# Latent-model robustness in structural measurement error models

By XIANZHENG HUANG, LEONARD A. STEFANSKI
AND MARIE DAVIDIAN

*Department of Statistics, North Carolina State University, Raleigh,*
*North Carolina 27695-8203, U.S.A.*

xhuang@ncsu.edu   stefanski@ncsu.edu   davidian@ncsu.edu

## SUMMARY

We present methods for diagnosing the effects of model misspecification of the true-predictor distribution in structural measurement error models. We first formulate latent-model robustness theoretically. Then we provide practical techniques for examining the adequacy of an assumed latent predictor model. The methods are illustrated via analytical examples, application to simulated data and with data from a study of coronary heart disease.

*Some key words*: Bias; Latent variable; Measurement error; Remeasurement method; Simulation extrapolation; Structural modelling.

## 1. INTRODUCTION

In a structural measurement error model (Carroll et al., 1995, § 1·2) the true predictor is a latent variable measured with error. Parametric modelling requires specification of a distribution for the unobserved latent variable, and issues of model misspecification naturally arise. We propose a framework for studying robustness to misspecification of the latent-variable distribution. We call this latent-model robustness and argue that latent-model robustness means lack of asymptotic bias in estimators for the parameters of interest. Model fitting with data measured with error is subject to the same pitfalls as model fitting without measurement error, and other aspects of robustness also come into play. However, in this paper we focus on misspecification of the latent model and refer to latent-model robustness to distinguish it from other aspects of robustness. Robustness to asymptotic bias does not imply asymptotic efficiency. Two models can be equally robust to bias, yet one might be more efficient. Our methods are useful for screening out those models that are nonrobust. Efficiency will generally be attained by the most parsimonious robust model.

## 2. STRUCTURAL MEASUREMENT ERROR MODEL

We consider the classical measurement error model. Let $Y$ be the response, let $X_{q \times 1}$ and $W_{q \times 1}$ be the true and observed predictor, respectively, and let $W = X + U$, where $U_{q \times 1}$ is a multivariate normal $N(0, \sigma_U^2)$ random vector with $q \times q$ covariance matrix $\sigma_U^2$. For now we assume that $\sigma_U^2$ is known. We assume that the conditional density of $Y$ given $X = x$ is $f_{Y|X}(y|x; \theta)$, known up to $\theta$, and that the joint density of $(Y, W)$ given $X = x$

is $f_{Y,W|X}(y, w|x; \theta) = f_{Y|X}(y|x; \theta) f_{W|X}(w|x; \sigma_U^2)$, where $f_{W|X}(w|x; \sigma_U^2)$ is the $N(x, \sigma_U^2)$ density. Inference about $\theta$ is of central interest.

Let $\{Y_j, W_j\}_{j=1}^n$ denote independent realisations from the measurement error model. Two ways of viewing $\{X_j\}_{j=1}^n$ lead to two types of measurement error model (Carroll et al., 1995, § 1·2). In a functional model, $\{X_j\}_{j=1}^n$ are viewed as unknown parameters, and the likelihood of the observed data is $L(\theta, X_1, \ldots, X_n) = \prod_{j=1}^n f_{Y,W|X}(Y_j, W_j|X_j; \theta)$. In a structural model, $\{X_j\}_{j=1}^n$ are regarded as random variables. Under the assumption that the density of $X$ is $f_X^{(a)}(x; \tau^{(a)})$, depending on a parameter vector $\tau^{(a)}$, the likelihood is

$$L(\theta, \tau^{(a)}) = \prod_{j=1}^n f_{Y,W}(Y_j, W_j; \theta, \tau^{(a)}) = \prod_{j=1}^n \int f_{Y|X}(Y_j|x; \theta) f_{W|X}(W_j|x; \sigma_U^2) f_X^{(a)}(x; \tau^{(a)}) dx. \quad (1)$$

Functional modelling makes minimal assumptions about the set of unobserved predictors and thus is generally applicable. However, maximising $L(\theta, X_1, \ldots, X_n)$ with respect to $\theta$ and $\{X_j\}_{j=1}^n$ is often difficult and seldom results in a consistent estimator for $\theta$. Consequently, many functional-model inference methods are moment methods or conditional likelihood methods (Fuller, 1987; Carroll et al., 1995). Much recent emphasis has been on structural models and methods. Inference based on (1) is often simpler than that in functional modelling, and maximum likelihood estimation offers the attraction of asymptotic efficiency when $f_X^{(a)}(x; \tau^{(a)})$ is a correct model for $X$.

A reason often cited for avoiding parametric structural modelling is that misspecification of the model for $X$ can result in inconsistent estimators for $\theta$. With regard to the issue of robustness of inference on $\theta$ to misspecification of this model, semiparametric modelling methods (Roeder et al., 1996; Schafer, 2001) and flexible-parametric modelling methods (Carroll et al., 1999; Richardson, 2002) provide some solutions. However, it is usually more difficult to implement them and their efficiency is not practically greater than carefully chosen parametric structural methods when the latter are robust. Hence, parametric structural modelling is appealing in practice because of its simplicity and potential efficiency, provided inferences are robust to misspecification of the distributional model.

## 3. Theoretical robustness
### 3·1. *Full latent-model robustness*

Henceforth we take the true predictor $X$ and its measurement $W$ to be scalars ($q = 1$), which suffices to illustrate the proposed methods. Asymptotic bias is possible only if $\sigma_U^2 > 0$ and the latent model for $X$ is misspecified. Thus, according to our definition of robustness, nonrobustness can be regarded as an interaction between two factors, measurement error variance and the model for $X$. We first define full latent-model robustness. Consider the structural-model likelihood (1). The maximum likelihood estimators for $(\theta, \tau^{(a)})$, denoted by $(\hat{\theta}, \hat{\tau}^{(a)})$, under this assumed structural model are the values maximising (1). Denote by $\theta^*$ the true value of $\theta$ in $f_{Y|X}(y|x; \theta)$. Let

$$\psi_S(y, w, \theta, \tau^{(a)}) = \{\partial/\partial(\theta, \tau^{(a)})\} \log\{f_{Y,W}(y, w; \theta, \tau^{(a)})\}$$

and define $\theta(.)$ and $\tau^{(a)}(.)$ as functions of $\sigma_U$ implicitly via

$$E[\psi_S\{Y, W, \theta(\sigma_U), \tau^{(a)}(\sigma_U)\}] = 0. \quad (2)$$

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

$$f_{Y,W}^*(y, w; \theta) = \int f_{Y|X}(y|x; \theta) f_{W|X}(w|x; \sigma_U^2) f_X^*(x) dx,$$

where $f_X^*(x)$ is the true density of $X$. The structural model maximum likelihood estimator for $\theta$ is robust provided

$$\theta(\sigma_U) \equiv \theta^* \quad (\sigma_U \geqslant 0). \tag{3}$$

It is worth pointing out that the model for $X$ does not have to be correctly specified for robustness of $\hat{\theta}$ to obtain. For certain regression models, if the assumed model for $X$ is sufficiently flexible that the moments of the true model, on which $\theta(\sigma_U)$ depends, are estimated consistently, then full latent-model robustness is possible. Two examples given next illustrate the consequence of using models with different degrees of flexibility for $X$.

*Example* 1: *Y given X follows a normal distribution with mean* $\beta_0 + \beta_1 X$. Assume that $Y|X = x \sim N(\beta_0 + \beta_1 x, \sigma_\varepsilon^2)$, so that $\theta = (\beta_0, \beta_1, \sigma_\varepsilon)^\mathrm{T}$. If one assumes $X$ to be normal, then normality of $X$ is not necessary for consistency of $\hat{\theta}$ (Fuller, 1987, p. 17). The explanation lies in the facts that the regression coefficients are functions of the first two moments and that the population moments are consistently estimated regardless of the true distribution of $X$. The key to this positive finding is that the normal model assumed for $X$ is flexible enough to permit consistent estimation of all required moments.

We now consider a less flexible normal model. Suppose that the distribution of $X$ is assumed to be normal $N(\tau^{(a)}, \tau^{(a)})$, that is with mean equal to the variance. The functions $\theta(.)$ and $\tau^{(a)}(.)$ defined through (2) give the probability limits, as $n \to \infty$, of $\hat{\theta}$ and $\hat{\tau}^{(a)}$. If the true distribution of $X$ is not normal with mean and variance equal, then the assumed model is incorrect and is too restrictive to permit consistent estimation of the first two moments of the true distribution of $X$. This will lead to potential bias in $\hat{\theta}$, whose magnitude is expected to increase with the magnitude of $\sigma_U$.

Figures 1(a) and (b) display $\beta_0(.)$ and $\beta_1(.)$ against $\sigma_U$ for three true distributions of $X$, $N(1, 1)$, $N(0{\cdot}5, 1)$ and $N(1{\cdot}5, 1)$, when $\theta^* = (0, 1, 1)^\mathrm{T}$. In the latter two cases, the assumed
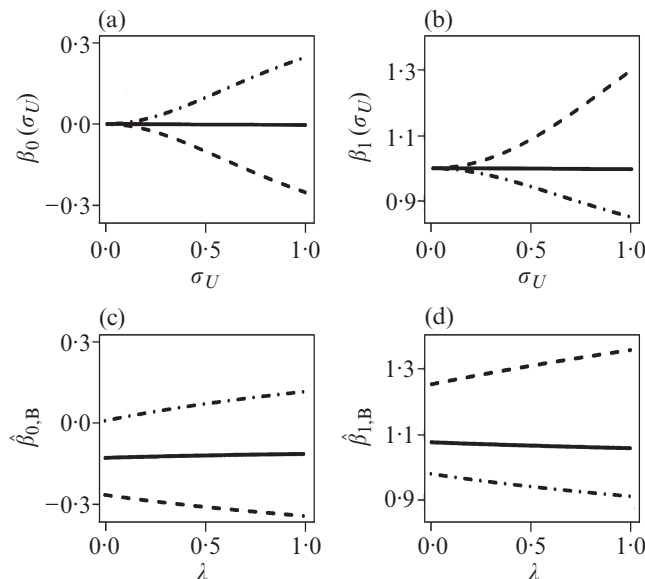


Fig. 1. (a) and (b) show $\beta_0(\sigma_U)$ and $\beta_1(\sigma_U)$, respectively, for assumed model $N(\tau^{(a)}, \tau^{(a)})$ and three true $X$ distributions, $N(1, 1)$, solid line, $N(0{\cdot}5, 1)$, dashed line, and $N(1{\cdot}5, 1)$, dashed-dotted line. (c) and (d) show single-sample remeasurement versions of (a) and (b) as described in Example 5.

model is incorrect and too restrictive compared to the true density. As shown in the plots, the misspecification and lack of flexibility in modelling $X$ result in asymptotic biases in $\hat{\beta}_0$ and $\hat{\beta}_1$ that increase in magnitude with $\sigma_U$. The plot of $\sigma_\varepsilon(.)$ is similar and is omitted.

In Example 1, we considered three true distributions of $X$ while fixing the assumed model for $X$ at a very restrictive distribution. In the next example, we fix the true distribution of $X$ and compare $\theta(.)$ and $\hat{\theta}$ under several assumed models for $X$.

*Example 2: Y given X follows a Bernoulli distribution with mean* $\mathrm{probit}(\beta_0 + \beta_1 X)$. Assume that $Y$ is binary and that $\mathrm{pr}(Y=1|X=x) = \Phi(\beta_0 + \beta_1 x)$, $\theta = (\beta_0, \beta_1)^{\mathrm{T}}$, where $\Phi(.)$ is the standard normal cumulative distribution function, and the true distribution of $X$ is the mixture of two normals, $N(\mu_i, \sigma_i^2)$, for $i = 1, 2$, with mixing proportion $\alpha \in (0, 0.5)$. Suppose $\theta^* = (0, 1)^{\mathrm{T}}$ and $\tau^* = (2.35, 0.64, -0.26, 0.62, 0.1)^{\mathrm{T}}$, the true values of $(\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha)^{\mathrm{T}}$. The true density curve of $X$ is right skewed with a small secondary mode.

Three assumed models for $X$ are used to construct the likelihood in (1). First, assume that $X \sim N(\mu_x, \sigma_x^2)$. Secondly, assume that $X$ has the second-order seminonparametric density

$$\frac{1}{\eta} \phi\left(\frac{x-\xi}{\eta}\right) \left\{ a_0 + a_1\left(\frac{x-\xi}{\eta}\right) + a_2\left(\frac{x-\xi}{\eta}\right)^2 \right\}^2, \tag{4}$$

where $\phi(.)$ is the standard normal density function, $(\xi, \eta, a_0, a_1, a_2)$ are unknown parameters, and $(a_0, a_1, a_2)$ are constrained so that (4) integrates to one (Zhang & Davidian, 2001). Thirdly, assume that $X$ follows a normal mixture distribution. Compared to the true model for $X$, the first assumed model is incorrect and probably too restrictive, the second assumed model is also incorrect but much more flexible than the first one, and the third one is correct.

Denote by $f_{Y,W}^{(n)}(y, w; \theta, \tau^{(n)}, \sigma_U)$, $f_{Y,W}^{(s)}(y, w; \theta, \tau^{(s)}, \sigma_U)$ and $f_{Y,W}^{(m)}(y, w; \theta, \tau^{(m)}, \sigma_U)$ the joint densities of $(Y, W)$ when $X$ is assumed to follow the normal, seminonparametric and normal mixture distributions, where $\tau^{(n)} = (\mu_x, \sigma_x)^{\mathrm{T}}$, $\tau^{(s)} = (\xi, \eta, a_0, a_1, a_2)^{\mathrm{T}}$ and $\tau^{(m)} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha)^{\mathrm{T}}$. Similarly, denote by $\theta^{(n)}(\sigma_U)$, $\theta^{(s)}(\sigma_U)$ and $\theta^{(m)}(\sigma_U)$ the $\theta(\sigma_U)$'s defined by (2) under these assumed models for $X$. Theoretically, $\theta^{(m)}(\sigma_U) \equiv \theta^*$, as it results from the correct modelling. Hence we use it as the gold standard to which $\theta^{(n)}$ and $\theta^{(s)}$ are compared. The differences, $\theta^{(n)} - \theta^{(m)}$ and $\theta^{(s)} - \theta^{(m)}$, are plotted against $\sigma_U$ in Figs 2(a) and (b). The plots indicate that $\theta^{(s)}(\sigma_U)$ is much more robust than $\theta^{(n)}(\sigma_U)$ and closely matches $\theta^{(m)}(\sigma_U)$.

Figures 2(c) and (d) are Monte Carlo estimated finite-sample versions of Figs 2(a) and (b). In the simulation study, 100 datasets each of size 500 were generated from the true structural measurement error model with the same parameter values as given above. For each dataset, $\hat{\theta}$ was computed by maximising (1), depending on the assumed model for $X$. The expectations, $E(\hat{\theta}^{(\cdot)})$, are estimated by the corresponding Monte Carlo averages. Clearly, no procedure can do better than the true-model estimator, $\hat{\theta}^{(m)}$, and we use it as the gold standard to which $\hat{\theta}^{(n)}$ and $\hat{\theta}^{(s)}$ are compared. The Monte Carlo averages of the differences, $\hat{\theta}^{(n)} - \hat{\theta}^{(m)}$ and $\hat{\theta}^{(s)} - \hat{\theta}^{(m)}$, plotted against $\sigma_U$ in Figs 2(c) and (d), indicate the robustness of $\hat{\theta}^{(s)}$ and the nonrobustness of $\hat{\theta}^{(n)}$.
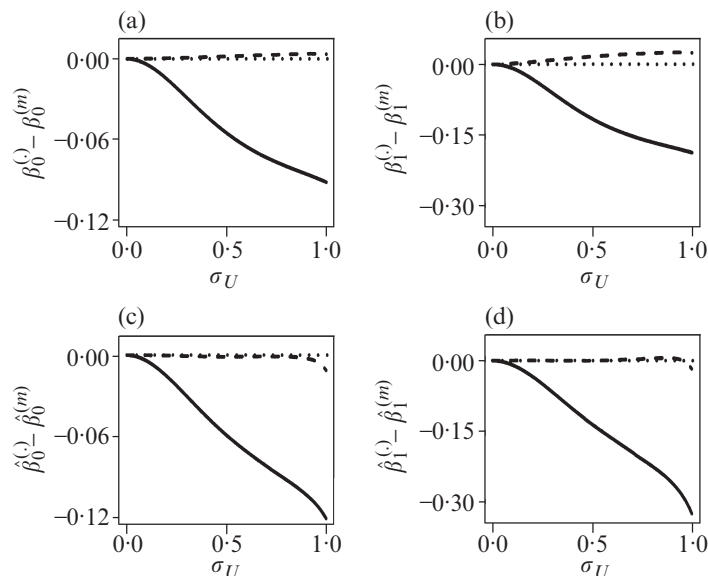
Fig. 2: Plots for Example 2 with $Y|X$ linear-probit. (a) and (b) show $\theta^{(n)}(\sigma_U) - \theta^{(m)}(\sigma_U)$ and $\theta^{(s)}(\sigma_U) - \theta^{(m)}(\sigma_U)$. (c) and (d) show Monte Carlo estimates based on 100 replicates, of finite-sample, $n = 500$, version of (a) and (b). The solid line and the dashed line correspond to the normal modelling and seminonparametric modelling, respectively. The dotted line is the reference line.

### 3·2. *First-order latent-model robustness*

The condition for full latent-model robustness (3) is not easily verified except in very simple models. Also, it is not obvious that it can be satisfied in general without making some assumptions about the true distribution of $X$, except in simple models. Thus its utility is limited.

However, note that, if (3) is satisfied, then the derivatives of $\theta(\sigma_U)$ with respect to $\sigma_U$ of any order are identically 0. More generally, whether (3) is satisfied or not, $\theta(\sigma_U)$ has the MacLaurin series expansion $\theta(\sigma_U) = \theta^* + \sigma_U^2 \theta''(0)/2 + o(\sigma_U^2)$. Thus, a necessary, first-order condition for robustness is that $\theta''(0) = 0$. This condition is somewhat easier to verify than (3). The required derivatives $\theta''(0)$ can be obtained by implicit differentiation as in Stefanski (1985). The following two examples illustrate this condition.

*Example* 3: *First-order latent-model robustness of location-scale models in simple linear regression.* Consider the simple linear regression model in which $Y$ given $X$ is $N(\beta_0 + \beta_1 X, \sigma_\varepsilon^2)$, so that $\theta = (\beta_0, \beta_1, \sigma_\varepsilon)^{\mathrm{T}}$. Suppose that the distribution of $X$ is modelled with a location-scale family, $f_X(x; \tau) = \tau_2 h(\tau_1 + \tau_2 x)$, for some fixed, known, but otherwise arbitrary density $h(.)$. For this model it can be shown that $\theta''(0)$ is a nonsingular matrix multiple of the vector

$$\left( \tau_2^* \beta_1^* E\left\{ \frac{h'(\tau_1^* + \tau_2^* X)}{h(\tau_1^* + \tau_2^* X)} \right\}, \; \beta_1^* + \tau_2^* \beta_1^* E\left\{ \frac{X h'(\tau_1^* + \tau_2^* X)}{h(\tau_1^* + \tau_2^* X)} \right\}, \; 0 \right)^{\mathrm{T}}, \tag{5}$$

where $\beta_1^*$ is the true value of $\beta_1$, and $\tau_1^*$ and $\tau_2^*$ are the probability limits of the maximum likelihood estimators for $\tau_1$ and $\tau_2$ when $\sigma_U = 0$. Regardless of whether or not the true density of $X$ is in the assumed location-scale family, $\tau_1^*$ and $\tau_2^*$ satisfy the asymptotic

location-scale likelihood equations

$$E\left\{\frac{h'(\tau_1^* + \tau_2^* X)}{h(\tau_1^* + \tau_2^* X)}\right\} = 0, \quad E\left\{\frac{Xh'(\tau_1^* + \tau_2^* X)}{h(\tau_1^* + \tau_2^* X)}\right\} + \frac{1}{\tau_2^*} = 0. \tag{6}$$

Note that equations (6) imply that (5) is equal to $0_{3\times 1}$. Thus, $\hat{\theta}$ is first-order robust for arbitrary location-scale models for $X$. The robustness of the normal distribution assumption in Example 1 is a special case of the first-order latent-model robustness of location-scale families.

*Example* 4: *First-order latent-model robustness of the normal distribution model in quadratic regression.* Consider the quadratic regression model in which $Y$ given $X \sim N(\beta_0 + \beta_1 X + \beta_2 X^2, \sigma_\varepsilon^2)$, so that $\theta = (\beta_0, \beta_1, \beta_2, \sigma_\varepsilon)^T$. Suppose that $X$ is modelled as $N(\tau_1, \tau_2)$. For this model it can be shown that $\theta''(0)$ is a nonsingular matrix multiple of the vector

$$\begin{pmatrix} -2\beta_2^* \tau_2^* + \beta_1^* E(X) + 2\beta_2^* E(X^2) - 2\beta_2^* \tau_1^* E(X) - \beta_1^* \tau_1^* \\ -\beta_1^* \tau_2^* - 4\beta_2^* \tau_2^* E(X) - \beta_1^* \tau_1^* E(X) - 2\beta_2^* \tau_1^* E(X^2) + \beta_1^* E(X^2) + 2\beta_2^* E(X^3) \\ -6\tau_2^* \beta_2^* E(X^2) - 2\tau_2^* \beta_1^* E(X) + \beta_1^* E(X^3) + 2\beta_2^* E(X^4) - \beta_1^* \tau_1^* E(X^2) - 2\beta_2^* \tau_1^* E(X^3) \\ 0 \end{pmatrix}, \tag{7}$$

where $\beta_1^*$ and $\beta_2^*$ are the true values of $\beta_1$ and $\beta_2$, and $\tau_1^*$ and $\tau_2^*$ are the probability limits of the maximum likelihood estimators for $\tau_1$ and $\tau_2$ when $\sigma_U = 0$. Thus $\tau_1^* = E(X)$ and $\tau_2^* = E(X^2) - \{E(X)\}^2$. Further simplification shows that the first component of (7) is identically 0, the second reduces to $2\beta_2^* \sigma_X^3 \kappa_{X,3}$, and the third equals

$$\beta_1^* \sigma_X^3 \kappa_{X,3} + 2\beta_2^* \{\sigma_X^4(\kappa_{X,4} - 3) + 3\mu_X \sigma_X^3 \kappa_{X,3}\},$$

where $\kappa_{X,3}$ and $\kappa_{X,4}$ are the skewness and kurtosis of $X$. Thus, estimation of the coefficients in the quadratic model with an assumed normal model for $X$ is first-order robust in general only if the true distribution of $X$ has $\kappa_{X,3} = 0$ and $\kappa_{X,4} = 3$. These conditions are satisfied for $X$ normal, but not in general.

Full latent-model robustness and first-order latent-model robustness are useful analytical constructs for understanding sensitivity of inference to model misspecification. The latter is easier to assess but still quite involved for many models. The more relevant problem for data analysis is assessing the robustness to a choice of model for $X$ in a particular application.

## 4. Empirically checking robustness
### 4·1. *The remeasurement method*

Our discussion of theoretical robustness shows that, when an inadequate model for $X$ is assumed, the bias in $\hat{\theta}$ is manifested by a nonconstant plot of $\theta(\sigma_U)$ that generally increases in absolute value as $\sigma_U$ increases. We now show how to construct empirical versions of these plots and how to check for lack of robustness, using remeasured data generated as in the simulation step of the simulation-extrapolaton, SIMEX, method (Cook & Stefanski, 1994; Stefanski & Cook, 1995; Carroll et al., 1995, Ch. 4). Remeasurement is

a simulation-based technique for determining the effects of measurement error, such as bias and variance, on a statistic. The idea is that the effects of measurement error from a particular dataset are determined by computing the statistic on simulated 'remeasured' datasets, in which the variables measured with error are further contaminated with Monte-Carlo-generated pseudo-measurement errors. We use remeasurement to reveal nonrobustness in the assumed model for $X$.

Our method exploits the fact that, if $\int f_{Y|X}(y|x;\theta)f_{W|X}(w|x;\sigma_U^2)f_X^{(a)}(x;\tau^{(a)})dx$ is a correct model for $(Y, W)$, then $\int f_{Y|X}(y|x;\theta)f_{W|X}\{w|x;(1+\lambda)\sigma_U^2\}f_X^{(a)}(x;\tau^{(a)})dx$ is a correct model for $(Y, W + \lambda^{\frac{1}{2}}\sigma_U Z)$ for all $\lambda > 0$, where $Z \sim N(0, 1)$ independently of $(Y, W)$. Consequently, if the assumed model for $X$ is correct or robust in the sense defined in § 3, an estimator for $\theta$ derived from the latter model fitted to remeasured data $\{Y_j, W_j + \lambda^{\frac{1}{2}}\sigma_U Z_j\}_{j=1}^n$ should be consistent regardless of the size of $\lambda$, and therefore should exhibit no dependence on $\lambda$. Conversely, if the model is incorrect and nonrobust, then absolute bias will tend to increase with increasing measurement error, and this will be manifested by a dependence on $\lambda$.

In our method, the $b$th $\lambda$-remeasured dataset ($b = 1, 2, \ldots, B$) is constructed by replacing $W_j$ with $W_{b,j}(\lambda) = W_j + \lambda^{\frac{1}{2}}\sigma_U Z_{b,j}$, for $j = 1, \ldots, n$, where the $Z_{b,j}$ are independent and identically distributed as $N(0, 1)$. Based on the $b$th $\lambda$-remeasured dataset $\{Y_j, W_{b,j}(\lambda)\}_{j=1}^n$, compute the $b$th $\lambda$-contaminated, pseudo-estimate, $\hat{\theta}_b(\lambda)$. Averaging over $b$ gives $\hat{\theta}_B(\lambda) = \sum_{b=1}^B \hat{\theta}_b(\lambda)/B$, which is an estimator for $\theta$ when the measurement error variance of the remeasured data is $(1 + \lambda)\sigma_U^2$. Nonconstancy in the plot of $\hat{\theta}_B(\lambda)$ indicates lack of robustness in the assumed model for $X$. For simulation-extrapolation estimation, Carroll et al. (1995) recommend taking $\lambda \in [0, \lambda_{\max}]$ with $1 \leqslant \lambda_{\max} \leqslant 3$. For our diagnostic purposes, we take $\lambda_{\max} = 1$ or 3. Note that the added variance is $\lambda\sigma_U^2$. Thus, if $\sigma_U^2$ is small, the amount of added noise will also be small provided $\lambda$ is not extremely large.

Note that our method is not specific to parametric likelihood estimation. For example, if $\sum \psi(Y_j, W_j, \theta, \sigma_U^2)$ is a correct or robust estimating equation for $\theta$, the same is true for $\sum \psi\{Y_j, W_{b,j}(\lambda), \theta, (1 + \lambda)\sigma_U^2\}$, and robustness of $\psi$ can be checked as described above.

## 4·2. *Simulation studies*

*Example* 5: *Y given X follows a normal distribution with mean* $\beta_0 + \beta_1 X$. This example complements Example 1. For each of the true distributions for $X$, $N(1, 1)$, $N(0·5, 1)$ and $N(1·5, 1)$, a random sample $\{Y_j, W_j\}_{j=1}^{500}$ was generated according to $Y|X \sim N(\beta_0 + \beta_1 X, \sigma_\varepsilon^2)$ with $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_\varepsilon = 1$, and $W|X \sim N(X, \sigma_U^2)$ with $\sigma_U^2 = 0·5$. In all three cases the assumed model for $X$ is $N(\tau^{(a)}, \tau^{(a)})$. For each fixed $\lambda$ varying from 0 to 1, $B = 500$ $\lambda$-remeasured datasets were generated. Figures 1(c) and (d) display plots of $\hat{\theta}_B(\lambda)$ versus $\lambda$. Note that, as $\lambda$ ranges from 0 to 1, $\text{var}\{W(\lambda)|X\}$ ranges from 0·5 to 1, where $W(\lambda) = W + \lambda^{\frac{1}{2}}\sigma_U Z$.

For the case $X \sim N(1, 1)$ when the assumed model is correct, the curves of $\hat{\beta}_{0,B}(\lambda)$ and $\hat{\beta}_{1,B}(\lambda)$ are almost horizontal lines, as expected. For the other two cases, the assumed model is incorrect and too restrictive, which in general should result in nonhorizontal curves of $\hat{\beta}_{0,B}(\lambda)$ and $\hat{\beta}_{1,B}(\lambda)$. These cases are also readily identified in the plots of the regression parameters.

*Example* 6: *Y given X follows a Bernoulli distribution with mean* $\text{probit}(\beta_0 + \beta_1 X)$. A random sample $\{Y_j, W_j\}_{j=1}^{2000}$ was generated from the true measurement error model defined in Example 2 with $\sigma_U^2 = 0·16$. For each fixed $\lambda$ ranging from 0 to 3, $B = 100$ $\lambda$-remeasured datasets were generated. The three assumed models chosen in Example 2 were used to construct (1).
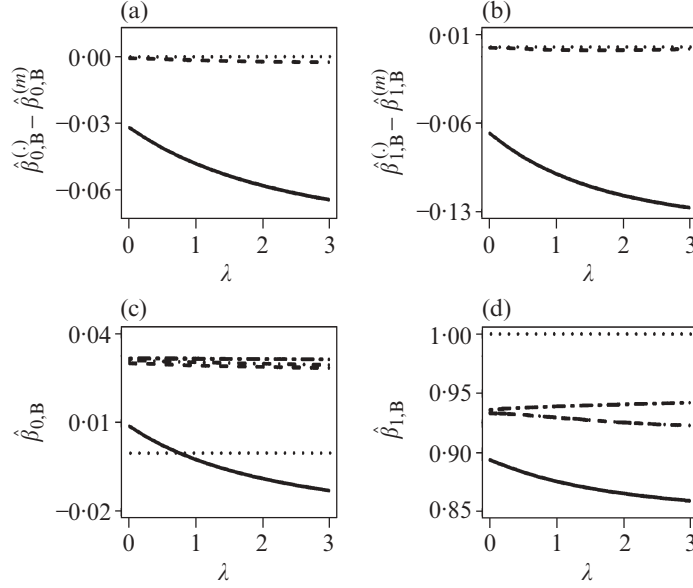
Fig. 3. (a) and (b) show plots for Example 6 with $Y|X$ linear-probit, $\hat{\theta}_{B}^{(n)} - \hat{\theta}_{B}^{(m)}$, solid line, and $\hat{\theta}_{B}^{(s)} - \hat{\theta}_{B}^{(m)}$, dashed line, resulting from the remeasurement method with $B = 100$. (c) and (d) show plots for Example 7 with $Y|X$ linear-logistic, $\hat{\theta}_{B}^{(n)}$, solid line, $\hat{\theta}_{B}^{(s)}$, dashed line, $\hat{\theta}_{B}^{(m)}$, dashed-dotted line, and $\hat{\theta}_{B}^{(c)}$, long-short-dashed line, resulting from the remeasurement method with $B = 100$. The dotted lines in plots (a)–(d) are the reference lines.

With the superscript notation introduced in Example 2, Figs 3(a) and (b) depict $\hat{\theta}_{B}^{(n)}(\lambda) - \hat{\theta}_{B}^{(m)}(\lambda)$ and $\hat{\theta}_{B}^{(s)}(\lambda) - \hat{\theta}_{B}^{(m)}(\lambda)$. The implications with regard to robustness of estimates resulting from different modellings agree with those suggested by Fig. 2. Moreover, the estimated moments, not shown, from seminonparametric modelling are virtually unbiased for the moments of the true model up to high orders, implying that the estimated seminonparametric density approximates the true mixture normal density very well.

*Example* 7: *Y given X follows a Bernoulli distribution with mean equal to the logistic transformation of $\beta_0 + \beta_1 X$.* The specifications of the measurement error model and the three assumed models for $X$ are identical to those in Example 6, except that now $\mathrm{pr}(Y = 1|X = x) = \{1 + \exp(-\beta_0 - \beta_1 x)\}^{-1}$.

In addition to $\hat{\theta}_{B}^{(n)}(\lambda)$, $\hat{\theta}_{B}^{(s)}(\lambda)$ and $\hat{\theta}_{B}^{(m)}(\lambda)$, the conditional score estimator for $\theta$ (Stefanski & Carroll, 1987), denoted by $\hat{\theta}_{B}^{(c)}(\lambda)$, was also computed. The conditional score equation for this linear-logistic measurement error model is derived from the conditional density of $Y|\Delta$, where $\Delta = W + Y\sigma_U^2\beta_1$. The conditional distribution of $Y|\Delta$ does not depend on $X$, and thus conditional score estimators satisfy our definition of full latent-model robustness. We include it here to show that our proposed methods also suggest its robustness.

From the same simulation set-up as in Example 6 but with the logistic model, $\hat{\theta}_{B}^{(n)}(\lambda)$, $\hat{\theta}_{B}^{(s)}(\lambda)$, $\hat{\theta}_{B}^{(m)}(\lambda)$ and $\hat{\theta}_{B}^{(c)}(\lambda)$ were computed for the generated $\lambda$-remeasured datasets. Figures 3(c) and (d) display these estimates as functions of $\lambda$. As expected, $\hat{\theta}_{B}^{(c)}(\lambda)$ appears to be robust; moreover, $\hat{\theta}_{B}^{(s)}(\lambda)$ and $\hat{\theta}_{B}^{(m)}(\lambda)$ are very similar to $\hat{\theta}_{B}^{(c)}(\lambda)$, while $\hat{\theta}_{B}^{(n)}(\lambda)$ is the least robust.

### 4·3. *Test of latent-model robustness*

So far, the claims of robustness or lack of robustness in the examples are based on visual assessment of the plots of $\hat{\theta}_B(\lambda) = \sum_{b=1}^B \hat{\theta}_b(\lambda)/B$. In practice, an objective assessment of robustness is required. The statistic $t(\lambda_*) = \{\hat{\theta}_B(0) - \hat{\theta}_B(\lambda_*)\}/\hat{v}_*$, where $\hat{v}_*^2$ is an estimator of $\text{var}\{\hat{\theta}_B(0) - \hat{\theta}_B(\lambda_*)\}$, provides a large-sample test of $H_0: E\{\hat{\theta}_B(0) - \hat{\theta}_B(\lambda_*)\} = 0$, with large values of $|t(\lambda_*)|$ indicating lack of robustness. The variance estimator $\hat{v}_*^2$ is derived in the Appendix. The choice of $\lambda_*$ is partly related to $B$. Ideally, $\lambda_*$ should be small to be consistent with the notion of first-order latent-model robustness in § 3·2. However, for $\lambda_*$ too small, the Monte Carlo variation associated with $B < \infty$ would render the statistic powerless. For the choices of $B$ used in this paper, we have investigated $\lambda_* = 1, 2, 3$ and found very little difference in performance of the test statistics. Hence we report results for only $\lambda_* = 3$.

Corresponding to the simulation presented in Figs 3(a) and (b) for Example 6, the upper half of Table 1 gives the $t(\lambda_*)$ statistics for evaluating the evidence regarding the changes in $\hat{\beta}_{0,B}$ and $\hat{\beta}_{1,B}$ for each assumed model for $X$. Only the statistics for the normal model exceed the asymptotic critical value of 1·96. The lower half of Table 1 gives the $t(\lambda_*)$ statistics for the simulation presented in Figs 3(c) and (d) for Example 7. The statistics are consistent with the claims made previously regarding robustness.

Table 1. *Statistics, $t(\lambda_*)$, assessing robustness of the parameter estimators for $\lambda_* = 3$ under three ways of modelling for the simulated data from the linear-probit measurement error model in Example 6, and for the simulated data from the linear-logistic measurement error model in Example 7. Example 7 also considers the conditional score estimator.*

|  | Normal | Seminonparametric | Normal mixture | Conditional score |
|---|---|---|---|---|
| | | Example 6 | | |
| Change in $\hat{\beta}_{0,B}$ | 6·91 | 0·12 | −0·83 | – |
| Change in $\hat{\beta}_{1,B}$ | 4·99 | −0·20 | −0·32 | – |
| | | Example 7 | | |
| Change in $\hat{\beta}_{0,B}$ | 5·26 | 0·65 | 0·61 | 0·12 |
| Change in $\hat{\beta}_{1,B}$ | 3·45 | 0·72 | 0·73 | −0·66 |

Under the model and parameter settings in Example 2, we examined operating characteristics of $t(3)$ via a Monte Carlo simulation study. In the simulation, 100 datasets of size 1000 were generated from the measurement error model with $\sigma_U^2 = 0·16$. For each dataset, the remeasurement method was applied with $B = 50$ and $\lambda$ varying from 0 to 3, and $t(3)$ statistics were computed for testing the robustness in $\hat{\beta}_{0,B}$ and $\hat{\beta}_{1,B}$. From the simulation, when $X$ is modelled as normal, seminonparametric and normal mixture, the proportions of 100 datasets with $|t(3)| > 1·96$ are 1·00, 0·08, 0·07 for $\hat{\beta}_{0,B}$ and 0·99, 0·1, 0·1 for $\hat{\beta}_{1,B}$. The maximum standard error for these estimated rejection rates was 0·03. The results indicate that the proposed statistic has the desired operating characteristics.

## 5. Coronary heart disease and blood pressure

We apply the remeasurement method and testing procedure to data on 1615 individuals from the Framingham study (Kannel et al., 1986). This study followed subjects for development of coronary heart disease over several examination periods, with one of the objectives being to characterise the relationship between a response $Y$, an indicator of evidence of

coronary heart disease at the end of an eight-year follow-up period after the second examination visit, and unobservable long-term systolic blood pressure $X$. We take $W$ to be the systolic blood pressure from examination 2 with $\sigma_U^2$ estimated using examination 2 and examination 3 measurements as in Carroll et al. (1995).

We assume that $\mathrm{pr}(Y=1|X)=\Phi(\beta_0+\beta_1 X)$, with $X$ modelled in the same three ways as in Example 2. For each fixed $\lambda$ varying from 0 to 3, $B=100$ $\lambda$-remeasured datasets were generated. Figure 4 presents the plots of $\hat{\beta}_{0,\mathrm{B}}(\lambda)$ and $\hat{\beta}_{1,\mathrm{B}}(\lambda)$ versus $\lambda$ for the three ways of modelling and shows that the seminonparametric modelling leads to the most robust estimates while the normal modelling results in the least robust estimates. It suggests a gain in robustness via flexible modelling for $X$.
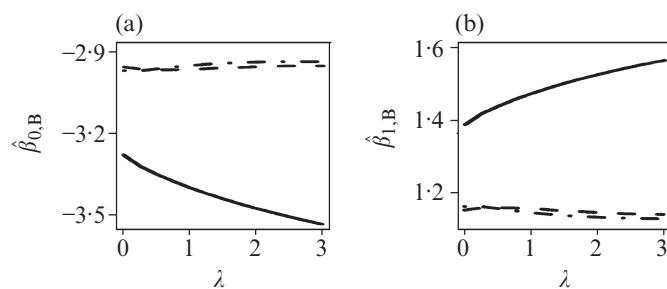


Fig. 4. Plots of $\hat{\theta}_\mathrm{B}^{(n)}$, solid line, $\hat{\theta}_\mathrm{B}^{(s)}$, dashed line, and $\hat{\theta}_\mathrm{B}^{(m)}$, dashed-dotted line, resulting from applying the remeasurement method with $B=100$ to data from the Framingham study.

The values of $t(3)$ for assessing robustness when $X$ is assumed to follow the normal, seminonparametric and normal mixture distributions are $3\cdot16$, $-0\cdot18$ and $-0\cdot81$ for $\hat{\beta}_{0,\mathrm{B}}$, and $-3\cdot21$, $0\cdot19$ and $0\cdot81$ for $\hat{\beta}_{1,\mathrm{B}}$. That the seminonparametric modelling gives the smallest $|t(3)|$ and normal modelling yields the largest $|t(3)|$ justifies the visual conclusions drawn from Fig. 4.

## 6. Discussion

In this paper we have proposed methods for diagnosis of robustness to the distributional specification of the true predictor in structural measurement error modelling, either by checking formal analytic robustness conditions or empirically via the remeasurement method and a testing procedure. Although we studied only the classical error model, the remeasurement method can be applied with any error model whose effects on the observed data can be simulated, such as multiplicative error models. When there are additional observable covariates $V$ in the model for $Y$, the relevant issue is about the specification of the model for $X|V$ instead of the model for $X$. Of course, the data analyst should always bear in mind that a finding of latent-model robustness using these methods does not preclude lack of robustness to other modelling assumptions.

The principles underlying both the theoretical and empirical methods are applicable to general structural latent variable models. One example, which can be viewed as a generalisation of the structural measurement error model, is that of so-called joint models, which link a longitudinal response and a primary endpoint through shared dependence on latent random effects (Henderson et al., 2000). We are currently studying adaptation of the proposed techniques to these more complex models, including joint models for an

error-prone longitudinal response and a primary endpoint that may be either a simple response (Li et al., 2004) or a censored time-to-event (Song et al., 2002; Tsiatis & Davidian, 2004).

## APPENDIX
### *Estimation of* $\mathrm{var}\{\hat{\theta}_B(0) - \hat{\theta}_B(\lambda_*)\}$ *in* § 4·3

Let $\theta$ be the vector of unknown parameters under the assumed measurement error model. We now derive an estimator for $\mathrm{var}\{\hat{\theta}_{\mathrm{B}}(0) - \hat{\theta}_{\mathrm{B}}(\lambda_*)\}$, where $\lambda_* > 0$. For brevity, this variance is denoted by $\mathrm{var}(\hat{\theta}^{(0)} - \hat{\theta}^{(\lambda_*)})$, where

$$\hat{\theta}^{(k)} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^{(k)}, \tag{A1}$$

and $\hat{\theta}_b^{(k)}$ solves $\sum_{j=1}^{n} \psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \theta) = 0$, for $k = 0, \lambda_*$ and $b = 1, \dots, B$. Here $\psi(.)$ is the score function under the assumed measurement error model. Approximating the estimator with the average of influence functions (Casella & Berger, 2002, p. 517) gives

$$\hat{\theta}_b^{(k)} \simeq \theta^{(k)} + \frac{1}{n} A(\theta^{(k)})^{-1} \sum_{j=1}^{n} \psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \theta^{(k)}) \quad (b = 1, \dots, B, k = 0, \lambda_*), \tag{A2}$$

where $\theta^{(k)}$ is an unknown parameter vector, and $A(\theta^{(k)}) = E\{-(\partial/\partial\theta^{\mathrm{T}})\psi(Y, W + k^{1/2}\sigma_U Z; \theta^{(k)})\}$, which is estimated by $\hat{A}(\hat{\theta}_b^{(k)}) = n^{-1}\sum_{j=1}^{n}\{-(\partial/\partial\theta^{\mathrm{T}})\psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(k)})\}$. Replacing $A(\theta^{(k)})$ with $\hat{A}(\hat{\theta}_b^{(k)})$ and the unknown $\theta^{(k)}$ in $\psi(.)$ with $\hat{\theta}_b^{(k)}$, and considering that the score function evaluated at $\hat{\theta}_b^{(k)}$ may not be exactly zero because of the numerical inaccuracy of the algorithm used to find the maximum likelihood estimator for $\theta^{(k)}$, yields an adjusted version of (A2), given by

$$\hat{\theta}_b^{(k)} \simeq \theta^{(k)} + \frac{1}{n} \sum_{j=1}^{n} \hat{A}(\hat{\theta}_b^{(k)})^{-1} \{\psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(k)}) - \bar{\psi}_b^{(k)}\}, \tag{A3}$$

where $\bar{\psi}_b^{(k)} = n^{-1}\sum_{j=1}^{n}\psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(k)})$, for $b = 1, \dots, B$ and $k = 0, \lambda_*$. It follows by (A1) and (A3) that $\hat{\theta}^{(k)} \simeq \theta^{(k)} + B^{-1}\sum_{b=1}^{B} n^{-1}\sum_{j=1}^{n} \hat{A}(\hat{\theta}_b^{(k)})^{-1}\{\psi(Y_j, W_j + k^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(k)}) - \bar{\psi}_b^{(k)}\}$, for $k = 0, \lambda_*$. Therefore,

$$\hat{\theta}^{(0)} - \hat{\theta}^{(\lambda_*)} \simeq \theta^{(0)} - \theta^{(\lambda_*)} + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{B}\sum_{b=1}^{B}[\hat{A}(\hat{\theta}_b^{(0)})^{-1}\{\psi(Y_j, W_j; \hat{\theta}_b^{(0)}) - \bar{\psi}_b^{(0)}\}$$
$$- \hat{A}(\hat{\theta}_b^{(\lambda_*)})^{-1}\{\psi(Y_j, W_j + \lambda_*^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(\lambda_*)}) - \bar{\psi}_b^{(\lambda_*)}\}]. \tag{A4}$$

Define

$$T_j = B^{-1}\sum_{b=1}^{B}[\hat{A}(\hat{\theta}_b^{(0)})^{-1}\{\psi(Y_j, W_j; \hat{\theta}_b^{(0)}) - \bar{\psi}_b^{(0)}\} - \hat{A}(\hat{\theta}_b^{(\lambda_*)})^{-1}\{\psi(Y_j, W_j + \lambda_*^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(\lambda_*)}) - \bar{\psi}_b^{(\lambda_*)}\}],$$

for $j = 1, \dots, n$. At $\lambda = 0$, the influence functions are equal for all $b = 1, \dots, B$, and $T_j$ simplifies to

$$T_j = \hat{A}(\hat{\theta}^{(0)})^{-1}\{\psi(Y_j, W_j; \hat{\theta}^{(0)}) - \bar{\psi}^{(0)}\} - B^{-1}\sum_{b=1}^{B}\hat{A}(\hat{\theta}_b^{(\lambda_*)})^{-1}\{\psi(Y_j, W_j + \lambda_*^{1/2}\sigma_U Z_{j,b}; \hat{\theta}_b^{(\lambda_*)}) - \bar{\psi}_b^{(\lambda_*)}\},$$

for $j = 1, \ldots, n$. Substituting $T_j$ in (A4) gives $\hat{\theta}^{(0)} - \hat{\theta}^{(\lambda_*)} \simeq \theta^{(0)} - \theta^{(\lambda_*)} + \sum_{i=1}^{n} T_i/n$. Since $\{T_j\}_{j=1}^{n}$ are independent, $\mathrm{var}(\hat{\theta}^{(0)} - \hat{\theta}^{(\lambda_*)}) \simeq \mathrm{var}(T_1)/n$. Finally, $\mathrm{var}(\hat{\theta}^{(0)} - \hat{\theta}^{(\lambda_*)}) \simeq S_T^2/n$, where $S_T^2$ denotes the sample variance-covariance matrix of $\{T_j\}_{j=1}^{n}$.

## References

Carroll, R. J., Roeder, K. & Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics* **55**, 44–54.

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

Casella, G. & Berger, R. L. (2002). *Statistical Inference*, 2nd ed. Pacific Grove, CA: Wadsworth & Cole.

Cook, J. & Stefanski, L. A. (1994). Simulation extrapolation estimation in parametric measurement error models. *J. Am. Statist. Assoc.* **89**, 1314–28.

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

Henderson, R., Diggle, P. & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **4**, 465–80.

Kannel, W. B., Neaton, J. D., Wentworth, D., Thomas, H. E., Stamler, J., Hulley, S. B. & Kjelsberg, M. O. (1986). Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for MRFIT. *Am. Heart J.* **112**, 825–36.

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

Richardson, S. (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *J. R. Statist. Soc. A* **165**, 549–66.

Roeder, K., Carroll, R. J. & Lindsay, B. G. (1996). A semiparametric mixture approach to case-control studies with errors in covariables. *J. Am. Statist. Assoc.* **91**, 722–32.

Schafer, D. W. (2001). Semiparametric maximum likelihood for measurement error model regression. *Biometrics*, **57**, 53–61.

Song, X., Davidian, M. & Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–53.

Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72**, 583–92.

Stefanski, L. A. & Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703–16.

Stefanski, L. A. & Cook, J. (1995). Simulation extrapolation: The measurement error jackknife. *J. Am. Statist. Assoc.* **90**, 1247–56.

Tsiatis, A. A. & Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statist. Sinica* **14**, 793–18.

Zhang, D. & Davidian, M. (2001). Linear mixed model with flexible distribution of random effects for longitudinal data. *Biometrics* **57**, 795–02.