - w^4 + 6w^2\sigma_u^2 + 3\sigma_u^4\}/4n \\ \{h(w, \boldsymbol{v}, \Lambda) - w\}v_2/n \\ \{h^2(w, \boldsymbol{v}, \Lambda) - w^2 + \sigma_u^2\}v_2/2n \\ \vdots \\ \{h(w, \boldsymbol{v}, \Lambda) - w\}v_K/n \\ \{h^2(w, \boldsymbol{v}, \Lambda) - w^2 + \sigma_u^2\}v_K/2n \end{bmatrix}.$$

Based on the adjusted data we fit the regression model $\widehat{Y} = \beta_0 + \boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_X h(W, \boldsymbol{V}, \widehat{\Lambda})$. The least squares estimator $\widehat{\boldsymbol{\beta}}$ solves $n^{-1} \sum_{i=1}^{n} \Psi_\beta(W_i, Y_i, \boldsymbol{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\Lambda})$ for

$$\Psi_\beta(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) = \begin{bmatrix} y - \beta_0 - \boldsymbol{\beta}_Z^T \boldsymbol{z} - \beta_X h(w, \boldsymbol{v}, \Lambda) \\ \{y - \beta_0 - \boldsymbol{\beta}_Z^T \boldsymbol{z} - \beta_X h(w, \boldsymbol{v}, \Lambda)\}\boldsymbol{z} \\ \{y - \beta_0 - \boldsymbol{\beta}_Z^T \boldsymbol{z} - \beta_X h(w, \boldsymbol{v}, \Lambda)\}h(w, \boldsymbol{v}, \Lambda) \end{bmatrix}.$$

Combining these estimating equations, we get $\Psi(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) =$

$(\Psi_\Lambda^T(w, \boldsymbol{v}, \sigma_u^2, \Lambda), \Psi_\beta^T(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda))^T$. The derivative of the estimating equations with respect to the parameters is

$$\dot{\Psi}(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) = \begin{bmatrix} (1, h, h^2, h^3, v_2, v_2 h, \ldots, v_K, v_K h)^T \partial h/\partial \Lambda^T & \mathbf{0} \\ (\beta_X, \beta_X \boldsymbol{z}^T, -(y - \beta_0 - \boldsymbol{\beta}_Z^T \boldsymbol{z} - 2h\beta_x))^T \partial h/\partial \Lambda^T & \begin{matrix} 1 & \boldsymbol{z}^T & h \\ \boldsymbol{z} & \boldsymbol{z}\boldsymbol{z}^T & \boldsymbol{z}h \\ h & h\boldsymbol{z}^T & h^2 \end{matrix} \end{bmatrix}$$

where $\mathbf{0}$ is an $(\sum_{k=1}^K M_k \times K)$ matrix of zeros and

$$\partial h/\partial \Lambda^T = \frac{-(1, h, h^2, h^3, v_2, v_2 h, \ldots, v_K, v_K h)}{1 + \sum_{k=1}^K \sum_{r=2}^{M_k} \lambda_{rk}(r-1) h^{r-2} v_k}$$

.

A.3.4 Logistic Regression

Our regression model for the relationship between a binary outcome, Y, latent variable, X, and error-free covariates, $\boldsymbol{Z}$, is the logistic model $P(Y = 1|X, \boldsymbol{Z}) = F(\beta_0 + \boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_X X)$ where $F(v) = 1 + \exp(-v)^{-1}$. In place of X we observe W, and we obtain adjusted values by MAI matching four moments of X and two cross products with the components of $\boldsymbol{V}$, where $\boldsymbol{V}$=$\boldsymbol{(1, Y, Z^*)}$ as in the previous section.

The consideration of a new outcome model requires no change to the adjusted data. The objective function A.5 is still applicable as is $\Psi_\Lambda(w, \boldsymbol{v}, \sigma_u^2, \Lambda)$. Based on the adjusted data we fit the regression model $\widehat{P}(Y = 1|X, Z) = F\{\beta_0 + \boldsymbol{\beta}_Z^T \boldsymbol{Z} + \beta_X h(W, \boldsymbol{V}, \widehat{\Lambda})\}$.

The maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ solves $n^{-1} \sum_{i=1}^n \Psi_\beta(W_i, Y_i, \boldsymbol{Z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\Lambda})$ for

$$\Psi_\beta(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) = \begin{bmatrix} y - F \\ (y - F)\boldsymbol{z} \\ (y - F)h(w, \boldsymbol{v}, \Lambda) \end{bmatrix}$$

where $F = F\{\beta_0 - \boldsymbol{\beta}_Z^T \boldsymbol{z} - \beta_X h(w, \boldsymbol{v}, \Lambda)\}$. The parameter estimate $\widehat{\theta} = (\widehat{\boldsymbol{\beta}}, \widehat{\Lambda})^T$ solves $n^{-1} \sum_{i=1}^n \Psi(W_i, Y_i, \widehat{\boldsymbol{\beta}}, \widehat{\Lambda})$=0 for $\Psi(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) = \{\Psi_\Lambda^T(w, \boldsymbol{v}, \sigma_u^2, \Lambda), \Psi_\beta^T(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda)\}^T$.

The derivative of the estimating equations with respect to the parameters is

$$\dot{\Psi}(w, y, \boldsymbol{z}, \boldsymbol{\beta}, \Lambda) =$$

$$\begin{bmatrix} (1, h, h^2, h^3, v_2, v_2 h, \ldots, v_K, v_K h)^T \partial h / \partial \Lambda^T & \mathbf{0} \\ (F^2 - F)\beta_X \partial h / \partial \Lambda^T & \\ (F^2 - F)\beta_X \boldsymbol{z} \partial h / \partial \Lambda^T & (F^2 - F) \begin{pmatrix} 1 & \boldsymbol{z}^T & h \\ \boldsymbol{z} & \boldsymbol{z}\boldsymbol{z}^T & \boldsymbol{z}h \\ h & h\boldsymbol{z}^T & h^2 \end{pmatrix} \\ \{(y - F) + (F^2 - F)h\beta_x\} \partial h / \partial \Lambda^T & \end{bmatrix}$$

where

$$\partial h / \partial \Lambda^T = \frac{-(1, h, h^2, h^3, v_2, v_2 h, \ldots, v_K, v_K h)}{1 + \sum_{k=1}^{K} \sum_{r=2}^{M_k} \lambda_{rk}(r-1) h^{r-2} v_k}.$$

A.4 Simulations in Logistic Regression

Table A.4: *Estimation of β_X for $P(Y=1|X,Z)=F(\beta_0+\beta_X X+\beta_Z Z)$, where X is normally distributed, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.05 and 0.12). Adjusted data $\widehat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_0,\beta_X,\beta_Z)=(-1.5,1,1)$						$(\beta_0,\beta_X,\beta_Z)=(-.6,.3,.3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.29	0.00	0.04	0.04	0.05	0.06	−0.09	0.00	0.00	0.00	0.00	0.00
		SD	0.15	0.22	0.24	0.24	0.25	0.26	0.11	0.16	0.16	0.16	0.16	0.16
		MSE-R	1.00	2.24	1.72	1.81	1.62	1.53	1.00	0.81	0.78	0.78	0.78	0.78
	1000	B	−0.43	−0.03	0.02	0.02	0.03	0.04	−0.09	0.00	0.00	0.00	0.00	0.00
		SD	0.16	0.14	0.17	0.16	0.18	0.18	0.06	0.09	0.09	0.09	0.09	0.09
		MSE-R	1.00	9.87	7.44	7.73	6.59	6.17	1.00	1.48	1.43	1.43	1.43	1.43
	2000	B	−0.31	−0.04	0.00	−0.01	0.00	0.01	−0.09	0.00	0.00	0.00	0.00	0.00
		SD	0.06	0.09	0.10	0.09	0.10	0.10	0.05	0.07	0.07	0.07	0.07	0.07
		MSE-R	1.00	11.18	10.87	11.01	10.38	10.12	1.00	2.24	2.18	2.18	2.18	2.18
0.50	300	B	−0.58	−0.05	0.04	0.02	0.08*	0.24*	−0.16	0.01	0.02	0.02	0.02	0.03
		SD	0.11	0.29	0.37	0.35	0.49*	3.56*	0.09	0.21	0.22	0.22	0.22	0.22
		MSE-R	1.00	4.07	2.48	2.74	1.37*	0.03*	1.00	0.78	0.70	0.70	0.68	0.65
	1000	B	−0.57	−0.05	0.01	0.00	0.02	0.03	−0.16	0.00	0.01	0.01	0.01	0.01
		SD	0.07	0.16	0.19	0.19	0.20	0.21	0.05	0.11	0.12	0.12	0.12	0.12
		MSE-R	1.00	11.67	8.85	9.25	8.26	7.42	1.00	2.30	2.14	2.14	2.11	2.11
	2000	B	−0.58	−0.07	−0.01	−0.02	0.00	0.01	−0.16	0.00	0.01	0.01	0.01	0.01
		SD	0.05	0.11	0.13	0.13	0.14	0.14	0.03	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	19.91	18.98	19.44	17.25	15.89	1.00	4.63	4.33	4.34	4.35	4.35

Table A.5: *Estimation of β_Z for $P(Y = 1|X, Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$, where X is normally distributed, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_Z(W)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.02 and 0.10). Adjusted data $\hat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$						$(\beta_0, \beta_X, \beta_Z) = (-.6, .3, .3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.08	−0.04	0.00	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.00	0.00
		SD	0.20	0.21	0.21	0.21	0.21	0.22	0.14	0.15	0.15	0.15	0.15	0.15
		MSE-R	1.00	1.05	1.03	1.03	1.02	1.00	1.00	0.96	0.96	0.96	0.96	0.96
	1000	B	0.10	−0.06	0.00	−0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		SD	0.11	0.12	0.12	0.12	0.12	0.12	0.08	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	1.35	1.65	1.67	1.63	1.58	1.00	1.08	1.08	1.08	1.08	1.08
	2000	B	0.07	−0.04	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
		SD	0.07	0.08	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	0.05	0.05
		MSE-R	1.00	1.42	1.72	1.72	1.71	1.68	1.00	1.39	1.38	1.38	1.38	1.38
0.50	300	B	0.15	−0.07	0.00	0.00	−0.01	$> 10^*$	0.07	0.00	0.01	0.01	0.01	0.01
		SD	0.20	0.22	0.23	0.23	0.24	$> 10^*$	0.13	0.15	0.15	0.15	0.15	0.15
		MSE-R	1.00	1.12	1.14	1.18	1.01	0.00^*	1.00	1.01	1.00	1.00	0.99	0.99
	1000	B	0.14	−0.07	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00
		SD	0.11	0.12	0.13	0.13	0.13	0.13	0.07	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	1.49	1.92	1.93	1.88	1.79	1.00	1.45	1.45	1.45	1.44	1.44
	2000	B	0.12	−0.08	−0.01	−0.01	−0.01	0.00	0.06	−0.01	−0.01	−0.01	−0.01	−0.01
		SD	0.07	0.08	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	0.05	0.05
		MSE-R	1.00	1.64	3.05	3.06	3.02	2.89	1.00	1.85	1.89	1.89	1.88	1.88

Table A.6: *Estimation of β_X for $P(Y = 1|X, Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$, where X is chi-square df=4, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.06 and 0.14). Adjusted data $\widehat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score.Values impacted by outliers marked by *.*

			$(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$						$(\beta_0, \beta_X, \beta_Z) = (-.6, .3, .3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.25	0.06	0.09	0.10	0.04	0.05	−0.07	0.02	0.02	0.02	0.02	0.02
		SD	0.16	0.23	0.25	0.26	0.24	0.24	0.12	0.17	0.18	0.18	0.18	0.18
		MSE-R	1.00	1.56	1.19	1.13	1.50	1.40	1.00	0.67	0.63	0.63	0.65	0.65
	1000	B	−0.27	0.02	0.05	0.06	0.00	0.00	−0.08	0.01	0.01	0.01	0.00	0.00
		SD	0.09	0.12	0.13	0.14	0.13	0.13	0.06	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	5.12	3.98	3.66	4.80	4.82	1.00	1.46	1.41	1.41	1.46	1.46
	2000	B	−0.27	0.02	0.05	0.06	0.01	0.01	−0.08	0.00	0.01	0.01	0.00	0.00
		SD	0.06	0.09	0.10	0.10	0.10	0.10	0.04	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	8.32	6.02	5.41	8.02	8.21	1.00	2.29	2.21	2.21	2.28	2.29
0.50	300	B	−0.52	0.07	0.14	0.18	0.08	0.30*	−0.16	0.01	0.02	0.02	0.02	−0.04*
		SD	0.13	0.31	0.39	0.41	0.46	1.90*	0.10	0.23	0.24	0.24	0.24	1.30*
		MSE-R	1.00	2.84	1.65	1.43	1.32	0.08*	1.00	0.69	0.61	0.62	0.60	0.02*
	1000	B	−0.52	0.06	0.10	0.13	0.02	0.03	−0.16	0.01	0.01	0.01	0.00	0.00
		SD	0.06	0.16	0.19	0.19	0.18	0.18	0.05	0.11	0.11	0.11	0.11	0.11
		MSE-R	1.00	9.85	6.38	4.96	8.81	7.93	1.00	2.30	2.15	2.14	2.27	2.27
	2000	B	−0.53	0.05	0.08	0.12	0.01	0.02	−0.16	0.00	0.00	0.00	0.00	0.00
		SD	0.05	0.12	0.14	0.15	0.14	0.15	0.04	0.08	0.09	0.09	0.09	0.09
		MSE-R	1.00	16.38	10.19	7.47	14.13	12.94	1.00	3.90	3.66	3.65	3.83	3.84

Table A.7: *Estimation of β_Z for $P(Y=1|X,Z)=F(\beta_0+\beta_X X+\beta_Z Z)$, where X is chi-square df=4, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_Z(W)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score.Values impacted by outliers marked by *.*

			$(\beta_0,\beta_X,\beta_Z)=(-1.5,1,1)$						$(\beta_0,\beta_X,\beta_Z)=(-.6,.3,.3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.08	−0.04	0.01	0.00	0.02	0.02	0.04	0.00	0.01	0.01	0.01	0.01
		SD	0.19	0.20	0.20	0.20	0.20	0.20	0.14	0.15	0.15	0.15	0.15	0.15
		MSE-R	1.00	1.05	1.05	1.05	1.06	1.04	1.00	0.95	0.95	0.95	0.95	0.95
	1000	B	0.07	−0.05	0.00	−0.01	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00
		SD	0.10	0.10	0.10	0.11	0.10	0.10	0.07	0.07	0.07	0.07	0.07	0.07
		MSE-R	1.00	1.08	1.31	1.30	1.32	1.31	1.00	1.10	1.10	1.10	1.10	1.10
	2000	B	0.06	−0.06	−0.01	−0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		SD	0.08	0.08	0.08	0.08	0.08	0.08	0.05	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.01	1.46	1.44	1.49	1.48	1.00	1.21	1.21	1.21	1.22	1.22
0.50	300	B	0.13	−0.11	−0.01	−0.02	0.00	$>10^*$	0.07	0.00	0.00	0.00	0.00	0.02*
		SD	0.19	0.22	0.22	0.22	0.23	$>10^*$	0.13	0.16	0.16	0.16	0.16	0.44*
		MSE-R	1.00	0.86	1.06	1.03	1.00	0.00*	1.00	0.92	0.91	0.91	0.91	0.11*
	1000	B	0.12	−0.12	−0.02	−0.03	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
		SD	0.10	0.11	0.11	0.11	0.11	0.12	0.07	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	0.94	1.85	1.77	1.93	1.83	1.00	1.36	1.37	1.37	1.38	1.39
	2000	B	0.11	−0.12	−0.02	−0.03	−0.01	0.00	0.06	0.00	0.00	0.00	0.00	0.00
		SD	0.07	0.08	0.08	0.08	0.08	0.08	0.05	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	0.85	2.62	2.40	2.86	2.80	1.00	1.98	2.01	2.00	2.03	2.03

Table A.8: *Estimation of β_X for $P(Y = 1|X, Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$, where X is bimodal normal, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.05 and 0.11). Adjusted data $\widehat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$						$(\beta_0, \beta_X, \beta_Z) = (-.6, .3, .3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.20	0.11	0.14	0.15	0.04	0.04	−0.07	0.02	0.03	0.03	0.02	0.02
		SD	0.16	0.22	0.25	0.25	0.21	0.21	0.12	0.16	0.17	0.17	0.16	0.16
		MSE-R	1.00	1.07	0.82	0.78	1.50	1.49	1.00	0.66	0.64	0.64	0.70	0.70
	1000	B	−0.22	0.08	0.11	0.12	0.02	0.01	−0.07	0.01	0.01	0.01	0.01	0.01
		SD	0.08	0.12	0.13	0.13	0.11	0.11	0.06	0.09	0.09	0.09	0.09	0.09
		MSE-R	1.00	2.66	1.96	1.80	4.54	4.67	1.00	1.22	1.19	1.19	1.30	1.30
	2000	B	−0.23	0.07	0.10	0.10	0.01	0.00	−0.08	0.01	0.01	0.01	0.00	0.00
		SD	0.06	0.09	0.10	0.10	0.08	0.08	0.04	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	4.33	3.08	2.77	8.22	8.61	1.00	2.38	2.33	2.33	2.52	2.53
0.50	300	B	−0.49	0.12	0.23	0.25	0.08	0.12*	−0.16	0.01	0.02	0.02	0.00	0.00
		SD	0.13	0.31	0.40	0.40	0.34	0.93*	0.09	0.20	0.21	0.21	0.20	0.20
		MSE-R	1.00	2.33	1.21	1.13	2.11	0.28*	1.00	0.83	0.76	0.76	0.86	0.86
	1000	B	−0.49	0.11	0.19	0.21	0.04	0.01	−0.16	0.02	0.02	0.02	0.01	0.01
		SD	0.07	0.16	0.20	0.20	0.17	0.16	0.05	0.11	0.12	0.12	0.11	0.11
		MSE-R	1.00	6.35	3.29	2.86	8.16	9.98	1.00	2.03	1.92	1.90	2.18	2.19
	2000	B	−0.49	0.10	0.17	0.19	0.02	0.00	−0.16	0.01	0.01	0.01	0.00	0.00
		SD	0.05	0.11	0.13	0.14	0.12	0.10	0.04	0.08	0.09	0.09	0.08	0.08
		MSE-R	1.00	10.65	5.22	4.33	17.29	23.70	1.00	3.81	3.60	3.57	4.10	4.12

Table A.9: *Estimation of β_Z for $P(Y=1|X,Z)=F(\beta_0+\beta_X X+\beta_Z Z)$, where X is bimodal normal, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.01); SD, standard deviation; MSE-R, $MSE_{\beta_Z(W)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MR, moment reconstruction; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_0,\beta_X,\beta_Z)=(-1.5,1,1)$						$(\beta_0,\beta_X,\beta_Z)=(-.6,.3,.3)$					
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{MR}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.09	−0.03	0.01	0.01	0.02	0.03	0.04	0.00	0.00	0.00	0.00	0.00
		SD	0.21	0.21	0.22	0.22	0.22	0.22	0.14	0.15	0.15	0.15	0.15	0.15
		MSE-R	1.00	1.16	1.09	1.09	1.10	1.08	1.00	0.96	0.96	0.96	0.96	0.96
	1000	B	0.07	−0.05	−0.01	−0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		SD	0.11	0.11	0.11	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.07	0.07
		MSE-R	1.00	1.11	1.22	1.22	1.26	1.24	1.00	1.10	1.09	1.10	1.10	1.10
	2000	B	0.07	−0.04	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.00	0.00	0.00
		SD	0.08	0.08	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	0.05	0.05
		MSE-R	1.00	1.42	1.68	1.69	1.70	1.67	1.00	1.28	1.27	1.27	1.28	1.27
0.50	300	B	0.16	−0.08	0.00	0.00	0.01	0.03*	0.08	0.01	0.01	0.01	0.02	0.02
		SD	0.20	0.21	0.23	0.23	0.22	0.36*	0.14	0.15	0.16	0.16	0.15	0.15
		MSE-R	1.00	1.21	1.20	1.20	1.26	0.48*	1.00	1.05	1.03	1.03	1.03	1.03
	1000	B	0.14	−0.09	−0.02	−0.02	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
		SD	0.11	0.12	0.12	0.13	0.12	0.12	0.07	0.08	0.08	0.08	0.08	0.08
		MSE-R	1.00	1.32	1.94	1.93	2.01	2.01	1.00	1.37	1.37	1.37	1.37	1.37
	2000	B	0.14	−0.09	−0.01	−0.02	0.00	0.01	0.06	0.00	0.00	0.00	0.00	0.00
		SD	0.07	0.08	0.08	0.08	0.08	0.08	0.05	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.65	3.51	3.47	3.66	3.61	1.00	1.87	1.88	1.88	1.89	1.89

Table A.10: *Standard deviation of $\hat{\beta}_X$ for $P(Y = 1|X, Z) = F(\beta_0 + \beta_X X + \beta_Z Z)$, three distributions of X, two reliability ratios (RR), two sample sizes (n). SD, monte carlo standard deviation; SE, sandwich standard deviation. Adjusted data $\widehat{X}$: MAI with M=(2,1,1) and M=(4,2,2).*

			$(\beta_0, \beta_X, \beta_Z) = (-1.5, 1, 1)$		$(\beta_0, \beta_X, \beta_Z) = (-.6, .3, .3)$	
RR	n	Stat.	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$
Normally distributed X						
0.75	300	SD	0.25	0.24	0.16	0.16
		SE	0.26	0.25	0.17	0.17
	2000	SD	0.10	0.09	0.07	0.07
		SE	0.10	0.09	0.06	0.06
0.50	300	SD	0.49	0.35	0.22	0.22
		SE	0.48	0.36	0.23	0.22
	2000	SD	0.14	0.13	0.08	0.08
		SE	0.14	0.13	0.08	0.08
Chi-square X						
0.75	300	SD	0.24	0.26	0.18	0.18
		SE	0.24	0.26	0.17	0.17
	2000	SD	0.10	0.10	0.06	0.06
		SE	0.09	0.10	0.06	0.06
0.50	300	SD	0.46	0.41	0.24	0.24
		SE	0.45	0.40	0.23	0.23
	2000	SD	0.14	0.15	0.09	0.09
		SE	0.14	0.14	0.08	0.08
Bi-modal X						
0.75	300	SD	0.21	0.25	0.16	0.17
		SE	0.21	0.24	0.16	0.16
	2000	SD	0.08	0.10	0.06	0.06
		SE	0.08	0.09	0.06	0.06
0.50	300	SD	0.34	0.40	0.20	0.21
		SE	0.39	0.40	0.20	0.21
	2000	SD	0.12	0.14	0.08	0.09
		SE	0.13	0.14	0.08	0.08

A.5 Simulations in Survival Analysis

Table A.11: *Estimation of β_X for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is normally distributed, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.04 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X, \beta_Z) = (0.7, 0.7)$					$(\beta_X, \beta_Z) = (0.3, 0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.22	−0.03	−0.04	−0.02	0.01	−0.08	0.01	0.01	0.01	0.02
		SD	0.08	0.12	0.12	0.13	0.14	0.09	0.12	0.12	0.12	0.13
		MSE-R	1.00	3.71	3.56	3.24	2.63	1.00	0.91	0.90	0.86	0.84
	1000	B	−0.22	−0.03	−0.04	−0.02	0.01	−0.09	0.00	0.00	0.00	0.00
		SD	0.05	0.06	0.06	0.07	0.08	0.05	0.07	0.07	0.07	0.07
		MSE-R	1.00	10.06	8.44	9.24	8.13	1.00	2.20	2.19	2.07	2.06
	2000	B	−0.22	−0.03	−0.04	−0.02	0.01	−0.09	0.00	0.00	0.00	0.00
		SD	0.03	0.05	0.05	0.05	0.06	0.04	0.05	0.05	0.05	0.05
		MSE-R	1.00	17.19	13.14	17.31	15.31	1.00	3.62	3.62	3.47	3.42
0.50	300	B	−0.40	−0.03	−0.06	0.00*	0.11*	−0.16	0.01	0.02	0.03	0.04
		SD	0.07	0.17	0.17	0.25*	0.54*	0.07	0.17	0.17	0.19	0.23
		MSE-R	1.00	5.42	5.22	2.57*	0.55*	1.00	1.07	1.05	0.86	0.56
	1000	B	−0.41	−0.06	−0.08	−0.04	0.03	−0.16	0.00	0.00	0.01	0.01
		SD	0.04	0.09	0.08	0.11	0.19	0.04	0.08	0.08	0.09	0.09
		MSE-R	1.00	16.16	12.03	13.11	4.72	1.00	4.04	4.05	3.56	3.36
	2000	B	−0.41	−0.06	−0.08	−0.04	0.01	−0.16	0.00	0.00	0.00	0.00
		SD	0.03	0.06	0.06	0.07	0.10	0.03	0.06	0.06	0.07	0.07
		MSE-R	1.00	24.88	15.72	23.13	17.36	1.00	7.30	7.32	6.53	6.30

Table A.12: *Estimation of β_Z for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is normally distributed, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_Z(X)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X, \beta_Z) = (0.7, 0.7)$					$(\beta_X, \beta_Z) = (0.3, 0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.06	−0.02	0.01	0.00	0.02	0.02	−0.01	−0.01	−0.01	−0.01
		SD	0.11	0.11	0.11	0.11	0.12	0.11	0.12	0.12	0.12	0.12
		MSE-R	1.00	1.23	1.22	1.21	1.10	1.00	0.93	0.93	0.93	0.93
	1000	B	0.04	−0.03	−0.01	−0.01	0.00	0.03	0.00	0.00	0.00	0.00
		SD	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.19	1.53	1.48	1.48	1.00	1.19	1.18	1.18	1.18
	2000	B	0.04	−0.03	−0.01	−0.01	0.00	0.03	0.00	0.00	0.00	0.00
		SD	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
		MSE-R	1.00	1.23	1.95	1.83	1.95	1.00	1.52	1.52	1.52	1.51
0.50	300	B	0.09	−0.06	−0.02	−0.03	−0.01*	0.06	−0.01	−0.01	−0.01	−0.01
		SD	0.10	0.12	0.12	0.13	0.18*	0.10	0.12	0.12	0.12	0.13
		MSE-R	1.00	1.04	1.28	1.08	0.57*	1.00	1.02	1.03	1.01	0.88
	1000	B	0.09	−0.05	−0.02	−0.02	0.00	0.06	0.00	0.00	0.00	0.00
		SD	0.05	0.06	0.06	0.07	0.07	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.53	2.58	2.27	2.00	1.00	1.85	1.85	1.81	1.78
	2000	B	0.09	−0.05	−0.02	−0.02	0.00	0.06	−0.01	0.00	0.00	0.00
		SD	0.04	0.05	0.04	0.05	0.05	0.04	0.04	0.04	0.05	0.05
		MSE-R	1.00	1.85	4.16	3.54	3.59	1.00	2.42	2.46	2.39	2.35

Table A.13: *Estimation of β_X for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is chi-square with df=4, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.06 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X, \beta_Z) = (0.7, 0.7)$					$(\beta_X, \beta_Z) = (0.3, 0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.18	0.04	0.00	−0.01	0.02	−0.07	0.02	0.01	0.00	0.00
		SD	0.08	0.12	0.11	0.11	0.12	0.08	0.12	0.12	0.11	0.11
		MSE-R	1.00	2.52	2.99	3.06	2.51	1.00	0.86	0.93	1.01	0.99
	1000	B	−0.18	0.03	−0.01	−0.02	0.01	−0.07	0.02	0.01	0.00	0.00
		SD	0.04	0.06	0.06	0.06	0.07	0.05	0.07	0.06	0.06	0.06
		MSE-R	1.00	7.03	9.03	7.94	7.41	1.00	1.58	1.77	1.99	1.95
	2000	B	−0.18	0.03	−0.01	−0.02	0.00	−0.07	0.02	0.01	0.00	0.00
		SD	0.03	0.04	0.04	0.04	0.05	0.03	0.05	0.04	0.04	0.04
		MSE-R	1.00	12.73	18.06	14.08	15.46	1.00	2.32	2.77	3.27	3.20
0.50	300	B	−0.37	0.03	−0.03	−0.04	0.08*	−0.15	0.03	0.02	0.00	0.01
		SD	0.06	0.17	0.16	0.19	0.46*	0.07	0.17	0.17	0.16	0.16
		MSE-R	1.00	4.82	5.65	3.95	0.65*	1.00	0.95	1.02	1.18	1.11
	1000	B	−0.37	0.03	−0.04	−0.05	0.02	−0.15	0.03	0.01	0.00	0.00
		SD	0.04	0.09	0.08	0.09	0.12	0.04	0.08	0.08	0.07	0.07
		MSE-R	1.00	14.64	16.53	12.17	9.69	1.00	3.09	3.70	4.51	4.38
	2000	B	−0.37	0.03	−0.04	−0.06	0.01	−0.15	0.03	0.02	0.00	0.00
		SD	0.02	0.06	0.06	0.06	0.08	0.03	0.06	0.06	0.05	0.05
		MSE-R	1.00	29.34	28.14	20.06	22.41	1.00	4.82	6.18	7.87	7.74

Table A.14: *Estimation of β_Z for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is chi-square df=4, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_Z(X)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.10). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X, \beta_Z) = (0.7, 0.7)$					$(\beta_X, \beta_Z) = (0.3, 0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.04	−0.05	−0.01	−0.01	0.00	0.03	−0.01	0.00	0.00	0.00
		SD	0.09	0.09	0.09	0.09	0.10	0.09	0.10	0.10	0.10	0.10
		MSE-R	1.00	0.87	1.07	1.06	0.99	1.00	1.00	1.00	1.01	1.00
	1000	B	0.04	−0.05	−0.02	−0.01	0.00	0.03	−0.01	0.00	0.00	0.00
		SD	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
		MSE-R	1.00	0.72	1.22	1.24	1.24	1.00	1.19	1.21	1.22	1.22
	2000	B	0.03	−0.05	−0.02	−0.01	0.00	0.03	−0.01	−0.01	0.00	0.00
		SD	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
		MSE-R	1.00	0.64	1.50	1.61	1.73	1.00	1.27	1.33	1.37	1.37
	300	B	0.09	−0.08	−0.02	−0.02	0.01*	0.06	−0.02	−0.01	0.00	0.00
		SD	0.09	0.11	0.11	0.11	0.16*	0.09	0.11	0.11	0.11	0.11
		MSE-R	1.00	0.79	1.27	1.17	0.57*	1.00	0.97	1.00	1.04	1.01
	1000	B	0.07	−0.09	−0.03	−0.03	0.00	0.06	−0.02	−0.01	0.00	0.00
		SD	0.05	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
		MSE-R	1.00	0.62	1.81	1.86	1.93	1.00	1.56	1.71	1.78	1.78
	2000	B	0.07	−0.09	−0.03	−0.03	0.00	0.06	−0.02	−0.01	−0.01	0.00
		SD	0.04	0.04	0.04	0.04	0.05	0.03	0.04	0.04	0.04	0.04
		MSE-R	1.00	0.61	2.26	2.53	3.03	1.00	2.18	2.51	2.66	2.67

Table A.15: *Estimation of β_X for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is bimodal normal, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\hat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.05 and 0.13). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X,\beta_Z)=(0.7,0.7)$					$(\beta_X,\beta_Z)=(0.3,0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	−0.18	0.02	0.00	0.00	0.01	−0.08	0.01	0.01	0.01	0.01
		SD	0.08	0.11	0.11	0.11	0.12	0.09	0.12	0.12	0.12	0.12
		MSE-R	1.00	3.08	3.27	3.05	2.68	1.00	0.89	0.91	0.94	0.93
	1000	B	−0.19	0.01	−0.01	−0.01	0.00	−0.08	0.01	0.00	0.00	0.00
		SD	0.05	0.06	0.06	0.07	0.07	0.05	0.06	0.06	0.06	0.06
		MSE-R	1.00	8.43	8.97	8.45	7.39	1.00	2.01	2.07	2.14	2.10
	2000	B	−0.19	0.01	−0.01	−0.01	0.00	−0.08	0.01	0.01	0.00	0.00
		SD	0.03	0.04	0.04	0.04	0.05	0.03	0.05	0.05	0.04	0.04
		MSE-R	1.00	19.13	19.23	18.46	17.33	1.00	3.18	3.31	3.51	3.49
0.50	300	B	−0.37	0.03	−0.01	0.03	0.06	−0.15	0.03	0.02	0.02	0.03
		SD	0.06	0.15	0.15	0.22	0.26	0.07	0.17	0.16	0.16	0.42
		MSE-R	1.00	5.66	6.13	2.93	1.98	1.00	1.00	1.02	1.09	0.16
	1000	B	−0.37	0.02	−0.03	−0.01	0.01	−0.16	0.02	0.01	0.00	0.00
		SD	0.04	0.08	0.08	0.10	0.14	0.04	0.09	0.08	0.08	0.08
		MSE-R	1.00	18.93	19.60	14.89	7.44	1.00	3.41	3.55	3.67	3.58
	2000	B	−0.37	0.02	−0.03	−0.02	0.00	−0.16	0.02	0.01	0.00	0.00
		SD	0.02	0.06	0.06	0.07	0.08	0.03	0.06	0.06	0.06	0.06
		MSE-R	1.00	34.36	33.29	27.06	20.41	1.00	6.96	7.41	8.08	8.00

Table A.16: *Estimation of β_Z for $\lambda(t|X,Z) = \lambda_0(t)\exp(\beta_X X + \beta_Z Z)$, where X is bimodal normal, two reliability ratios (RR), three sample sizes (n). B, bias (standard error approximately 0.005); SD, standard deviation; MSE-R, $MSE_{\beta_Z(X)}/MSE_{\beta_Z(\hat{X})}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.09). Adjusted data $\widehat{X}$: RC, regression calibration; MAI with M=(2,1,1) and M=(4,2,2); CS, conditional score. Values impacted by outliers marked by *.*

			$(\beta_X, \beta_Z) = (0.7, 0.7)$					$(\beta_X, \beta_Z) = (0.3, 0.3)$				
RR	n	Stat.	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$	W	$\hat{X}_{RC}$	$\hat{X}_{2,1,1}$	$\hat{X}_{4,2,2}$	$\hat{X}_{CS}$
0.75	300	B	0.05	−0.03	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.00
		SD	0.11	0.11	0.11	0.11	0.12	0.10	0.11	0.11	0.11	0.11
		MSE-R	1.00	1.09	1.13	1.10	1.05	1.00	1.00	1.00	1.00	0.98
	1000	B	0.04	−0.04	−0.01	−0.01	0.00	0.03	0.00	0.00	0.00	0.00
		SD	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.05	1.34	1.33	1.32	1.00	1.23	1.22	1.23	1.22
	2000	B	0.05	−0.03	−0.01	−0.01	0.01	0.03	0.00	0.00	0.00	0.00
		SD	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
		MSE-R	1.00	1.46	2.23	2.20	2.14	1.00	1.47	1.46	1.46	1.46
0.50	300	B	0.09	−0.07	−0.02	−0.03	0.00	0.07	0.00	0.00	0.00	0.00
		SD	0.11	0.13	0.13	0.14	0.15	0.10	0.12	0.12	0.12	0.16
		MSE-R	1.00	0.99	1.20	1.02	0.93	1.00	1.07	1.07	1.06	0.57
	1000	B	0.08	−0.07	−0.02	−0.03	0.00	0.06	0.00	0.00	0.00	0.00
		SD	0.06	0.06	0.06	0.06	0.07	0.05	0.06	0.06	0.06	0.06
		MSE-R	1.00	1.17	2.25	1.99	2.03	1.00	1.86	1.86	1.85	1.85
	2000	B	0.08	−0.07	−0.02	−0.03	0.00	0.06	0.00	0.00	0.00	0.00
		SD	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.04
		MSE-R	1.00	1.38	3.79	3.17	3.70	1.00	2.95	2.94	2.93	2.94

A.6 OPTIMIZE-HF supplement

Table A.17: *Covariates included in Model 1.5 and regarded as error-free*

age
black race
heart rate
diastolic blood pressure
sodium level
serum creatinine level
hemoglobin level
primary cause of hospital admission
prior cerebrovascular accident or transient ischemic attack
hyperlipidemia
hypertension
liver disease
smoker within past year
chronic obstructive pulmonary disease
peripheral vascular disease
known heart failure prior to this admission
rales
LVSD

Table A.18: *Covariates included in Model 1.5 and regarded as error-free*

serum creatinine level lower than 4 mg/dL at admission
age
reactive airway disease
weight
lower extremity edema
lipid lowering agent at discharge
sodium level
depression
any F-blocker use at discharge
systolic blood pressure at discharge
serum creatinine level at discharge
liver disease

Appendix B

Supplement to Chapter 2

This Appendix provides a supplement to the comparison in Chapter 2 of the three methods of implementing moment adjustment: sequential, unweighted joint minimization (2.2), and weighted joint minimization (2.1). In addition, we assess the importance of order in adjusting data sequentially. In Chapter 2, we propose that sequential adjustment be performed in the order of smallest to largest measurement error. The adjusted data obtained this way will be denoted $\widehat{\boldsymbol{X}}_{Seq-C}$, indicating "correct" adjustment order. When the order is switched, the adjusted data are denoted $\widehat{\boldsymbol{X}}_{Seq-W}$, for "wrong" order.

B.1 Comparison of Adjustment: Scenario 1

In this section, we compare the adjusted data, $\widehat{\boldsymbol{X}}$, and estimation of logistic regression model coefficients, based on a single data set. The data generation is similar to Chapter 2, Section 2.4. However, the measurement error in W_1 and W_2 is extremely different. W_1 has large measurement error, with a reliability ratio of 0.5, and W_2 has virtually no measurement error, with a reliability ratio greater than 0.99. The scenario is the following:

- Model: $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- $n = 1000$ observations of (Y, X_1, X_2)
- $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$
- $\Sigma_X = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$

- $(\beta_0, \beta_1, \beta_2) = (1, .5, .5)$ (moderately significant for $n = 1000$, 70% event rate)
- $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$
- $\Sigma_U = \begin{pmatrix} 1 & 0.8(.01) \\ 0.8(.01) & .0001 \end{pmatrix}$
- $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$

Here, we really have univariate mis-measured data, W_1, which should be adjusted, and W_2 should not. If an adjustment procedure is implemented for $\boldsymbol{W}$, including both W_1 and W_2, we would hope that the impact on W_2 would be negligible and that parameter estimation would be equivalent to a univariate adjustment of W_1. In Figure B.1 we compare the adjusted $\widehat{X}_2$ to W_2. The correctly ordered sequential adjustment and weighted joint adjustment do not meaningfully alter the W_2. The other two approaches do alter W_2, even though it was error free. If the adjusted data are of particular interest, this is undesirable.

Despite differences in the adjusted data themselves, these various methods produce virtually identical logistic regression parameter estimates, even under these extreme circumstances (Table B.1). This is consistent with the simulation results in Chapter 2.

Table B.1: *Coefficient estimates. Methods: X, true covariates; W, mis-measured covariates; Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; J, joint minimization (2.2); JW, joint minimization (2.1);* $\widehat{X}_1$*, univariate adjustment of* W_1

Method	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
X	1.048	0.557	0.454
W	1.032	0.238	0.592
Seq-W	1.043	0.541	0.475
Seq-C	1.043	0.540	0.474
J	1.043	0.541	0.474
JW	1.043	0.540	0.474
$\widehat{X}_1$	1.043	0.539	0.473

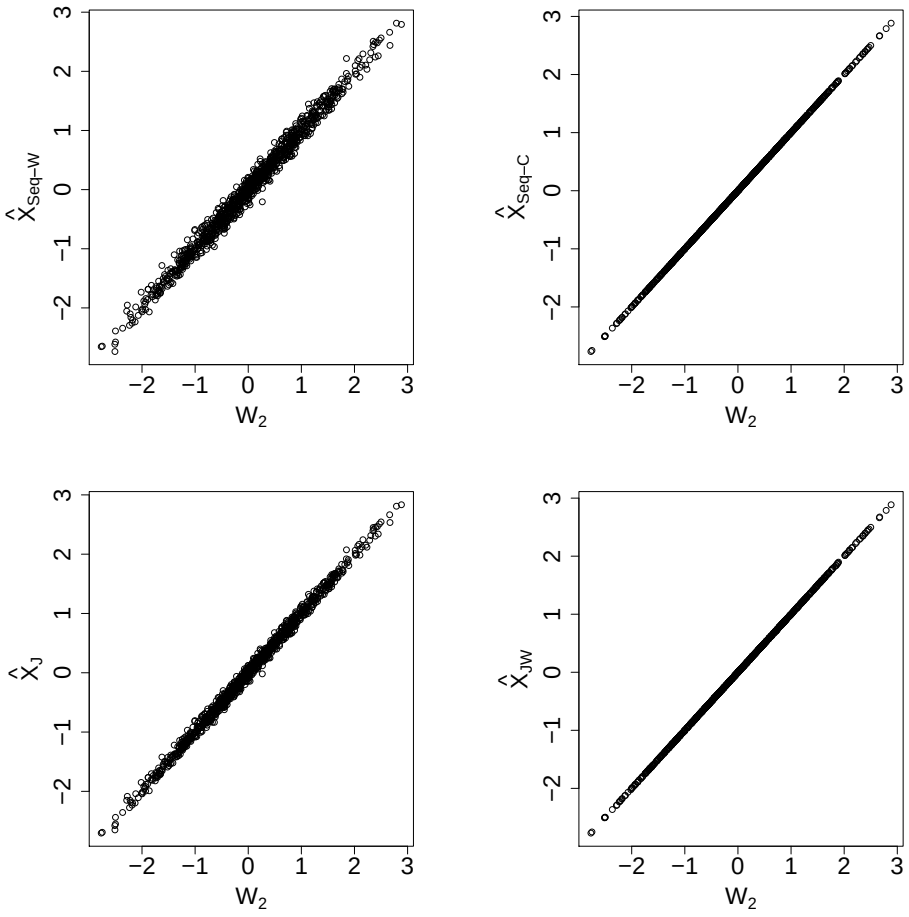

Figure B.1: *Direct Comparison of $\widehat{X}_2$ to W_2 for a single data set. $\widehat{X}_2$: Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; J, joint minimization (2.2); JW, joint minimization (2.1)*

B.2 Comparison of Adjustment: Scenario 2

Here we make comparisons for a slightly different data scenario. In this case both variables are measured with error, although the amount of measurement error differs. More importantly, the measurement error is highly correlated. The data are generated under the following scenario:

- Model: $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- $n = 1000$ observations of (Y, X_1, X_2)
- $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$
- $\Sigma_X = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
- $(\beta_0, \beta_1, \beta_2) = (1.5, .5, .5)$ (moderately significant for $n = 1000$, 70% event rate)
- $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$
- $\Sigma_U = \begin{pmatrix} 1 & 0.8(.6) \\ 0.8(.6) & .36 \end{pmatrix}$
- $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$

The sequential adjustment accounts for correlated measurement error only in the moment estimation, but not in the minimization. Thus, it will differ from the weighted joint adjustment. Here, we evaluate the extent of that difference. In Figure B.2 we see that the sequentially adjusted data are more similar to the weighted joint minimization when the adjustment occurs in the "wrong" order. In terms of re-creating the joint adjustment, the "wrong" order is preferable under these circumstances. However, in Table B.2 we see that none of the methods differ in terms of logistic regression parameter estimation.

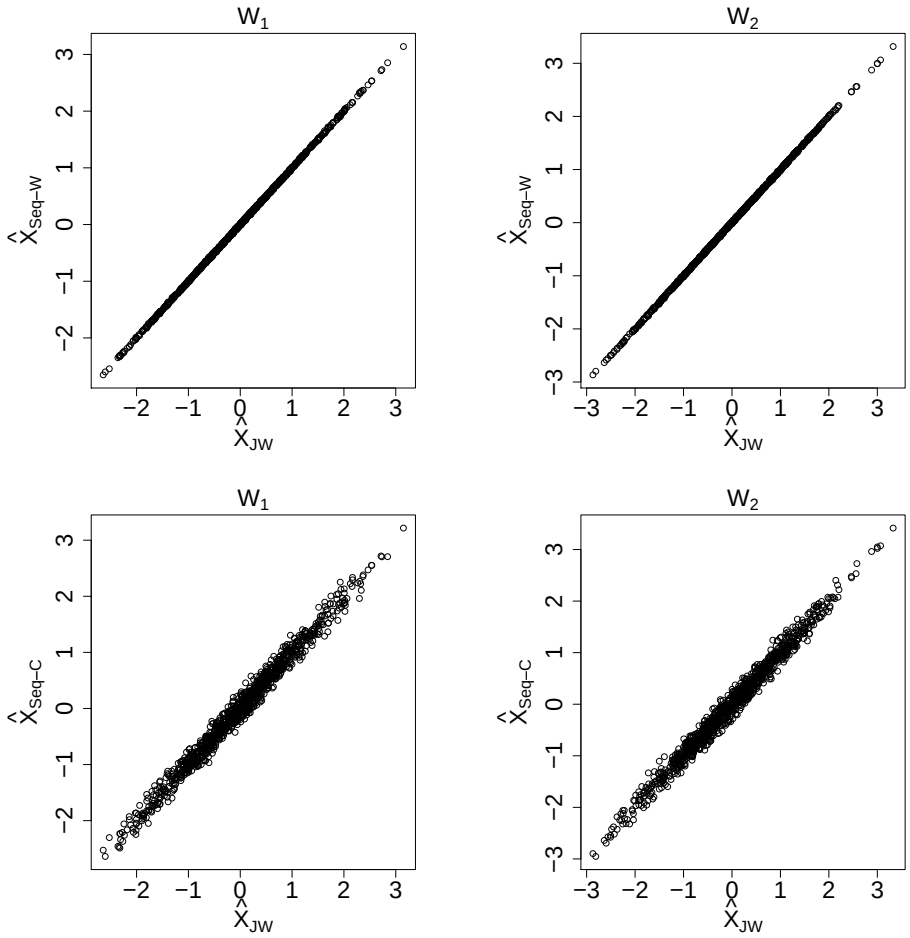

Figure B.2: *Direct Comparison of* $\widehat{\boldsymbol{X}}$ *for a single data set.* $\widehat{X}$*: Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; JW, joint minimization (2.1)*

Table B.2: *Coefficient estimates. Methods: X, true covariates; W, mis-measured covariates; Seq-W, wrongly ordered sequential; Seq-C, correctly ordered sequential; J, joint minimization (2.2); JW, joint minimization (2.1)*

Method	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
X	0.964	0.535	0.512
W	0.902	0.182	0.392
Seq-W	0.970	0.561	0.481
Seq-C	0.969	0.557	0.483
J	0.970	0.559	0.482
JW	0.970	0.560	0.481

B.3 Additional Simulations

Finally, we include an additional simulation to compare logistic regression parameter estimation for the various methods. The data are generated under the following scenario.

- $B = 100$ data sets
- Model: $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- $n = 1000$ observations of (Y, X_1, X_2)
- $\boldsymbol{X} = (X_1, X_2)^T \sim MVN(0, \Sigma_X)$
- $\Sigma_X = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$
- $(\beta_0, \beta_1, \beta_2) = (-2.5, .25, .25)$ (weakly significant for $n = 1000$, 7% event rate)
- $\boldsymbol{U} = (U_1, U_2)^T \sim MVN(0, \Sigma_U)$
- $\Sigma_U = \begin{pmatrix} 1 & 0.8(.01) \\ 0.8(.01) & .0001 \end{pmatrix}$
- $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$

There are two differences between this scenario and that of the simulation in Chapter 2. The coefficient magnitudes and event rate differ. These coefficients are weakly significant and the event rate is very low. In addition, only a single variable is actually measured with error, but both are adjusted. As in other cases, we see no meaningful difference in the parameter estimation, regardless of how the data were adjusted (Table B.3).

Table B.3: *Estimation of $\boldsymbol{\beta}$ for $P(Y = 1|X_1, X_2) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$. True value of $\boldsymbol{\beta} = (-2.5, .25, .25)$; B, bias; SD, standard deviation; M, $MSE_W/MSE_{\widehat{\boldsymbol{X}}}$. Adjusted data $\widehat{\boldsymbol{X}}$: Seq-W, wrongly ordered sequential; J, joint adjustment (2.2); JW, joint adjustment (2.1).*

		Matching 2 moments			Matching 4 moments		
Stat.	W	$\widehat{\boldsymbol{X}}_{Seq-W}$	$\widehat{\boldsymbol{X}}_{J}$	$\widehat{\boldsymbol{X}}_{JW}$	$\widehat{\boldsymbol{X}}_{Seq-W}$	$\widehat{\boldsymbol{X}}_{J}$	$\widehat{\boldsymbol{X}}_{JW}$
		Estimation of β_1; true value is 0.25					
B	-0.135	0.031	0.03	0.03	0.032	0.037	0.031
SD	0.098	0.235	0.235	0.235	0.236	0.236	0.238
M	0.028	0.497	0.498	0.498	0.490	0.491	0.485
		Estimation of β_2; true value is 0.25					
B	0.073	-0.006	-0.006	-0.006	-0.006	-0.007	-0.002
SD	0.133	0.164	0.164	0.164	0.163	0.170	0.166
M	0.023	0.860	0.860	0.860	0.864	0.706	0.835

Appendix C

Supplement to Chapter 3

C.1 Conditional score for mean and variance predictors

The conditional score approach relies on obtaining sufficient statistics for mis-measured covariates. When we are interested in the simple logistic regression model that includes only the inverse of the subject-specific variance as a predictor this method is easily derived. However, the problem becomes quite complicated when the model involves other mis-measured covariates.

In this section we show another case where a sufficient statistic can be obtained, but the disadvantages to this approach become obvious. Let $P(Y = 1|\sigma^2; \boldsymbol{\beta}) = F\{\beta_0 + \beta_{\gamma_0}\left(\gamma_0/\sigma^2\right) + \beta_\sigma(1/\sigma^2)\}$ for $\boldsymbol{\beta} = (\beta_0, \beta_{\gamma_0}, \beta_\sigma)$. The corresponding density of Y is

$$f_Y(y; \gamma_0, \sigma^2, \boldsymbol{\beta}) = \exp\left[y\left(\beta_0 + \beta_{\gamma_0}\frac{\gamma_0}{\sigma^2} + \beta_\sigma\frac{1}{\sigma^2}\right) + \log\left\{1 - F\left(\beta_0 + \beta_{\gamma_0}\frac{\gamma_0}{\sigma^2} + \beta_\sigma\frac{1}{\sigma^2}\right)\right\}\right], \tag{C.1}$$

where γ_0 and σ^2 are regarded as unknown constants. The observed variance estimators $\widehat{\sigma}^2$ have density

$$f_{\widehat{\sigma}^2}(\widehat{\sigma}^2; \sigma^2) = \frac{\left(\frac{\nu}{\sigma^2}\right)}{2^{\nu/2}\Gamma(\nu/2)}\left(\frac{\nu\widehat{\sigma}^2}{\sigma^2}\right)^{\nu/2-1}\exp\left(\frac{-\nu\widehat{\sigma}^2}{2\sigma^2}\right)I(\widehat{\sigma}^2 > 0). \tag{C.2}$$

The estimators $\widehat{\gamma}_0$ have density

$$f_{\widehat{\gamma}_0}(\widehat{\gamma}_0; \gamma_0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2 c}}\exp\left\{-\frac{1}{2\sigma^2 c}(\widehat{\gamma}_0 - \gamma_0)^2\right\}. \tag{C.3}$$

$$f_{Y,\widehat{\sigma}^2,\widehat{\gamma}_0}(y,\widehat{\sigma}^2,\widehat{\gamma}_0;\gamma_0,\sigma^2,\boldsymbol{\beta}) = g(Y,\widehat{\sigma}^2,\widehat{\gamma}_0)h(\sigma^2,\gamma_0)\times$$
$$\exp\left[\frac{-1}{2\sigma^2}\left\{(\gamma_0 - y\beta_{\gamma_0} - \widehat{\gamma}_0)^2 - (y\beta_{\gamma_0} + \widehat{\gamma}_0)^2 + \widehat{\gamma}_0^2 + \nu\widehat{\sigma}^2 - 2y\beta_\sigma\right\}\right]. \qquad \text{(C.4)}$$

By factorization, the two dimensional statistic $\{Y\beta_{\gamma_0}+\widehat{\gamma}_0, -(Y\beta_{\gamma_0}+\widehat{\gamma}_0)^2+\widehat{\gamma}_0^2+\nu\widehat{\sigma}^2-2Y\beta_\sigma\}$ is sufficient for (γ_0,σ^2).

This factorization relies on the specific form $P(Y=1|\sigma^2;\boldsymbol{\beta}) = F\{\beta_0+\beta_{\gamma_0}\left(\gamma_0/\sigma^2\right)+\beta_\sigma(1/\sigma^2)\}$. Otherwise, we are unable to factor the joint density of observed data. In practice, it may be difficult to justify the use of covariates γ_0/σ^2 and $1/\sigma^2$, especially when interpretable parameters are desired. For a logistic model that includes $\boldsymbol{\gamma} = (\gamma_0,\gamma_1)^T$ and $T(\sigma^2)$ as covariates, our factorization requires that the model be defined $P(Y=1|\boldsymbol{\gamma},\sigma^2;\boldsymbol{\beta}) = F(\beta_0 + (1/\sigma^2)\boldsymbol{\gamma}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{\beta}_\gamma + T(\sigma^2)\beta_\sigma)$. In this case, the model involves linear combinations of γ_0 and γ_1, which does not have a reasonable interpretation. For this reason, we do not pursue this method any further.

C.2 Derivations for the approximately unbiased score

The approximately unbiased score function (3.7) is based on estimators for the quantities $\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$, $\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$, $(\sigma^2)^{-\beta_\sigma}$, and $\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. In this section, we derive the estimators defined in Chapter 3.

We want to obtain an estimator, $T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta})$ for $\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. Begin with a naive estimator, $\exp(-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma)$. We have $\widehat{\boldsymbol{\gamma}}|\boldsymbol{\gamma} \sim N\{\boldsymbol{\gamma},\sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\}$. For notational convenience let $\Sigma = \sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}$. The naive estimator is biased since

$$E\left\{\exp(-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma)|\boldsymbol{\gamma},\Sigma\right\} =$$
$$c\int \exp(-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma)\exp\left\{-\frac{1}{2}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})^T\Sigma^{-1}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})\right\}d\widehat{\boldsymbol{\gamma}}$$
$$= c\int \exp\left\{-\frac{1}{2}(\widehat{\boldsymbol{\gamma}}-b)^T\Sigma^{-1}(\widehat{\boldsymbol{\gamma}}-b)\right\}\exp\left\{-\frac{1}{2}\left(\boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma} - b^T\Sigma^{-1}b\right)\right\}d\widehat{\boldsymbol{\gamma}},$$
$$\text{where } b = (\boldsymbol{\gamma}-\Sigma\boldsymbol{\beta}_\gamma)$$
$$= \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \boldsymbol{\beta}_\gamma^T\Sigma\boldsymbol{\beta}_\gamma/2).$$

Hence, $\exp(-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma - \boldsymbol{\beta}_\gamma^T\Sigma\boldsymbol{\beta}_\gamma/2)$ is unbiased for $\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. However, this includes the unknown σ^2 in $\Sigma{=}\sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}$. We substitute $\widehat{\sigma}^2$ to define the estimator

$$T_1(\widehat{\boldsymbol{\gamma}},\widehat{\sigma}^2,\boldsymbol{\beta}) = \exp\{-\widehat{\boldsymbol{\gamma}}^T\boldsymbol{\beta}_\gamma - \boldsymbol{\beta}_\gamma^T\widehat{\sigma}^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{\beta}_\gamma/2\}. \qquad \text{(C.5)}$$

Let $T_2(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})$ denote the estimator for $\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. Note that $\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$ $= -d/d\boldsymbol{\beta}_\gamma \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. We already have $E\left\{T_1(\widehat{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\beta})\right\} = \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. Taking the derivative on both sides gives $-d/d\boldsymbol{\beta}_\gamma E\left\{T_1(\widehat{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\beta})\right\} = \boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. By interchanging differentiation and expectation, we have $E\left\{-d/d\boldsymbol{\beta}_\gamma T_1(\widehat{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\beta})\right\} = \boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. Our unbiased estimator is $-d/d\boldsymbol{\beta}_\gamma T_1(\widehat{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\beta}) = \left\{\widehat{\boldsymbol{\gamma}} + \sigma^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{\beta}_\gamma\right\} T_1(\widehat{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\beta})$. Because this depends on the unknown σ^2, we substitute $\widehat{\sigma}^2$ to obtain an approximately unbiased estimator

$$T_2(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta}) = \left\{\widehat{\boldsymbol{\gamma}} + \widehat{\sigma}^2(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{\beta}_\gamma\right\} T_1(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta}). \tag{C.6}$$

Let $g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})$ denote the estimator for $(\sigma^2)^{-\beta_\sigma}$. Begin with a naive estimator $(\widehat{\sigma}^2)^{-\beta_\sigma}$. This is biased, as

$$\begin{aligned} E\left\{(\widehat{\sigma}^2)^{-\beta_\sigma}|\sigma^2\right\} &= \frac{\left(\frac{\nu}{\sigma^2}\right)}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty (\widehat{\sigma}^2)^{-\beta_\sigma} \left(\frac{\nu\widehat{\sigma}^2}{\sigma^2}\right)^{\nu/2-1} \exp\left(\frac{-\nu\widehat{\sigma}^2}{2\sigma^2}\right) d(\widehat{\sigma}^2) \\ &= (\sigma^2)^{-\beta_\sigma}\,(\nu/2)^{\beta_\sigma}\,\Gamma(\nu/2-\beta_\sigma)/\Gamma(\nu/2). \end{aligned}$$

Correcting for the bias, we define

$$g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta}) = (\widehat{\sigma}^2)^{-\beta_\sigma}\,(\nu/2)^{-\beta_\sigma}\,\Gamma(\nu/2)/\Gamma(\nu/2-\beta_\sigma). \tag{C.7}$$

Let $g_2(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})$ denote the estimator for $\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. Note that $\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$ $= -d/d\beta_\sigma(\sigma^2)^{-\beta_\sigma}$. We already have $E\left\{g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})\right\} = (\sigma^2)^{-\beta_\sigma}$. Taking the derivative on both sides gives $-d/d\beta_\sigma E\left\{g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})\right\} = \log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. By interchanging differentiation and expectation, we have $E\left\{-d/d\beta_\sigma g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})\right\} = \log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. Our unbiased estimator is

$$\begin{aligned} g_2(\widehat{\sigma}^2, \nu, \boldsymbol{\beta}) &= -d/d\beta_\sigma g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta}) \\ &= \Gamma(\nu/2)/\Gamma(\nu/2-\beta_\sigma)\left(2/\nu\widehat{\sigma}^2\right)^{\beta_\sigma}\left\{\log\left(2/\nu\widehat{\sigma}^2\right) + D(\nu/2-\beta_\sigma)\right\}. \end{aligned} \tag{C.8}$$

The approximately unbiased score function (3.7) is defined by substituting the estimators $T_1(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})$ for E_1, $T_2(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})g_1(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})$ for E_2, and $T_1(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})g_2(\widehat{\sigma}^2, \nu, \boldsymbol{\beta})$ for E_3.

C.3 Corrected score equation

We require unbiased estimators of $E_1(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) = \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma}$, $E_2(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta})$ $= \boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)(\sigma^2)^{-\beta_\sigma}$, and $E_3(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\beta}) = \exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)\log(\sigma^2)(\sigma^2)^{-\beta_\sigma}$. Let $T_1^*(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})$

and $T_2^*(\widehat{\boldsymbol{\gamma}}, \widehat{\sigma}^2, \boldsymbol{\beta})$ be unbiased estimators for $\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$ and $\boldsymbol{\gamma}\exp(-\boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma)$. These can be defined by the methods of Stefanski, Novick, and Devanarayan (2005). Randomly partition the replicate data into two halves. Estimate $\widehat{\boldsymbol{\gamma}}_{1i}$ and $\widehat{\sigma}^2_{1i}$ using the first r_{1i} replicates and $\widehat{\boldsymbol{\gamma}}_{2i}$ and $\widehat{\sigma}^2_{2i}$ using the second r_{2i} replicates, where $r_{1i} + r_{2i} = r_i$. The estimates $\widehat{\boldsymbol{\gamma}}_{1i}$, $\widehat{\sigma}^2_{1i}$, $\widehat{\boldsymbol{\gamma}}_{2i}$ and $\widehat{\sigma}^2_{2i}$ are mutually independent. Define the following estimators

$$\begin{aligned}
\widehat{E}_1(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) &= \tfrac{1}{2}T_1^*(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \boldsymbol{\beta})g_1(\widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) + \tfrac{1}{2}T_1^*(\widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \boldsymbol{\beta})g_1(\widehat{\sigma}^2_1, \nu_1, \boldsymbol{\beta}),\\
\widehat{E}_2(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) &= \tfrac{1}{2}T_2^*(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \boldsymbol{\beta})g_1(\widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) + \tfrac{1}{2}T_2^*(\widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \boldsymbol{\beta})g_1(\widehat{\sigma}^2_1, \nu_1, \boldsymbol{\beta}),\\
\widehat{E}_3(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) &= \tfrac{1}{2}T_1^*(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \boldsymbol{\beta})g_2(\widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta}) + \tfrac{1}{2}T_1^*(\widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \boldsymbol{\beta})g_2(\widehat{\sigma}^2_1, \nu_1, \boldsymbol{\beta}).
\end{aligned}$$

Then an unbiased score function is $\Psi_*(Y, \widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{Z}, \boldsymbol{\beta}) =$

$$\begin{bmatrix}
(Y-1) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_Z)\widehat{E}_1(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta})\\
(Y-1)\widehat{\boldsymbol{\gamma}} + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_Z)\widehat{E}_2(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_{2i}, \nu_2, \boldsymbol{\beta})\\
(Y-1)\widehat{\log}(\sigma^2) + Y\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_Z)\widehat{E}_3(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_{2i}, \nu_2, \boldsymbol{\beta})\\
(Y-1)\boldsymbol{Z} + Y\boldsymbol{Z}\exp(-\beta_0 - \boldsymbol{Z}^T\boldsymbol{\beta}_Z)\widehat{E}_1(\widehat{\boldsymbol{\gamma}}_1, \widehat{\sigma}^2_1, \nu_1, \widehat{\boldsymbol{\gamma}}_2, \widehat{\sigma}^2_2, \nu_2, \boldsymbol{\beta})
\end{bmatrix}.$$

The corrected score estimator $\widehat{\boldsymbol{\beta}}_*$ solves $\sum_{i=1}^n \Psi_*(Y_i, \widehat{\boldsymbol{\gamma}}_{1i}, \widehat{\sigma}^2_{1i}, \nu_{1i}, \widehat{\boldsymbol{\gamma}}_{2i}, \widehat{\sigma}^2_{2i}, \nu_{2i}, \boldsymbol{Z}_i, \widehat{\boldsymbol{\beta}}_*) = 0$. This score function is more complicated than the approximate score. Numerical algorithms that solve for $\widehat{\boldsymbol{\beta}}_*$ frequently fail to converge or result in outlying values. Our simulations demonstrate negligible bias in the approximate estimator, $\widehat{\boldsymbol{\beta}}_A$, and this is easier to obtain numerically. So we implement the Approximately Corrected Score approach in all analyses.

C.4 Moments of log transformed variance

Suppose that our primary outcome model depends on $T(\sigma_i^2) = \log(\sigma_i^2)$, so we are interested in adjusting $\log(\widehat{\sigma}_i^2)$. This corresponds to the scenario for which we have developed a corrected score approach. Estimators for the moments of $\log(\sigma_i^2)$ are more complicated, but can still be defined in a closed form. We first take the conditional expectation of $\{\log(\widehat{\sigma}_i^2)\}^r$, given σ_i^2, and find that is has additive bias. We then subtract off a method of moments estimator of this bias. The resulting estimators for m_{r3}=$E\{\log(\sigma_i^2)^r\}$ are provided below after defining some preliminary functions. The polygamma function is denoted $\psi_n(z) = \frac{d^{n+1}}{dz^{n+1}}\log[\Gamma(z)]$. Polygamma functions can be computed using mathematical software, such as the psigamma() function in R. Solutions to the following integrals are

given in terms of polygamma functions

$$I_1\left(\frac{\nu}{2},2\right)=\int_0^\infty \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left\{\log(u)\right\}u^{\frac{\nu}{2}-1}e^{-u/2}du=\left\{\psi_0\left(\frac{\nu}{2}\right)+\log(2)\right\}$$

$$I_2\left(\frac{\nu}{2},2\right)=\int_0^\infty \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left\{\log(u)\right\}^2u^{\frac{\nu}{2}-1}e^{-u/2}du=\left\{I_1\left(\frac{\nu}{2},2\right)\right\}^2+\psi_1\left(\frac{\nu}{2}\right)$$

$$I_3\left(\frac{\nu}{2},2\right)=\int_0^\infty \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left\{\log(u)\right\}^3u^{\frac{\nu}{2}-1}e^{-u/2}du=\left\{I_1\left(\frac{\nu}{2},2\right)\right\}^3+3\left\{I_1\left(\frac{\nu}{2},2\right)\right\}\psi_1\left(\frac{\nu}{2}\right)+\psi_2\left(\frac{\nu}{2}\right)$$

$$I_4\left(\frac{\nu}{2},2\right)=\int_0^\infty \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left\{\log(u)\right\}^4u^{\frac{\nu}{2}-1}e^{-u/2}du=\left\{I_1\left(\frac{\nu}{2},2\right)\right\}^4+6\left\{I_1\left(\frac{\nu}{2},2\right)\right\}^2\psi_1\left(\frac{\nu}{2}\right)+4\left\{I_1\left(\frac{\nu}{2},2\right)\right\}\psi_2\left(\frac{\nu}{2}\right)+3\left\{I_1\left(\frac{\nu}{2},2\right)\right\}^2+\psi_3\left(\frac{\nu}{2}\right).$$

First, we show the complete derivation of the estimator for m_{13}=$E\{\log(\sigma_i^2)\}$. To simplify notation, let $W=\widehat{\sigma}_i^2$ and $X=\sigma_i^2$ and $W|X\sim\frac{X}{\nu}\chi_\nu$. Then $E\{\log(\widehat{\sigma}_i^2)|\sigma_i^2\}$=

$$\begin{aligned}E\{\log(W)|X\}&=\frac{(\nu/X)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty\log(w)\left(\frac{\nu w}{X}\right)^{\frac{\nu}{2}-1}e^{-kw/2X}dw\\&=\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty\log\left(\frac{Xu}{k}\right)(u)^{\frac{\nu}{2}-1}e^{-u/2}du\\&=\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\int_0^\infty\log(u)(u)^{\frac{\nu}{2}-1}e^{-u/2}du+\log(X)-\log(\nu)\\&=I_1\left(\frac{\nu}{2},2\right)+\log(X)-\log(\nu)\end{aligned}$$

Taking the expectation on both sides of the equation, we see that $\log(W)$ is biased for $E\{\log(X)\}$. Subtracting off the bias, we have $E\left[E\left\{\log(W)-I_1\left(\frac{\nu}{2},2\right)+\log(\nu)|X\right\}\right]=E\{\log(X)\}$. In terms of σ_i^2 we can define the estimator $\widehat{m}_{13}$=$n^{-1}\sum_{i=1}^n\log(\sigma_i^2)-I_1\left(\frac{\nu}{2},2\right)+\log(\nu)$, which is unbiased for m_{13}=$E\{\log(\sigma_i^2)\}$.

We can perform similar integrations to obtain $E[\{\log(W)\}^r|X]$. For $r=2,3,4$

these are

$$
\begin{aligned}
E[\{\log(W)\}^2|X] &= \log(X)^2 + I_2\left(\frac{\nu}{2},2\right) + 2I_1\left(\frac{\nu}{2},2\right)\{\log(X)-\log(\nu)\} \\
&\quad - 2\log(X)\log(\nu) + \log(\nu)^2 \\
E[\{\log(W)\}^3|X] &= \log(X)^3 + I_3\left(\frac{\nu}{2},2\right) + 3I_2\left(\frac{\nu}{2},2\right)\{\log(X)-\log(\nu)\} \\
&\quad + 3I_1\left(\frac{\nu}{2},2\right)\left\{\log(X)^2 - 2\log(X)\log(\nu) + \log(\nu)^2\right\} \\
&\quad - 3\log(\nu)\log(X)^2 + 3\log(\nu)^2\log(X) - \log(\nu)^3 \\
E[\{\log(W)\}^4|X] &= \log(X)^4 + I_4\left(\frac{\nu}{2},2\right) + 4I_3\left(\frac{\nu}{2},2\right)\{\log(X)-\log(\nu)\} \\
&\quad + 6I_2\left(\frac{\nu}{2},2\right)\left\{\log(X)^2 - 2\log(X)\log(\nu) + \log(\nu)^2\right\} \\
&\quad + 4I_1\left(\frac{\nu}{2},2\right)\left\{\log(X)^3 - 3\log(\nu)\log(X)^2 + 3\log(\nu)^2\log(X) - \log(\nu)^3\right\} \\
&\quad - 4\log(X)^2\log(\nu) + 6\log(X)^2\log(\nu)^2 - 4\log(X)\log(\nu)^3 + \log(\nu)^4
\end{aligned}
$$

The quantity $\{\log(W)\}^r$ has additive bias for $E[\{\log(X)\}^r]$. When $r = 1$ the bias is constant and is subtracted off to obtain an unbiased estimator. For $r > 1$ the bias involves lower order moments. We replace the unknown, lower order moments with unbiased estimators and subtract off the resulting estimator of bias. In terms of σ_i^2, unbiased estimators for m_{r1}=$E[\{\log(\sigma_i^2)\}^r]$ are $\widehat{m}_{r1}$ =$n^{-1}\sum_{i=1}^n H_r\{\nu_i, \log(\widehat{\sigma}_i^2)\}$ where $H_r(\nu, w)$ is defined as follows:

$$
\begin{aligned}
H_1(\nu,w) &= w - [I_1\left(\frac{\nu}{2},2\right) - \log(\nu)] \\
H_2(\nu,w) &= w^2 - [I_2\left(\frac{\nu}{2},2\right) + 2I_1\left(\frac{\nu}{2},2\right)\{\widehat{m}_{11} - \log(\nu)\} - 2\widehat{m}_{11}\log(\nu) + \log(\nu)^2] \\
H_3(\nu,w) &= w^3 - [I_3\left(\frac{\nu}{2},2\right) + 3I_2\left(\frac{\nu}{2},2\right)\{\widehat{m}_{11} - \log(\nu)\} \\
&\quad + 3I_1\left(\frac{\nu}{2},2\right)\{\widehat{m}_{21} - 2\widehat{m}_{11}\log(\nu) + \log(\nu)^2\} \\
&\quad - 3\widehat{m}_{21}\log(\nu) + 3\widehat{m}_{11}\log(\nu)^2 + \log(\nu)^3] \\
H_4(\nu,w) &= w^4 - [I_4\left(\frac{\nu}{2},2\right) + 4I_3\left(\frac{\nu}{2},2\right)\{\widehat{m}_{11} - \log(\nu)\} \\
&\quad + 6I_2\left(\frac{\nu}{2},2\right)\{\widehat{m}_{21} - 2\widehat{m}_{11}\log(\nu) + \log(\nu)^2\} \\
&\quad + 4I_1\left(\frac{\nu}{2},2\right)\{\widehat{m}_{31} - 3\widehat{m}_{21}\log(\nu) + 3\widehat{m}_{11}\log(\nu)^2 - \log(\nu)^3\} \\
&\quad - 4\widehat{m}_{31}\log(\nu) + 6\widehat{m}_{21}\log(\nu)^2 - 4\widehat{m}_{11}\log(\nu)^3 + \log(\nu)^4].
\end{aligned}
$$

We have $E\{H_r(\nu_i, \log(\widehat{\sigma}_i^2))|\sigma_i^2\} = \{\log(\sigma_i^2)\}^r$ so that the estimators $\widehat{m}_{r1}$ = $n^{-1}\sum_{i=1}^n H_r\{\nu_i, \log(\widehat{\sigma}_i^2)\}$ are unbiased for $E[\{\log(\sigma_i^2)\}^r]$=$m_{r1}$ and $r = 1, \ldots, 4$. Assuming

that V_{ik} are independent of $\widehat{\sigma}_i^2$, given σ_i^2, the cross products with V_{ik} can be estimated by $\widehat{m}_{k1}^V = n^{-1}\sum_{i=1}^n H_1(\nu_i, \log(\widehat{\sigma}_i^2))V_{ik}$, for $k = 1, \ldots, K$. Since $\widehat{\boldsymbol{\gamma}}_i$ are conditionally independent of $\widehat{\sigma}_i^2$ given $(\boldsymbol{\gamma}_i, \sigma_i^2)$, the estimators $\widehat{m}_{12}^* = \widehat{m}_{21}^* = n^{-1}\sum_{i=1}^n H_1(\nu_i, \log(\widehat{\sigma}_i^2))\widehat{\gamma}_{0i}$ and $\widehat{m}_{13}^* = \widehat{m}_{31}^* = n^{-1}\sum_{i=1}^n H_1(\nu_i, \log(\widehat{\sigma}_i^2))\widehat{\gamma}_{1i}$ are also unbiased for m_{12}^* and m_{13}^*.

C.5 Additional Simulation Results

C.5.1 Model (3.1) with slope, intercept and variance covariates

This appendix provides additional simulation results for the following logistic regression model

$$P(Y_i = 1|\boldsymbol{\gamma}_i, \sigma_i^2; \boldsymbol{\beta}_\gamma, \beta_\sigma) = F\{\beta_0 + \boldsymbol{\gamma}_i^T\boldsymbol{\beta}_\gamma + \log(\sigma_i^2)\beta_\sigma\} \tag{C.9}$$

where $F(t) = 1/(1 + e^{-t})$. These results are similar to the case where $T(\sigma^2)$=log(σ^2), and are discussed in Section 3.4.1.

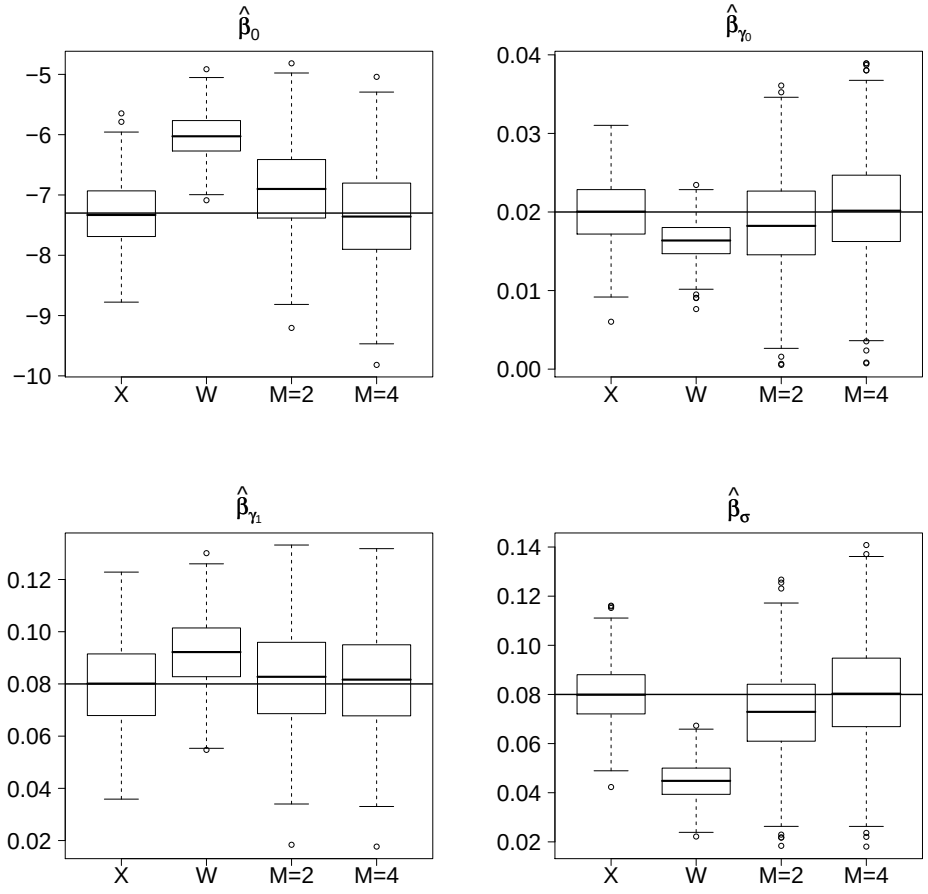

Figure C.1: *Boxplots of* $\widehat{\boldsymbol{\beta}}$*, from* $B = 500$ *simulated data sets where* $P(Y = 1|\boldsymbol{\gamma}, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \boldsymbol{\gamma}^T\boldsymbol{\beta}_\gamma + \sigma\beta_\sigma)$*, with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma) = (-7.3, 0.02, 0.08, 0.08)$*,* 10*, and* 5000*. Methods: X, true covariates; W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

Table C.1: *Estimation of* $\boldsymbol{\beta}$*, where* $P(Y = 1|\beta_{\gamma_0}, \beta_{\gamma_1}, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \gamma_0\beta_{\gamma_0} + \gamma_1\beta_{\gamma_1} + \sigma\beta_\sigma)$*, with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma)$=$(-7.3, 0.02, 0.08, 0.08)$*, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.03 for n=1000, 0.01 for n=5000); SD, Monte Carlo standard deviation of* $\widehat{\beta}$ *multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; P, power; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R,* $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ *(standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.10). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

			β_{γ_0}			β_{γ_1}			β_σ		
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
r=10	n=1000	RB	−0.18	−0.02	0.05	0.16	0.06	0.09	−0.44	−0.08	0.04
		SD	0.65	1.43	1.60	3.22	4.16	4.49	1.77	4.11	5.00
		SE/SD	1.00	0.90	0.88	1.00	0.97	0.92	1.00	0.95	0.93
		CP	0.87	0.94	0.94	0.94	0.95	0.94	0.46	0.95	0.96
		MSE-R	1.00	0.27	0.22	0.00	0.69	0.59	0.00	0.91	0.63
	n=5000	RB	−0.17	0.00	0.05	0.14	0.02	0.04	−0.44	−0.08	0.02
		SD	0.27	0.57	0.65	1.38	1.74	1.97	0.73	1.66	1.99
		SE/SD	1.00	0.97	0.92	1.00	1.02	0.91	1.00	1.02	0.98
		CP	0.73	0.94	0.93	0.89	0.95	0.94	0.00	0.94	0.95
		MSE-R	1.00	0.56	0.43	0.00	1.03	0.80	0.00	4.12	3.30
r=40	n=1000	RB	−0.04	−0.01	0.02	0.07	0.05	0.07	−0.15	−0.02	0.02
		SD	0.76	0.95	1.04	3.66	3.89	4.03	2.40	3.03	3.29
		SE/SD	1.00	1.02	0.96	1.00	0.94	0.91	1.00	0.97	0.95
		CP	0.96	0.96	0.95	0.93	0.93	0.92	0.91	0.94	0.94
		MSE-R	1.00	0.64	0.53	0.00	0.90	0.83	0.00	0.79	0.67
	n=5000	RB	−0.03	−0.01	0.01	0.03	0.01	0.02	−0.15	−0.02	0.02
		SD	0.33	0.42	0.48	1.51	1.60	1.80	1.04	1.31	1.44
		SE/SD	1.00	1.03	0.93	1.00	1.01	0.90	1.00	0.98	0.95
		CP	0.95	0.95	0.95	0.95	0.96	0.95	0.80	0.94	0.95
		MSE-R	1.00	0.66	0.51	0.00	0.91	0.72	0.00	1.45	1.21

Table C.2: *Estimation of* $\boldsymbol{\beta}$*, where* $P(Y = 1|\beta_{\gamma_0}, \beta_{\gamma_1}, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \gamma_0\beta_{\gamma_0} + \gamma_1\beta_{\gamma_1} + \sigma\beta_\sigma)$*, with true values* $(\beta_0, \beta_{\gamma_0}, \beta_{\gamma_1}, \beta_\sigma)=(-5, 0.01, 0.04, 0.04)$*, two levels of replication (r), two sample sizes (n). RB, relative bias (standard error approximately 0.05 for n=1000, 0.02 for n=5000); SD, Monte Carlo standard deviation of* $\widehat{\beta}$ *multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R,* $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ *(standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.06). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

			β_{γ_0}			β_{γ_1}			β_σ		
r	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	−0.21	−0.11	−0.06	0.15	0.06	0.10	−0.40	0.00	0.08
		SD	0.57	1.27	1.34	3.57	4.45	4.57	1.92	4.25	4.66
		SE/SD	1.00	1.06	1.03	1.00	0.93	0.92	1.00	0.98	0.97
		CP	0.96	0.96	0.96	0.93	0.93	0.93	0.87	0.95	0.95
		M	1.00	0.23	0.20	0.00	0.66	0.62	0.00	0.35	0.29
	5000	RB	−0.17	−0.02	0.00	0.12	−0.01	0.00	−0.42	−0.05	0.01
		SD	0.27	0.56	0.58	1.48	1.81	1.83	0.82	1.76	1.91
		SE/SD	1.00	1.03	1.04	1.00	1.01	1.00	1.00	1.02	1.02
		CP	0.90	0.96	0.96	0.94	0.95	0.95	0.48	0.95	0.95
		MSE-R	1.00	0.32	0.31	0.00	0.73	0.72	0.00	1.12	0.97
40	1000	RB	−0.06	−0.05	0.00	0.04	0.01	0.04	−0.15	−0.02	−0.01
		SD	0.81	1.01	1.04	3.53	3.75	3.85	2.39	3.02	3.15
		SE/SD	1.00	0.98	0.97	1.00	0.99	0.97	1.00	1.02	1.01
		CP	0.95	0.94	0.94	0.96	0.96	0.96	0.95	0.95	0.95
		MSE-R	1.00	0.64	0.61	0.00	0.89	0.84	0.00	0.66	0.61
	5000	RB	−0.05	−0.03	−0.01	0.02	0.00	0.01	−0.14	−0.01	0.01
		SD	0.36	0.45	0.47	1.52	1.62	1.70	1.06	1.34	1.40
		SE/SD	1.00	1.00	0.96	1.00	1.02	0.98	1.00	1.02	1.01
		CP	0.96	0.96	0.95	0.95	0.95	0.94	0.92	0.95	0.94
		MSE-R	1.00	0.65	0.58	0.00	0.89	0.80	0.00	0.79	0.73

C.5.2 Model (3.1) with mean and variance covariates

Here we consider the case where a primary outcome is related to the subject-specific means and variances of a continuous response by the logistic regression model

$$P(Y_i = 1|\mu_i, \sigma_i^2; \beta_\mu, \beta_\sigma) = F\{\beta_0 + \mu_i\beta_\mu + T(\sigma^2)\beta_\sigma\}, \tag{C.10}$$

where $F(t) = 1/(1 + e^{-t})$. This is appropriate when the subject-specific data are repeated measures, rather than longitudinal data. Even with longitudinal data, the slopes may be zero, and the replicate data can be described by an intercept-only model in which the intercept represents the mean of the data. Model (C.10) is a special case of model (3.1) where μ=γ_0 and there is no subject-specific slope.

Table C.3: *Estimation of $\boldsymbol{\beta}$, where $P(Y = 1|\mu, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \mu\beta_\mu + \sigma\beta_\sigma)$, with true values $(\beta_0, \beta_\mu, \beta_\sigma)$=$(-7.2, 0.02, 0.08)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.02 for 1000, 0.01 for 5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.07). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_μ			β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	0.12	0.01	0.00	−0.47	−0.07	0.07
		SD	0.74	0.97	1.01	1.78	3.39	4.13
		SE/SD	1.00	0.94	0.93	1.00	0.98	0.97
		CP	0.93	0.94	0.94	0.44	0.95	0.96
		MSE-R	1.00	0.65	0.60	1.00	1.47	1.00
	5000	RB	0.14	0.04	0.03	−0.48	−0.08	0.02
		SD	0.31	0.39	0.41	0.78	1.47	1.72
		SE/SD	1.00	1.02	1.01	1.00	0.99	0.99
		CP	0.84	0.96	0.96	0.00	0.93	0.94
		MSE-R	1.00	1.08	1.02	1.00	5.78	5.02
40	1000	RB	0.05	−0.01	−0.02	−0.16	−0.01	0.05
		SD	0.74	0.80	0.81	2.21	2.68	2.88
		SE/SD	1.00	1.01	1.02	1.00	1.01	1.01
		CP	0.96	0.95	0.95	0.93	0.97	0.96
		MSE-R	1.00	0.88	0.86	1.00	0.92	0.78
	5000	RB	0.07	0.01	0.00	−0.17	−0.02	0.02
		SD	0.32	0.35	0.35	1.00	1.20	1.28
		SE/SD	1.00	1.05	1.05	1.00	1.00	1.00
		CP	0.94	0.96	0.96	0.72	0.95	0.95
		MSE-R	1.00	1.01	0.99	1.00	1.98	1.74

Table C.4: *Estimation of $\boldsymbol{\beta}$, where $P(Y=1|\mu,\sigma^2;\boldsymbol{\beta}) = F(\beta_0+\mu\beta_\mu+\sigma\beta_\sigma)$, with true values $(\beta_0,\beta_\mu,\beta_\sigma)$=$(-5, 0.01, 0.04)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.04 for 1000, 0.02 for 5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.06). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_μ			β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	0.10	−0.02	−0.03	−0.46	−0.05	0.04
		SD	0.74	0.97	0.99	2.02	3.83	4.20
		SE/SD	1.00	0.98	0.98	1.00	0.96	0.95
		CP	0.96	0.94	0.95	0.86	0.94	0.95
		MSE-R	1.00	0.60	0.57	1.00	0.51	0.42
	5000	RB	0.12	0.00	0.00	−0.46	−0.06	0.00
		SD	0.33	0.43	0.43	0.88	1.65	1.79
		SE/SD	1.00	1.00	1.00	1.00	0.98	0.97
		CP	0.93	0.96	0.96	0.42	0.95	0.95
		MSE-R	1.00	0.67	0.65	1.00	1.51	1.31
40	1000	RB	0.11	0.06	0.05	−0.19	−0.04	−0.01
		SD	0.82	0.89	0.89	2.46	2.98	3.08
		SE/SD	1.00	0.96	0.96	1.00	0.98	0.98
		CP	0.94	0.94	0.94	0.94	0.95	0.95
		MSE-R	1.00	0.87	0.86	1.00	0.75	0.70
	5000	RB	0.07	0.01	0.01	−0.18	−0.03	0.00
		SD	0.36	0.38	0.38	1.05	1.27	1.31
		SE/SD	1.00	1.01	1.00	1.00	1.04	1.04
		CP	0.94	0.96	0.96	0.91	0.96	0.96
		MSE-R	1.00	0.90	0.89	1.00	0.99	0.93

Table C.5: *Estimation of $\boldsymbol{\beta}$, where $P(Y = 1|\mu, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \mu\beta_\mu + log(\sigma^2)\beta_\sigma)$, with true values $(\beta_0, \beta_\mu, \beta_\sigma)$=$(-10.3, 0.02, 0.80)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.02 for 1000, 0.01 for 5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.03 and 0.08). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_μ			β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	0.11	0.00	0.00	−0.46	−0.05	0.03
		SD	0.70	0.91	0.95	18.39	35.63	40.64
		SE/SD	1.00	0.98	0.98	1.00	0.96	0.94
		CP	0.94	0.95	0.95	0.42	0.93	0.94
		M	1.00	0.64	0.59	1.00	1.31	1.02
	5000	RB	0.11	0.01	0.01	−0.48	−0.10	−0.02
		SD	0.32	0.42	0.43	7.99	15.29	17.07
		SE/SD	1.00	0.95	0.95	1.00	0.98	0.99
		CP	0.88	0.95	0.95	0.00	0.90	0.93
		MSE-R	1.00	0.89	0.83	1.00	5.18	5.14
40	1000	RB	0.04	−0.01	−0.02	−0.16	−0.03	0.01
		SD	0.73	0.79	0.80	23.87	28.56	30.24
		SE/SD	1.00	1.02	1.02	1.00	0.95	0.95
		CP	0.95	0.96	0.96	0.89	0.94	0.94
		M	1.00	0.87	0.85	1.00	0.90	0.81
	5000	RB	0.06	0.01	0.01	−0.17	−0.04	0.00
		SD	0.32	0.34	0.35	9.75	11.64	12.30
		SE/SD	1.00	1.04	1.04	1.00	1.04	1.04
		CP	0.95	0.97	0.97	0.72	0.94	0.95
		MSE-R	1.00	1.00	0.98	1.00	1.94	1.86

Table C.6: *Estimation of $\boldsymbol{\beta}$, where $P(Y = 1|\mu, \sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \mu\beta_\mu + log(\sigma^2)\beta_\sigma)$, with true values $(\beta_0, \beta_\mu, \beta_\sigma)$=$(-6.4, 0.01, 0.40)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.04 for 1000, 0.02 for 5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.01 and 0.06). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_μ			β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	0.16	0.03	0.03	−0.45	−0.03	0.02
		SD	0.67	0.88	0.89	17.32	32.74	34.81
		SE/SD	1.00	1.02	1.02	1.00	0.99	0.99
		CP	0.95	0.96	0.96	0.82	0.95	0.95
		M	1.00	0.62	0.60	1.00	0.58	0.51
	5000	RB	0.15	0.03	0.03	−0.44	−0.05	0.00
		SD	0.30	0.38	0.39	7.81	14.54	15.39
		SE/SD	1.00	1.03	1.03	1.00	0.98	0.98
		CP	0.93	0.96	0.96	0.38	0.94	0.94
		MSE-R	1.00	0.74	0.72	1.00	1.76	1.60
40	1000	RB	0.13	0.08	0.07	−0.15	−0.02	0.01
		SD	0.78	0.84	0.84	22.41	26.58	27.37
		SE/SD	1.00	0.95	0.95	1.00	0.96	0.96
		CP	0.94	0.93	0.94	0.92	0.93	0.94
		M	1.00	0.88	0.87	1.00	0.76	0.72
	5000	RB	0.06	0.01	0.01	−0.15	−0.02	0.00
		SD	0.33	0.35	0.35	9.83	11.64	11.99
		SE/SD	1.00	1.01	1.01	1.00	0.98	0.98
		CP	0.94	0.95	0.95	0.90	0.94	0.94
		MSE-R	1.00	0.90	0.89	1.00	0.99	0.94

C.5.3 Model (3.1) with variance covariate only

This appendix provides additional simulation results for the following logistic regression model

$$P(Y_i = 1|\sigma_i^2; \beta_\sigma) = F\{\beta_0 + T(\sigma_i^2)\beta_\sigma\}, \tag{C.11}$$

where $F(t) = 1/(1 + e^{-t})$. These results are similar to the case where $T(\sigma^2)$=log(σ^2), and are discussed in Section 3.4.2.

Table C.7: *Estimation of $\boldsymbol{\beta}$, where $P(Y = 1|\sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \sigma\beta_\sigma)$, with true values (β_0, β_σ)=$(-4, 0.08)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.01 for 1000, 5000) ; SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.04 and 0.08). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	−0.43	−0.06	0.02
		SD	1.57	2.60	3.07
		SE/SD	1.00	1.00	0.99
		CP	0.41	0.95	0.95
		MSE-R	1.00	2.04	1.51
	5000	RB	−0.44	−0.08	0.00
		SD	0.70	1.15	1.34
		SE/SD	1.00	0.99	0.98
		CP	0.00	0.91	0.95
		MSE-R	1.00	7.49	7.04
40	1000	RB	−0.15	−0.02	0.01
		SD	1.86	2.14	2.27
		SE/SD	1.00	1.04	1.04
		CP	0.90	0.96	0.96
		MSE-R	1.00	1.06	0.94
	5000	RB	−0.16	−0.03	0.00
		SD	0.87	1.00	1.06
		SE/SD	1.00	0.99	0.99
		CP	0.69	0.94	0.94
		MSE-R	1.00	2.23	2.13

Table C.8: *Estimation of $\boldsymbol{\beta}$, where $P(Y = 1|\sigma^2; \boldsymbol{\beta}) = F(\beta_0 + \sigma\beta_\sigma)$, with true values (β_0, β_σ)=$(-3.4, 0.04)$, two levels of replication (R), two sample sizes (n). RB, relative bias (standard error approximately 0.03 for 1000, 0.01 for 5000); SD, Monte Carlo standard deviation of $\widehat{\beta}$ multiplied by 100; SE/SD ratio of average sandwich standard deviation to Monte Carlo standard deviation; CP, coverage probability of nominal* 95% *Wald confidence interval; MSE-R, $MSE_{\beta_W}/MSE_{\beta_{\widehat{X}}}$ (standard error in MSE-R divided by MSE-R ranges between 0.02 and 0.06). Methods: W, naive approach; M=2 and M=4, MAI matching 2 and 4 moments.*

				β_σ	
R	n	Stat.	W	$\widehat{X}_{M=2}$	$\widehat{X}_{M=4}$
10	1000	RB	−0.45	−0.10	−0.04
		SD	1.70	2.80	3.03
		SE/SD	1.00	1.00	1.00
		CP	0.85	0.95	0.95
		MSE-R	1.00	0.77	0.67
	5000	RB	−0.43	−0.06	−0.02
		SD	0.79	1.28	1.38
		SE/SD	1.00	0.96	0.96
		CP	0.41	0.94	0.94
		MSE-R	1.00	2.10	1.89
40	1000	RB	−0.19	−0.07	−0.05
		SD	2.05	2.36	2.43
		SE/SD	1.00	0.99	0.99
		CP	0.94	0.94	0.94
		MSE-R	1.00	0.85	0.80
	5000	RB	−0.14	−0.01	0.00
		SD	0.94	1.08	1.12
		SE/SD	1.00	0.97	0.97
		CP	0.89	0.94	0.94
		MSE-R	1.00	1.03	0.98