PSYCHOMETRIKA
2012
DOI: 10.1007/s11336-012-9285-1

IMPROVED REGRESSION CALIBRATION

ANDERS SKRONDAL

DIVISION OF EPIDEMIOLOGY, NORWEGIAN INSTITUTE OF PUBLIC HEALTH

JOUNI KUHA

DEPARTMENT OF STATISTICS, LONDON SCHOOL OF ECONOMICS

The likelihood for generalized linear models with covariate measurement error cannot in general be expressed in closed form, which makes maximum likelihood estimation taxing. A popular alternative is regression calibration which is computationally efficient at the cost of inconsistent estimation. We propose an improved regression calibration approach, a general pseudo maximum likelihood estimation method based on a conveniently decomposed form of the likelihood. It is both consistent and computationally efficient, and produces point estimates and estimated standard errors which are practically identical to those obtained by maximum likelihood. Simulations suggest that improved regression calibration, which is easy to implement in standard software, works well in a range of situations.

Key words: covariate measurement error, measurement model, generalized linear model, pseudo maximum likelihood estimation, regression calibration.

1. Introduction

Generalized linear models (e.g., McCullagh & Nelder, 1989) are the workhorses in many applications of statistical methods. A tacit assumption in these models is that all covariates are perfectly measured without error. Violation of this assumption will produce inconsistent estimators unless the measurement error problem is addressed. A body of research has hence evolved to allow at least approximate inference in generalized linear models with covariate measurement error (see Carroll, Ruppert, Stefanski, & Crainiceanu, 2006, and Buonaccorsi, 2010, for comprehensive overviews; we will discuss some of this literature in more detail later).

In this article, we consider *structural* covariate measurement error models, where a parametric distribution is specified for the erroneously measured covariates. An obvious approach to estimation is then maximum likelihood which produces consistent estimates if the model is correctly specified (e.g., Schafer, 1987; Schafer & Purdy, 1986; Higdon & Schafer, 2001). Unfortunately, the joint likelihood of the response and the measures cannot in general be expressed in closed form and computationally intensive methods based on numerical integration or simulation must be used. The computational burden involved in a full likelihood analysis is, therefore, often considerable.

Regression calibration has been proposed as a computationally efficient approach to estimating generalized linear models with covariate measurement error (e.g., Armstrong, 1985; Rosner, Willett, & Spiegelman, 1989; Rosner, Spiegelman, & Willett, 1990; Carroll & Stefanski, 1990). It is based on an approximation of the likelihood function where the basic idea is to plug in "best" predictions for the covariates measured with error and proceed in estimating the generalized linear model as if the predictions were covariates measured *without* error. Unfortunately, estimates of the regression parameters from regression calibration are, in general, inconsistent.

Requests for reprints should be sent to Anders Skrondal, Division of Epidemiology, Norwegian Institute of Public Health, P.O. Box 4404, Nydalen, 0403 Oslo, Norway. E-mail: anders.skrondal@fhi.no

Published online: 18 September 2012

The inconsistency is typically small when the true effects of the covariates measured with error are moderate and/or the measurement error variances are small, but more pronounced when these conditions do not hold.

In this article, we propose a pseudo maximum likelihood approach, called improved regression calibration (IRC), which simultaneously addresses the computational challenge in likelihood analysis and the inconsistency problem in conventional regression calibration. The basic idea is to consider a decomposed form of the likelihood where one component is expressed in closed form and trivial to maximize, and the second component is accurately maximized using crude and fast numerical integration. In contrast to conventional regression calibration, where predicted covariates measured with error are treated as fixed in point estimation, the stochastic nature of the predictions is handled by using predictive densities of the covariates measured with error as mixing distributions.

2. Generalized Linear Models with Covariate Measurement Error

Let y_i be the outcome variable for unit i, $i = 1, \ldots, N$, $\mathbf{x}_i$ an $m \times 1$ vector of covariates or "exposures" measured with error by the measures $\mathbf{w}_i$, and $\mathbf{z}_i$ a vector of perfectly measured covariates, including a constant 1.

Following Clayton (1992), we can view a generalized linear model with covariate measurement error as composed of three parts: (i) an outcome model $g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$, (ii) a measurement model $g(\mathbf{w}_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{M}})$, and (iii) an exposure model $g(\mathbf{x}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{E}})$, where $g(\cdot|\cdot)$ are conditional density functions and $\boldsymbol{\vartheta}_{\mathsf{O}}$, $\boldsymbol{\vartheta}_{\mathsf{M}}$, and $\boldsymbol{\vartheta}_{\mathsf{E}}$ the corresponding parameter vectors. We define the complete parameter vector as $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_{\mathsf{O}}, \boldsymbol{\vartheta}'_{\mathsf{M}}, \boldsymbol{\vartheta}'_{\mathsf{E}})'$. Throughout, we make the standard assumption of "nondifferential measurement error" that y_i and $\mathbf{w}_i$ are independent conditional on $(\mathbf{x}_i, \mathbf{z}_i)$.

2.1. *Outcome Model* $g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$

The outcome model is a generalized linear model (e.g., McCullagh & Nelder, 1989) with three parts: (i) a linear predictor, which in the present context takes the form $\eta_i \equiv \mathbf{z}'_i\boldsymbol{\beta}_z + \mathbf{x}'_i\boldsymbol{\beta}_x$, (ii) a link function $g(\cdot)$ that links the linear predictor to the conditional expectation of the response, given the covariates, $\mathrm{E}(y_i|\mathbf{x}_i, \mathbf{z}_i) = g^{-1}(\eta_i)$, and (iii) a conditional distribution for the response, given the covariates, taken from the exponential family,

$$g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\}.$$

Here, $\theta_i = \theta_i(\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$ is the canonical or natural parameter, $\phi = \phi(\boldsymbol{\vartheta}_{\mathsf{O}})$ is the scale or dispersion parameter, and $b(\cdot)$ and $c(\cdot)$ are functions depending on the member of the exponential family. The most common nonlinear instance of this is the binary logistic model where y_i follows a Bernoulli distribution and $\theta_i = \eta_i = \log\{\mathrm{E}(y_i)/[1 - \mathrm{E}(y_i)]\}$. For this model, $\phi = 1$ and $\boldsymbol{\vartheta}_{\mathsf{O}} = \boldsymbol{\beta} = (\boldsymbol{\beta}'_z, \boldsymbol{\beta}'_x)'$. Due to its popularity, we will consider a logistic outcome model in our simulations and data analysis.

2.2. *Measurement Model* $g(\mathbf{w}_i|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}})$

The form of the measurement model depends on the nature of the available data. Here we focus on the case of replication data, where at least a subsample of subjects provides several measures for each fallibly measured covariate. The main alternative is validation data where both $\mathbf{x}_i$ and $\mathbf{w}_i$ are observed for a subsample, in which case the proposed estimation procedures can be modified in a straightforward manner.

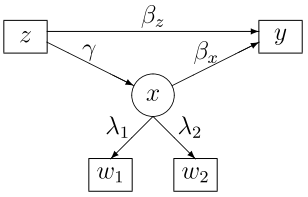

FIGURE 1.
Graph of generalized linear model with covariate measurement error.

In general, the measurements $\mathbf{w}_i$ may depend on the covariates $\mathbf{z}_i$ measured without error as well as on $\mathbf{x}_i$, similarly to differential item functioning in item response theory. This would be straightforward to handle in our suggested approach, but here we omit $\mathbf{z}_i$ for simplicity and consider measurement models of the form $g(\mathbf{w}_i|\mathbf{x}_i;\boldsymbol{\vartheta}_{\mathsf{M}})$.

The vector $\mathbf{x}_i$ is measured by fallible measures $\mathbf{w}_i = (\mathbf{w}'_{1i}, \ldots, \mathbf{w}'_{mi})'$, where each $\mathbf{w}_{li} = (w_{li1}, \ldots, w_{lin_{li}})'$ is a vector of n_{li} replicate measurements. For the moment, consider balanced data where $n_{li} = n_l$. A general multidimensional measurement model for m sets of congeneric measures (e.g., Jöreskog, 1971) can be expressed as

$$\mathbf{w}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{x}_i + \boldsymbol{\delta}_i, \quad \boldsymbol{\delta}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Theta}) \tag{1}$$

where $\boldsymbol{\Psi} \equiv \mathrm{Cov}(\mathbf{x}_i)$, $\boldsymbol{\Theta} \equiv \mathrm{Cov}(\boldsymbol{\delta}_i)$, and it is assumed that $\mathrm{Cov}(\mathbf{x}_i, \boldsymbol{\delta}_i) = \mathbf{0}$. The matrix $\boldsymbol{\Lambda}$ is partitioned as

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\lambda}_m \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\lambda}_l$ is a vector of scale parameters for the measures of covariate l. Further constraints are often imposed on the parameters of the measurement model, e.g., to obtain tau-equivalent or parallel models.

2.3. *Exposure Model* $g(\mathbf{x}_i|\mathbf{z}_i;\boldsymbol{\vartheta}_{\mathsf{E}})$

The dependence between the exposures measured with error $\mathbf{x}_i$ and the covariates measured without error $\mathbf{z}_i$ is specified as

$$\mathbf{x}_i = \boldsymbol{\Gamma}\mathbf{z}_i + \boldsymbol{\zeta}_i, \tag{3}$$

where $\boldsymbol{\Gamma}$ is a regression parameter matrix, $\boldsymbol{\zeta}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi})$, and $\mathrm{Cov}(\mathbf{z}_i, \boldsymbol{\zeta}_i) = \mathbf{0}$. As the scale of $\mathbf{x}_i$ is not identifiable from (1) and (3), some standard identification restrictions are imposed on the parameters. The parameter vector $\boldsymbol{\vartheta}_{\mathsf{M}}$ then consists of the unique elements of $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$, and $\boldsymbol{\vartheta}_{\mathsf{E}}$ of the unique elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$.

A generalized linear model with covariate measurement error is shown graphically in Figure 1 for the simple case of an exposure x_i fallibly measured by two measures w_{i1} and w_{i2}, and a covariate z_i measured without error. A common identifiability constraint for this case is to assume $\nu_1 = \nu_2 = 0$ and $\lambda_1 = \lambda_2 = 1$, which give the "classical" measurement error model $w_{ij} = x_i + \delta_{ij}$.

The method that we propose below is not dependent on this specific combination of measurement and outcome models, but applies also more generally. Looking ahead to the rest of the paper, other study designs, and corresponding changes to measurement and outcome models,

affect only Stage 1 of our two-stage estimation. For example, a situation where the number of replicate measurements is not the same for all units i is accounted for by the selection matrix $\mathbf{C}_i$ included in Equation (8), and the case where y_i is not observed for some units by omitting these from the log-likelihood component $\ell_2(\boldsymbol{\vartheta}_{\mathsf{O}}, \boldsymbol{\vartheta}_{\mathsf{ME}})$ in (5). If a validation sample rather than replication data are available, Stage 1 of the estimation could be done by modeling the conditional moments of $\mathbf{x}_i$ given $\mathbf{w}_i$ and $\mathbf{z}_i$ (Equations (11) and (12)) directly rather than via the exposure and measurement models; in this case, the formulas of the variance estimation in the appendix would also be simplified.

3. Estimation Methods

We now consider different approaches to estimation of generalized linear models with covariate measurement error. We start by briefly describing maximum likelihood (ML) estimation, then proceed by developing our suggested approach of improved regression calibration (IRC) before contrasting this with conventional regression calibration (RC). We then conclude this section by a discussion of previous literature on these approaches to measurement error modeling. Throughout, we consider likelihoods for the response y_i and the measures $\mathbf{w}_i$ conditional on the perfectly measured covariates $\mathbf{z}_i$.

3.1. Maximum Likelihood (ML) Estimation

The likelihood contribution for a single unit i is

$$g(y_i, \mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}) = \int g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}) g(\mathbf{w}_i|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}}) g(\mathbf{x}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{E}})\, \mathrm{d}\mathbf{x}_i, \tag{4}$$

the log-likelihood contribution is $\ell_i(\boldsymbol{\vartheta}) = \log g(y_i, \mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta})$, and the log-likelihood $\ell(\boldsymbol{\vartheta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\vartheta})$. When $\boldsymbol{\Theta}$ is diagonal, as is often assumed, $g(\mathbf{w}_i|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}}) = \prod_{l=1}^{m} \prod_{j=1}^{n_{li}} g(w_{lij}|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}})$. The ML estimator $\widehat{\boldsymbol{\vartheta}}$ is obtained by maximizing $\ell(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$.

Unfortunately, the joint likelihood of generalized linear models with covariate measurement error cannot generally be expressed in closed form and requires integration, typically accomplished by Gaussian quadrature. In general, the performance of Gaussian quadrature depends on the smoothness of the integrand. According to the fundamental theorem of Gaussian quadrature (e.g., Davis & Rabinowitz, 1984; Thisted, 1988, Theorem 5.3-1), ordinary Gaussian quadrature is *exact* if the function in the integrand is a $2R - 1$ order polynomial (where R is the number of quadrature points). However, a likelihood component including a product of conditional response distributions for continuous responses, such as $\prod_{l=1}^{m} \prod_{j=1}^{n_{li}} g(w_{lij}|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}})$ above, tends to produce a peaked integrand in the marginal likelihood (a tendency exacerbated as the number of measures and their intraclass correlation increases). Such likelihood contributions are poorly approximated by low-degree polynomials, and ordinary Gauss–Hermite quadrature does not work well for this situation (e.g., Albert & Follmann, 2000; Lesaffre & Spiessens, 2001). This is illustrated in the left panel of Figure 2 where we see that all quadrature points completely miss the integrand.

Therefore, more computationally demanding *adaptive* Gaussian quadrature methods that align the quadrature points under the integrand are recommended when continuous responses are involved (e.g., Rabe-Hesketh, Skrondal, & Pickles, 2005). A limitation of the full likelihood approach is, hence, that it becomes computationally intensive.

3.2. *Improved Regression Calibration (IRC)*

As an alternative to full ML we propose to break the estimation problem into two parts, allocating as many parameters as possible to a likelihood component that is easy to maximize. This is an instance of a general two-stage approach to estimation known as pseudo maximum likelihood (PML) estimation (Gong & Samaniego, 1981).

Letting $\boldsymbol{\vartheta}_{\mathsf{ME}} = (\boldsymbol{\vartheta}'_{\mathsf{M}}, \boldsymbol{\vartheta}'_{\mathsf{E}})'$, we first re-express $g(\mathbf{w}_i|\mathbf{x}_i; \boldsymbol{\vartheta}_{\mathsf{M}})g(\mathbf{x}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{E}})$ in (4) as $g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})g(\mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})$, and the log-likelihood as

$$\ell(\boldsymbol{\vartheta}) = \sum_{i=1}^{N} \log g(y_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}, \boldsymbol{\vartheta}_{\mathsf{ME}}) + \sum_{i=1}^{N} \log g(\mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}}) \equiv \ell_2(\boldsymbol{\vartheta}_{\mathsf{O}}, \boldsymbol{\vartheta}_{\mathsf{ME}}) + \ell_1(\boldsymbol{\vartheta}_{\mathsf{ME}}) \quad (5)$$

where

$$g(y_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}, \boldsymbol{\vartheta}_{\mathsf{ME}}) = \int g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})\, g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})\, \mathrm{d}\mathbf{x}_i. \quad (6)$$

In Stage 1 of IRC, we estimate the combined measurement and exposure model $g(\mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})$ by maximizing just $\ell_1(\boldsymbol{\vartheta}_{\mathsf{ME}})$, to obtain estimates $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$. These are not full ML estimates because they omit the typically small amount of information about of $\boldsymbol{\vartheta}_{\mathsf{ME}}$ contained in y_i. In Stage 2, these estimates from Stage 1 are then treated as known, and estimates $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$ for the parameters of primary interest $\boldsymbol{\vartheta}_{\mathsf{O}}$ are obtained by maximizing $\ell_2(\boldsymbol{\vartheta}_{\mathsf{O}}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$. A detailed description of the two stages is provided in the next section.

The basic idea of IRC is that maximizing the approximate decomposed likelihood is considerably less demanding than maximizing the joint likelihood. In Stage 1, the component $g(\mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})$ is in closed form and trivial to maximize. In Stage 2, the mixing distribution in the integral (6) is the *predictive density* $g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$ of the covariates measured with error, given their observed measures and covariates measured without error, which is also trivial to obtain.

The dimensionality of integration (the number of covariates measured with error) in Stage 2 is the same as for full ML. At first glance, there does, hence, not appear to be any computational benefits to be reaped from using IRC. However, the integrand is now the single logistic function $g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$, which due to its smoothness is well approximated by a low order polynomial. For instance, the seminal work on nonlinear factor analysis by McDonald (1967) demonstrated that a cubic function sufficed for approximating the normal ogive (which is very close to the logistic function). We therefore expect that crude and fast ordinary Gauss–Hermite quadrature, using just a few quadrature points, would work well for IRC. This is illustrated in the right panel of Figure 2, where all three quadrature points nicely cover the logistic integrand, in contrast to the case for the likelihood in the left panel.

It is likely that direct maximization of the full likelihood expressed as (5) could also be based on more crude Gauss–Hermite quadrature than what is required for the standard form (4). In this article, however, we focus on the two-stage approach to (5), since it is straightforward to implement in publicly available software.

The savings compared to ML are especially pronounced in three settings and their combinations: (i) large datasets, (ii) when the relative number of parameters allocated to the easily maximized likelihood component is large (a large number of measures and/or realistically complex measurement models), and (iii) when the same predictive distributions can be used in several models, so that the Stage-1 likelihood components need only be maximized once.

3.3. *Conventional Regression Calibration (RC)*

Conventional regression calibration is also a two-stage method which can be seen as an approximation of pseudo-ML (IRC) estimation. Stage 1 is the same as for IRC, but estimation in

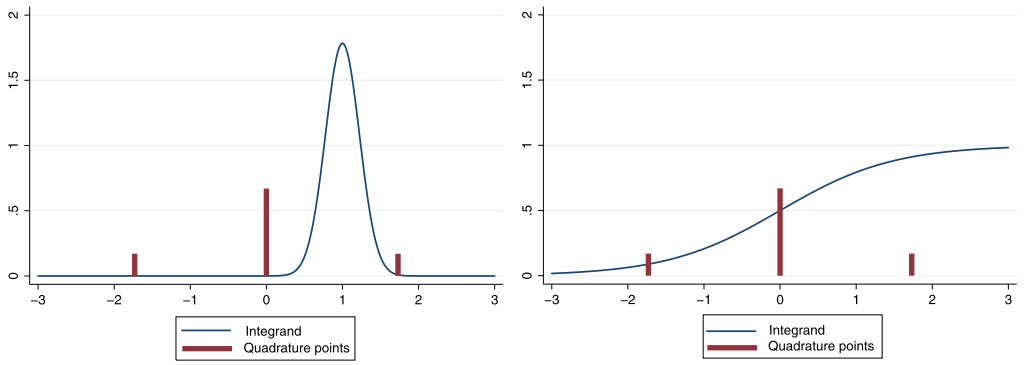

FIGURE 2.
Illustration of integrand and quadrature points (locations and weights) for 3-point ordinary Gauss–Hermite quadrature. Maximum likelihood in *left panel* and improved regression calibration in *right panel*.

Stage 2 is based on the further approximation

$$g(y_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}) \approx g(y_i|\widetilde{\boldsymbol{\xi}}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}) \tag{7}$$

where $g(y_i|\widetilde{\boldsymbol{\xi}}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$ is of the same form as the outcome model $g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$, now with the "predictive mean" $\widetilde{\boldsymbol{\xi}}_i = \mathrm{E}(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$ used in the place of $\mathbf{x}_i$. RC thus carries only $\widetilde{\boldsymbol{\xi}}_i$ forward from Stage 1 to Stage 2 of the estimation, whereas IRC takes the whole predictive density $g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$ into account in Stage 2. In contrast to IRC, RC is generally inconsistent because it employs the approximation (7) of the likelihood function (6).

3.4. ML, PML, and RC in the Measurement Error Literature

The books by Carroll et al. (2006) and Buonaccorsi (2010) provide excellent summaries of methods of estimation in measurement error modeling. The use of full ML estimation has been advocated in a series of papers by Daniel Schafer and coauthors. Schafer (1993), for binary probit models, and Schafer and Purdy (1986), for normal linear models, consider cases where the likelihood can be evaluated in a closed form. For cases where this is not possible, such as binary logistic regression, Schafer (1987) uses a closed-form approximation to avoid numerical integration, while Higdon and Schafer (2001) employ ordinary Gauss–Hermite quadrature to evaluate the likelihood. Rabe-Hesketh, Skrondal, and Pickles (2003) propose using more accurate adaptive quadrature in this setting. Another possibility is to estimate the models in a Bayesian framework, using simulation-based MCMC methods (e.g., Stephens & Dellaportas, 1992; Richardson & Gilks, 1993; Kuha, 1997; Gustafson, 2004).

Key references for regression calibration include Armstrong (1985), Rosner et al. (1989, 1990), Carroll and Stefanski (1990), and Gleser (1990), and the overview in Carroll et al. (2006). Buonaccorsi (2010) points out that regression calibration is also a "pseudo-type" two-stage method, which can be viewed as an approximation of PML estimation.

The possibility of PML estimation for regression models with covariates measured with error was noted early, for example, by Carroll, Spiegelman, Lan, Bailey, and Abbott (1984), who apply it for a binary probit model, and Armstrong (1985). PML estimation has been suggested for some specific models where its implementation is relatively straightforward, such as probit models with a single covariate (Burr, 1988), linear mixed models (Buonaccorsi, Demidenko, & Tosteson, 2000) and linear structural equation models with latent variables (Skrondal & Laake, 2001). For other models, however, the approach has not been developed, perhaps because of a perception that its implementation requires "specialized programming" (Buonaccorsi, 2010,

p. 227). The IRC method proposed here provides a general approach to PML for covariate measurement models which largely avoids such programming.

4. The Anatomy of Improved Regression Calibration

We will now have a closer look at each of the stages of IRC.

4.1. Stage 1: Estimation of the MIMIC Model $g(\mathbf{w}_i|\mathbf{z}_i;\boldsymbol{\vartheta}_{\mathsf{ME}})$

We can view (1) as representing the measurement model for a possibly hypothetical complete set of replicate measurements $\mathbf{w}_i$, where the numbers of measurements in $\mathbf{w}_{li}$ are n_l for each unit i. The numbers of actually observed replicates may in fact be $n_{li} < n_l$ for some i, l, due to design and/or nonresponse. The most common case of unbalanced data by design is one where replicate measurements are only collected for a subsample, so that $n_{li} = 1$ outside the subsample. Defining $n_i = \sum_l n_{li}$ and $n = \sum_l n_l$, the model for such possibly incomplete measurements is obtained by multiplying the right-hand side of (1) by an $n_i \times n$ selection matrix $\mathbf{C}_i$. We will henceforth include $\mathbf{C}_i$ where appropriate in the formulae since this is required for obtaining correct results in the unbalanced case where the n_{li} are not constant.

Together, the measurement and exposure models constitute a multiple-indicator multiple-cause (MIMIC) model (e.g., Robinson, 1974; Jöreskog & Goldberger, 1975). To obtain $g(\mathbf{w}_i|\mathbf{z}_i;\boldsymbol{\vartheta}_{\mathsf{ME}})$, we substitute the exposure model into the measurement model, producing the reduced form MIMIC model

$$\mathbf{w}_i = \mathbf{C}_i(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{z}_i + \boldsymbol{\Lambda}\boldsymbol{\zeta}_i + \boldsymbol{\delta}_i), \tag{8}$$

for which the conditional first and second order moments are

$$\boldsymbol{\mu}_i \equiv \mathrm{E}(\mathbf{w}_i|\mathbf{z}_i) = \mathbf{C}_i(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{z}_i) \quad \text{and} \tag{9}$$

$$\boldsymbol{\Sigma}_i \equiv \mathrm{COV}(\mathbf{w}_i|\mathbf{z}_i) = \mathbf{C}_i\big(\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\big)\mathbf{C}_i'. \tag{10}$$

The density for the measures, given the perfectly measured covariates, becomes $\mathbf{w}_i|\mathbf{z}_i \sim \mathrm{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, and the log-likelihood $\ell_1(\boldsymbol{\vartheta}_{\mathsf{ME}})$ for the combined measurement and exposure model can be expressed in closed form.

The estimates $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$ that maximize $\ell_1(\boldsymbol{\vartheta}_{\mathsf{ME}})$ can be obtained in a very computationally efficient manner using standard methods for moment structure modeling (e.g., Bentler, 1983). The estimates are consistent as $N \to \infty$ for fixed n_i under mild regularity conditions, not requiring the normality assumptions imposed above (e.g., Shapiro, 2007). They remain consistent also when measurements are missing at random (MAR) in the sense of Rubin (1976), although MAR is slightly more restrictive here than for full ML since y_i is not a part of the Stage-1 likelihood.

4.2. Stage 2: Estimation of the Model $g(y_i|\mathbf{w}_i, \mathbf{z}_i;\boldsymbol{\vartheta}_{\mathsf{O}}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$

Under the models (1) and (3) assumed in Stage 1, the predictive density of the covariates measured with error given their observed measures and the covariates measured without error becomes $\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i \sim \mathrm{N}(\boldsymbol{\xi}_i, \boldsymbol{\Omega}_i)$, with the conditional mean and variance matrix

$$\boldsymbol{\xi}_i \equiv \mathrm{E}(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i;\boldsymbol{\vartheta}_{\mathrm{ME}}) = \boldsymbol{\Gamma}\mathbf{z}_i + \boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_i) \quad \text{and} \tag{11}$$

$$\boldsymbol{\Omega}_i \equiv \mathrm{Cov}(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i;\boldsymbol{\vartheta}_{\mathrm{ME}}) = \boldsymbol{\Psi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{C}_i\boldsymbol{\Lambda}\boldsymbol{\Psi}, \tag{12}$$

where we note the role of the selection matrix $\mathbf{C}_i$. Substituting estimates $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$ for the parameters in (11) and (12), we obtain empirical Bayes (EB) predictions $\widetilde{\boldsymbol{\xi}}_i$ for $\mathbf{x}_i$ for each unit i, and their predictive variances $\widehat{\boldsymbol{\Omega}}_i$ (e.g., Skrondal & Rabe-Hesketh, 2004, Chap. 6, and 2009). The EB predictions are identical to the empirical best linear unbiased predictions (EBLUP), which do not hinge on distributional assumptions (e.g., Robinson, 1991).

We finally estimate the parameters of primary interest $\boldsymbol{\vartheta}_{\mathsf{O}}$. Note that, conditional on $(\mathbf{w}_i, \mathbf{z}_i)$ and given the estimates $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$, we can write $\mathbf{x}_i = \widetilde{\boldsymbol{\xi}}_i + \mathbf{u}_i$ where $\mathbf{u}_i \sim \mathrm{N}(\mathbf{0}, \widehat{\boldsymbol{\Omega}}_i)$, independent of $\mathbf{w}_i$ and $\mathbf{z}_i$. Substituting this into (6) gives

$$
\begin{aligned}
g(y_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}) &= \int g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})\, g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})\, \mathrm{d}\mathbf{x}_i \\
&= \int g^*(y_i|\widetilde{\boldsymbol{\xi}}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\vartheta}_{\mathsf{O}})\, g(\mathbf{u}_i; \widehat{\boldsymbol{\Omega}}_i)\, \mathrm{d}\mathbf{u}_i \qquad (13)
\end{aligned}
$$

where $g^*(y_i|\widetilde{\boldsymbol{\xi}}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$ is a generalized linear model of the same kind as $g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})$, but with the linear predictor

$$
\eta_i = \mathbf{z}_i'\boldsymbol{\beta}_z + (\widetilde{\boldsymbol{\xi}}_i + \mathbf{u}_i)'\boldsymbol{\beta}_x = \mathbf{z}_i'\boldsymbol{\beta}_z + \widetilde{\boldsymbol{\xi}}_i'\boldsymbol{\beta}_x + \mathbf{u}_i'\boldsymbol{\beta}_x, \qquad (14)
$$

which includes the vector of random effects $\mathbf{u}_i$. For the case of a single covariate x_i measured with error, the linear predictor can be expressed as $\eta_i = \mathbf{z}_i'\boldsymbol{\beta}_z + \beta_x\widetilde{\xi}_i + \beta_x u_i$, where $u_i \sim \mathrm{N}(0, \hat{\omega}_i)$ and $\hat{\omega}_i = \widehat{\boldsymbol{\Omega}}_i$ is a scalar.

Model (13) is a special case of a generalized linear latent and mixed model (GLLAMM), see, for instance, Rabe-Hesketh, Skrondal, and Pickles (2004a) and Skrondal and Rabe-Hesketh (2004, 2007). It differs from a conventional generalized linear mixed model (GLMM) in several regards. First, the model is for single-level data instead of multilevel or clustered data. The model is identified because the covariance matrix $\widehat{\boldsymbol{\Omega}}_i$ of $\mathbf{u}_i$ is treated as known from Stage 1, and $\boldsymbol{\beta}_x$ is constrained to be equal to the coefficients of $\widetilde{\boldsymbol{\xi}}_i$ (a model simply introducing level-1 random effects with a free variance matrix, without any parameter restriction, is not identified). Second, the mixing distribution is the *predictive density* of the unobserved $\mathbf{x}_i$. Third, the random effects are multiplied by unknown parameters. An important practical merit of IRC is that model (13) can be estimated using the gllamm program (e.g., Rabe-Hesketh, Skrondal, & Pickles, 2004b; Rabe-Hesketh & Skrondal, 2012).

5. Properties of Improved Regression Calibration

The IRC estimator $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$ is the value of $\boldsymbol{\vartheta}_{\mathsf{O}}$ which maximizes the second-stage log-likelihood $\ell_2(\boldsymbol{\vartheta}_{\mathsf{O}}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$, where $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$ is a consistent estimator of $\boldsymbol{\vartheta}_{\mathsf{ME}}$ obtained by maximizing $\ell_1(\boldsymbol{\vartheta}_{\mathsf{ME}})$ in the first stage. This is an instance of a general approach to estimation where the parameters of a model are divided into two sets, one of which contains the parameters of interest and the other involves only nuisance parameters. The nuisance parameters are first estimated by some consistent and computationally convenient estimators, and the parameters of interest are then estimated by maximizing an appropriate objective function with the estimates of the nuisance parameters from the first step treated as known. This is known as pseudo maximum likelihood (PML) estimation when, as here, the second-stage objective function is a likelihood (Gong & Samaniego, 1981), and more generally as quasi generalized extremum estimation (Gourieroux & Monfort, 1995).

It is well known that such two-stage estimators are consistent and asymptotically normally distributed under very general regularity conditions. The conditions and a proof of the consistency are given by Gourieroux and Monfort (1995, Sects. 24.2.4 and 24.2.2). In the notation of

our problem, denote the true parameter value by $\boldsymbol{\vartheta}^* = (\boldsymbol{\vartheta}_{\mathsf{O}}^{*\prime}, \boldsymbol{\vartheta}_{\mathsf{ME}}^{*\prime})'$. Then $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$ is consistent for $\boldsymbol{\vartheta}_{\mathsf{O}}^*$ if, first, standard regularity conditions hold so that the ML estimator of the whole of $\boldsymbol{\vartheta}$ is itself consistent for $\boldsymbol{\vartheta}^*$ and, second, if (i) $\boldsymbol{\vartheta}_{\mathsf{O}}$ and $\boldsymbol{\vartheta}_{\mathsf{ME}}$ can vary independently of each other, and (ii) $\widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}$ is consistent for $\boldsymbol{\vartheta}_{\mathsf{ME}}^*$. All of these conditions are satisfied in the case considered here.

Let $\mathbf{u}(\boldsymbol{\vartheta}) = \partial \ell(\boldsymbol{\vartheta})/\partial \boldsymbol{\vartheta}$ be the score function, partitioned as

$$\mathbf{u}(\boldsymbol{\vartheta}) = \left(\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}'},\ \frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_{\mathsf{ME}}'} \right)' = \left(\mathbf{u}_{\vartheta_{\mathsf{O}}}(\boldsymbol{\vartheta})',\ \mathbf{u}_{\vartheta_{\mathsf{ME}}}(\boldsymbol{\vartheta})' \right)',$$

and define the mean score as $\bar{\mathbf{u}}(\boldsymbol{\vartheta}) = (\bar{\mathbf{u}}_{\vartheta_{\mathsf{O}}}(\boldsymbol{\vartheta})',\ \bar{\mathbf{u}}_{\vartheta_{\mathsf{ME}}}(\boldsymbol{\vartheta})')' = N^{-1}\,\mathbf{u}(\boldsymbol{\vartheta})$. Define the Fisher information matrix

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\vartheta}^*) = \lim_{N\to\infty} \mathrm{E}_{\boldsymbol{\vartheta}^*}\left[-\frac{\partial \bar{\mathbf{u}}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'}\bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}^*} \right] = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}} & \\ \boldsymbol{\mathcal{I}}_{\mathsf{ME},\mathsf{O}} & \boldsymbol{\mathcal{I}}_{\mathsf{ME},\mathsf{ME}} \end{bmatrix}$$

with partitions corresponding to $\boldsymbol{\vartheta}_{\mathsf{O}}$ and $\boldsymbol{\vartheta}_{\mathsf{ME}}$. For the asymptotic normality of $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$, it is further supposed that

$$N^{1/2} \begin{bmatrix} \bar{\mathbf{u}}_{\vartheta_{\mathsf{O}}}(\boldsymbol{\vartheta}_{\mathsf{O}}^*, \boldsymbol{\vartheta}_{\mathsf{ME}}^*) \\ \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}} - \boldsymbol{\vartheta}_{\mathsf{ME}}^* \end{bmatrix} \xrightarrow{\mathcal{L}} \mathsf{N}\left(\mathbf{0},\ \begin{bmatrix} \boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}} & \\ \mathbf{V}_{\mathsf{ME},\mathsf{O}} & \mathbf{V}_{\mathsf{ME},\mathsf{ME}} \end{bmatrix} \right). \tag{15}$$

Then

$$N^{1/2}\left(\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}} - \boldsymbol{\vartheta}_{\mathsf{O}}^* \right) \xrightarrow{\mathcal{L}} \mathsf{N}(\mathbf{0}, \boldsymbol{\Xi}) \tag{16}$$

where

$$\boldsymbol{\Xi} = \boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}}^{-1} + \boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}}^{-1} \boldsymbol{\mathcal{I}}_{\mathsf{ME},\mathsf{O}}' \mathbf{V}_{\mathsf{ME},\mathsf{ME}} \boldsymbol{\mathcal{I}}_{\mathsf{ME},\mathsf{O}} \boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}}^{-1}. \tag{17}$$

The relatively simple form of (17) follows from the fact that for PML estimators $\mathbf{V}_{\mathsf{ME},\mathsf{O}} = \mathbf{0}$ in general, so terms involving $\mathbf{V}_{\mathsf{ME},\mathsf{O}}$ disappear from the expression (Parke, 1986). The asymptotic covariance matrix of the IRC estimator, which also takes into account the uncertainty of the Stage-1 estimates, is then given as $\mathrm{ACOV}(\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}) = N^{-1}\,\boldsymbol{\Xi}$.

In (17), $N^{-1}\boldsymbol{\mathcal{I}}_{\mathsf{O},\mathsf{O}}^{-1}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$ if $\boldsymbol{\vartheta}_{\mathsf{ME}}$ were known. An estimate of it is obtained as a by-product of fitting model (13), and an estimate of $N^{-1}\mathbf{V}_{\mathsf{ME},\mathsf{ME}}$ similarly from fitting (8). The remaining part of (17) is $\boldsymbol{\mathcal{I}}_{\mathsf{ME},\mathsf{O}}$, which we estimate by

$$\widehat{\boldsymbol{\mathcal{I}}}_{\mathsf{ME},\mathsf{O}} = N^{-1} \sum_{i=1}^{N} \mathbf{u}_{\vartheta_{\mathsf{ME}},i}\left(\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}} \right) \mathbf{u}_{\vartheta_{\mathsf{O}},i}\left(\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}} \right)' \tag{18}$$

where $\mathbf{u}_{\vartheta_{\mathsf{O}},i}(\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}})$ and $\mathbf{u}_{\vartheta_{\mathsf{ME}},i}(\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}})$ are the gradients of the log-likelihood $\ell_i(\boldsymbol{\vartheta})$ for unit i, evaluated at the parameter estimates $\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}} = (\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}\prime}, \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}}')'$. How to obtain the required gradients is demonstrated in the appendix.

In summary, the difference between ML and IRC does not concern consistency, as both estimators are consistent. Rather, the difference is the loss of efficiency, compared to ML, which is incurred by IRC when it discards the data on y_i in estimating $\boldsymbol{\vartheta}_{\mathsf{ME}}$ in the first stage. However, we would expect this inefficiency to be slight because very little information about $\boldsymbol{\vartheta}_{\mathsf{ME}}$ is contained in the y_i in the sample. This is examined further in the next section.

6. Simulations

We use a simulation study to compare the performance of maximum likelihood (ML), improved regression calibration (IRC) and conventional regression calibration (RC) estimators. This is done in two parts, comparing first ML and IRC—which turn out to be virtually identical—and then IRC with RC.

For the exposure model we simulate a covariate measured with error as $x_i = 0.3z_i + \zeta_i$, with $z_i \sim N(0, 1)$, independently distributed of $\zeta_i \sim N(0, \psi)$, where $\psi = 1$. For the measurement model we consider $n_i = 2$ measures w_{ij} of x_i for each i, and simulate from a parallel or classical linear measurement model $w_{ij} = x_i + \delta_{ij}$, where $\delta_{ij} \sim \mathrm{N}(0, \theta)$. Finally, for the outcome model we simulate from the logistic regression model $\mathrm{logit}\{\Pr(y_i = 1|x_i, z_i)\} = \beta_0 + \beta_z z_i + \beta_x x_i$.

Three values of the coefficient β_x of the fallibly measured covariate are considered: a moderate magnitude $\beta_x = 0.5$, a high magnitude $\beta_x = 1$, and a very high magnitude $\beta_x = 1.5$, which correspond respectively to odds ratios of 1.65, 2.72, and 4.48 for one standard deviation change in x. The very high magnitude case is included in the spirit of Buzas and Stefanski (1995, p. 546) to provide a tough test. For the measurement error variance θ, we use values $\theta = 1$ and $\theta = 0.33$. These give two different values for the reliabilities $\rho = \psi/(\psi + \theta)$, a moderate reliability case where $\rho = 0.5$ and a high reliability case where $\rho = 0.75$. The parameters β_z and β_0 are fixed at 0.5 and -2, respectively, throughout all simulations. We consider the sample sizes $N = 200$, $N = 1000$, and $N = 5000$. For each setting, 1000 replications of datasets are simulated.

ML estimation was carried out using numerical integration with 8 point adaptive quadrature. For IRC we used 3 point ordinary Gaussian quadrature, motivated by the earlier discussion of crude and fast quadrature approximation in this setting. There were, however, a handful of cases where the latter was not accurate enough, indicated by clearly divergent estimates from ML and IRC. To rectify this, we re-estimated the models using adaptive quadrature whenever the IRC estimate of β_x or β_z was larger than 3 in absolute value, which was required for only four data sets in one simulation setting. This decision rule is straightforward to apply also in the analysis of real data, since the ML estimates need not be known.

We first compare ML and IRC estimators, and also assess the performance of estimators of the variance (17) of the IRC estimator. These results are reported in Tables 1 and 2. It is clear that the estimates of the regression coefficients from IRC are almost identical to those from ML, regardless of the sample size and the parameter values. This is the case not only on average, but also for nearly every individual data set. As a result, the simulation standard deviations of the estimators are also very similar. There thus appears to be virtually no loss of efficiency from the two-stage method of estimation employed by IRC.

On the other hand, computing times for the two approaches can be very different. On a desktop PC with a 2.4 GHz Intel Core 2 processor and 2 GB RAM, estimation for one dataset of sample sizes 200, 1000, and 5000, respectively, took around 15, 45, and 360 seconds for ML, and around 1, 3, and 15 seconds for IRC. It thus appears that the relative advantage in computing time of IRC over ML increases as the sample sizes increase. The same is true when the number of replicate measurements w_{ij} is increased. In tests with $n_i = 3$ replicates (not shown here), the computing times for IRC were essentially unchanged, while the times for ML increased to about 17, 55, and 520 seconds for $N = 200$, 1000, and 5000, respectively.

The estimated standard errors of the IRC estimates, taking into account uncertainty from both stages of the estimation, are obtained by estimating (17) as shown in the appendix. It can be seen that this approach performs well. In the most difficult cases, with small sample size, large effects and low reliability of measurement, the standard errors somewhat underestimate the true sampling variation. This is mainly due to right-skewed sampling distributions of the estimates in these cases, which is also reflected in a small upward bias of both ML and IRC estimates. The tails of the sampling distribution do not affect the coverage of the Wald-based 95 % confidence intervals for the parameters, which is 93.6–97.1 % across all the simulations.

TABLE 1.

Simulation results for maximum likelihood (ML) and improved regression calibration (IRC) estimators of regression parameter β_x for covariate measured with error under different measurement reliabilities ρ, true values of β_x, and sample sizes N. In each case, the true value of the other regression coefficient β_z is 0.5. The results are based on 1000 replications. The table shows the simulation mean and standard deviation (SD) of the point estimates $\hat{\beta}_x$, mean of their estimated standard errors (m(SE)), and coverage percentage of 95 % confidence intervals (C95). For IRC estimates, also shown are coverage of 95 % intervals based on a naive estimated standard error which ignores the first-stage uncertainty (C95-2), and the average percentage that this uncertainty contributes to the full standard errors (%-1).

β_x	N	ML				IRC					
		Mean	SD	m(SE)	C95	Mean	SD	m(SE)	C95	C95-2	%-1
$\rho = 0.75$											
0.5	200	0.520	0.253	0.248	96.0	0.520	0.253	0.253	96.2	96.0	2.0
	1000	0.507	0.110	0.107	94.9	0.507	0.110	0.107	95.0	94.9	0.5
	5000	0.500	0.047	0.047	95.5	0.500	0.047	0.047	95.6	95.5	0.2
1.0	200	1.051	0.294	0.284	96.0	1.051	0.294	0.289	96.2	95.9	2.2
	1000	1.018	0.122	0.121	95.3	1.018	0.123	0.121	95.4	95.3	0.8
	5000	1.001	0.053	0.053	94.9	1.001	0.053	0.053	94.8	94.8	0.5
1.5	200	1.592	0.371	0.353	97.0	1.592	0.371	0.359	97.1	96.8	2.4
	1000	1.519	0.144	0.147	96.5	1.519	0.144	0.148	96.6	96.4	1.3
	5000	1.502	0.064	0.065	94.8	1.502	0.064	0.065	94.7	94.5	1.0
$\rho = 0.5$											
0.5	200	0.533	0.310	0.296	96.7	0.533	0.310	0.301	97.0	96.2	2.8
	1000	0.509	0.130	0.124	94.0	0.509	0.130	0.124	94.2	93.6	1.3
	5000	0.500	0.054	0.055	95.5	0.500	0.054	0.055	95.5	95.4	1.0
1.0	200	1.088	0.409	0.368	96.9	1.089	0.411	0.375	96.9	96.6	4.7
	1000	1.006	0.148	0.146	95.9	1.007	0.148	0.147	95.9	95.4	3.3
	5000	1.005	0.065	0.065	95.4	1.005	0.065	0.065	95.5	95.0	3.0
1.5	200	1.666	0.586	0.519	96.7	1.664	0.584	0.523	96.9	96.5	6.4
	1000	1.527	0.189	0.193	96.4	1.528	0.190	0.194	96.4	95.1	5.4
	5000	1.509	0.083	0.084	95.5	1.510	0.083	0.085	95.5	94.0	5.1

The last two columns of Tables 1 and 2 examine a simplified estimate of the standard errors of the IRC estimates that is obtained by using only the first term on the right-hand side of (17), and omitting the second. In other words, this simply ignores the uncertainty in the estimated parameters of the exposure and measurement models from the first stage. Such an approach would be very convenient in practice because it entails using the estimated standard errors from the second-stage model directly, without any further adjustment. In the cases considered here, this simplification would do us little harm since the coverage of the confidence intervals (shown in the column "C95-2" of the tables) is still quite satisfactory. The reason for this is indicated by the last column of the tables, which shows the average percentage that the second term of (17) contributes to the full estimated standard error. This is mostly around 2 %, rising to 6.4 % in the most challenging configuration considered here.

Tables 3 and 4 compare the simulation results for IRC and RC estimators, omitting the full ML estimators because they are so similar to IRC. The focus here is on the finite-sample means and variabilities of the estimators, to examine their relative performances in different settings. We note also that computing times for IRC and RC were very similar, typically around 10 % higher for IRC.

The results show that best performances occur in different circumstances for the two estimators. IRC (and ML) estimators have an upward bias in small samples, due to the right-skewness of their sampling distributions, but the bias disappears in larger samples because these estimators

TABLE 2.
Simulation results for maximum likelihood (ML) and improved regression calibration (IRC) estimators of regression parameter β_z for perfectly measured covariate under different measurement reliabilities ρ, true values of the other regression coefficient β_x, and sample sizes N. In each case, the true value of β_z is 0.5. The results are based on 1000 replications. The columns of the table are the same as in Table 1.

β_x	N	ML				IRC					
		Mean	SD	m(SE)	C95	Mean	SD	m(SE)	C95	C95-2	%-1
$\rho = 0.75$											
0.5	200	0.508	0.237	0.236	95.8	0.508	0.237	0.239	95.8	95.8	1.2
	1000	0.509	0.105	0.103	94.1	0.509	0.105	0.103	94.1	94.1	0.3
	5000	0.498	0.045	0.045	95.8	0.498	0.045	0.045	95.9	95.8	0.1
1.0	200	0.514	0.234	0.236	96.2	0.514	0.234	0.239	96.5	96.2	1.4
	1000	0.511	0.104	0.102	94.4	0.511	0.104	0.103	94.6	94.3	0.4
	5000	0.497	0.044	0.045	95.8	0.497	0.044	0.045	95.8	95.8	0.2
1.5	200	0.513	0.255	0.244	96.1	0.513	0.255	0.247	96.1	95.9	1.5
	1000	0.507	0.109	0.105	94.4	0.507	0.109	0.105	94.7	94.4	0.6
	5000	0.499	0.047	0.047	94.1	0.499	0.047	0.047	94.1	93.9	0.4
$\rho = 0.5$											
0.5	200	0.507	0.242	0.241	96.1	0.507	0.242	0.244	96.3	95.7	1.6
	1000	0.508	0.107	0.104	93.6	0.508	0.107	0.105	93.6	93.5	0.6
	5000	0.497	0.045	0.046	95.2	0.497	0.045	0.046	95.2	95.2	0.3
1.0	200	0.514	0.246	0.247	97.0	0.514	0.246	0.250	96.9	96.3	2.4
	1000	0.504	0.108	0.105	93.6	0.504	0.108	0.105	93.6	93.4	1.4
	5000	0.500	0.047	0.047	95.4	0.500	0.047	0.047	95.4	94.9	1.1
1.5	200	0.514	0.281	0.266	96.4	0.514	0.280	0.269	96.7	96.0	3.5
	1000	0.506	0.114	0.111	94.6	0.506	0.114	0.111	94.9	94.1	2.4
	5000	0.501	0.048	0.049	95.1	0.502	0.048	0.049	95.1	94.5	2.2

are consistent. In contrast, RC estimators have a bias due to their approximate nature, which is largest when the reliability of measurement is low or when the regression coefficients are large. Taking into account both the biases and sampling variances, root mean squared errors tend to be smaller for RC when the sample size is small or moderate, and for IRC when the sample size is reasonably large. The bias of RC means that in the most difficult cases the coverage of confidence intervals based on them is substantially below the nominal level, while for IRC the coverage levels are always adequate.

In summary, the simulation study suggests, first, that we can generally replace ML with pseudo-ML (IRC) estimation, with essentially no loss in efficiency of estimation but with a substantial gain in computational speed. Second, when comparing IRC with RC, we find that the preferred estimator can depend on the circumstances of the analysis. RC tends to perform best with smaller samples and relatively mild measurement error problems, whereas IRC does best when the sample sizes are large, measurement error is severe or the effects being estimated are strong. The choice between RC and IRC is not informed by speed of computation, which is essentially the same for both of them.

7. Empirical Illustration: Ability and High Earnings

To illustrate covariate measurement error modeling in practice, we apply the investigated methods to a dataset on 935 non-black men from the 1980 wave of the Young Men's Cohort

TABLE 3.

Simulation results for improved regression calibration (IRC) and conventional regression calibration (RC) estimators of regression parameter β_x for covariate measured with error under different measurement reliabilities ρ, true values of β_x, and sample sizes N. In each case, the true value of the other regression coefficient β_z is 0.5. The results are based on 1000 replications. The table shows the simulation mean, % bias, and root mean squared error (RMSE) of the point estimates of β_x, and coverage percentage of 95 % confidence intervals (C95).

β_x	N	IRC				RC			
		Mean	% Bias	RMSE	C95	Mean	% Bias	RMSE	C95
$\rho = 0.75$									
0.5	200	0.520	4.0	0.254	96.2	0.515	3.0	0.247	96.2
	1000	0.507	1.4	0.111	95.0	0.504	0.8	0.109	94.9
	5000	0.500	0.0	0.047	95.6	0.497	−0.5	0.046	95.6
1.0	200	1.051	5.1	0.299	96.2	1.020	2.0	0.268	95.3
	1000	1.018	1.8	0.124	95.4	0.993	−0.7	0.114	94.2
	5000	1.001	0.1	0.053	94.8	0.978	−2.2	0.054	92.4
1.5	200	1.592	6.2	0.382	97.1	1.492	−0.6	0.301	95.3
	1000	1.519	1.3	0.145	96.6	1.439	−4.1	0.137	91.9
	5000	1.502	0.1	0.064	94.7	1.426	−4.9	0.092	72.5
$\rho = 0.5$									
0.5	200	0.533	6.6	0.312	97.0	0.518	3.7	0.288	95.9
	1000	0.509	1.9	0.131	94.2	0.502	0.4	0.125	93.8
	5000	0.500	0.1	0.054	95.5	0.494	−1.1	0.053	95.5
1.0	200	1.089	8.9	0.421	96.9	1.005	0.5	0.308	94.9
	1000	1.007	0.7	0.148	95.9	0.954	−4.6	0.135	92.4
	5000	1.005	0.5	0.065	95.5	0.954	−4.6	0.072	85.3
1.5	200	1.664	11.0	0.607	96.9	1.415	−5.7	0.341	92.4
	1000	1.528	1.9	0.192	96.4	1.354	−9.7	0.198	79.0
	5000	1.510	0.7	0.084	95.5	1.345	−10.3	0.166	26.8

of the U.S. National Longitudinal Survey (NLS), previously analyzed by Griliches (1976) and Blackburn and Neumark (1992), among others.

The binary outcome y_i we consider here is being a high earner, defined as having a salary above the 90 % percentile of the sample distribution. The covariate of main interest is ability x_i, also denoted [Ability], which is measured with error. Three covariates which are assumed measured without error are also included: working experience in years z_{i1} [Exper] (sample mean 11.6, s.d. 4.4), a dummy variable for living in an urban area z_{i2} [Urban] (71.8 % of the sample), and a dummy variable for being black z_{i3} [Black] (12.8 %).

Under the standard assumptions previously stated, the outcome model is

$$\text{logit}\big\{\Pr(y_i = 1|x_i, z_{i1}, z_{i2}, z_{i3})\big\} = \beta_{z_0} + \beta_{z_1} z_{i1} + \beta_{z_2} z_{i2} + \beta_{z_3} z_{i3} + \beta_x x_i,$$

and the exposure model is

$$x_i = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \zeta_i, \quad \zeta_i \sim \text{N}(0, \psi).$$

The mens' abilities are measured by two fallible measures. The first measure is an IQ test w_{i1} [IQ], collected as part of a survey of the respondents' schools conducted in 1968. Since a wide variety of IQ tests were used in different states, these were recoded into "IQ equivalents" by the Center for Human Resources Research at the Ohio State University which administers the NLS. The second measure is a test of "Knowledge of World of Work" w_{i2} [Know], which examines respondents' knowledge of the labor market, covering the duties, educational attainment, and

TABLE 4.
Simulation results for improved regression calibration (IRC) and conventional regression calibration (RC) estimators of regression parameter β_z for perfectly measured covariate under different measurement reliabilities ρ, true values of the other regression coefficient β_x, and sample sizes N. In each case, the true value of β_z is 0.5. The results are based on 1000 replications. The columns of the table are the same as in Table 3.

β_x	N	IRC				RC			
		Mean	% Bias	RMSE	C95	Mean	% Bias	RMSE	C95
$\rho = 0.75$									
0.5	200	0.508	1.5	0.237	95.8	0.505	0.9	0.236	95.9
	1000	0.509	1.8	0.105	94.1	0.506	1.2	0.105	94.1
	5000	0.498	−0.5	0.045	95.9	0.495	−1.0	0.045	95.7
1.0	200	0.514	2.9	0.234	96.5	0.502	0.3	0.228	96.4
	1000	0.511	2.2	0.104	94.6	0.499	−0.2	0.102	94.1
	5000	0.497	−0.5	0.044	95.8	0.486	−2.7	0.045	94.5
1.5	200	0.513	2.7	0.256	96.1	0.485	−3.1	0.241	95.6
	1000	0.507	1.4	0.109	94.7	0.481	−3.8	0.105	94.2
	5000	0.499	−0.3	0.047	94.1	0.473	−5.3	0.052	90.3
$\rho = 0.5$									
0.5	200	0.507	1.4	0.242	96.3	0.500	−0.1	0.240	95.9
	1000	0.508	1.7	0.107	93.6	0.502	0.4	0.106	93.8
	5000	0.497	−0.5	0.045	95.2	0.492	−1.7	0.046	95.0
1.0	200	0.514	2.9	0.247	96.9	0.485	−3.1	0.232	96.1
	1000	0.504	0.8	0.108	93.6	0.479	−4.3	0.106	93.3
	5000	0.500	−0.0	0.047	95.4	0.475	−5.0	0.051	91.2
1.5	200	0.514	2.8	0.281	96.7	0.451	−9.9	0.248	95.1
	1000	0.506	1.3	0.115	94.9	0.450	−9.9	0.114	90.5
	5000	0.502	0.3	0.048	95.1	0.447	−10.6	0.068	78.8

relative earnings of ten occupations. It is intended to reflect both the quantity and quality of schooling, intelligence, and motivation (curiosity about the outside world). The seminal paper by Griliches (1976) provides a lucid discussion of the data, variables and specification issues.

We use versions of the two fallible measures standardized to have sample mean 0 and variance 1. Denoting these standardized variables by w_{i1} and w_{i2}, we consider the classical measurement model

$$w_{ij} = x_i + \delta_{ij}, \quad \delta_{ij} \sim \mathrm{N}(0, \theta), \ j = 1, 2.$$

This is obtained from the general model (1) for a scalar x_i by assuming $\lambda_1 = \lambda_2 = 1$, and then setting $\nu_1 = \nu_2 = 0$ and $\theta_1 = \theta_2 = \theta$ because the marginal means and variances of w_{i1} and w_{i2} are equal. Note that for identifiability the model thus specifies that the two measures have equal loadings, i.e., that on the scale of the standardized measures they are equally discriminating measures of ability. This assumption could be relaxed if more than two fallible measures were available.

Estimates from ML, IRC, and RC are shown in Table 5. The parameter estimates for the outcome model are practically identical for ML and IRC, whereas the estimates from RC are smaller, as expected. In particular, the estimate for the parameter of main interest β_x from IRC, $\hat{\beta}_x = 2.50$, is essentially identical to the ML estimate, whereas the estimate from RC is $\hat{\beta}_x = 2.35$.

The estimated standard errors of estimates of $\boldsymbol{\beta}$ are practically identical for ML and IRC, apart from numerical differences. This indicates that the loss of efficiency in estimating the pa-

TABLE 5.

Ability and high earnings: Estimates for logistic regression with covariate measurement error based on maximum likelihood (ML), improved regression calibration (IRC), and conventional regression calibration (RC). For IRC, SE are estimated standard errors based on asymptotic covariance matrix derived in this article and SE-2 are naive estimated standard errors ignoring uncertainty in Stage-1 estimates.

Parameter	Covariate	ML		IRC			RC	
		Est	(SE)	Est	(SE)	(SE-2)	Est	(SE)
Outcome model:								
β_{z_0}		−3.68	(0.57)	−3.68	(0.56)	(0.55)	−3.29	(0.45)
β_{z_1}	[Exper]	0.02	(0.03)	0.02	(0.03)	(0.03)	0.02	(0.03)
β_{z_2}	[Urban]	0.50	(0.34)	0.50	(0.33)	(0.33)	0.45	(0.31)
β_{z_3}	[Black]	0.52	(0.76)	0.52	(0.74)	(0.73)	0.48	(0.68)
β_x	[Ability]	2.49	(0.50)	2.50	(0.51)	(0.47)	2.35	(0.42)
Exposure model:								
γ_0		0.20	(0.08)	0.20	(0.08)	(0.08)	0.20	(0.08)
γ_1	[Exper]	−0.02	(0.01)	−0.02	(0.01)	(0.01)	−0.02	(0.01)
γ_2	[Urban]	0.20	(0.06)	0.20	(0.06)	(0.06)	0.20	(0.06)
γ_3	[Black]	−1.00	(0.07)	−1.00	(0.07)	(0.07)	−1.00	(0.07)
ψ		0.29	(0.03)	0.29	(0.03)	(0.03)	0.29	(0.03)
Measurement model:								
θ		0.58	(0.03)	0.59	(0.03)	(0.03)	0.59	(0.03)
Log-likelihood		$\ell = -2738.38$		$\ell = -2738.41$				

rameters of the exposure and measurement models from only Stage 1 of IRC is effectively nill; indeed, estimates of these parameters and associated estimated standard errors are identical to the full ML results to at least three decimal places. Uncertainty from Stage 1, i.e., the second term of the variance matrix (17), contributes around 8 % of the estimated standard error of $\hat{\beta}_x$ for IRC. We also note that the sum of the maximized log-likelihood components for IRC of $\ell = -2738.41$ is very close to the maximum of the log-likelihood $\ell = -2738.38$.

From the estimated exposure model, the ability measure is significantly associated with urbanity, race and working experience. Its conditional variance given these covariates is $\hat{\psi} = 0.29$. The estimated measurement error variance is $\hat{\theta} = 0.58$, and the conditional reliability of the measures (given the covariates) is thus $\hat{\psi}/(\hat{\psi} + \hat{\theta}) = 0.33$.

Regarding the outcome model, there is a strong estimated association between the ability measure and high earnings when controlling for working experience, urbanity, and race. The estimated coefficient of $\hat{\beta}_x = 2.50$ translates to an odds ratio of 3.8 for being a high earner corresponding to an increase of one conditional standard deviation in ability. The other covariates are retained in the model, but they could possibly also have been omitted because they do not have statistically significant associations with high earnings at the 5 % level. It is worth noting that if the model was simplified by omitting some control variables, we could still choose to use the predicted values $\widetilde{\xi}_i$ and variances $\hat{\omega}_i$ conditional on all of them, without re-calculating these predictions. This only requires the modification that in the calculation of the standard errors (as shown in the appendix) the corresponding elements of $\boldsymbol{\beta}_z$ are set to 0.

8. Discussion

In this article, we have proposed an improved regression calibration approach to the estimation of generalized linear models with covariate measurement error, a pseudo maximum likelihood method that simultaneously addresses the computational challenge of maximum likelihood

and the inconsistency of conventional regression calibration. A decomposed form of the likelihood was exploited, where the component for the measurement and exposure models is in closed form and trivial to maximize, and the component for the outcome model is accurately maximized using crude and fast numerical integration.

Our simulations show that improved regression calibration produces parameter estimates that are practically indistinguishable from those produced by maximum likelihood. Interval estimation based on the asymptotic covariance matrix for improved regression calibration that was derived in this article has excellent performance. Even interval estimation based on the naive estimator of the asymptotic covariance matrix (ignoring the uncertainty incurred in the first step) usually performs well. Compared to conventional regression calibration, improved regression calibration offers little or no advantage when sample sizes are small, but performs best when samples are reasonably large and especially when the measurement error or the effects are not small.

Both the fallibly measured covariates and their measures are continuous in the models considered here. Improved regression calibration can also be used when the observed measures are categorical, in which case categorical factor models would be used as measurement models. Since the predictive distributions are then no longer normal, it is not obvious that improved regression calibration would work well. If both the fallibly measured covariates and their measures are categorical, the problem is one of misclassification where integration is replaced by summation and maximum likelihood estimation becomes computationally straightforward.

Acknowledgements

We are grateful to H.K. Gjessing for helpful discussions and three anonymous reviewers for constructive comments.

Appendix: Obtaining $\widehat{\mathcal{I}}_{\mathsf{ME,O}}$ in (18)

Here we describe the calculation of the estimate (18) of the matrix $\mathcal{I}_{\mathsf{ME,O}}$, which is used in the calculation of the variance matrix (17) of $\widehat{\boldsymbol{\vartheta}}_{\mathsf{O}}^{\mathsf{IRC}}$. Let us first introduce some convenient shorthand notation for the logarithm of the likelihood contribution (6):

$$\log \underbrace{g(y_i, \mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta})}_{\equiv g_i} = \log \underbrace{\int \overbrace{g(y_i|\mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{O}})}^{\equiv g_{yi}} \overbrace{g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})}^{\equiv g_{xi}} \, \mathrm{d}\mathbf{x}_i}_{\equiv g_{1i}} + \log \underbrace{g(\mathbf{w}_i|\mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})}_{\equiv g_{2i}}.$$

Here g_{xi} and g_{2i} are multivariate normal density functions with parameters $\boldsymbol{\theta}_{1i} = (\boldsymbol{\xi}_i', \mathrm{vec}(\boldsymbol{\Omega}_i)')'$ and $\boldsymbol{\theta}_{2i} = (\boldsymbol{\mu}_i', \mathrm{vec}(\boldsymbol{\Sigma}_i)')'$ respectively, as defined by (11)–(12) and (9)–(10). These in turn are functions of the parameters $\boldsymbol{\chi} = (\boldsymbol{\nu}', \mathrm{vec}(\boldsymbol{\Lambda})', \mathrm{vec}(\boldsymbol{\Theta})', \mathrm{vec}(\boldsymbol{\Gamma})', \mathrm{vec}(\boldsymbol{\Psi})')'$, and $\boldsymbol{\vartheta}_{\mathsf{ME}}$ are the distinct, unknown elements of $\boldsymbol{\chi}$.

The required gradients for (18) are

$$\mathbf{u}_{\vartheta_{\mathsf{O}},i}(\boldsymbol{\vartheta}) = \frac{\partial \log g_i}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}} = \frac{1}{g_{1i}} \frac{\partial g_{1i}}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}} \quad \text{and} \tag{A.1}$$

$$\mathbf{u}_{\vartheta_{\mathsf{ME}},i}(\boldsymbol{\vartheta}) = \frac{\partial \log g_i}{\partial \boldsymbol{\vartheta}_{\mathsf{ME}}} = \frac{1}{g_{1i}} \left(\frac{\partial g_{1i}}{\partial \boldsymbol{\theta}_{1i}'} \frac{\partial \boldsymbol{\theta}_{1i}}{\partial \boldsymbol{\chi}'} \frac{\partial \boldsymbol{\chi}}{\partial \boldsymbol{\vartheta}_{\mathsf{ME}}'} \right)' + \left(\frac{\partial \log g_{2i}}{\partial \boldsymbol{\theta}_{2i}'} \frac{\partial \boldsymbol{\theta}_{2i}}{\partial \boldsymbol{\chi}'} \frac{\partial \boldsymbol{\chi}}{\partial \boldsymbol{\vartheta}_{\mathsf{ME}}'} \right)', \tag{A.2}$$

where

$$g_{1i} = \int g_{yi}\, g_{xi}\, \mathrm{d}\mathbf{x}_i\,, \tag{A.3}$$

$$\frac{\partial g_{1i}}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}} = \int \frac{\partial g_{yi}}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}}\, g_{xi}\, \mathrm{d}\mathbf{x}_i\,, \quad \text{and} \tag{A.4}$$

$$\frac{\partial g_{1i}}{\partial \boldsymbol{\theta}'_{1i}} = \int g_{yi}\, \frac{\partial g_{xi}}{\partial \boldsymbol{\theta}'_{1i}}\, \mathrm{d}\mathbf{x}_i\,. \tag{A.5}$$

Estimated values for these quantities, and thus for the estimated matrix $\widehat{\mathcal{I}}_{\mathsf{ME,O}}$ given by (18), are obtained by substituting estimates $\widehat{\boldsymbol{\vartheta}}^{\mathsf{IRC}}$ of the parameters.

Starting with (A.2), we note that each element of $\boldsymbol{\chi}$ is either a known constant or equal to a single element of $\boldsymbol{\vartheta}_{\mathsf{ME}}$; for illustration, consider $\boldsymbol{\Lambda}$ as shown in (2). Suppose that $\boldsymbol{\chi}$ is of length t and $\boldsymbol{\vartheta}_{\mathsf{ME}}$ of length u. Then $\partial \boldsymbol{\chi}/\partial \boldsymbol{\vartheta}'_{\mathsf{ME}}$ is a $t \times u$ matrix whose (i, j)th element is 1 if the ith element of $\boldsymbol{\chi}$ is equal to the jth element of $\boldsymbol{\vartheta}_{\mathsf{ME}}$, and 0 otherwise.

Next, the elements of $\partial \boldsymbol{\theta}_{2i}/\partial \boldsymbol{\chi}'$ in (A.2) are

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\nu}'} &= \mathbf{C}_i\,,\\
\frac{\partial \boldsymbol{\mu}_i}{\partial \mathrm{vec}(\boldsymbol{\Lambda})'} &= (\boldsymbol{\Gamma}\mathbf{z}_i)' \otimes \mathbf{C}_i\,,\\
\frac{\partial \boldsymbol{\mu}_i}{\partial \mathrm{vec}(\boldsymbol{\Theta})'} &= \mathbf{0}\,,\\
\frac{\partial \boldsymbol{\mu}_i}{\partial \mathrm{vec}(\boldsymbol{\Gamma})'} &= \mathbf{z}'_i \otimes (\mathbf{C}_i \boldsymbol{\Lambda})\,,\\
\frac{\partial \boldsymbol{\mu}_i}{\partial \mathrm{vec}(\boldsymbol{\Psi})'} &= \mathbf{0}\,,\\
\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{\nu}'} &= \mathbf{0}\,,\\
\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathrm{vec}(\boldsymbol{\Lambda})'} &= \left[(\mathbf{C}_i \boldsymbol{\Lambda}\boldsymbol{\Psi}) \otimes \mathbf{C}_i\right] + \left[\mathbf{C}_i \otimes (\mathbf{C}_i \boldsymbol{\Lambda}\boldsymbol{\Psi})\right]\mathbf{K}_{rm}\,,\\
\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathrm{vec}(\boldsymbol{\Theta})'} &= \mathbf{C}_i \otimes \mathbf{C}_i\,,\\
\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathrm{vec}(\boldsymbol{\Gamma})'} &= \mathbf{0}\,,\\
\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)}{\partial \mathrm{vec}(\boldsymbol{\Psi})'} &= (\mathbf{C}_i \boldsymbol{\Lambda}) \otimes (\mathbf{C}_i \boldsymbol{\Lambda})\,,
\end{aligned}$$

and the elements of $\partial \boldsymbol{\theta}_{1i}/\partial \boldsymbol{\chi}'$ are

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \boldsymbol{\nu}'} = \frac{\partial \boldsymbol{\xi}_i}{\partial \boldsymbol{\mu}'_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\nu}'}\,,$$

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Lambda})'} = \left\{\left[\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_i)\right]' \otimes \boldsymbol{\Psi}\right\}\mathbf{K}_{rm} + \frac{\partial \boldsymbol{\xi}_i}{\partial \boldsymbol{\mu}_i'}\frac{\partial \boldsymbol{\mu}_i}{\partial \operatorname{vec}(\boldsymbol{\Lambda})'} + \frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Lambda})'},$$

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Theta})'} = \frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Theta})'},$$

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Gamma})'} = \mathbf{z}_i' \otimes \mathbf{I}_m + \frac{\partial \boldsymbol{\xi}_i}{\partial \boldsymbol{\mu}_i'}\frac{\partial \boldsymbol{\mu}_i}{\partial \operatorname{vec}(\boldsymbol{\Gamma})'},$$

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Psi})'} = \left[\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_i)\right]' \otimes \mathbf{I}_m + \frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Psi})'},$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \boldsymbol{\nu}'} = \mathbf{0},$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Lambda})'} = -\left\{\left[\left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{C}_i\right) \otimes \boldsymbol{\Psi}\right]\mathbf{K}_{rm} + \left[\boldsymbol{\Psi} \otimes \left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{C}_i\right)\right]\right\} + \frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Lambda})'},$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Gamma})'} = \mathbf{0},$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Theta})'} = \frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Theta})},$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Psi})'} = \mathbf{I}_{m^2} - (\mathbf{I}_{m^2} + \mathbf{K}_{mm})\left[\left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{C}_i\boldsymbol{\Lambda}\right) \otimes \mathbf{I}_m\right] + \frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)}{\partial \operatorname{vec}(\boldsymbol{\Psi})},$$

where

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \boldsymbol{\mu}_i'} = -\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\boldsymbol{\Sigma}_i^{-1},$$

$$\frac{\partial \boldsymbol{\xi}_i}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'} = (\mathbf{w}_i - \boldsymbol{\mu}_i)' \otimes \left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\right),$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Omega}_i)}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})'} = -\left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\right) \otimes \left(\boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{C}_i'\right),$$

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i^{-1})}{\partial \operatorname{vec}(\boldsymbol{\Sigma}_i)'} = -\boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1},$$

and vec(·) denotes the column-by-column vectorization operator, $\otimes$ the Kronecker product, $\mathbf{I}_m$ an $m \times m$ identity matrix, and $\mathbf{K}_{rm}$ an $rm \times rm$ commutation matrix. The formulas are obtained through repeated application of rules of matrix differentiation (see, e.g., Lütkepohl, 1996).

In the second term of (A.2), the elements of $\partial \log g_{2i}/\partial \boldsymbol{\theta}'_{2i}$ are $\partial \log g_{2i}/\partial \boldsymbol{\mu}'_i = (\mathbf{w}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}$ and $\partial \log g_{2i}/\partial \mathrm{vec}(\boldsymbol{\Sigma}_i)' = \mathrm{vec}[\boldsymbol{\Sigma}_i^{-1}(\mathbf{w}_i - \boldsymbol{\mu}_i)(\mathbf{w}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}]'/2$.

The remaining elements of (A.1) and (A.2) depend also on the outcome model for y_i. For the logistic model, which is predominant in applications of generalized linear models with covariate measurement error, and which is also used in our simulations and example, $g_{yi} = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$ where $\pi_i = \exp(\eta_i)/[1+\exp(\eta_i)]$ and $\eta_i = \mathbf{z}'_i\boldsymbol{\beta}_z + \mathbf{x}'_i\boldsymbol{\beta}_x$. For this model we employ the well-known closed-form approximation $g_{1i} \approx (\pi_i^*)^{y_i}(1-\pi_i^*)^{1-y_i}$, where $\pi_i^* = \exp(\eta_i^*)/[1+\exp(\eta_i^*)]$, $\eta_i^* = \eta_{1i}\eta_{2i}^{-1/2}$, $\eta_{1i} = \mathbf{z}'_i\boldsymbol{\beta}_z + \boldsymbol{\xi}'_i\boldsymbol{\beta}_x$, $\eta_{2i} = 1 + d\boldsymbol{\beta}'_x\boldsymbol{\Omega}_i\boldsymbol{\beta}_x$, and $d = 1/1.7^2$ (e.g., Liang & Liu, 1991). For this approximation,

$$\frac{\partial g_{1i}}{\partial \boldsymbol{\vartheta}_{\mathsf{O}}} = (-1)^{1-y_i}\pi_i^*\big(1-\pi_i^*\big)\eta_{2i}^{-1/2}\big(\mathbf{z}'_i, \boldsymbol{\xi}_i^{*\prime}\big)' \quad \text{and}$$

$$\frac{\partial g_{1i}}{\partial \boldsymbol{\theta}'_{1i}} = (-1)^{1-y_i}\pi_i^*\big(1-\pi_i^*\big)\eta_{2i}^{-1/2}\big[\boldsymbol{\beta}'_x, -(d/2)\eta_{1i}\eta_{2i}^{-1}\big(\boldsymbol{\beta}'_x \otimes \boldsymbol{\beta}'_x\big)\big],$$

where $\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_i - \eta_{1i}\eta_{2i}^{-1}d\boldsymbol{\Omega}_i\boldsymbol{\beta}_x$. These formulas complete explicit expressions for (A.1) and (A.2).

In our data analysis, we also apply a similar idea for the conventional regression calibration estimate of $\boldsymbol{\vartheta}_{\mathsf{O}}$, which uses the first-order approximation $g_{1i} \approx (\pi_i^{\mathrm{RC}})^{y_i}(1-\pi_i^{\mathrm{RC}})^{1-y_i}$ where $\pi_i^{\mathrm{RC}} = \exp(\eta_{1i})/[1+\exp(\eta_{1i})]$. We estimate its variance matrix analogously to (17)–(18), using in (A.1) and (A.2) $\partial g_{1i}/\partial \boldsymbol{\vartheta}_{\mathsf{O}} = (\partial g_{1i}/\partial \eta_{1i})(\mathbf{z}'_i, \boldsymbol{\xi}')'$ and $\partial g_{1i}/\partial \boldsymbol{\theta}'_{1i} = (\partial g_{1i}/\partial \eta_{1i})[\boldsymbol{\beta}'_x, \mathbf{0}']$, where $\partial g_{1i}/\partial \eta_{1i} = (-1)^{1-y_i}\pi_i^{\mathrm{RC}}(1-\pi_i^{\mathrm{RC}})$.

For other, less popular models, we must evaluate the integrals involved in (A.3)–(A.5). Note first that the partial derivatives $\partial g_{xi}/\partial \boldsymbol{\theta}'_{1i}$ are given by

$$\frac{\partial g_{xi}}{\partial \boldsymbol{\xi}'_i} = (\mathbf{x}_i - \boldsymbol{\xi}_i)'\boldsymbol{\Omega}_i^{-1}\, g_{xi} \quad \text{and}$$

$$\frac{\partial g_{xi}}{\partial \mathrm{vec}(\boldsymbol{\Omega}_i)'} = (1/2)\mathrm{vec}\big[\boldsymbol{\Omega}_i^{-1}(\mathbf{x}_i - \boldsymbol{\xi}_i)(\mathbf{x}_i - \boldsymbol{\xi}_i)'\boldsymbol{\Omega}_i^{-1} - \boldsymbol{\Omega}_i^{-1}\big]'\, g_{xi}.$$

Substituting these into (A.5), we see that each of the integrals there, and also in (A.3) and (A.4), are of the form $\int h_i(\mathbf{x}_i)g_{xi}\,\mathrm{d}\mathbf{x}_i$ for some function $h_i(\mathbf{x}_i)$ of $\mathbf{x}_i$, integrated over the multivariate normal density $g_{xi} = g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\vartheta}_{\mathsf{ME}})$. This suggests that the integrals can be evaluated through Monte Carlo integration, by first generating M independent draws $\mathbf{x}_{ij}$, $j = 1, \ldots, M$, from $g(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \widehat{\boldsymbol{\vartheta}}_{\mathsf{ME}})$, and then approximating the integrals by the averages $M^{-1}\sum_{j=1}^{M} h_i(\mathbf{x}_{ij})$ for each of the $h_i(\cdot)$. Only one set of random draws is needed for all the observations i, if we first generate M uncorrelated m-vectors $\mathbf{u}_j$ of standard normal random variates and then calculate $\mathbf{x}_{ij} = \widetilde{\boldsymbol{\xi}}_i + \mathbf{B}_i\mathbf{u}_j$, where $\widehat{\boldsymbol{\Omega}}_i = \mathbf{B}_i\mathbf{B}'_i$.

References

Albert, P.S., & Follmann, D.A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, *56*, 667–677.

Armstrong, B. (1985). Measurement error in generalized linear models. *Communications in Statistics. Series B*, *16*, 529–544.

Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures. *Psychometrika*, *48*, 493–517.

Blackburn, M., & Neumark, D. (1992). Unobserved ability, efficiency wages, and interindustry wage differentials. *Quarterly Journal of Economics*, *107*, 1421–1436.

Buonaccorsi, J., Demidenko, E., & Tosteson, T. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica*, *10*, 885–903.

Buonaccorsi, J. (2010). *Measurement error: models, methods and applications*. Boca Raton: Chapman & Hall/CRC.

Burr, D. (1988). On errors-in-variables in binary regression—Berkson case. *Journal of the American Statistical Association*, *83*, 739–743.

Buzas, J.S., & Stefanski, L.A. (1995). Instrumental variable estimation in generalized linear measurement error models. *Journal of the American Statistical Association*, *91*, 999–1006.

Carroll, R.J., Ruppert, D., Stefanski, L.A., & Crainiceanu, C.M. (2006). *Measurement error in nonlinear models* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Carroll, R.J., Spiegelman, C.H., Lan, K.G., Bailey, K.T., & Abbott, R.D. (1984). On errors-in-variables for binary regression models. *Biometrika*, *71*, 19–25.

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

Clayton, D.G. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In J.H. Dwyer, M. Feinlieb, P. Lippert, & H. Hoffmeister (Eds.), *Statistical models for longitudinal studies on health* (pp. 301–331). New York: Oxford University Press.

Davis, P.J., & Rabinowitz, P. (1984). *Methods of numerical integration* (2nd ed.). New York: Academic Press.

Gleser, L.J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In P.J. Brown & W.A. Fuller (Eds.), *Statistical analysis of measurement error models and applications* (pp. 99–114). Providence: American Mathematical Society.

Gong, G., & Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics*, *9*, 861–869.

Gourieroux, C., & Monfort, A. (1995). *Statistics and econometric models* (Vol. 2). Cambridge: Cambridge University Press.

Griliches, Z. (1976). Wages of very young men. *Journal of Political Economy*, *85*, S69–S86.

Gustafson, P. (2004). *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Boca Raton: Chapman & Hall/CRC.

Higdon, R., & Schafer, D.W. (2001). Maximum likelihood computations for regression with measurement error. *Computational Statistics & Data Analysis*, *35*, 283–299.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109–133.

Jöreskog, K.G., & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631–639.

Kuha, J. (1997). Estimation by data augmentation in regression models with continuous and discrete covariates measured with error. *Statistics in Medicine*, *16*, 189–202.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Journal of the Royal Statistical Society. Series C*, *50*, 325–335.

Liang, K.-Y., & Liu, X.-H. (1991). Estimating equations in generalized linear models with measurement error. In V.P. Godambe (Ed.), *Estimating functions* (pp. 47–63). Oxford: Oxford University Press.

Lütkepohl, H. (1996). *Handbook of matrices*. Chichester: Wiley.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McDonald, R.P. (1967). *Nonlinear factor analysis* (Psychometric Monograph No. 15). Richmond: Psychometric Corporation.

Parke, W.R. (1986). Pseudo maximum likelihood estimation: the asymptotic distribution. *Annals of Statistics*, *14*, 355–357.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2003). Maximum likelihood estimation of generalized linear models with covariate measurement error. *The Stata Journal*, *3*, 385–410.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004a). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004b). *Gllamm manual* (Technical report 160). U.C. Berkeley Division of Biostatistics. Downloadable from http://www.bepress.com/ucbbiostat/paper160/.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*, 301–323.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata, vol. II: categorical responses, counts, and survival* (3rd ed.). College Station: Stata Press.

Richardson, S., & Gilks, W.S. (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine*, *12*, 1703–1722.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, *6*, 15–51.

Robinson, P.M. (1974). Identification, estimation, and large sample theory for regressions containing unobservable variables. *International Economic Review*, *15*, 680–692.

Rosner, B., Spiegelman, D., & Willett, W.C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*, *132*, 734–745.

Rosner, B., Willett, W.C., & Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, *8*, 1031–1040.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.
Schafer, D.W. (1987). Covariate measurement error in generalized linear models. *Biometrika*, *74*, 385–391.
Schafer, D.W. (1993). Likelihood analysis for probit regression with measurement error. *Biometrika*, *80*, 899–904.
Schafer, D.W., & Purdy, K.G. (1986). Likelihood analysis for errors-in-variables regression with replicate measurements. *Biometrika*, *83*, 813–824.
Shapiro, A. (2007). Statistical inference of moment structures. In S.Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 229–259). Amsterdam: Elsevier.
Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–575.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. Boca Raton: Chapman & Hall/CRC.
Skrondal, A., & Rabe-Hesketh, S. (2007). Latent variable modelling: a survey. *Scandinavian Journal of Statistics*, *34*, 712–745.
Skrondal, A., & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear mixed models. *Journal of the Royal Statistical Society. Series A*, *172*, 659–687.
Stephens, D.A., & Dellaportas, P. (1992). Bayesian analysis of generalised linear models with covariate measurement error. In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (Eds.), *Bayesian statistics* (Vol. 4, pp. 813–820). Oxford: Oxford University Press.
Thisted, R.A. (1988). *Elements of statistical computing*. London: Chapman & Hall.

Manuscript Received: 4 JUL 2011
Final Version Received: 3 FEB 2012