Harvard University
Harvard University Biostatistics Working Paper Series

Year 2011 *Paper* 113

A Regularization Corrected Score Method for Nonlinear Regression Models with Covariate Error

David M. Zucker* Malka Gorfine†
Yi Li‡ Donna Spiegelman**

*Hebrew University, mszucker@mscc.huji.ac.il
†Technion - Israel Institute of Technology
‡University of Michigan
**Harvard School of Public Health, stdls@hsph.harvard.edu

A Regularization Corrected Score Method for Nonlinear Regression Models with Covariate Error

David M. Zucker

Department of Statistics, Hebrew University

Mount Scopus, Jerusalem, Israel

email: mszucker@mscc.huji.ac.il

Malka Gorfine

Faculty of Industrial Engineering and Management

Technion - Israel Institute of Technology, Haifa, Israel

Yi Li

Department of Biostatistics

University of Michigan, Ann Arbor MI, USA

and

Donna Spiegelman

Departments of Epidemiology and Biostatistics

Harvard School of Public Health, Boston MA, USA

September 21, 2011

Summary

Many regression analyses involve explanatory variables that are measured with error, and failing to account for this error is well known to lead to biased estimates of the regression coefficients. We present here a new general method for adjusting for covariate error. Our method consists of an approximate version of the Stefanski-Nakamura corrected score approach, using the method of regularization for approximate solution of integral equations. We develop the theory in the setting of classical likelihood models, covering linear regression, nonlinear regression, logistic regression, and Poisson regression. The method is extremely general in terms of the types of measurement error models covered, and is a functional method in the sense of not requiring information on the distribution of the true covariate. We discuss the theoretical properties of the method and present simulation results in the logistic regression setting (univariate and multivariate). For illustration, we apply the method to data from the Harvard Nurses' Health Study concerning the relationship between physical activity and breast cancer death among patients with diagnosed breast cancer.

Key words: Errors in variables, nonlinear models, logistic regression, integral equations

1. Introduction

Many regression analyses involve explanatory variables that are measured with error. It is well known that failing to account for the covariate error can lead to biased estimates of the regression coefficients, and there is a large literature on correcting for covariate measurement error. Fuller (1987) provides an authoritative treatment for linear models. For nonlinear models, work on the covariate error problem began in the early 1980's. Carroll et al. (2006) summarizes the major developments in the area. Currently the covariate error problem for nonlinear models continues to be an active research area, bearing on such common statistical models as nonlinear regression with a continuous response, logistic regression for binary responses, Poisson regression for count data, and the Cox proportional hazards regression for survival data. This paper presents a flexible new method for nonlinear regression problems with covariate error, built on earlier work.

Three basic study designs are of interest: (1) the replicate measures design, where repeat covariate measurements are available (either for all individuals or for a subsample), (2) the internal validation design, where the true covariate values are available on a sample of individuals in the main study, and (3) the external validation design, where the key parameters of the measurement error distribution are estimated (assuming reasonable transportability) from an external study, independent of the main study, with paired measurements of the true and surrogate covariate. Also, two types of methods are of interest: structural methods, which make use of a distributional model for the true covariates, and functional methods, which do not make use of such a model.

A wide variety of approaches have appeared in the literature. We focus here on the SIMEX and corrected score approaches, bothe of which are functional modeling approaches. These are general approaches that can handle both internal and external validation designs as well as, with slight adaptation, the replicate measures design. Our proposed method is based on the corrected score approach.

The SIMEX method of Cook and Stefanski (1995) involves simulating new covariate

values with various levels of artificially added measurement error, carrying out a naive model fit for each of the resulting new data sets, and then back-extrapolating to zero measurement error. While some success has been achieved with this approach, obviously the back-extrapolation process is uncertain. Moreover, the SIMEX approach can be difficult to apply in certain non-classical settings. One challenging setting is when the measurement error variance depends on the true covariate value. A version of SIMEX that can handle this setting under the replicate measures design has been described by Devanarayan and Stefanski (2002), but it does not appear that SIMEX can handle this setting under internal or external validation study designs.

The corrected score approach, advanced by Stefanski (1989) and Nakamura (1990, 1992), involves replacing the likelihood score in the conventional likelihood-based analysis with a function of the surrogate covariates that serves as an "unbiased" substitute. In Section 2, we present the exact definition. For the independent additive error model, Nakamura (1990) showed that this approach works for normal linear regression, Poisson regression, Gamma regression, and inverse Gaussian regression. Nakamura (1992) presented an approximate corrected score method for the Cox regression model, which Kong and Gu (1999) later showed to yield consistent estimates. Novick and Stefanski (2002) presented a corrected score method that is aimed at the independent additive error model with normal errors, and is valid when the likelihood score function is an entire function in the complex plane. Stefanski, Novick, and Devanarayan (2005) elaborate on this approach. When the error-prone covariate is discrete, a corrected score can be formulated easily; the relevant theory was developed for generalized linear models by Akazawa, Kinukawa, and Nakamura (1998), and extended to the Cox regression model by Zucker and Spiegelman (2008).

On the other hand, for logistic regression with additive normal error, Stefanski (1989) showed that an exact corrected score method does not exist. Huang and Wang (2001) presented an exact modified corrected score method for logistic regression, in which the

terms in the original score function are cleverly reweighted in such a way that an exact corrected score for the modified score function can be found. This reweighting leads to some loss in efficiency. Moreover, the method of Huang and Wang is designed only for the case of independent additive measurement error. Buzas (2009) presents an approximate corrected score method for the logistic regression model with high efficiency when the covariate effect is moderate, but this method is designed only for the case of independent additive normal error.

The basic problem with the corrected score approach in the logistic regression model and other cases with a continuous error-prone covariate X is that obtaining the corrected score requires solving a challenging integral equation. The equation involved falls into the class of Fredholm integral equations of the first kind, which are discussed by Delves and Mohamed (1985, Ch. 12) and Kress (1989, Ch. 16). Such equations do not always have an exact solution; the logistic regression problem is one case of this. Moreover, even when an exact solution exists, the problem can be ill-conditioned. We attempted to tackle the case of a continuous covariate by discretizing the covariate to various degrees of fineness and applying the methodology for the discrete case. This attempt, however, met with only limited success. We had promising results under the classical normal error model, but in more general cases we ran into difficulties. A major problem was that the classification matrix tended to be ill-conditioned even with a modest degree of fineness, such as six categories.

In this paper, we develop a new approach. The idea is to handle the integral equation using the method of regularization (Delves and Mohamed, 1985, Sec. 12.3; Kress, 1989, Ch. 16), which involves minimizing a penalized distance function to obtain an approximate solution. In contrast with the original integral equation problem, the regularized problem always has a solution, and is reasonably well conditioned provided that the weight α on the penalty term is not too small. As α tends to infinity, the estimation procedure tends to a naive analysis in which we ignore the covariate error, and simply substitute

the surrogate covariate value for the true value. Conversely, under suitable conditions, as α tends to zero the procedure approaches an exact corrected score procedure. The idea is to push α as close as possible to zero to get good estimates of the model parameters. We call our approach the regularized corrected score (RECS) approach.

The advantage of RECS is that it is extremely flexible. Its formulation is very general, and it is a functional method in that it does not involve the distribution of the true covariate, but only the conditional density of the surrogate covariate given the true covariate. The method can handle both internal and external validation designs. It can handle the replicate measures design as well, with the overall surrogate measurement defined as the sample mean (or other summary measure) of the available measurements on the individual. Moreover, the method can handle measurement error structures of an arbitrary nature, not just independent additive measurement error. Differential measurement error, where the measurement error depends on the response, is also covered.

The technique of regularization is an established one which has been used in various statistical applications, for example nonparametric regression based on spline smoothing (Wahba, 1990). Recently, Carrasco and Florens (2011) have used the technique in a measurement error context, to attack the problem of deconvolving a density. However, our use of regularization in the corrected score context is novel.

The goal of this paper is to develop the RECS method in detail for the classical likelihood setting. Section 2 lays out the setting and background. Section 3 presents the proposed procedure and its theoretical properties. Section 4 presents simulation results under the logistic regression model. Section 5 presents a real-data illustration of the method in the logistic regression setting. Section 6 presents a brief summary and discussion.

2. Setting and Background

We assume a typical setup with n independent units whose response values Y_i, $i = 1, \ldots n$,

follow a regression model involving several covariates. We assume for now that only one of the covariates is subject to error; later we will generalize to the case of multiple error-prone covariates. We denote by X_i the true value of the error-prone covariate, and by W_i the measured value. We let $\mathbf{Z}_i$ denote the vector of error-free covariates, which may include an arbitrary number of discrete and continuous components. We denote the conditional density or mass function of Y_i given $(X_i, \mathbf{Z}_i)$ by $f(y|X_i, \mathbf{Z}_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a p-vector of unknown parameters, including regression coefficients and auxiliary parameters such as error variances. In contrast with the SIMEX method, the Huang and Wang (2001) method, and most other methods in the literature, we provide the option of allowing measurement error to depend on X_i, $\mathbf{Z}_i$, and the outcome Y_i (differential error). We denote by $a_i(x, w)$ the conditional density of W_i given $X_i = x$, with the subscript i signifying possible dependence on $\mathbf{Z}_i$ and Y_i. To ease the presentation of the theoretical results, we assume that $(X_i, \mathbf{Z}_i)$ are i.i.d. random vectors. We stress, however, that our method does not involve any modeling (either parametric or nonparametric) of the distribution of $(X_i, \mathbf{Z}_i)$. The theoretical results can, in principle, be extended to the case where $(X_i, \mathbf{Z}_i)$ are non-random values satisfying suitable ergodicity conditions.

Define $\mathbf{u}(y, x, \mathbf{z}, \boldsymbol{\theta}) = [\partial/\partial\boldsymbol{\theta}] \log f(y|x, \mathbf{z}, \boldsymbol{\theta})$ and $\mathbf{u}_i(x, \boldsymbol{\theta}) = \mathbf{u}(Y_i, x, \mathbf{Z}_i, \boldsymbol{\theta})$. The classical normalized likelihood score function when there is no covariate error is then given by $\mathbf{U}(\boldsymbol{\theta}) = n^{-1} \sum_i \mathbf{u}_i(X_i, \boldsymbol{\theta})$, and the maximum likelihood estimate (MLE) is obtained by solving the equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$.

The idea of the Stefanski-Nakamura corrected score approach is to find a function $\bar{\mathbf{u}}(y, w, \mathbf{z}, \boldsymbol{\theta})$ such that

$$E[\bar{\mathbf{u}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\theta})|X_i, \mathbf{Z}_i, Y_i] = \mathbf{u}(Y_i, X_i, \mathbf{Z}_i, \boldsymbol{\theta}). \tag{1}$$

We put $\bar{\mathbf{u}}_i(w, \boldsymbol{\theta}) = \bar{\mathbf{u}}(Y_i, w, \mathbf{Z}_i, \boldsymbol{\theta})$, and then use the modified likelihood score function $\bar{\mathbf{U}}(\boldsymbol{\theta}) = n^{-1} \sum_i \bar{\mathbf{u}}_i(W_i, \boldsymbol{\theta})$ in place of $\mathbf{U}(\boldsymbol{\theta})$ as the basis for estimation. The estimation equation thus becomes $\bar{\mathbf{U}}(\boldsymbol{\theta}) = \mathbf{0}$.

In the present setting, the equation (1) for the corrected score function takes the form

$$\int a_i(x,w)\bar{u}_{ij}(w)dw = u_{ij}(x), \tag{2}$$

where $u_{ij}(x)$ and $\bar{u}_{ij}(w)$, respectively, denote the j-th component of $\mathbf{u}_i(x,\boldsymbol{\theta})$ and $\bar{\mathbf{u}}_i(w,\boldsymbol{\theta})$ (suppressing the argument $\boldsymbol{\theta}$ in the definitions), and the integral is over the entire range of W. As indicated in the introduction, we do not seek an exact solution to (2), but instead use the method of regularization to find an approximate solution.

Define the integral operator

$$A_i g(x) = \int a_i(x,w)g(w)dw.$$

Write $\Delta_{ij}(x) = u_{ij}(x) - A_i u_{ij}(x)$ and $\bar{\Delta}_{ij}(w) = \bar{u}_{ij}(w) - u_{ij}(w)$. We can then write (2) as $A_i\bar{\Delta}_{ij} = \Delta_{ij}$. We seek the $\bar{\Delta}_{ij}(\cdot,\boldsymbol{\theta},\alpha)$ that minimizes the penalized loss function

$$\mathcal{L}_{ij}(\bar{\Delta}_{ij}) = \|A_i\bar{\Delta}_{ij} - \Delta_{ij}\|^2 + \alpha\|\bar{\Delta}_{ij}\|^2, \tag{3}$$

where $\|g\|^2$ denotes the weighted squared L^2 norm $\|g\|^2 = \int c(v)g(v)^2dv$ and $\alpha > 0$ is a penalty factor. After obtaining $\bar{\Delta}_{ij}(\cdot,\boldsymbol{\theta},\alpha)$, we use $\bar{u}_{ij}(w,\boldsymbol{\theta},\alpha) = u_{ij}(w) + \bar{\Delta}_{ij}(w,\boldsymbol{\theta},\alpha)$ as a corrected score term.

For the weight function $c(v)$, we propose as a generic choice the standard normal density, i.e., $c(v) = \varphi(v)$ with $\varphi(v) = \exp(-v^2/2)/\sqrt{2\pi}$, after standardizing W to mean 0 and variance 1. The weight function is designed to put emphasis on the region of the covariate space where the bulk of the data lie. One could consider the possibility of tailoring the choice of the weight function to the pattern of the observed distribution of W, but we do not discuss this here.

We formulate the minimization problem in terms of $\bar{\Delta}_{ij}(\cdot,\boldsymbol{\theta},\alpha)$ rather than $\bar{u}_{ij}(\cdot,\boldsymbol{\theta},\alpha)$ in order to anchor the procedure at $u_{ij}(w)$, which corresponds to the naive analysis in which

we ignore the covariate error, and simply substitute W_i for X_i. See Hansen (1994, Sec. 2, second paragraph) for the idea of centering the regularization process around an initial estimate of the desired solution to the integral equation. As $\alpha \to \infty$, the loss function $\mathcal{L}_{ij}(\bar{\Delta}_{ij})$ puts increasingly heavy weight on $\|\bar{\Delta}_{ij}\|^2$, which causes the minimizer $\bar{\Delta}_{ij}$ to tend to the zero function, leading to the naive estimates based on $u_{ij}(w)$. At the other extreme, as $\alpha \to 0$, the problem of minimizing $\mathcal{L}_{ij}(\bar{\Delta}_{ij})$ approaches the problem of solving $A\bar{\Delta}_{ij} = \Delta_{ij}$. In presenting our method, we first describe the procedure for a fixed α, and then discuss the selection of the value of α.

By working with the L^2 norm, we ensure that the problem of minimizing the loss function always has a unique solution, and the solution has a mathematically convenient form. Delves and Mohamed (1985, Sec. 12.3) and Kress (1989, Theorem 16.1) present the relevant theory. Let

$$A_i^* h(w) = \int a_i(x, w) h(x) dx$$

denote the adjoint operator corresponding to the operator A_i. Then, for any L^2 function δ, the minimizer of $\mathcal{L}(\bar{\delta}; A_i, \delta, \alpha) = \|A_i\bar{\delta} - \delta\| + \alpha\|\bar{\delta}\|^2$ is given by

$$\bar{\delta} = (A_i^* A_i + \alpha\mathcal{I})^{-1} A_i^* \delta, \tag{4}$$

where $\mathcal{I}$ is the identity operator.

In the next section, we present a numerical scheme for finding the solution that leads to a simple linear system of equations. We thus obtain a procedure that is easily implemented.

3. The Procedure

3.1. *The Procedure for a Given α*

To numerically determine the minimizer of $\mathcal{L}_{ij}(\bar{\Delta}_{ij})$, we use a Galerkin-type basis expansion approach, in the spirit of (though not identical to) Delves and Mohamed (1985,

Sec. 12.4). Specifically, we represent the solution $\bar{\Delta}_{ij}(\cdot, \boldsymbol{\theta}, \alpha)$ in a basis expansion of the form

$$\bar{\Delta}_{ij}(x, \boldsymbol{\theta}, \alpha) = \sum_{m=1}^{M} d_{ijm}(\alpha)\psi_m(x), \tag{5}$$

where the ψ_m are specified basis functions. In our numerical work, we use the "probabilists'" Hermite polynomials, which are orthonormal with respect to the weight function φ. One has to choose the number M of basis functions to include. We found that $M = 6$ yields good performance; the results with $M = 4$ are inferior to those with $M = 6$, while the results with $M = 10$ are similar to those with $M = 6$ but with more outliers.

Denote $\phi_{im}(x) = A_i\psi_m(x)$ and $\mathbf{d}_{ij} = [d_{ij1} \dots d_{ijM}]^T$ (suppressing the argument α in d_{ijm} for the time being). We then can express the objective function $\mathcal{L}_{ij}(\bar{\Delta}_{ij})$ as

$$\mathcal{L}_{ij}(\bar{\Delta}_{ij}) = \| \sum_{m=1}^{M} d_{ijm}\phi_{im} - \Delta_{ij}\|^2 + \alpha\, \mathbf{d}_{ij}^T\mathbf{d}_{ij},$$

where $\|\bar{\Delta}_{ij}\|^2 = \mathbf{d}_{ij}^T\mathbf{d}_{ij}$ because of the orthonormality of the ψ_m functions. We now approximate the L^2 norm in the first term on the right side via the quadrature approximation

$$\int \phi(v)g(v)dv \doteq \sum_{k=1}^{K} q_k g(x_k), \tag{6}$$

where x_k and q_k are the classical Gauss-Hermite quadrature points and weights (modified slightly to account for our use of the weight function $\exp(-v^2/2)/\sqrt{2\pi}$ as opposed to the standard Hermite weight function $\exp(-v^2)$). Given the approximation (6), we can express the objective function as

$$\begin{aligned}
\mathcal{L}_{ij}(\bar{\Delta}_{ij}) &= \sum_{k=1}^{K} q_k \left[\sum_{m=1}^{M} d_{ijm}\phi_{im}(x_k) - \Delta_{ij}(x_k)\right]^2 + \alpha\, \mathbf{d}_{ij}^T\mathbf{d}_{ij} \\
&= \sum_{k=1}^{K} \left[\sum_{m=1}^{M} d_{ijm}\tilde{\phi}_{im}(x_k) - \tilde{\Delta}_{ij}(x_k)\right]^2 + \alpha\, \mathbf{d}_{ij}^T\mathbf{d}_{ij},
\end{aligned}$$

where $\tilde{\phi}_{im}(x) = \sqrt{q_k}\phi_m(x_k)$ and $\tilde{\Delta}_{ij}(x) = \sqrt{q_k}\Delta_{ij}(x)$. Next, define the matrix $\tilde{\boldsymbol{\Phi}}_i$ by

$\tilde{\Phi}_{ikm} = \tilde{\phi}_{im}(x_k)$ and the vector $\tilde{\boldsymbol{\Delta}}_{ij} = [\tilde{\Delta}_{ij}(x_1) \ldots \tilde{\Delta}_{ij}(x_K)]$. We obtain $\mathcal{L}_{ij}(\bar{\Delta}_{ij}) = (\tilde{\boldsymbol{\Phi}}_i \mathbf{d}_{ij} - \tilde{\boldsymbol{\Delta}}_{ij})^T (\tilde{\boldsymbol{\Phi}}_i \mathbf{d}_{ij} - \tilde{\boldsymbol{\Delta}}_{ij}) + \alpha\, \mathbf{d}_{ij}^T \mathbf{d}_{ij}$. We then find, by standard least squares theory, that the vector $\mathbf{d}_{ij}(\alpha)$ that minimizes the above quantity is given by $\mathbf{d}_{ij}(\alpha) = \mathbf{C}(\alpha)\tilde{\boldsymbol{\Delta}}_{ij}$, where $\mathbf{C}(\alpha) = (\tilde{\boldsymbol{\Phi}}^T \tilde{\boldsymbol{\Phi}} + \alpha \mathbf{I})^{-1}$. Note that $\mathbf{C}(\alpha)$ does not depend on $\boldsymbol{\theta}$. Finally, we define $\bar{u}_{ij}(w, \boldsymbol{\theta}, \alpha) = u_{ij}(w, \boldsymbol{\theta}) + \bar{\Delta}_{ij}(w, \boldsymbol{\theta}, \alpha)$, where $\bar{\Delta}_{ij}(w, \boldsymbol{\theta}, \alpha)$ is given by (5) with $d_{ijm}(\alpha)$ obtained as just described. Then, as indicated in the preceding section, we put $\bar{\mathbf{U}}(\boldsymbol{\theta}, \alpha) = n^{-1} \sum_i \bar{\mathbf{u}}_i(W_i, \boldsymbol{\theta}, \alpha)$ and define the estimator $\hat{\boldsymbol{\theta}}^{(\alpha)}$ to be the solution to $\bar{\mathbf{U}}(\boldsymbol{\theta}, \alpha) = \mathbf{0}$.

In the course of the foregoing procedure, we have to evaluate integrals of the form

$$E[g(W_i)|X_i = x] = A_i g(x) = \int a_i(x, w) g(w) dw. \tag{7}$$

Integrals of this type appear in $\phi_{im}(x) = A_i \psi_m(x)$ and in $\Delta_{ij}(x) = u_{ij}(x) - A_i u_{ij}(x)$. These integrals can be evaluated by K'-point numerical quadrature for suitable K'. Appendix 1 presents the details.

In regard to the choice of K and K', in our numerical work we generally used $K = K' = 20$; we reran selected simulations with $K = K' = 30$ and obtained similar results. In a data analysis, the analyst can try a succession of increasing values of K and K', and stop when there is no further change in the results.

In practice, $a_i(x, w)$ has to be estimated, using data from a replicate measures study or an internal or external validation study. We assume that $a_i(x, w)$ follows a known parametric model depending on parameters $\boldsymbol{\xi}$ (distinct from $\boldsymbol{\theta}$) which are estimated from the relevant data. Accordingly, we write $a_i(x, w, \boldsymbol{\xi})$. The parametric model is allowed, however, to be of any specified form. Thus, in addition to the classical independent additive error model, we allow for models with dependence between the error and the true covariate value, and models with differential error. Moreover, in our numerical studies,

we have examined the effect of misspecifying the parametric form.

3.2. *Theoretical Properties*

In general, $\hat{\boldsymbol{\theta}}^{(\alpha)}$ will not converge to the true value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$, but rather to a limiting value $\tilde{\boldsymbol{\theta}}^{(\alpha)}$ that is close to $\boldsymbol{\theta}_0$ when α is small. In practice, we cannot make α arbitrarily small, but we can try to make it small enough to obtain estimates whose bias is small, and the numerical studies presented in the next section indicate that this goal can be achieved. Thus, our method does not produce an exactly consistent estimator, but it does produce an approximately consistent estimator. Moreover, under standard regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(\alpha)})$ is asymptotically normal. The foregoing properties are formalized in the following theorem, which is proved in Appendix 2.

Theorem: Assume the following regularity conditions.

A1. The parameter space $\boldsymbol{\Theta}$ is compact with a nonempty interior which includes the true value $\boldsymbol{\theta}_0$.

A2. The function $\mathbf{u}_i(x, \boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ over $\boldsymbol{\Theta}$ for every x, with derivative that is bounded over x by an L^2 function of x.

A3. The matrix $\mathbf{D}_E(\boldsymbol{\theta})$ defined by $(D_E)_{rs}(\boldsymbol{\theta}) = -E[(\partial^2/\partial\theta_r\partial\theta_s)\log f(Y|X, \mathbf{Z}, \boldsymbol{\theta})]$ is positive definite over $\boldsymbol{\Theta}$.

A4. The null space $\mathcal{N}(A_i^*)$ consists only of the zero function, i.e., the only solution to $A_i^* h = 0$ is $h = 0$.

Define $\bar{\mathbf{u}}_E(\boldsymbol{\theta}, \alpha) = E[\bar{\mathbf{u}}_i(W_i, \boldsymbol{\theta}, \alpha)]$. Let $\bar{\mathbf{D}}^{(\alpha)}(\boldsymbol{\theta})$ denote -1 times the Jacobian matrix of $\bar{\mathbf{U}}(\boldsymbol{\theta}, \alpha)$. The following results then hold.

a. We have

$$E[\bar{u}_{ij}(W_i, \boldsymbol{\theta}, \alpha)|X_i, \mathbf{Z}_i, Y_i] = u_{ij}(X_i, \boldsymbol{\theta}) + r_{ij}(X_i, \boldsymbol{\theta}, \alpha)$$

with $\lim_{\alpha\to 0}\sup_{\boldsymbol{\theta}}\|r_{ij}(\cdot,\boldsymbol{\theta},\alpha)\| = 0$.

b. Similarly, with $u_{ijs}(x,\boldsymbol{\theta}) = [\partial/\partial\theta_s]u_{ij}(x,\boldsymbol{\theta})$ and $\bar{u}_{ijs}(w,\boldsymbol{\theta},\alpha) = [\partial/\partial\theta_s]\bar{u}_{ijs}(w,\boldsymbol{\theta},\alpha)$, we have

$$E[\bar{u}_{ijs}(W_i,\boldsymbol{\theta},\alpha)|X_i,\mathbf{Z}_i,Y_i] = u_{ijs}(X_i,\boldsymbol{\theta}) + r_{ijs}(X_i,\boldsymbol{\theta},\alpha)$$

with $\lim_{\alpha\to 0}\sup_{\boldsymbol{\theta}}\|r_{ijs}(\cdot,\boldsymbol{\theta},\alpha)\| = 0$.

c. For all α sufficiently small, the equation $\bar{\mathbf{u}}_E(\boldsymbol{\theta},\alpha) = \mathbf{0}$ has a unique solution, which we denote by $\tilde{\boldsymbol{\theta}}^{(\alpha)}$.

d. For fixed α, we have $\hat{\boldsymbol{\theta}}^{(\alpha)} \to \tilde{\boldsymbol{\theta}}^{(\alpha)}$ almost surely as $n\to\infty$.

e. We have $\tilde{\boldsymbol{\theta}}^{(\alpha)} \to \boldsymbol{\theta}_0$ as $\alpha\to 0$.

f. If $a_i(x,w)$ is known, $\sqrt{n}(\hat{\boldsymbol{\theta}}^{(\alpha)} - \tilde{\boldsymbol{\theta}}^{(\alpha)})$ is asymptotically mean-zero normal with covariance matrix that can be estimated using the sandwich estimator

$$\mathbf{V}^{(\alpha)}(\hat{\boldsymbol{\theta}}^{(\alpha)}) = \bar{\mathbf{D}}^{(\alpha)}(\hat{\boldsymbol{\theta}}^{(\alpha)})^{-1}\mathbf{F}^{(\alpha)}(\hat{\boldsymbol{\theta}}^{(\alpha)})\bar{\mathbf{D}}^{(\alpha)}(\hat{\boldsymbol{\theta}}^{(\alpha)})^{-1}, \tag{8}$$

with

$$\mathbf{F}^{(\alpha)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{u}}_i(\boldsymbol{\theta},\alpha)\bar{\mathbf{u}}_i(\boldsymbol{\theta},\alpha)^T. \tag{9}$$

Under a parametric model for $a_i(x,w)$ with estimated parameters $\boldsymbol{\xi}$, a similar result holds, with a suitable adjustment to the estimated covariance matrix to account for the estimation of $\boldsymbol{\xi}$, as described in Appendix 3.

Remark 1: The expectation in the definition of $\bar{\mathbf{u}}_E(\boldsymbol{\theta},\alpha)$ is an unconditional expectation over all random variables in the model. In Result (f), the covariance is an unconditional covariance matrix, paralleling that in Nakamura's (1990) Eqn. 2.

Remark 2: It is natural to ask about the rate of convergence of $\tilde{\boldsymbol{\theta}}^{(\alpha)}$ to $\boldsymbol{\theta}_0$. This convergence rate depends on the rate of convergence of $\|A_i\bar{\Delta}_{ij} - \Delta_{ij}\|$ to zero as α tends

to zero. The discussion in Delves and Mohamed (1985, pp. 308-309) indicates that the latter convergence rate depends on the rate of decay in the Fourier coefficients of Δ_{ij} with respect to a basis defined by the eigenfunctions of the operator A_i. In our setting, this rate of decay is hard to characterize, making precise convergence rate results hard to obtain.

Remark 3: Assumptions A1-A3 are typical assumptions made in asymptotic theory; see, for example, van der Vaart (1998, p. 46, bottom). Assumption A4 is a modest assumption that holds in many cases of interest. For example, suppose (X, W) follows the independent additive error model $W = X + \sigma\varepsilon$, where ε is a random variable with density f_ε, independent of X. We then have $a(x, w) = \sigma^{-1} f_\varepsilon((w - x)/\sigma) = \sigma^{-1}\tilde{f}_\varepsilon((x - w)/\sigma)$, with $\tilde{f}_\varepsilon(u) = f_\varepsilon(-u)$. The assumption thus will be satisfied provided that the location-scale family of densities $f(x; w, \sigma) = \sigma^{-1}\tilde{f}_\varepsilon((x - w)/\sigma)$ is a complete family of densities with respect to the parameters (w, σ). This condition certainly holds if f_ε is a density of exponential family form; see Lehmann (1986, p. 142). Next, consider the extended model $W = X + \sigma(X, \boldsymbol{\gamma})\varepsilon$, which we examine in our numerical studies, where $\boldsymbol{\gamma}$ is a vector of parameters. Assumption A4 will hold in this setting if the family of densities

$$f(x; w, \boldsymbol{\gamma}) = \frac{1}{\sigma(x, \boldsymbol{\gamma})}\tilde{f}_\varepsilon\left(\frac{x - w}{\sigma(x, \boldsymbol{\gamma})}\right) \Big/ \int \frac{1}{\sigma(x', \boldsymbol{\gamma})}\tilde{f}_\varepsilon\left(\frac{x' - w}{\sigma(x', \boldsymbol{\gamma})}\right) dx'$$

is a complete family of densities with respect to the parameters $(w, \boldsymbol{\gamma})$. Again, this condition will hold if f_ε is a density of exponential family form.

3.3. *Choice of the Penalty Parameter Function* α

The issue of how to choose the penalty parameter in a regularization problem has been investigated in previous literature. Hansen (1994, 2007) describes three leading criteria: the L-curve criterion, the GCV criterion, and the quasi-optimality criterion. We tried all three, and found the GCV criterion to be the most satisfactory. The GCV criterion is

defined as

$$\mathrm{GCV}(\alpha) = \frac{(\tilde{\mathbf{\Phi}}_i \mathbf{d}_{ij}(\alpha) - \tilde{\mathbf{\Delta}}_{ij})^T (\tilde{\mathbf{\Phi}}_i \mathbf{d}_{ij}(\alpha) - \tilde{\mathbf{\Delta}}_{ij})}{[\mathrm{trace}(\mathbf{I} - \mathbf{C}(\alpha))]^2},$$

and α is chosen to minimize the value of this quantity. In our setting, we have a separate value $\mathrm{GCV}_{ij}(\alpha)$ for each i and j. We work with the summary criterion

$$\mathrm{GCV}^*(\alpha) = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} \mathrm{GCV}_{ij}(\alpha),$$

and choose α to minimize this quantity. In implementing this rule, we evaluate $\tilde{\mathbf{\Delta}}_{ij}$ at the naive estimate of $\boldsymbol{\theta}$, and then keep α fixed at the resulting value for the remainder of the estimation process.

3.5. *Multiple Error-Prone Covariates*

The method can be readily extended to the case of two error-prone covariates X_1 and X_2. For the basis functions, we use the tensor product of the univariate basis functions. The integrals involved in quantities of the form $A_i g(x)$ become double integrals, which are evaluated by bivariate quadrature. In the L^2 norm appearing in the objective function, we take the weight function to be $c(w_1, w_2) = \varphi(w_1)\varphi(w_2)$, and evaluate the integral using the bivariate version of (6).

In the case of three or more error-prone covariates, the situation becomes more complicated. Taking the basis function set to be the tensor product of the univariate basis functions will typically produce too large a basis function set, so some reduction will be necessary. One could, for example, take the basis function set to include all the univariate basis functions for the individual covariates plus the cross-products of the linear terms. In the evaluation of the integrals $A_i g(x) = E[g(\mathbf{W}_i)|X_i = x]$, a Monte-Carlo procedure will probably be more workable than a classical quadrature procedure. The L^2 norms can be computed using a multivariate version of (6), but the computational load may be demanding.

The simulation work presented in the next section includes results for the case of two

error-prone covariates. We have not attempted a numerical study of the case of three error-prone covariates. On a practical level, it appears that it would be challenging to apply our method in the setting of three or more error-prone covariates. However, many applications involve only one or two error-prone covariates, and thus can be handled by our method in a reasonable way. An arbitrary number of error-free covariates can be handled without difficulty.

4. The Logistic Regression Model: Simulation Studies

4.1. *Simulation Study Designs*

To investigate the finite sample performance of our method, we conducted a series of simulation studies. This subsection describes the simulation study designs; the next subsection describes the results. The studies were conducted in the setting of the logistic regression model. The response variable Y_i equals either 0 or 1. Defining $\mathbf{T}_i = (\mathbf{X}_i, \mathbf{Z}_i)$ and $T_{i0} \equiv 1$, the model is

$$\text{logit}\,\Pr(Y_i = 1 | \mathbf{X}_i = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \sum_{j=0}^{p} \beta_j t_j$$

and the likelihood score function $\mathbf{u}(y, x, z, \boldsymbol{\theta})$ is given by

$$u_j(y, x, z, \boldsymbol{\theta}) = t_j \left[y - \text{expit} \left(\sum_{j=0}^{p} \beta_j t_j \right) \right],$$

with $\text{expit}(a) = e^a/(1 + e^a)$.

In the simulations we examined the following methods:

1. Naive analysis ignoring measurement error
2. RECS
3. The Novick and Stefanski (2001) complex variable corrected score method (N&S)
4. SIMEX, with linear, quadratic, and nonlinear extrapolation
5. The Huang and Wang (2001) nonparametric corrected score method (H&W)

In the simulations on the Huang and Wang method, which requires replicate measurements of W, we took two replicates per individual and doubled the error variance for comparability with the other methods. Note that H&W is designed to provide accurate estimates only for the slope parameter, not for the the intercept parameter. All simulation results are based on 1,000 simulation replications.

The SIMEX method is designed for independent additive normal measurement error. The N&S method is designed for the case where the measurement error is independent additive normal and the likelihood score function is an entire function in the complex plane. The latter condition does not hold for logistic regression, and thus the N&S method is not designed to handle any of the simulation scenarios we have studied. The H&W nonparametric method is designed for independent additive measurement error with an arbitrary distribution, which is not modeled in any way. The RECS method is designed for parametric measurement error models of arbitrary form, including non-normal, heteroscedastic, and differential error.

For each estimator considered, we summarize the bias in terms of the mean and median of the difference between the estimated and true parameter values, and the dispersion in terms of the empirical standard deviation and standardized interquartile range (dividing the raw interquartile range by 1.349, which is the ratio between the interquartile range and the standard deviation for a normal distribution). The median and the interquartile distance are more robust to outliers, and therefore provide further insight in addition to that provided by the mean and the standard deviation. In addition, for the naive estimator and the RECS estimator, we present the coverage rates of the 95% Wald confidence interval based on the asymptotic normal theory with the relevant variance estimators.

In Simulation Sets A and B, we considered the case of a single continuous error-prone covariate X_i and no other covariates. In these two simulation sets, we worked in the setting of a main study / external validation study design, involving a main study sample with data on W and Y and an external validation sample with data on W and

X. The main study sample size was 200 and the external validation sample size was 70. The measurement error parameters were estimated by maximum likelihood. The RECS method was implemented with $M = 6$ basis functions and $K = K' = 20$ quadrature points. The true values of the regression parameters were set at $\beta_0 = \beta_1 = 1$.

Simulation Set A involved measurement error models of the form $W_i = X_i + \epsilon_i$, where ϵ_i is normally distributed, but with error variance possibly depending on X_i and Y_i. We examined three simulation scenarios, as follows:

Scenario A1: $X_i \sim N(0,1),\ \epsilon_i|(X_i, Y_i) \sim N(0, \gamma)$

Scenario A2: $X_i \sim N(0,1),\ \epsilon_i|(X_i, Y_i) \sim N(0, \gamma_1 + \gamma_2|X_i|)$

Scenario A3: $X_i \sim N(0,1),\ \epsilon_i|(X_i, Y_i) \sim N(0, \gamma_1 + \gamma_2|X_i| + \gamma_3|Y_i|)$

For each of the above scenarios, we examined two sets of measurement error parameters. In Scenario A1, we took $\gamma = 0.5$ or 1. In Scenarios A2 and A3, the two sets of measurement error parameters were chosen such that the unconditional variance of ϵ_i was approximately equal to 0.5 or 1, respectively.

Scenario A1 is the classical additive error model, which is theoretically covered by RECS, SIMEX, and H&W (for N&S, the measurement error model assumption is satisfied but the entire function condition is not satisfied). Scenarios A2 and A3 involve heteroscedastic error models that are theoretically covered only by RECS.

Simulation Set B involved measurement error models of the form $W_i = X_i + \epsilon_i$, with non-normal ϵ_i. We considered two distributions for ϵ_i, the double-exponential distribution (DBLEXP(γ), with γ denoting the variance) and a modified chi-square distribution MODCHI which Huang and Wang (2001) used in their simulation work. Specifically, the MODCHI(γ) is defined to be the distribution of a χ^2_1 variate truncated at the value 5, recentered to mean zero, and then rescaled to a variance of γ. In Scenario B3, we also take X to have a MODCHI distribution. The DBLEXP distribution is similar to the

normal, but with heavier tails. The MODCHI distribution is highly skewed. The specific scenarios examined were as follows:

Scenario B1: $X_i \sim N(0,1),\ \epsilon_i|(X_i, Y_i) \sim \text{DBLEXP}(\gamma)$

Scenario B2: $X_i \sim N(0,1),\ \epsilon_i|(X_i, Y_i) \sim \text{MODCHI}(\gamma)$

Scenario B3: $X_i \sim \text{MODCHI}(1),\ \epsilon_i|(X_i, Y_i) \sim \text{MODCHI}(\gamma)$

These scenarios are theoretically covered by RECS and H&W, but not by SIMEX or N&S.

For each of these scenarios, we ran simulations for $\gamma = 0.5$ and $\gamma = 1$. For the MODCHI distribution, integrals of the form $A_i g(x)$ were evaluated using the representation

$$A_i g(x) = \frac{2}{\Pr(\chi^2_1 \leq 5)} \int_0^{\sqrt{5}} \varphi(v) g\left(x + \gamma\left[\frac{v^2 - \mu_{MC}}{\sigma_{MC}}\right]\right) dv$$

in conjunction with Gauss-Hermite quadrature, where μ_{MC} and σ_{MC} denote, respectively, the mean of the chi-square distribution truncated at 5.

Note that the MODCHI(γ) is an non-regular distributional family: it has support that depends on γ. As a result, classical asymptotic theory for MLE's does not apply to the MLE of γ, but, at the same time, because of the restricted range of γ values that are compatible with a given dataset due to the definition of the support, with an external validation sample of 70 the value of γ is estimated with virtually no error.

Simulation Set C examined, in the setting of Simulation Sets A and B, the effect of misspecifying the error distribution. We generated the errors according to one of two possible non-normal distributions, but implemented our method assuming the errors are normal. The non-normal error distributions used were the MODCHI(γ) distribution and a modified version of Azzalini's (1985) skewed normal distribution. Azzalini's skewed normal distribution $SN(\lambda)$ has density $2\phi(y)\Phi(\lambda y)$, $y \in \mathbb{R}$, where ϕ and Φ denote the

standard normal density and distribution function, respectively, and λ is a parameter that regulates the skewness ($\lambda = 0$ gives the standard normal). Our modified version recenters to mean zero and then rescales to the specified variance. We took $W_i = X_i + \epsilon_i$, with X_i taken to be either $N(0,1)$, $SN(50)$, or MODCHI, and the distribution of ϵ_i taken to be either the MODCHI(γ) or the the modified $SN(\lambda)$ with $\lambda = 50$. The skewness of the $SN(50)$ distribution is 1, and that of the MODCHI is 1.7.

Simulation Set D considered the case of two error-prone covariates X_1, X_2 and one error-free covariate Z. For this simulation set, we took the sample size to be 500, and we took the measurement error parameters as known. For the RECS method, the basis function set was taken to be the tensor product of the univariate basis functions sets with $M = 6$, and in the quadrature calculations we took $K' = 20$ and $K = 10$. The true regression coefficients were taken to be $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$. The scenarios examined were as follows.

Scenario D1: X_1, X_2, Z i.i.d. $N(0,1)$, $\epsilon_1; \epsilon_2$ i.i.d. $N(0,\gamma)$, independent of X_1, X_2, Z

Scenario D2: X_1, X_2, Z i.i.d. $N(0,1)$, $\epsilon_1; \epsilon_2$ conditionally independent given X_1, X_2, Z and distributed as $N(0, \gamma_1 + \gamma_2(|X_1| + |X_2| + |Z|))$

Scenario D3: $X_1, X_2 \sim$ MODCHI(1); $Z \sim N(0,1)$; $\epsilon_1, \epsilon_2 \sim$ MODCHI(γ), all random variables independent of each other

In Scenarios D1 and D3, we took $\gamma = 1$, while in Scenario D2 we took $\gamma_1 = 0.4$ and $\gamma_2 = 0.25$, so that the unconditional variance of ϵ_1 and ϵ_2 was about 1. Scenario S1 is the classical additive error model, which is theoretically covered by RECS, SIMEX, and H&W (for N&S, the measurement error model assumption is satisfied but the entire function condition is not satisfied). Scenario A2 involves a heteroscedastic error model that is theoretically covered only by RECS. Scenario A3 involves a non-normal error model that is theoretically covered by RECS and H&W, but not SIMEX and N&S.

The R code for the simulations is posted on the first author's website, as follows:

R code for Simulation Studies A, B, and C

`http://pluto.huji.ac.il/~mszucker/recs-siml-univ.zip`

R code for Simulation Study D

`http://pluto.huji.ac.il/~mszucker/recs-siml-multiv.zip`

4.2. *Simulation Results*

Tables 1, 2, and 3, present, respectively, the results of Simulation Studies A, B, and C, while Table 4-6 present results of Simulation Study D. Here we discuss the findings, focusing on estimation of the slope parameter. Overall, the RECS method performs very well in terms of bias and confidence interval coverage. Below we discuss how RECS compares with the competing methods.

We begin with the Simulation Study A, the setting of normal measurement error with possible heteroscedasticity. RECS showed low bias throughout, especially when we look at the median of the estimates. N&S also generally showed low bias, although the bias was greater than that of RECS in a number of cases. SIMEX-NL also showed low bias in most cases (the last panel of Table 1 being a notable exception), but estimation procedure was unsuccessful 6-10% of the time due to failure of the extrapolation process. SIMEX-L, SIMEX-Q, and H&W performed markedly less well in terms of bias. The variability of the RECS estimates was comparable to that of the SIMEX-NL estimates and generally higher than that of the other methods.

We turn now to Simulation Study B, with non-normal measurement error. Again, RECS showed low bias throughout. N&S had low bias with double exponential error, and showed the best performance among the various methods in this case. With MODCHI error, RECS had low bias, while all the other methods had substantial bias.

Next, we discuss Simulation Study C, which considered the performance of RECS when the error model is misspecified (assumed normal but actually skewed normal or

MODCHI). Aside from the case with both X and ϵ distributed MODCHI, the mean bias was in the range of 15-35% and the median bias was in the range 10-20%. Thus, from the standpoint of median bias, the RECS method performed reasonably – certainly much better than the naive analysis with no measurement error correction. In the case with both X and ϵ distributed MODCHI, RECS performed poorly. However, this result is not too disturbing – because of the great difference between the MODCHI and normal distributions, it is unlikely that an analyst would mistakenly fit a normal model to MODCHI measurement errors. In principle, added robustness can be incorporated by using a flexible distributional form for the error model, such as the "semi-nonparametric normal" model of Gallant and Nychka (1987).

Finally, we turn to Simulation Study D, with two error-prone covariates and one error-free covariate. The SIMEX and H&W methods performed poorly in estimating the slope parameters of the two error-prone covariates. The SIMEX-NL method performed especially poorly, producing estimates way off in the wrong direction. The RECS method performed well. In Scenarios D1 (normal homoscedastic error) and D3 (MODCHI error), RECS performed markedly better than N&S in terms of producing estimates with low bias, while in Scenario D3 (normal heteroscedastic), the peformance was similar, with RECS overestimating and N&S underestimating.

The degree of penalization in the RECS method tended to be very low throughout; the mean α value was less than 0.01 in all simulation scenarios studied.

5. Practical Illustration

We illustrate the method on data from the Nurses' Health Study (NHS). The NHS began in 1976 when 121,700 female nurses aged 30-55 returned a questionnaire about their lifestyle and their health. Here, we analyze the relationship between physical activity and mortality among women diagnosed with breast cancer during the course of the NHS follow-up. This relationship was previously examined by Holmes et al. (2005). The

present analysis involves a subset of the nurses included in the analysis of Holmes et al. Specifically, we consider the group of 1660 nurses who were diagnosed with breast cancer at least 10 years before the administrative end of the study, June 2002. This is the main study. The endpoint is the binary variable defined as breast cancer death with the first 10 years following diagnosis. The total number of such deaths was 188. In NHS, physical activity was assessed by a questionnaire in which women were asked how much time they spent on average during the past year on each of the most common forms of leisure time physical activity. The questionnaire results were then converted into metabolic equivalent task hours per week (METS). Validation data were available from 149 women from the NHS II study (Wolf et al., 1994), a study begun in 1989 which involved a cohort of U.S. female nurses similar to that of the NHS, and in which the same physical activity questions were asked. In our analysis, these data are regarded as arising from an external validation study. In the validation study, METS was assessed using both the questionnaire and a detailed activity diary, with the diary regarded as the gold standard. We denote the METS value based on diary data by X and the METS value based on questionnaire data by W. The degree of measurement error is considerable, with the correlation between X and W in the validation study being 0.47.

Wolf et al. reported that the distribution of METS was skewed, and a square-root transformation yielded a distribution closer to the normal. Initially we considered using the square-root transformation in the measurement error model in our analysis, but we later decided that it would be better to use a transformation that maps the positive half line into the entire real line, in order to facilitate modeling the measurement error using the normal distribution or another distribution with support on the entire real line. Denoting the true METS value by X, the transformation we ultimately used was $X^* = \log(1 + X)$.

We analyzed the main cohort data using two measurement error models developed from the validation data. We refer to these models as MEM1 and MEM2. Both models were of the form $W_i^* = \omega_0 + \omega_1 X_i^* + \epsilon_i$. In MEM1, ϵ_i was taken to have the $N(0, \sigma^2)$

distribution, independent of X_i^*. MEM2 incorporated dependence between ϵ_i and X_i^*, with the conditional distribution of ϵ_i given X_i^* taken to be $N(0, \gamma_1 + \gamma_2|X_i^* - \mu|)$, with $\mu = E[X_i^*]$. The parameters ω_0, ω_1 were estimated by simple linear regression in both models (weighted regression based on the MEM2 model for ϵ_i produced virtually identical estimates). For MEM1, σ^2 was estimated by the regression MSE in the standard manner. For MEM2, γ_1 and γ_2 were estimated via regression analysis of the squares of the residuals obtained from MEM1; in this regression analysis, the t-test on γ_2 was borderline significant (p=0.0614), suggesting some evidence of heteroscedasticity. The estimates obtained were $\omega_0 = 1.2271, \omega_1 = 0.5653, \sigma^2 = 0.8181, \gamma_1 = 0.5883$, and $\gamma_2 = 0.3497$. Graphical inspection of the normalized residuals based on the MEM2 model for ϵ_i showed reasonable conformity to a normal distribution.

A preliminary analysis of the main study data indicated that the log odds of breast cancer could be reasonably expressed as a linear function of $\log(1 + \text{METS})$, so we took this as the analysis variable in our logistic regression model. The preliminary analysis also examined a wide range of potential confounding variables, and revealed that adjusting for these variables had little effect on the results, and so in our analysis we deal only with the variable METS.

We implemented the RECS method with 6 basis functions and $K = K' = 20$ quadrature points in the quadrature calculations. The R code for running this example is posted on the first author's website (`http://pluto.huji.ac.il/~mszucker/recs-example.zip`). We also applied the SIMEX and N&S methods to the data. The H&W method is not relevant, since we have an external validation design rather than a replicate measures design. Standard errors for the SIMEX and N&S methods were computed using the bootstrap with 100 replications.

Table 7 presents the results for the various methods. The naive method was applied in two forms: (1) using W_i^* as is (Naive1) and (2) using $\tilde{W}_i^* = (W_i^* - \omega_0)/\omega_1$, thus correcting for location-scale bias but not for measurement error (Naive2). SIMEX-L

yielded a slope estimate similar to that yielded larger than Naive2, while SIMEX-Q gave a slightly larger estimate. RECS-MEM1, N&S, and SIMEX-NL yielded slope estimates differing markedly from the Naive2 estimate, and the estimates yielded by these three methods were comparable. RECS-MEM2 yielded a slope estimate differing substantially from RECS-MEM1, showing the impact of a more refined error model.

6. Discussion

We have presented a new "regularized corrected score" (RECS) approach to adjusting for covariate error in nonlinear regression problems. The approach builds on the corrected score method developed by Nakamura and colleagues. In the case of a continuous error-prone covariate, the corrected score approach involves solving an integral equation. In many problems, an exact solution to this integral equation does not exist or cannot be practically obtained, and so we have proposed using an approximate solution obtained using the method of regularization. In the setting of logistic regression, a series of simulation studies showed that the method performs well in general, and offers an advantage over existing methods in terms of superior performance in certain situations. In particular, the RECS method showed a marked advantage over competing methods in the case of a single error-prone covariate with MODCHI error and in the case of two error-prone and one error-free covariates under two of the error models studied.

We have developed the theory in the general setting of classical likelihood models, which covers, in particular, generalized linear models such as nonlinear regression, logistic regression, and Poisson regression. It is possible to extend the development to other settings. In particular, it is of interest to extend the method to the setting of Cox regression for survival data, using the work of Zucker and Spiegelman (2008) on corrected score analysis for the Cox models with a discrete error-prone covariate as a starting point. We plan to develop this extension in future work.

The computational complexity and load of the method is modest. For example, the data analysis described in the preceding section finished in about 1 minute of real time,

when run in R in batch mode on a VMware virtual machine configured with one AMD 2700 MHz processor and 1GB memory, installed on a physical machine SUN FIRE X4240.

The method presented here is a functional method in the sense of not requiring information on the distribution of the true covariate. This is in contrast to many other methods in the measurement error literature, such as regression calibration and likelihood-based methods. We do rely on a parametric model for the conditional distribution of the surrogate variable given the true variable, but our simulation results suggest that the performance of the estimates is robust to the parametric model except when there is an extreme discrepancy between the assumed and actual error model. In addition, it is possible in principle to use a flexible parametric model such as on Gallant and Nychka's (1987) "semi-nonparametric" model, which makes the reliance on parametric modeling less of a restriction. The method is extremely general in terms of the types of measurement error models covered. It allows the measurement error to depend on the true covariate value and on other covariates. It also allows differential error, where the measurement error depends on the outcome. This flexibility is a distinct advantage relative to other methods in the literature.

Acknowledgments

This work has been supported by grant 2R01CA050597 from the U.S. National Cancer Institute. We thank the Editor, Associate Editor, and referees for their very helpful comments, which led to major improvements in the paper. We also thank Yijian (Eugene) Huang for providing code for the Huang and Wang (2001) method.

References

Akazawa, K., Kinukawa, N., and Nakamura, T. (1998). A note on the corrected score function corrected for misclassification. *Journal of the Japan Statistical Society* **28**:115–123.

Buzas, J. (2009). A note on corrected scores for logistic regression. *Statistics and Probability Letters* **79**:2351–2358

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

Carrasco, M., and Florens, J.P. (2011). Spectral method for deconvolving a density. *Econometric Theory* **27**:546–581.

Cook, J.R., and Stefanski, L.A. (1995). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89**:1314–1328.

Delves, L.M., and Mohamed, J.L. (1985). *Computational Methods for Integral Equations.* Cambridge University Press: Cambridge.

Devanarayan, V., and Stefanski, L.A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics and Probability Letters* **59**:219–225.

Fuller, W.A. (1987). *Measurement Error Models.* Wiley: New York.

Hansen, P.C. (1994). Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms* **6**:1-35.

Hansen, P.C. (2007). Regularization Tools Version 4.0 for Matlab 7.3. *Numerical Algorithms* **46**:189-194.

Holmes, M.D., Chen, W.Y., Feskanich, D., Kroenke, C.H., and Colditz, G.A. (2005). Physical activity and survival after breast cancer diagnosis. *Journal of the American Medical Association* **293**:2479–2486.

Huang, Y., and Wang, C.Y. (2001). Consistent function methods for logistic regression with errors in covariates. *Journal of the American Statistical Association* **95**:1209–1219.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 221–233. University of California Press: Berkeley.

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

Kress, R. (1989). *Linear Integral Equations.* Springer: Berlin.

Kong, F.H., and Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* **9**:953-969.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wadsworth and Brooks/Cole: Pacific Grove, CA.

Nakamura, T. (1990). Corrected score function of errors-in-variables models: methodology and application to generalized linear models. *Biometrika* **77**:127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**:829–838.

Novick, S.J., and Stefanski, L.A. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association* **97**:472–481.

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

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

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press: Cambridge.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM: Philadelphia.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**:1–25.

Wolf, A.M., Hunter, D.J., Colditz, G.A., Manson, J.E., Stampfer, M.J., Corsano, K.A., Rosner, B., Krlska, A., and Willett, W.C. (1994) Reproducibility and validity of a self-administered physical activity questionnaire. *International Journal of Epidemiology* **23**:991–999.

Zucker, D.M., and Spiegelman, D. (2008). Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in Medicine* **27**:1911–1933.

Appendix 1

Evaluation of Integrals of the Form $A_i g(x)$

As noted in the main text, the RECS procedure requires evaluation of integrals of the form

$$E[g(W_i)|X_i = x] = A_i g(x) = \int a_i(x, w) g(w) dw. \tag{10}$$

In our simulation studies, we work with measurement error models of the form $W = X + \sigma(X)\varepsilon$, where ε is a random variable with density f_ε, independent of X. In this case, we can write

$$A_i g(x) = \int f_\varepsilon(v) g(x + \sigma(x)v) dv.$$

This integral can be evaluated using a quadrature approximation of the form

$$A_i g(x) \doteq \sum_{k=1}^{K'} r_k g(x + \sigma(x) v_k)$$

for suitable quadrature points and weights v_k and r_k. One case we consider is $\varepsilon \sim N(0,1)$, and in this case we use (modified) Gauss-Hermite points and weights. Another case we consider is the case where ε is the double-exponential (Laplace) distribution with variance 1, and in this case we use Gauss-Laguerre type points and weights (with the points and weights on the negative side of the real axis being the mirror image of those on the positive side).

In cases where specialized methods of the above sort are not readily available, a more general numerical approach can be used. In particular, if we let $F_{W|X}$ be the conditional distribution function of W given X, then $A_i g(x)$ can be evaluated as

$$A_i g(x) \doteq \frac{1}{K'} \sum_{k=1}^{K'} g(F_{W|X}^{-1}(v_k|x))$$

with $v_k = (k - \frac{1}{2})/K'$.

Appendix 2

Proof of the Theorem

In proving Part (a) of the theorem, we rely on the L^2 theory of integral operators as set forth, for example, in Kress (1989). We recall that, for a general integral operator B, the range of B is defined as $\mathcal{R}(B) = \{h \in L^2 : h = Bg \text{ for some } g \in L^2\}$ and the null space of B is defined $\mathcal{N}(B) = \{g \in L^2 : Bg = 0\}$. We use a superscript $\perp$ to denote orthogonal complement and the notation $cl(C)$ to denote the L^2 closure of a set $C \subset L^2$.

Kress (1989, Theorem 15.8) states that for a bounded linear operator $B : L^2 \to L^2$ with adjoint B^*, we have $\mathcal{N}(B^*)^\perp = cl(\mathcal{R}(B))$. Since $a_i(x,w)$ is a conditional density,

the operator A_i is a bounded linear operator with norm 1. Assumption A4 specifies that $\mathcal{N}(A_i^*)$ consists only of the zero function. It follows that the L^2 closure of $\mathcal{R}(A_i)$ is equal to the whole of L^2. This, in turn, implies that $\inf_{\bar{\delta}\in L^2} \|A_i\bar{\delta} - g\| = 0 \; \forall \, g \in L^2$, although the infimum is not necessarily attained, which, in our setting, corresponds to the fact that an exact corrected score may not exist.

Now, for a given L^2 function δ, let $\bar{\delta}(\alpha)$ denote the minimizer of $\mathcal{L}(\bar{\delta}; A_i, \delta, \alpha) = \|A_i\bar{\delta} - \delta\|^2 + \alpha\|\bar{\delta}\|^2$. We claim that $\lim_{\alpha\to 0} \|\bar{\delta}(\alpha) - \delta\|=0$. The proof is simple, and is implicit in Kress (1989, Chapter 16), but we give it for completeness. We have

$$\|A_i\bar{\delta}(\alpha) - \delta\|^2 \leq \|A_i\bar{\delta}(\alpha) - \delta\|^2 + \alpha\|\bar{\delta}(\alpha)\|^2 \leq \|A_ig - \delta\|^2 + \alpha\|g\|^2$$

for any $g \in L^2$. Letting $\alpha \to 0$, we get $\lim_{\alpha\to 0} \|A_i\bar{\delta}(\alpha) - \delta\|^2 \leq \|A_ig - \delta\|^2$. Since g was arbitrary, we get $\lim_{\alpha\to 0} \|A_i\bar{\delta}(\alpha) - \delta\|^2 \leq \inf_g \|A_ig - \delta\|^2$, but the infimum on the right side, as we just saw, is equal to zero. We have thus proved the claim.

In the context of our RECS estimator, we obtain the following result: defining $r_{ij}(x, \boldsymbol{\theta}, \alpha) = A_i\bar{u}_{ij}(x, \boldsymbol{\theta}, \alpha) - u_{ij}(x, \boldsymbol{\theta})$, we have $\|r_{ij}(\cdot, \boldsymbol{\theta}, \alpha)\| \to 0$ as $\alpha \to 0$. At this point, we have this convergence only at a fixed value of $\boldsymbol{\theta}$. However, since $\boldsymbol{\Theta}$ is assumed compact and $u_{ij}(x, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, pointwise convergence implies uniform convergence. This yields the desired result.

Let us now turn to Part (b). Define $\Delta_{ijs}(x) = u_{ijs}(x, \boldsymbol{\theta}) - A_iu_{ijs}(x, \boldsymbol{\theta})$ and $\bar{\Delta}_{ijs}(w, \boldsymbol{\theta}, \alpha) = \bar{u}_{ijs}(w, \boldsymbol{\theta}, \alpha) - u_{ijs}(w, \boldsymbol{\theta})$. We see from (4) that the solution to $\mathcal{L}(\bar{\delta}; A_i, \delta, \alpha)$ is a linear function of δ. It follows that, just as the function $\bar{\Delta}_{ij}(\cdot, \boldsymbol{\theta}, \alpha)$ is the minimizer of $\|A_i\bar{\delta} - \Delta_{ij}\|^2 + \alpha\|\bar{\delta}\|^2$, so, too, the function $\bar{\Delta}_{ijs}(\cdot, \boldsymbol{\theta}, \alpha)$ is the minimizer of $\|A_i\bar{\delta} - \Delta_{ijs}\|^2 + \alpha\|\bar{\delta}\|^2$. We can therefore apply the arguments just used in the proof of Part (a) to prove Part (b).

Given the results in Parts (a) and (b) of the theorem, the results in Parts (c), (d), and (f) follow from standard estimating equations theory, as in, for example, Huber (1967),

White (1982), and van der Vaart (1998, Ch. 5).

Finally, we turn to Part (e). From the development in Part (a), we have

$$\bar{\mathbf{u}}_E(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \alpha) = E[\bar{\mathbf{u}}_i(W_i, \tilde{\boldsymbol{\theta}}^{(\alpha)})] = E[\mathbf{u}_i(X_i, \tilde{\boldsymbol{\theta}}^{(\alpha)})] + E[\mathbf{r}_i(X_i, \tilde{\boldsymbol{\theta}}^{(\alpha)}, \alpha)].$$

The left side of the above equation, by definition, is zero. Expanding the first term of the right side in a first-order Taylor series around $\boldsymbol{\theta}_0$, we get

$$-\mathbf{D}_E(\boldsymbol{\theta}^{\#})(\tilde{\boldsymbol{\theta}}^{(\alpha)} - \boldsymbol{\theta}_0) + E[\mathbf{r}_i(X_i, \tilde{\boldsymbol{\theta}}^{(\alpha)}, \alpha)] = 0.$$

where $\mathbf{D}_E(\boldsymbol{\theta})$ is as defined in Assumption A3 $\boldsymbol{\theta}^{\#}$ is some value between $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}^{(\alpha)}$. Part (a) implies that the second term on the right side tends to zero as $\alpha \to 0$. Given this, along with the nonsingularity condition A3, we obtain the desired conclusion $\tilde{\boldsymbol{\theta}}^{(\alpha)} \to \boldsymbol{\theta}_0$ as $\alpha \to 0$.

Appendix 3

Correcting the Covariance for Estimation of $\mathbf{A}^{(i)}$

We describe here how to correct the covariance of $\hat{\boldsymbol{\theta}}$ for estimation of $\mathbf{A}^{(i)}$. In the development below, we generally suppress the dependence of various quantities on the penalty parameter α.

The parameter $\boldsymbol{\xi}$ entering into $a_i(x, w, \boldsymbol{\xi})$ is estimated on the basis of an external or internal validation sample, or a replicate measures study, of size denoted by m. Let $\boldsymbol{\xi}_0$ denote the true value of $\boldsymbol{\xi}$. We assume that the estimator $\hat{\boldsymbol{\xi}}$ has an approximate normal distribution with mean $\boldsymbol{\xi}_0$ and covariance matrix $m^{-1}\boldsymbol{\Gamma}$, along with an estimator of the matrix $\boldsymbol{\Gamma}$. This setup is a typical one in practice. For the asymptotics we assume that m and n are of the same order of magnitude, i.e., $m/n \to \zeta$ for some constant ζ as $n \to \infty$. Otherwise the error in $\hat{\boldsymbol{\xi}}$ will either be dominated by or will dominate the error

in $\boldsymbol{\theta}$ due to the variation in the outcome data. Typically ζ will be between 0 and 1. The asymptotic covariance matrix of $\boldsymbol{\xi}$ may then be expressed as $n^{-1}\zeta^{-1}\boldsymbol{\Gamma}$. To emphasize the dependence of the corrected score on $\boldsymbol{\xi}$, we write $\bar{\mathbf{U}}(\boldsymbol{\theta}, \boldsymbol{\xi})$. The estimated asymptotic covariance matrix of $\bar{\mathbf{U}}(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)$ is $n^{-1}\mathbf{F}(\hat{\boldsymbol{\theta}})$ where $\mathbf{F}(\boldsymbol{\theta})$ is as defined in (9). We denote the asymptotic covariance between $\bar{\mathbf{U}}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\xi}}$ by $n^{-1}\boldsymbol{\Upsilon}$. The form of $\boldsymbol{\Upsilon}$ depends on the type of data used to estimate $\boldsymbol{\xi}$, and will be discussed shortly.

Let $\bar{\mathbf{U}}'(\boldsymbol{\theta}, \boldsymbol{\xi})$ denote the matrix whose (r, ν) element is the partial derivative of $\bar{U}_r(\boldsymbol{\theta}, \boldsymbol{\xi})$ with respect to ξ_ν. By Taylor expansion, we have

$$\mathbf{0} = \bar{\mathbf{U}}(\hat{\boldsymbol{\theta}}^{(\alpha)}, \hat{\boldsymbol{\xi}}) = \bar{\mathbf{U}}(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)\bar{\mathbf{D}}(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)(\hat{\boldsymbol{\theta}}^{(\alpha)} - \tilde{\boldsymbol{\theta}}^{(\alpha)}) + \bar{\mathbf{U}}'(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) + o_p(1/n),$$

leading to

$$-\sqrt{n}(\hat{\boldsymbol{\theta}}^{(\alpha)} - \tilde{\boldsymbol{\theta}}^{(\alpha)}) = \bar{\mathbf{D}}(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)^{-1}\left[\{\sqrt{n}\,\bar{\mathbf{U}}(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)\} + \bar{\mathbf{U}}'(\tilde{\boldsymbol{\theta}}^{(\alpha)}, \boldsymbol{\xi}_0)\{\sqrt{n}\,(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)\}\right] + o_p(1/n).$$

Accordingly, the estimated asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}^{(\alpha)})$ is $\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}})$, where $\hat{\mathbf{V}}(\boldsymbol{\theta}, \boldsymbol{\xi})$ is now defined as

$$\hat{\mathbf{V}}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \bar{\mathbf{D}}(\boldsymbol{\theta}, \boldsymbol{\xi})^{-1}\left[\mathbf{F}(\boldsymbol{\theta}) + \zeta^{-1}\bar{\mathbf{U}}'(\boldsymbol{\theta}, \boldsymbol{\xi})\hat{\boldsymbol{\Gamma}}\bar{\mathbf{U}}'(\boldsymbol{\theta}, \boldsymbol{\xi})^T + \hat{\boldsymbol{\Upsilon}}\bar{\mathbf{U}}'(\boldsymbol{\theta}, \boldsymbol{\xi})^T\right]\bar{\mathbf{D}}(\boldsymbol{\theta}, \boldsymbol{\xi})^{-1}.$$

If $\boldsymbol{\xi}$ is estimated from an external validation or replicate measures study, then $\hat{\boldsymbol{\xi}}$ is obviously independent of the corrected score function $\bar{\mathbf{U}}(\boldsymbol{\theta})$, and thus $\boldsymbol{\Upsilon} = \mathbf{0}$. When $\boldsymbol{\xi}$ is estimated from an internal validation or replicate measures study, $\boldsymbol{\Upsilon}$ is nonzero and must be estimated. We consider the setting where the validation/replicate data are i.i.d. across individuals, and $\boldsymbol{\xi}$ is estimated by maximum likelihood. Let $g_i(\boldsymbol{\xi})$ denote the log likelihood function for the validation/replicate data on individual i. The overall normalized log likelihood for $\boldsymbol{\xi}$ is then $g(\boldsymbol{\xi}) = m^{-1}\sum_{i\in\mathcal{R}} g_i(\boldsymbol{\xi})$, where $\mathcal{R}$ denotes the set of individuals in the internal validation/replicate sample. Let $\mathbf{g}'(\boldsymbol{\xi})$ and $\mathbf{g}''(\boldsymbol{\xi})$ denote the gradient vector and Hessian matrix, respectively, of $g(\boldsymbol{\xi})$, and let $\mathbf{g}_i'(\boldsymbol{\xi})$ denote the

gradient of $g_i(\boldsymbol{\xi})$. We can then express $\hat{\boldsymbol{\xi}}$ in terms of the classic asymptotic approximation $\hat{\boldsymbol{\xi}} \doteq -\mathbf{g}''(\boldsymbol{\xi}_0)^{-1}\mathbf{g}'(\boldsymbol{\xi}_0)$. Define $\boldsymbol{\Omega} = \mathrm{Cov}(\bar{\mathbf{u}}_i(\boldsymbol{\theta}_0, \boldsymbol{\xi}_0), \mathbf{g}_i'(\boldsymbol{\xi}_0))$. The matrix $\boldsymbol{\Omega}$ can be estimated empirically by

$$\hat{\boldsymbol{\Omega}} = \frac{1}{m}\sum_{i\in\mathcal{R}} \bar{\mathbf{u}}_i(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}})\mathbf{g}_i'(\hat{\boldsymbol{\xi}})^T. \tag{11}$$

We then estimate $\boldsymbol{\Upsilon}$ by $\hat{\boldsymbol{\Upsilon}} = \hat{\boldsymbol{\Omega}}\mathbf{g}''(\hat{\boldsymbol{\xi}})^{-1}$. In addition, in the present setup we have $\hat{\boldsymbol{\Gamma}} = \mathbf{g}''(\hat{\boldsymbol{\xi}})^{-1}$.

In principle, expressions can be worked out for the partial derivatives that make up the matrix $\bar{\mathbf{U}}'(\boldsymbol{\theta}, \boldsymbol{\xi})$, but the algebra is cumbersome. Therefore, in our practical implementation, we use numerical partial derivatives.

Table 1

Simulation Study A
$X \sim N(0,1)$ and $\epsilon \sim N(0, \gamma_1 + \gamma_2|X| + \gamma_3 Y)$
Sample Size n=200, Validation Sample Size m=70

	β_0					β_1					
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI	F
				$\gamma_1 = 0.5$	$\gamma_2 = 0$	$\gamma_3 = 0$					
Naive	0.9526	0.9459	0.1737	0.1695	0.93	0.6469	0.6379	0.1512	0.1520	0.36	0
RECS	1.0323	1.0217	0.2122	0.1983	0.96	1.0782	1.0314	0.3373	0.3013	0.96	0
N&S	1.0219	1.0110	0.2002	0.1905	-	1.0477	1.0330	0.2685	0.2876	-	0
SIMEX - Q	0.9942	0.9840	0.1890	0.1890	-	0.8940	0.8737	0.2245	0.2157	-	0
SIMEX - L	0.9718	0.9620	0.1800	0.1764	-	0.7619	0.7501	0.1803	0.1794	-	0
SIMEX - NL	1.0548	1.0252	0.3080	0.2261	-	1.0681	1.0161	0.3312	0.3032	-	61
H&W	0.9470	0.9752	0.6279	0.2387	-	1.0600	0.9514	0.5264	0.3818	-	4
				$\gamma_1 = 1$	$\gamma_2 = 0$	$\gamma_3 = 0$					
Naive	0.9231	0.9153	0.1694	0.1665	0.91	0.4713	0.4660	0.1262	0.1264	0.03	0
RECS	1.0613	1.0346	0.2683	0.2283	0.96	1.1157	1.0231	0.5421	0.4266	0.93	2
N&S	0.9944	0.9849	0.1962	0.1950	-	0.9499	0.9759	0.2263	0.2380	-	0
SIMEX - Q	0.9645	0.9515	0.1855	0.1809	-	0.7182	0.6989	0.2048	0.2039	-	0
SIMEX - L	0.9390	0.9270	0.1748	0.1675	-	0.5663	0.5595	0.1525	0.1534	-	0
SIMEX - NL	1.0564	1.0068	0.4722	0.2328	-	1.1431	0.9997	0.5593	0.4151	-	63
H&W	0.9065	0.9495	0.6558	0.2483	-	0.8840	0.7851	0.5249	0.3751	-	4
				$\gamma_1 = 0.3$	$\gamma_2 = 0.25$	$\gamma_3 = 0$					
Naive	0.9522	0.9446	0.1734	0.1728	0.93	0.6573	0.6494	0.1553	0.1573	0.39	0
RECS	1.0233	1.0121	0.2057	0.1953	0.96	1.0429	1.0054	0.3102	0.2784	0.95	0
N&S	1.0148	1.0058	0.1966	0.1942	-	1.1025	1.0868	0.2914	0.3217	-	0
SIMEX - Q	0.9933	0.9825	0.1884	0.1913	-	0.9169	0.8926	0.2368	0.2313	-	0
SIMEX - L	0.9713	0.9615	0.1796	0.1779	-	0.7750	0.7620	0.1857	0.1868	-	0
SIMEX - NL	1.0427	1.0199	0.2912	0.2209	-	1.1265	1.0607	0.3867	0.3321	-	74
H&W	0.9408	0.9680	0.5539	0.2150	-	1.0370	0.9679	0.4451	0.2980	-	1
				$\gamma_1 = 0.7$	$\gamma_2 = 0.35$	$\gamma_3 = 0$					
Naive	0.9238	0.9145	0.1695	0.1664	0.91	0.4812	0.4749	0.1284	0.1279	0.04	0
RECS	1.0363	1.0161	0.2423	0.2153	0.96	1.0658	0.9797	0.4673	0.3663	0.91	1
N&S	0.9854	0.9719	0.1940	0.1920	-	0.9924	1.0177	0.2351	0.2424	-	0
SIMEX - Q	0.9650	0.9536	0.1855	0.1861	-	0.7358	0.7167	0.2115	0.2053	-	0
SIMEX - L	0.9397	0.9285	0.1748	0.1668	-	0.5780	0.5685	0.1552	0.1542	-	0
SIMEX - NL	1.0484	1.0065	0.4375	0.2305	-	1.2113	1.0383	0.6183	0.4470	-	63
H&W	0.9388	0.9663	0.6524	0.2431	-	1.0310	0.9134	0.5109	0.3566	-	4
				$\gamma_1 = 0.15$	$\gamma_2 = 0.25$	$\gamma_3 = 0.25$					
Naive	0.9524	0.9433	0.1715	0.1717	0.94	0.6512	0.6437	0.1513	0.1488	0.37	0
RECS	1.0274	1.0114	0.2066	0.1964	0.96	1.0782	1.0316	0.3477	0.2921	0.95	0
N&S	1.0107	0.9963	0.1925	0.1890	-	1.0065	0.9783	0.2847	0.2847	-	0
SIMEX - Q	0.9891	0.9783	0.1838	0.1838	-	0.8718	0.8545	0.2241	0.2194	-	0
SIMEX - L	0.9704	0.9600	0.1769	0.1794	-	0.7587	0.7491	0.1799	0.1735	-	0
SIMEX - NL	1.0223	1.0131	0.2497	0.2053	-	1.0206	0.9698	0.3251	0.2758	-	102
H&W	1.0370	1.0350	0.4937	0.2231	-	1.0410	0.9682	0.4373	0.3121	-	2
				$\gamma_1 = 0.35$	$\gamma_2 = 0.25$	$\gamma_3 = 0.50$					
Naive	0.9292	0.9207	0.1676	0.1657	0.91	0.5091	0.5029	0.1282	0.1285	0.06	0
RECS	1.0439	1.0182	0.2389	0.2105	0.96	1.1123	0.9989	0.4765	0.3793	0.93	0
N&S	0.9972	0.9847	0.1913	0.1868	-	0.9089	0.9034	0.2590	0.2906	-	0
SIMEX - Q	0.9662	0.9564	0.1797	0.1757	-	0.7259	0.7125	0.1986	0.1868	-	0
SIMEX - L	0.9450	0.9362	0.1721	0.1720	-	0.6028	0.5954	0.1534	0.1520	-	0
SIMEX - NL	1.0190	0.9887	0.3647	0.2187	-	0.9651	0.8793	0.4155	0.2943	-	67
H&W	1.1080	1.0940	0.6033	0.2595	-	1.0600	0.9600	0.5004	0.3425	-	3

N&S - Novick and Stefanski (2002); H&W - Huang and Wang (2001)
SIMEX results are based on $B = 100$, $\lambda = (0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9)$ and the simex R library
L - linear extrapolation; Q - quadratic extrapolation; NL - non-linear extrapolation of Cook and Stefanski (1994)
M - empirical mean; MD - empirical median; Emp-SD - empirical standard deviation; IQ-SD - inter-quartile dispersion;
95% CI - empirical coverage rate of 95% Wald confidence interval; F - number of samples with no solution

Table 2

Simulation Study B
Non-Normal Measurement Error with Normal or Non-Normal True Covariate
Sample Size n=200, Validation Sample Size m=70

	β_0					β_1					
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI	F
	$X \sim N(0,1)$; $\epsilon \sim$ Double exponential with Var$(\epsilon) = 0.5$										
Naive	0.9467	0.9350	0.1775	0.1804	0.93	0.6378	0.6287	0.1609	0.1517	0.33	0
RECS	1.0288	1.0048	0.2192	0.1989	0.95	1.0684	0.9986	0.3611	0.3264	0.95	0
N&S	1.0138	0.9983	0.2027	0.1964	-	1.0553	1.0290	0.2964	0.3195	-	0
SIMEX - Q	0.9884	0.9763	0.1931	0.1883	-	0.8905	0.8641	0.2438	0.2276	-	0
SIMEX - L	0.9654	0.9529	0.1836	0.1772	-	0.7519	0.7381	0.1924	0.1831	-	0
SIMEX - NL	1.0471	1.0189	0.2754	0.2135	-	1.1111	1.0256	0.4207	0.3336	-	68
H&W	0.9980	0.9919	0.4298	0.2157	-	1.0770	0.9848	0.4523	0.3202	-	0
	$X \sim N(0,1)$; $\epsilon \sim$ Double exponential with Var$(\epsilon) = 1$										
Naive	0.9174	0.9078	0.1731	0.1787	0.91	0.4659	0.4598	0.1370	0.1349	0.04	0
RECS	1.0545	1.0167	0.2735	0.2291	0.96	1.1362	1.0089	0.5714	0.4335	0.93	6
N&S	0.9746	0.9617	0.1965	0.1898	-	0.9647	1.0078	0.2724	0.2683	-	0
SIMEX - Q	0.9583	0.9474	0.1891	0.1861	-	0.7205	0.6970	0.2289	0.2165	-	0
SIMEX - L	0.9327	0.9228	0.1782	0.1772	-	0.5599	0.5514	0.1659	0.1601	-	0
SIMEX - NL	1.0032	0.9982	0.5742	0.2513	-	1.2130	1.0485	0.7743	0.4804	-	106
H&W	0.9353	0.9704	0.6913	0.2543	-	0.9886	0.8985	0.4636	0.3336	-	1
	$X \sim N(0,1)$; $\epsilon \sim$ Modified χ^2_1 with Var$(\epsilon) = 0.5$										
Naive	0.9732	0.9745	0.1785	0.1692	0.93	0.6822	0.6793	0.1706	0.1672	0.47	0
RECS	1.0335	1.0295	0.2091	0.1929	0.95	1.0555	1.0348	0.2981	0.2775	0.96	0
N&S	1.0920	1.0855	0.2245	0.2142	-	1.1800	1.1821	0.3121	0.3269	-	0
SIMEX - Q	1.0340	1.0277	0.2044	0.1846	-	0.9678	0.9500	0.2624	0.2617	-	0
SIMEX - L	0.9980	0.9973	0.1878	0.1727	-	0.8080	0.8028	0.2052	0.2024	-	0
SIMEX - NL	1.1297	1.0882	0.3964	0.2483	-	1.2028	1.1617	0.3912	0.3573	-	45
H&W	0.7934	0.8692	0.5560	0.1964	-	1.0260	0.9809	0.3822	0.2861	-	2
	$X \sim N(0,1)$; $\epsilon \sim$ Modified χ^2_1 with Var$(\epsilon) = 1$										
Naive	0.9510	0.9502	0.1765	0.1667	0.93	0.5196	0.5146	0.1547	0.1552	0.13	0
RECS	1.0519	1.0315	0.2444	0.2116	0.96	1.0869	1.0345	0.3970	0.3546	0.97	0
N&S	1.1200	1.1172	0.2150	0.2039	-	1.1779	1.2122	0.2307	0.1927	-	0
SIMEX - Q	1.0262	1.0140	0.2135	0.1890	-	0.8320	0.8234	0.2711	0.2765	-	0
SIMEX - L	0.9734	0.9703	0.1857	0.1690	-	0.6269	0.6187	0.1880	0.1853	-	0
SIMEX - NL	1.1226	1.0980	1.0702	0.4804	-	1.6444	1.4262	0.9090	0.7413	-	59
H&W	0.5574	0.7494	0.8302	0.2506	-	0.9640	0.8949	0.4753	0.3069	-	3
	$X \sim$ Modified χ^2_1 with Var$(X) = 1$; $\epsilon \sim$ Modified χ^2_1 with Var$(\epsilon) = 0.5$										
Naive	0.9059	0.9015	0.1689	0.1632	0.90	0.5461	0.5348	0.1709	0.1662	0.25	0
RECS	1.0862	1.0356	0.2953	0.2626	0.97	1.1279	1.0479	0.4989	0.4414	0.94	0
N&S	1.2710	1.2686	0.3225	0.3558	-	1.3957	1.5027	0.4630	0.5263	-	0
SIMEX - Q	0.9703	0.9592	0.1968	0.1913	-	0.8175	0.7884	0.2870	0.2713	-	0
SIMEX - L	0.9253	0.9166	0.1750	0.1690	-	0.6492	0.6354	0.2056	0.1979	-	0
SIMEX - NL	1.0547	1.0083	0.9484	0.3625	-	1.2668	1.1121	0.6949	0.5137	-	43
H&W	0.6861	0.8618	0.8291	0.2053	-	0.9610	0.8409	0.5507	0.3729	-	6
	$X \sim$ Modified χ^2_1 with Var$(X) = 1$; $\epsilon \sim$ Modified χ^2_1 with Var$(\epsilon) = 1$										
Naive	0.8784	0.8740	0.1661	0.1604	0.87	0.3917	0.3824	0.1448	0.1402	0.03	0
RECS	1.1134	1.0370	0.3589	0.2897	0.97	1.1272	1.0319	0.5784	0.5316	0.93	0
N&S	1.1856	1.1649	0.2699	0.2898	-	1.1392	1.2540	0.3594	0.3506	-	0
SIMEX - Q	0.9337	0.9243	0.1926	0.1794	-	0.6396	0.6176	0.2602	0.2498	-	0
SIMEX - L	0.8923	0.8876	0.1709	0.1616	-	0.4720	0.4604	0.1749	0.1683	-	0
SIMEX - NL	0.7096	0.8059	0.8610	0.4040	-	1.4470	1.1876	1.0740	0.8718	-	146
H&W	0.6004	0.7683	0.7942	0.2090	-	0.8084	0.6995	0.5894	0.4255	-	13

N&S - Novick and Stefanski (2002); H&W - Huang and Wang (2001)
SIMEX results are based on $B = 100$, $\lambda = (0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9)$ and the simex R library
L - linear extrapolation; Q - quadratic extrapolation; NL - non-linear extrapolation of Cook and Stefanski (1994)
M - empirical mean; MD - empirical median; Emp-SD - empirical standard deviation; IQ-SD - inter-quartile dispersion;
95% CI - empirical coverage rate of 95% Wald confidence interval; F - number of samples with no solution

Table 3

Simulation Study C
Performance of RECS Under Misspecified Error Model
Assumed Error Model: $\epsilon \sim N(0, \sigma^2)$
Results for Estimate of β_1

		Naive Estimate				RECS Estimate			
Distn of X	True Distn of ϵ	Mean	Median	Emp-SE	IQ-SE	Mean	Median	Emp-SE	IQ-SE
Normal	Skewed Normal	0.6621	0.6569	0.1655	0.1668	1.1885	1.1028	0.4790	0.3466
Normal	MODCHI	0.6821	0.6791	0.1706	0.1672	1.2778	1.1592	0.5621	0.4267
Skewed Normal	Skewed Normal	0.5992	0.5895	0.1620	0.1589	1.1983	1.0923	0.5506	0.4355
Skewed Nomal	MODCHI	0.6058	0.6043	0.1669	0.1698	1.3466	1.1862	0.6752	0.5526
MODCHI	Skewed Normal	0.5250	0.5209	0.1573	0.1532	1.3280	1.1335	0.7406	0.5932
MODCHI	MODCHI	0.5460	0.5347	0.1710	0.1661	1.6553	1.3567	1.0210	0.9698

Table 4

Simulation Study D
Two Error-Prone Covariates and One Error-Free Covariate
Scenario D1: $X_1, X_2, Z \sim N(0, 1)$ and $\epsilon_1, \epsilon_2 \sim N(0, 1)$
Sample Size n=500

	β_1					β_2				
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI
Naive	0.4296	0.4242	0.0804	0.0781	0.000	0.4282	0.4235	0.0853	0.0857	0.000
RECS	1.0817	1.0341	0.3116	0.2595	0.972	1.0784	1.0305	0.3333	0.2764	0.956
N&S	0.8292	0.8299	0.1308	0.1299	-	0.8256	0.8280	0.1436	0.1434	-
SIMEX - Q	0.8113	0.8038	0.1377	0.1333	-	0.8096	0.8006	0.1450	0.1378	-
SIMEX - L	0.5235	0.5189	0.0917	0.0911	-	0.5219	0.5150	0.0969	0.0953	-
SIMEX - NL	-2.6768	-2.5286	1.1003	1.0482	-	-2.5988	-2.4974	1.1896	1.0356	-
H&W	0.6678	0.6675	0.5038	0.3360	-	0.6975	0.6744	0.5318	0.3407	-

	β_0					β_3					
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI	F
Naive	0.8587	0.8575	0.1163	0.1129	0.734	0.8626	0.8599	0.1201	0.1223	0.769	0
RECS	1.0500	1.0294	0.2112	0.1791	0.972	1.0525	1.0286	0.2079	0.1907	0.979	0
N&S	0.9435	0.9415	0.1371	0.1359	-	0.9476	0.9444	0.1384	0.1403	-	0
SIMEX - Q	0.9553	0.9502	0.1433	0.1394	-	0.9587	0.9524	0.1452	0.1519	-	0
SIMEX - L	0.8821	0.8803	0.1215	0.1214	-	0.8857	0.8819	0.1247	0.1278	-	0
SIMEX - NL	0.3396	0.3544	1.1798	0.5785	-	0.3642	0.4221	1.1742	0.5205	-	285
H&W	0.7535	0.8516	0.7885	0.2376	-	1.0277	0.9679	0.4669	0.3039	-	14

N&S - Novick and Stefanski (2002); H&W - Huang and Wang (2001)
SIMEX results are based on $B = 100$, $\lambda = (0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9)$ and the simex R library
L - linear extrapolation; Q - quadratic extrapolation; NL - non-linear extrapolation of Cook and Stefanski (1994)
M - empirical mean; MD - empirical median; Emp-SD - empirical standard deviation; IQ-SD - inter-quartile dispersion;
95% CI - empirical coverage rate of 95% Wald confidence interval; F - number of samples with no solution

Table 5

Simulation Study D
Two Error-Prone Covariates and One Error-Free Covariate
Scenario D2: $X_1, X_2, Z \sim N(0,1)$ and $\epsilon_1, \epsilon_2 \sim N(0, 0.4 + 0.25(|X_1| + |X_2| + |Z|))$
Sample Size n=500

	β_1					β_2				
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI
Naive	0.4386	0.4334	0.0824	0.0792	0.001	0.4364	0.4313	0.0868	0.0857	0.000
RECS	1.1884	1.1122	0.3800	0.3126	0.986	1.1822	1.1151	0.4044	0.3313	0.977
N&S	0.8588	0.8602	0.1339	0.1339	-	0.8536	0.8582	0.1472	0.1455	-
SIMEX - Q	0.8345	0.8242	0.1431	0.1394	-	0.8315	0.8231	0.1497	0.1439	-
SIMEX - L	0.5339	0.5265	0.0940	0.0903	-	0.5314	0.5257	0.0985	0.0973	-
SIMEX - NL	-2.3990	-2.2229	0.9623	0.9407	-	-2.3830	-2.1980	1.0071	0.9666	-
H&W	0.6165	0.6176	0.4809	0.3277	-	0.6455	0.6072	0.4809	0.3271	-

	β_0					β_3					
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI	F
Naive	0.8593	0.8586	0.1162	0.1153	0.740	0.8712	0.8694	0.1223	0.1223	0.791	0
RECS	1.0787	1.0475	0.2359	0.1896	0.981	1.0828	1.0482	0.2405	0.2007	0.983	1
N&S	0.9335	0.9310	0.1358	0.1391	-	0.9674	0.9648	0.1439	0.1472	-	0
SIMEX - Q	0.9569	0.9531	0.1434	0.1394	-	0.9793	0.9746	0.1516	0.1562	-	0
SIMEX - L	0.8827	0.8824	0.1214	0.1219	-	0.8961	0.8921	0.1276	0.1314	-	0
SIMEX - NL	0.3789	0.3653	1.1233	0.5524	-	0.3096	0.3920	0.9646	0.4150	-	185
H&W	0.7637	0.8325	0.9796	0.2280	-	0.9627	0.9279	0.4248	0.2837	-	9

N&S - Novick and Stefanski (2002); H&W - Huang and Wang (2001)
SIMEX results are based on $B = 100$, $\lambda = (0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9)$ and the simex R library
L - linear extrapolation; Q - quadratic extrapolation; NL - non-linear extrapolation of Cook and Stefanski (1994)
M - empirical mean; MD - empirical median; Emp-SD - empirical standard deviation; IQ-SD - inter-quartile dispersion;
95% CI - empirical coverage rate of 95% Wald confidence interval; F - number of samples with no solution

Table 6

Simulation Study D
Two Error-Prone Covariates and One Error-Free Covariate
Scenario D3: $X_1, X_2 \sim$ Modified χ^2_1, $\text{Var}(X_1) = \text{Var}(X_2) = 1$ $Z \sim N(0,1)$
and $\epsilon_1, \epsilon_2 \sim$ Modified χ^2_1, $\text{Var}(\epsilon_1) = \text{Var}(\epsilon_2) = 1$
Sample Size n=500

	β_1					β_2				
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI
Naive	0.3853	0.3814	0.0887	0.0915	0.000	0.3830	0.3811	0.0843	0.0856	0.000
RECS	1.0430	0.9841	0.3725	0.3328	0.942	1.0322	0.9850	0.3319	0.3016	0.946
N&S	0.8668	0.8692	0.2603	0.2901	-	0.8662	0.8805	0.2536	0.2885	-
SIMEX - Q	0.7463	0.7315	0.1586	0.1598	-	0.7431	0.7317	0.1493	0.1488	-
SIMEX - L	0.4672	0.4624	0.1009	0.1021	-	0.4647	0.4602	0.0958	0.0989	-
SIMEX - NL	-1.5004	-1.2992	0.7297	0.5271	-	-1.5046	-1.3341	0.6896	0.5078	-
H&W	0.4723	0.4276	0.2733	0.2108	-	0.4625	0.4216	0.2849	0.1992	-

	β_0					β_3					
	M	MD	Emp-SD	IQ-SD	95% CI	M	MD	Emp-SD	IQ-SD	95% CI	F
Naive	0.8007	0.7971	0.1144	0.1113	0.548	0.8883	0.8893	0.1206	0.1207	0.809	0
RECS	1.0595	1.0225	0.2447	0.2084	0.963	1.0413	1.0232	0.1856	0.1667	0.971	0
N&S	1.0419	1.0414	0.1563	0.1547	-	0.9247	0.9198	0.1399	0.1417	-	0
SIMEX - Q	0.9049	0.8978	0.1456	0.1383	-	0.9626	0.9588	0.1402	0.1399	-	0
SIMEX - L	0.8199	0.8146	0.1191	0.1133	-	0.9059	0.9073	0.1243	0.1244	-	0
SIMEX - NL	0.5453	0.5554	0.1777	0.1286	-	0.4842	0.5551	0.9577	0.3875	-	60
H&W	0.3332	0.4979	0.6451	0.2517	-	0.9155	0.9023	0.1798	0.1638	-	1

N&S - Novick and Stefanski (2002); H&W - Huang and Wang (2001)
SIMEX results are based on $B = 100$, $\lambda = (0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9)$ and the simex R library
L - linear extrapolation; Q - quadratic extrapolation; NL - non-linear extrapolation of Cook and Stefanski (1994)
M - empirical mean; MD - empirical median; Emp-SD - empirical standard deviation; IQ-SD - inter-quartile dispersion;
95% CI - empirical coverage rate of 95% Wald confidence interval; F - number of samples with no solution

Table 7

NHS Results

	β_0		β_1	
Method	Estimate	SE	Estimate	SE
Naive1	-1.7213	0.1436	-0.1802	0.0680
Naive2	-1.9425	0.0862	-0.1019	0.0384
RECS-MEM1	-1.7496	0.1328	-0.2733	0.1141
RECS-MEM2	-1.6186	0.1693	-0.3970	0.1758
N&S	-1.7501	0.1297	-0.2726	0.1270
SIMEX-L	-1.9186	0.0883	-0.1223	0.0463
SIMEX-Q	-1.8722	0.0945	-0.1630	0.0603
SIMEX-NL	-1.7256	0.1404	-0.2966	0.1408