International Statistical Review (2007), **75**, 2, 183–198 doi:10.1111/j.1751-5823.2007.00014.x

An Overview of Normal Theory Structural Measurement Error Models

Jeffrey R. Thompson[1] and Randy L. Carter[2]

[1]*Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA*
E-mail: thompson@stat.ncsu.edu
[2]*Department of Biostatistics, University of Buffalo, Buffalo, NY 14214, USA*
E-mail: rcarter@buffalo.edu

Summary

This paper gives an introduction and overview to the often under-used measurement error model. The purpose is to provide a simple summary of problems that arise from measurement error and of the solutions that have been proposed. We start by describing how measurement error models occur in real-world situations. Then we proceed with defining the measurement error model, initially introducing the multivariate form of the model, and then, starting with the simplest form of the model thoroughly discuss its features and solutions to the problems introduced due to measurement error. We discuss higher-dimensional and more advanced forms of the model and give a brief numerical illustration.

Key words: Instrumental variables; linear models; measurement error; nonlinear models; structural relationship.

1 Introduction

Measurement error models are models with at least one independent variable that is measured with error. Aside from the term *measurement error models*, this area of statistics is known by several other names: *errors in variables* and *regression with errors in x* are but two others (Stefanski, 2000). The measurement error problem occurs often in practice. The medical field, agriculture, and econometrics are just a few areas that contain problems involving predictor variables that are contaminated by measurement error.

One major impact of measurement error in independent variables, when it exists and is overlooked, is that standard estimation methods for model parameters result in biased estimates (see, for example, Carroll *et al.*, 1995). In the common simple linear regression model, for example, the ordinary least squares estimates of the regression parameters, in the presence of ignored measurement error, lead to a slope estimate that is attenuated towards zero and thus underrepresents the parameter of interest on average. There is a diverse collection of topics in the known measurement error literature, but no single, clear presentation of the basic results. The purpose of this paper is to shed further light on the under-studied and often under-used measurement error model. We present known, standard solutions to the effects of measurement error in normal theory structural models and promote the idea that among the solutions, the instrumental variable solution often provides the most tractable and realistic means of handling

measurement error problems. A data example is provided illustrating an instrumental variable's correction of attenuation in the estimate of a slope parameter.

1.1 Common Occurrences of the Measurement Error Problem

The general measurement error problem occurs often in medical research when the association between an unobservable predictor variable and a response variable is of interest. One of the many examples involves Alzheimer's disease. It is thought that the level of aluminium deposits, which may build up in the brain over time, have an effect on an evaluation score for diagnosing an individual developing Alzheimer's (Campbell, 2002). If one were to estimate the association between the likelihood of Alzheimer's disease and the level of aluminium deposits in the brain, problems would arise because a perfectly accurate measure of aluminium levels is simply not attainable. Measurement error models have also been used to assess the reliability and validity of a measuring method in the absence of a true "gold standard" method. Carter (1981), for example, assessed the reliability and validity of a measure of specific activity of the enzyme sucrase that was obtained from a homogenated sample of the small intestine of intestinal bypass patients.

Measurement error in exposure dose is often an issue in epidemiologic studies. In the study of late-term effects of atomic bomb radiation, for example, measurement error in dose could result in attenuated estimates of effects. This, in turn, would lead to higher thresholds for acceptable occupational exposures, which are based (in part) on atomic bomb survivors' data. Researchers at the Radiation Effects Research Foundation in Hiroshima, Japan, have recognized this problem and continue to address it (Pierce *et al.*, 1990, 1992).

Problems with measurement error in predictor variables often occur in economics. Klepper & Leamer (1984), for example, faced this problem when estimating the quantity of a good exported by a country as a linear function of the country's land, labour and capital. They could not apply usual linear regression techniques to correctly estimate the quantity exported, because these three explanatory variables were, in their words, "doubtlessly measured with error" (Klepper & Leamer, 1984, p. 180).

2 The Multivariate Measurement Error Model

Let $\{\underline{y}_i\}_{i=1}^{\infty}$ and $\{\underline{e}_i\}_{i=1}^{\infty}$ be sequences of p-dimensional random column vectors and let $\{\underline{x}_i\}_{i=1}^{\infty}$ be a sequence of k-dimensional independent column vectors. Further, let $\underline{\beta} \in \psi \subset R^{kp}$ be a $kp \times 1$ column vector of parameters and $\underline{f}$ be a p-dimensional vector whose components are real valued Borel measurable functions mapping $R^k \times \psi$ into R^1 and describing $E(\underline{Y}_i \mid \underline{x}_i)$ as a function of $\underline{x}_i$. Then,

$$\underline{Y}_i = \underline{f}(\underline{x}_i; \underline{\beta}) + \underline{e}_i, \ i = 1, 2, \ldots, \tag{1}$$

where $\underline{e}_i$ may be a combination of model error and measurement error on $\underline{Y}_i$, if it exists. Also, assume

$$E(e_{ij}) = E(x_{il}e_{i'j}) = 0 \tag{2}$$

for all $j = 1, 2, \ldots, p$, $l = 1, 2, \ldots, k$, $i = 1, 2, \ldots$, and $i' = 1, 2, \ldots$. That is, assume that model errors and measurement errors in $\underline{Y}_i$, if they exist, are uncorrelated with the $\underline{x}_i$. If the elements of $\underline{\beta}$ are not functionally related and

$$\underline{f}(\underline{x}_i; \underline{\beta}) = \underline{B}\underline{x}_i, \ i = 1, 2, \ldots, \tag{3}$$

where $\underline{B}$ is a $p \times k$ matrix of parameters formed by placing the first k elements of the $kp \times 1$ vector $\underline{\beta}$ in the first row of $\underline{B}$, the next k in the second row, and so on, then the model defined by Equations (1), (2) and (3) is called the *multivariate linear regression model*. Without the addition of Equation (3), the model defined by 1 and 2 is called the *multivariate nonlinear regression model*. Note that if the model contains an intercept, the first element of $\underline{x}_i$ is unity.

Sometimes the vectors $\underline{x}_i$, $i = 1, 2, \ldots$, are not directly observable. Instead, assume that one observes

$$\underline{X}_i = \underline{x}_i + \underline{u}_i, \quad i = 1, 2, \ldots, n \tag{4}$$

where

$$E(u_{il}) = E(u_{il}x_{i'l'}) = E(u_{il}y_{i'j}) = 0 \tag{5}$$

for all $j = 1, 2, \ldots, p$, $l = 1, 2, \ldots, k$, $l' = 1, 2, \ldots, k$, $i = 1, 2, \ldots, n$, and $i' = 1, 2, \ldots, n$. That is, assume that the $\underline{X}_i$ are unbiased measurements of the $\underline{x}_i$, that the measurement errors ($\underline{u}_i$) and true values ($\underline{x}_i$) are all uncorrelated, and that the measurement errors are not correlated with model errors nor measurement errors on $\underline{Y}_i$, if they exist. The model defined by Equations (1), (2), (4), and (5) is called the *general measurement error* (henceforth, ME) *model*. With the addition of Equation (3), the model is called the *linear measurement error model* (linear ME). When the unobservable $\underline{x}_i$ are fixed, the model is known as the *functional model* and when they are non-constant random vectors, the model is known as the *structural model*.

The term functional model arose from early distinctions of a special case of the functional model, as described above, where the $\underline{e}_i$ in Equation (1) is comprised of measurement error only (Kendall & Stuart 1979, pp. 375–379). That is, purely functional relationships between two sets of nonrandom variables hold,

$$\underline{y}_i = \underline{f}(\underline{x}_i; \underline{\beta})$$

but those variables are measured with error,

$$\underline{Y}_i = \underline{y}_i + \underline{e}_i$$
$$\underline{X}_i = \underline{x}_i + \underline{u}_i$$

where $E(e_{il}y_{i'j}) = 0$. In this case, no estimation of $\underline{\beta}$ would be required in the absence of measurement error, as the elements of $\underline{\beta}$ then could be obtained by solving the functional equations. If the $\underline{e}_i$ in Equation (1) is comprised of model error and measurement error, or comprises model error only, then $E(e_{il}y_{i'j}) \neq 0$ and Equation (1) defines a regression model with parameters that must be estimated even in the absence of measurement error in $\underline{y}_i$. In either case, the parameters of Equation (1) would be estimated by regression methods, if the values of the $\underline{x}_i$ were observed without error. The complicating factor is that they are not observed. In the functional model they are unknown constants, i.e. parameters that need to be estimated. Unfortunately, they cannot be estimated, because they increase in number with the sample size (Neyman & Scott, 1948). Neyman & Scott (1948) referred to parameters that enter the distribution of the observable random variables for only finitely many individuals in the population (e.g. parameters that are unique to individuals) as *incidental parameters*. Generally, it is only the structural parameters (e.g. $\underline{\beta}$ in the above discussion) that we can hope to estimate consistently.

Functional models arise naturally when the relationship of interest is deterministic in the absence of measurement error. For example, the relationship between pressure (P) and volume (V) in the condition of adiabatic expansion is a functional one of the form $P \times V^{\alpha} = \gamma$, where α and γ are constants. This functional relationship describes a modified version of Boyle's law.

Functional relationships also arise when the underlying model is a regression model, where the values of $\underline{x}_i$ are fixed. In the latter case, the $\underline{x}_i$, $i = 1, 2, \ldots, n$, are either the conditional values of a random vector (as would be the case if we condition on the aluminium values in an observed sample in the Alzheimer's example) or are the fixed values in a finite population of individuals (such as in the atomic bomb example). In either of the latter two cases, however, the models alternatively could be treated as structural by not conditioning on the $\underline{x}_i$, $i = 1, 2, \ldots, n$, but by treating them as random instead.

This paper is mainly devoted to the structural model, but it is nonetheless important to distinguish between the structural and functional models. Again, the $(\underline{x}_1, \underline{x}_2, \ldots)$ in Equation (1) may be observations on nonrandom variables in which case the model is defined to be the functional model. In this case, $\underline{x}_i$ can be thought of as a constant vector and they appear as parameters in the distribution function of $(\underline{Y}_i, \underline{X}_i)$. In the structural model with normal true values, $\underline{x}_i$ are assumed to be independent drawings from a $N(\underline{\mu}_x, \Sigma_{xx})$ distribution. The name "structural" comes from the fact that these models describe the structure of the specified relationship between the random variables $\underline{x}$ and $\underline{Y}$.

3 The Linear ME Model with Normal Errors and True Values

3.1 The Simple Linear ME Model

For illustrative purposes, we will begin with the simple linear ME model, which contains one dependent and one independent variable and has normally distributed model error. Consider the model defined by

$$Y_i = \beta_0 + \beta_1 x_i + e_i \tag{6}$$

$$X_i = x_i + u_i, \tag{7}$$

where $i = 1, 2, \ldots, n$, e_i are independent $N(0, \sigma_{ee})$ and may potentially be a combination of model and measurement error, and u_i is a $N(0, \sigma_{uu})$ random variable. So, at this point, we have the following assumptions:

$$\begin{bmatrix} x_i \\ e_i \\ u_i \end{bmatrix} \sim^{iid} N\left(\begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{xx} & \sigma_{xe} & \sigma_{xu} \\ \sigma_{xe} & \sigma_{ee} & \sigma_{eu} \\ \sigma_{xu} & \sigma_{eu} & \sigma_{uu} \end{bmatrix} \right), \tag{8}$$

where $\sim^{iid}$ means "independently and identically distributed" and N means "normally" distributed. For some estimation purposes, it may be assumed that x_i, e_i and u_i are independent for all i. When there is no measurement error (i.e. the classical regression model), it is well known that the least squares estimator of β_1 is

$$\widehat{\beta}_1 = \left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \right]^{-1} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y}).$$

This estimator is also the maximum likelihood estimator and is unbiased for β_1 in both the functional and structural models.

In addition to the assumptions in (8) that the $(x_1, x_2, \ldots, x_n)$ in Equations (6) and (7) are random (i.e. they are drawn from a $N(\mu_x, \sigma_{xx})$ distribution) we may in some applications also assume that x_i, e_i, and u_i are independent for all i. Then we have that the vector

$$(x_i, e_i, u_i)' \sim^{iid} N\left[(\mu_x, 0, 0)', \text{diag}(\sigma_{xx}, \sigma_{ee}, \sigma_{uu}) \right], \tag{9}$$

where diag() represents a diagonal matrix with the given elements on the diagonal. Then the vector $(Y_i, X_i)'$, where Y_i is defined by Equation (6) and X_i is defined by Equation (7), has a bivariate normal distribution with mean vector $E\{(Y_i, X_i)'\} = (\mu_Y, \mu_X)' = (\beta_0 + \beta_1\ \mu_x, \mu_x)'$ and variance–covariance matrix

$$\text{Var}\{(Y_i,\ X_i)'\} = \begin{bmatrix} \sigma_{YY} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{XX} \end{bmatrix} = \begin{bmatrix} \beta_1^2\sigma_{xx} + \sigma_{ee} & \beta_1\sigma_{xx} \\ \beta_1\sigma_{xx} & \sigma_{xx} + \sigma_{uu} \end{bmatrix}.$$

3.2 Non-Identifiability and Asymptotic Bias of the OLS Estimator

Under the model defined by Equations (6) and (7) and under the assumptions stated in (9), using the observed variables, one might naively estimate the regression coefficient, β_1, as

$$\widehat{\beta}_{1,OLS} = \left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{-1} \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}). \tag{10}$$

By the properties of the bivariate normal distribution, this naive estimator has expected value $E\{\widehat{\beta}_{1,OLS}\} = \sigma_{XX}^{-1}\sigma_{XY} = \beta_1(\sigma_{xx} + \sigma_{uu})^{-1}\sigma_{xx}$. One can see that, because the denominator is inflated by σ_{uu}, the least squares regression coefficient is biased towards zero, and that the bias does not vanish with increasing sample size. It is important to note that the key assumptions used to derive the properties of this estimator were that the measurement error, u_i, was independent of both the true values, x_i, and the model errors, e_i. This bias towards zero in the regression coefficient is also referred to as attenuation of the coefficient toward zero. The ratio $\kappa_{xx} = \sigma_{XX}^{-1}\sigma_{xx}$, which defines the degree of attenuation, is known as the reliability ratio, and measures the reliability of X as a measurement of x. Note that the ratio ranges from zero to 1.0 with larger values indicating greater reliability of measurement.

Another important concept that we will discuss is that of identifiability of a model. Fuller (1987, pp. 9–10) provides the following definition of *identifiability*: "Let $\mathbf{Z}$ be the vector of observable random variables and let $F_{\mathbf{Z}}(\mathbf{a} : \theta)$ be the distribution function of $\mathbf{Z}$, evaluated at $\mathbf{a}$, for the given parameter θ in the parameter space Θ. The parameter θ is identified if, for any $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$, $\theta_1 \neq \theta_2$ implies that $F_{\mathbf{Z}}(\mathbf{a} : \theta_1) \neq F_{\mathbf{Z}}(\mathbf{a} : \theta_2)$ for some $\mathbf{a}$." In addition, we say that if the vector θ is identified, then the "model" is identified (Fuller, 1987, p. 10). By the term "model", we mean a specification of the variables and parameters of interest, the relationships among the variables, and the assumptions about the stochastic properties of the random variables. For the structural model defined by Equations (6) and (7) the vector of unknown parameters, θ, is given by $(\mu_x, \sigma_{xx}, \sigma_{ee}, \sigma_{uu}, \beta_0, \beta_1)$. Under the assumptions of the stated model, the observations (Y_i, X_i) have a bivariate normal distribution. Properties of the bivariate normal distribution state that it will be completely characterized by the elements of its mean vector and variance–covariance matrix. Thus the distribution of (Y_i, X_i) is characterized by the five parameters in its mean vector and variance–covariance matrix, namely, $(\mu_Y, \mu_X, \sigma_{YY}, \sigma_{XX}, \sigma_{XY})$. Because the model contains six different parameters in the vector θ, the relationship between the elements of θ and the parameters of the distribution of the observable random variables is not invertable. That is, there are infinitely many sets of θ's that produce the same parameters of the distribution of the observable variables. In other words, there exist multiple parametric configurations (i.e. different θ) that would lead to the same distribution of the observations. So by applying the definition of identifiability, there exist vectors $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$, such that $\theta_1 \neq \theta_2$, but $F_{\mathbf{Z}}(\mathbf{a} : \theta_1) = F_{\mathbf{Z}}(\mathbf{a} : \theta_2)$, for all $\mathbf{a}$. Thus, the model is not identified.

It is possible, however, that certain individual parameters are identifiable. For this model, μ_x is identified since the mean of $\underline{x}$ is equal to the mean of $\underline{X}$. In order to construct a consistent

estimator for the entire vector θ, however, one must have additional information to identify the model.

3.3 Additional Information Required for Identifiability

Known parameters

Identifying information could come in the form of known parameters, such as known measurement error variance, σ_{uu}, or reliability coefficient, κ_{xx}, for the measured value of x_i.

If the reliability ratio, $\kappa_{xx} = \sigma_{XX}^{-1}\sigma_{xx}$, defined in Section 3.2, is assumed to be known, then the unbiased estimator for the regression coefficient β_1, in the structural model, is given by

$$\widehat{\beta}_1 = \kappa_{xx}^{-1}\widehat{\beta}_{1,OLS},$$

where $\widehat{\beta}_{1,OLS}$ is the least squares coefficient defined by Equation (10). Here, $\widehat{\beta}_1$ is referred to as the regression coefficient corrected for attenuation.

Now let us assume that σ_{uu} is known. Under the given model, $\mathbf{Z}_i = (Y_i, X_i)$ has a bivariate normal distribution and therefore the sample mean $\overline{\mathbf{Z}} = (\overline{Y},\ \overline{X})$ and the sample covariances (m_{YY}, m_{XY}, m_{XX}) form a set of sufficient statistics for parameter estimation. The sample covariances are computed in their usual way, e.g. $m_{XY} = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$. Fuller (1987, p. 14) refers to m_{YY}, m_{XY} and m_{XX} as the maximum likelihood estimators adjusted for degrees of freedom. This is due to the fact that when there are no parametric restrictions on the covariance matrix of $\mathbf{Z}_i$, $n^{-1}(n-1)m_{ZZ}$ is the maximum likelihood estimator of the covariance of $\mathbf{Z}_i$. When the parameter vector is identified, the maximum likelihood estimator will be a function of the sufficient statistics. Recall that the population moments of (Y_i, X_i) under the model defined by (6) and (7) satisfy

$$\begin{aligned}(\sigma_{YY},\ \sigma_{XY},\ \sigma_{XX}) &= \left(\beta_1^2\sigma_{xx} + \sigma_{ee},\ \beta_1\sigma_{xx},\ \sigma_{xx} + \sigma_{uu}\right)\\ (\mu_Y,\ \mu_X) &= (\beta_0 + \beta_1\mu_x,\ \mu_x).\end{aligned} \tag{11}$$

By substituting the sample estimators of the unknown population moments into the left-hand side of Equation (11), one creates a system of equations that can be solved to obtain the parameter estimates. Doing so results in the following estimators:

$$\begin{aligned}\widehat{\beta}_1 &= (m_{XX} - \sigma_{uu})^{-1}m_{XY},\\ (\widehat{\sigma}_{xx},\ \widehat{\sigma}_{ee}) &= (m_{XX} - \sigma_{uu},\ m_{YY} - \widehat{\beta}_1 m_{XY}),\\ (\widehat{\mu}_x,\ \widehat{\beta}_0) &= (\overline{X},\ \overline{Y} - \widehat{\beta}_1\overline{X}),\end{aligned}$$

where $\widehat{\sigma}_{xx}$ and $\widehat{\sigma}_{ee}$ can be negative with positive probability. Knowing σ_{uu} allows us to construct a one-to-one mapping of the minimal sufficient statistic to the vector $(\widehat{\mu}_x,\ \widehat{\sigma}_{xx},\ \widehat{\beta}_0,\ \widehat{\beta}_1,\ \widehat{\sigma}_{ee})$. In order for these estimators to be proper estimators, they must lie in the parameter space. So $\widehat{\sigma}_{xx}$ and $\widehat{\sigma}_{ee}$ must both be non-negative. These two estimators will both be positive as long as $m_{XX} - \sigma_{uu} > 0$ and $m_{YY} - \widehat{\beta}_1 m_{XY} > 0$, or equivalently, $m_{YY}(m_{XX} - \sigma_{uu}) - m_{XY}^2 > 0$.

SIMEX and Regression Calibration

Two other methods for making use of this type of additional information are described next. First, a method known as *simulation-extrapolation*, or SIMEX, was first introduced by Cook & Stefanski (1994) and is also discussed in detail in Carroll *et al.* (1995). This method is employed when there is some additional information available and is useful in the general ME model (Carroll *et al.*, 1995). The basic idea behind SIMEX is that in a simulation step, additional independent measurement error is simulated and added to the original measured

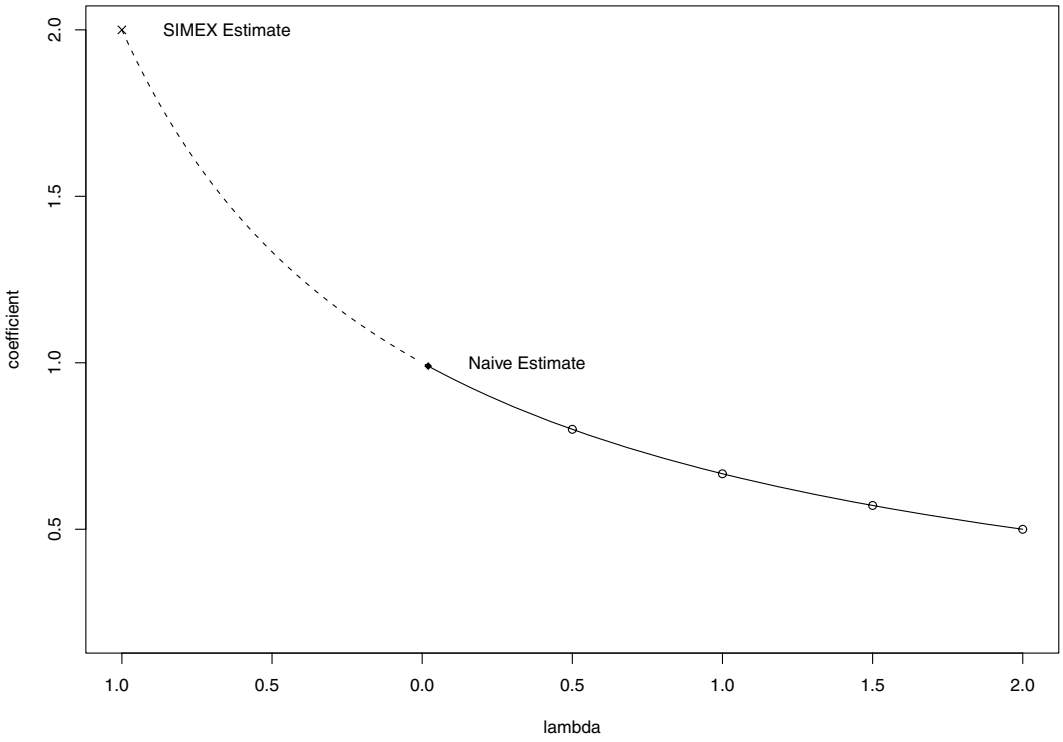

Figure 1. ***SIMEX Plot.*** *A generic SIMEX plot showing the effect of measurement error of size* $(1 + \lambda)\sigma_{uu}$ *on parameter estimates. Note that the SIMEX estimate occurs at* $\lambda = -1$ *and the Naive Estimate occurs at* $\lambda = 0$.

values in order to create additional data sets with successively more variable values of the independent variable. In the extrapolation step, the trend of the bias induced by the measurement error versus the variance of the additional measurement error is determined and then the trend is extrapolated back to the case of no measurement error to provide a nearly unbiased parameter estimate.

The SIMEX procedure is easiest to understand under the simple linear ME model defined by Equations (6)–(9). We still assume that σ_{uu} is known. First additional, independent measurement error with variance $\lambda_m \sigma_{uu}$, where $0 = \lambda_1 < \lambda_2 < \ldots < \lambda_M$, is generated and added to the original data. Cook & Stefanski (1994, p. 1317) recommend letting lambda range from zero to two and in most of their examples they use a rather coarse grid, namely, $\lambda \in \{0, 0.5, 1, 1.5, 2\}$. So the total measurement error in the m-th data set is $(1 + \lambda_m)\sigma_{uu}$ and the least squares estimator of the slope parameter in the m-th data set, $\widehat{\beta}_{1,m}$, would consistently estimate $\beta_1(\frac{\sigma_{xx}}{\sigma_{xx}+(1+\lambda_m)\sigma_{uu}})$. Finally, the problem is thought of as a nonlinear regression problem with $\widehat{\beta}_{1,m}$ as the dependent variable and λ_m as the independent variable, having a mean function of the form $G(\lambda) = (\frac{\beta_1\sigma_{xx}}{\sigma_{xx}+(1+\lambda)\sigma_{uu}})$, $\lambda \geq 0$, and extrapolation back to $\lambda = -1$, the "no-measurement error" case, yields the parameter estimate. See Figure 1 for a generic SIMEX plot.

A second general approach to incorporating additional identifying information into a measurement error analysis was first introduced by Prentice (1982) for the proportional hazards model and as a general approach by Carroll & Stefanski (1990). It is known as the *regression calibration* method. This method is discussed in detail by Carroll *et al.* (1995). Regression calibration, like SIMEX, cannot be implemented without some additional information. In fact,

these should be viewed as methods of incorporating the additional information that allows parameter identification.

The basic idea of regression calibration is to replace x_i by an estimate of $E(x_i \mid X_i)$ and then perform a standard analysis. One drawback of this method is that estimating $E(x_i \mid X_i)$ often requires information that is specific to the current problem only, and there is no general methodology to estimate this quantity. In the simple linear ME model with known κ_{xx} or σ_{uu}, the regression calibration method produces the usual correction for attenuation. This is easily seen with the present simple linear ME model under the current normality assumptions since it is known that $E(x_i \mid X_i) = \mu_x + \sigma_{XX}^{-1}\sigma_{xX}(X_i - \mu_x)$. Note that $\overline{X}$ is an unbiased estimate of μ_x and $\sigma_{XX}^{-1}\,\sigma_{xX} = \sigma_{xx}/(\sigma_{xx} + \sigma_{uu}) = \kappa_{xx}$ is known or can be estimated when κ_{xx} or σ_{uu} is known or estimated. So regression calibration, under the current model, amounts to needing to estimate or have knowledge about κ_{xx}. The resulting estimator using the regression calibration technique under the model defined by (6) and (7) and with assumptions defined in (9) with σ_{uu} known, for example, is $\widehat{\beta}_1 = (m_{XX} - \sigma_{uu})^{-1}m_{XY} = \widehat{\kappa}_{xx}^{-1}\widehat{\beta}_{1,OLS}$. When κ_{xx} is known the regression calibration estimator is $\widehat{\beta}_1 = \kappa_{xx}^{-1}\widehat{\beta}_{1,OLS}$.

Often in the measurement error literature, it is discussed how to include error-free covariates into the model. Using the notation from Carroll *et al.* (1995), we let Z_i represent covariates in the model, which are not contaminated by measurement error. Note that there are no such error-free covariates in the simple linear ME model. The reason a brief discussion of such covariates is given here is that much of the existing literature that specifically implements regression calibration includes such error-free covariates. The extension to include such terms in a regression calibration setting, for example, is straight forward. In the presence of error-free covariates, regression calibration replaces x_i by an estimate of $E(x_i \mid Z_i, X_i)$, rather than just an estimate of $E(x_i \mid X_i)$, and then runs the standard analysis. Other results from above remain substantively unchanged when error-free terms are included in the model. For example, the bias term in the slope parameter would be denoted as $(\sigma_{xx|z} + \sigma_{uu})^{-1}\sigma_{xx|z} = (\sigma_{XX|z})^{-1}\sigma_{xx|z}$ where $\sigma_{XX|z}$ is the residual variance of the regression of x_i on Z_i and $\sigma_{xx|z}$ is the residual variance of the regression of X_i on Z_i (see, for example, Carroll *et al.*, 1995, p. 25). If such error-free covariates exist in a measurement error model, the parameter estimates for their regression coefficients, in general, are biased as well.

Repeated measures of true values

Independent, repeated measures of the truth allows for an independent estimate of σ_{uu} in the simple linear ME model. Suppose there are two replicate observations, then we would have $X_{i1} = x_i + u_{i1}$ and $X_{i2} = x_i + u_{i2}$. So $(X_{i1} - X_{i2}) = (u_{i1} - u_{i2})$ and $S^2_{X_1 - X_2} = \widehat{\mathrm{Var}}(u_{i1} - u_{i2}) = 2\widehat{\sigma}_{uu}$. Therefore $\widehat{\sigma}_{uu} = S^2_{X_1 - X_2}/2$. Then a consistent estimate of κ_{xx}, the reliability ratio, is $\widehat{\kappa}_{xx} = (m_{XX} - \widehat{\sigma}_{uu})/m_{XX}$, since $\widehat{\mathrm{Var}}(X) = \widehat{\mathrm{Var}}(x) + \widehat{\mathrm{Var}}(u)$ under assumption (9), and the resulting estimator, derived from the method of moments, is $\widehat{\beta}_{1,OLS}/\widehat{\kappa}_{xx}$, the bias-adjusted version of $\widehat{\beta}_{1,OLS}$. As in the case with known measurement error, this estimator is also the regression calibration estimator. Usually in cases with repeated observations, the replicate means may be thought of as a better measure of x than a single observation (Carroll *et al.*, 1995, p. 13). Therefore these replicate means are used in place of the single observation throughout the analysis when repeated measures are available, i.e. rewrite the model as $Y_i = \beta_0 + \beta_1 x_i + e_i$ and $\overline{X}_{i\cdot} = x_i + \overline{u}_{i\cdot}$.

With an independent estimate of σ_{uu}, Carroll *et al.* (1995) state that either SIMEX or the regression calibration technique can be used to develop estimators for β_1. The regression calibration estimator is given there as $\widehat{\beta}_{1,OLS}m_{XX}/(m_{XX} - \widehat{\sigma}_{uu}) = \widehat{\kappa}_{xx}^{-1}\widehat{\beta}_{1,OLS}$.

Instrumental variables

Assume, again, that the model defined by (6), (7) and (8), the simple linear ME model, holds. In this context, the definition of an *instrumental variable* (IV), given by Carroll *et al.* (1995, p. 107), is a variable that is correlated with the true variable, x_i, but is uncorrelated with the measurement error, u_i, and the model error, e_i. One possible choice of an instrumental variable is a second measure of x_i from a conditionally independent (given x_i) method of measurement (see, for example, Carter, 1981), or similarly from a second response variable that is correlated to x_i but independent both of u_i and e_i. Greenland (2000) provides a nice introduction and overview of IV's for a nonstatistical audience by discussing their role in epidemiological studies. We will denote the instrumental variable as W_i. Fuller (1987, p. 51) provides a formal definition. According to his definition, under the model defined by (6) and (7), W_i is an instrumental variable if the following two conditions are met:

$$\begin{aligned} &(i) \quad E\left\{n^{-1}\sum_{i=1}^{n}(W_i-\overline{W})(e_i,\,u_i)\right\}=(0,\,0)\\ &(ii) \quad E\left\{n^{-1}\sum_{i=1}^{n}(W_i-\overline{W})x_i\right\}\neq 0 \end{aligned} \tag{12}$$

where $\overline{W}=n^{-1}\sum_{i=1}^{n}W_i$. For convenience, let us express the fact that x_i and W_i are related by using a parametric expression. To do this, we will denote the parameters of the population regression of x_i on W_i by π_2, for the slope, and π_1, for the intercept. Using these parameters we can write $x_i=\pi_1+\pi_2W_i+r_i$, $i=1,2,\ldots,n$. Here, r_i represents the failure of x_i to be perfectly linearly related to W_i, or simply, the model error in this regression equation. By the least squares regression method used to construct this equation, r_i has zero correlation with W_i, i.e. the "error" term is independent of the "independent" variable. By substituting the fact that $X_i=x_i+u_i$, we have $X_i=\pi_1+\pi_2W_i+(r_i+u_i)=\pi_1+\pi_2W_i+a_i$. Note that $E\{\sum_{i=1}^{n}W_ia_i\}=0$ by the assumption $E\{n^{-1}\sum_{i=1}^{n}(W_i-\overline{W})u_i\}=0$.

Under the model defined by (6), (7) and (8) and in the presence of an instrumental variable, there are now 12 independent unknown parameters, namely β_0, β_1, μ_W, π_1, π_2, σ_{xx}, σ_{xe}, σ_{ee}, σ_{xu}, σ_{eu}, σ_{uu}, and σ_{WW}. Here $\mu_x=\pi_1+\pi_2\mu_W$ and note that we do not assume that σ_{eu} or σ_{xu} is zero. Assuming that the IV is normally distributed, along with the assumptions from Equation (8), we have that the vector of the observed data, (Y_i,X_i,W_i), is normally distributed with mean $(\mu_Y,\mu_X,\mu_W)=(\beta_0+\beta_1\pi_1+\beta_1\pi_2\mu_W,\pi_1+\pi_2\mu_W,\mu_W)$ and variance–covariance matrix

$$\begin{bmatrix} \beta_1^2\sigma_{xx}+2\beta_1\sigma_{xe}+\sigma_{ee} & \beta_1\sigma_{xx}+\sigma_{xe}+\beta_1\sigma_{xu}+\sigma_{eu} & \beta_1\pi_2\sigma_{WW} \\ \beta_1\sigma_{xx}+\sigma_{xe}+\beta_1\sigma_{xu}+\sigma_{eu} & \sigma_{xx}+2\sigma_{xu}+\sigma_{uu} & \pi_2\sigma_{WW} \\ \beta_1\pi_2\sigma_{WW} & \pi_2\sigma_{WW} & \sigma_{WW} \end{bmatrix}.$$

We assume $\sigma_{xW}=\pi_2\sigma_{WW}\neq 0$. There are 9 sample statistics that make up the set of minimal sufficient statistics for a sample of n observations. They are from the sample mean vector and sample variance–covariance matrix of (Y_i,X_i,W_i). An estimator for β_1 can then be developed by noting that the ratio of covariances $\sigma_{XW}^{-1}\sigma_{YW}=(\pi_2\sigma_{WW})^{-1}\,\beta_1\,\pi_2\,\sigma_{WW}=\beta_1$. So, we estimate β_0 and β_1 by

$$\widehat{\beta}_1=m_{XW}^{-1}m_{YW}, \tag{13}$$

$$\widehat{\beta}_0=\overline{Y}-\widehat{\beta}_1\overline{X}, \tag{14}$$

where $\overline{Y}$ and $\overline{X}$ are the sample means and m_{XW} and m_{YW} are the sample covariances. Under the stated assumptions, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the maximum likelihood estimators as well as consistent estimators, since the sample moments are consistent estimators of the population moments. Note, however, that due to the number of minimal sufficient statistics we cannot hope to estimate all 12 population parameters, unless additional assumptions are made.

Fuller (1987, p. 53) provides a theorem and proof concerning the asymptotics of the estimators given in (13) and (14). The theorem does not depend on the normality assumption, but does require, along with finite fourth moments, that σ_{We} and σ_{Wu} are zero, $\sigma_{xW} \neq 0$, and $E\{(W_i - \mu_W)^2 v_i^2\} = \sigma_{WW}\sigma_{vv}$ and $E\{v_i^2 (W_i - \mu_W)\} = 0$, where $v_i = e_i - \beta_1 u_i$. If the stated conditions hold,

$$n^{1/2}\begin{bmatrix} \widehat{\beta}_0 - \beta_0 \\ \widehat{\beta}_1 - \beta_1 \end{bmatrix} \rightarrow^L N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{vv} + \mu_x^2 V_{22} & -\mu_x V_{22} \\ -\mu_x V_{22} & V_{22} \end{bmatrix}\right),$$

where $V_{22} = \sigma_{xW}^{-2}\sigma_{WW}\sigma_{vv}$. The reason for the assumptions of zero covariance between W_i and the error terms, e_i and u_i, and the non-zero covariance between x_i and W_i is to allow for the estimation of β_1 when measurement error and x_i are correlated.

Carter & Fuller (1980) discuss instrumental variable estimation and its properties in the simple ME model with slightly more restrictive assumptions. The reason for their more restrictive assumptions, as alluded to above, is to allow for the estimation of all unknown parameters in the model. They assumed that $(x_i, e_i, u_i, W_i)'$, $i = 1, 2, \ldots n$, are distributed as independent drawings from a multivariate normal distribution with a zero mean vector and a covariance matrix, Σ, given by

$$\Sigma = \begin{bmatrix} \sigma_{xx} & 0 & 0 & \sigma_{xW} \\ 0 & \sigma_{ee} & \sigma_{eu} & 0 \\ 0 & \sigma_{eu} & \sigma_{uu} & 0 \\ \sigma_{xW} & 0 & 0 & \sigma_{WW} \end{bmatrix}.$$

They derived restricted maximum likelihood estimators for the models where the error covariance, σ_{eu}, is known to be zero and where σ_{eu} is unknown. The consistency of the restricted maximum likelihood estimator of β_1 was also shown and it's asymptotic distribution derived. See Carter (1981) for details of similar work under these assumptions.

Recall in regression calibration an estimate of $E(x_i \mid X_i)$ is used to replace x_i and then a standard analysis can be run. In the presence of an unbiased IV W_i, (i.e. where $E(x_i \mid X_i) = E(W_i \mid X_i)$, which is the case, for example, when $x_i = \pi_0 + \pi_1 W_i + r_i$ and $\pi_0 = 0, \pi_1 = 1$) Carroll *et al.* (1995) point out that $\widehat{E}(W_i \mid X_i)$ is obtained from the regression of W_i on x_i and is an unbiased estimate of $E(x_i \mid X_i)$. When no such unbiased instrument exists, which is usually the case, and one wishes to estimate $E(x_i \mid X_i)$, recall under normality that this expectation is $\mu_x + \sigma_{XX}^{-1}\sigma_{xX}(X_i - \mu_x)$, where μ_x is estimated unbiasedly by $\overline{X}$ and one is left with the estimation of $\sigma_{XX}^{-1}\sigma_{xX} = \kappa_{xx}$. When an IV is present, it can be shown that an estimate of κ_{xx} may be obtained by regressing W on X. The resulting estimator is $\widehat{\kappa}_{xx} = m_{XY}m_{XW}(m_{YW}m_{XX})^{-1}$, which can be seen from the variance–covariance matrix of the observed data since $\sigma_{XY}\sigma_{XW}(\sigma_{YW}\sigma_{XX})^{-1} = \beta_1\sigma_{xx}\pi_2\sigma_{WW}(\beta_1\pi_2\sigma_{WW}(\sigma_{xx} + \sigma_{uu}))^{-1} = \sigma_{xx}(\sigma_{xx} + \sigma_{uu})^{-1}$, when σ_{xe}, σ_{xu}, and σ_{eu} are zero. In the presence of Z_i covariates measured without error, the extension is again straightforward. In such cases, for example, an estimate of $E(x_i \mid Z_i, X_i)$ may be used to replace x_i prior to running the analysis.

SIMEX estimators rely on knowing or estimating the measurement error variance and using this to generate data sets with successively larger measurement error variances and using these new

data sets then to extrapolate back to the no measurement error case. Because the SIMEX method uses measurement error variance as the identifying piece of information, SIMEX estimators could be used with instrumental variables by using the IV to estimate σ_{uu} and then using $\widehat{\sigma}_{uu}$ in the SIMEX procedure.

3.4 Extension to the Multivariable Linear ME Model

Building up in complexity, the model that follows the simple linear ME model is the multivariable linear ME model. The extension to the multivariable linear ME model is straightforward and thus only briefly mentioned here. A multivariable linear ME model contains more than one x variable and hence writing the multivariable model amounts to using matrix notation to extend the notation of the simple linear ME model. Using standard matrix notation (e.g. underlines indicating column vectors), in the presence of multiple independent variables measured with error, Equation (6) may be written as $Y_i = \beta_0 + \underline{x}_i'\underline{\beta}_1 + e_i$ and Equation (7) becomes $\underline{X}_i = \underline{x}_i + \underline{u}_i$. Other concepts discussed above in the context of the simple linear ME model extend naturally as well. For example, the bias in the naive ordinary least squares estimator of the coefficient vector is expressed as $(\Sigma_{xx} + \Sigma_{uu})^{-1}\Sigma_{xx}$, where Σ would represent the appropriate variance–covariance matrix of the subscripted variable. Another illustration of the extension in matrix notation is seen by providing the IV estimator of the vector of parameters: $\widehat{\underline{\beta}}_1 = (WX')^{-1}(W\underline{Y})$, where capital letters represent appropriate design matrices of observed random variables. More details on the multivariable linear ME model being expressed in matrix notation may be found in references such as Fuller (1980, 1987).

3.5 Numerical Illustration

We give a numerical example that illustrates the correction for attenuation in the simple linear ME model. We present an example that uses an IV as the identifying information. The data presented in Table 1 are measurements of blood glucose taken from three different measurement techniques on 16 "normal" patients. Of the three measurement techniques, one is a manual method, which is thought to be relatively accurate, but is also expensive. The other two techniques, which are less expensive are done by machines, labelled as machine A and machine B. The goal of such a study may be to determine which of the two less expensive techniques, machine A or B, could be used to measure blood glucose.

Assume the model defined in (6), (7) and (8), holds, and that further there exists a normally distributed IV, W_i as defined in Subsection 3.3. Let Y_i be the measurement on the i-th patient taken by machine B, X_i be the measurement on the i-th patient taken by the manual method, and W_i be the measurement on the i-th patient taken by machine A. Here, machine A may be considered an IV since it provides a measure of x_i taken from a conditionally independent method. True blood glucose is the variable contaminated by measurement error and the manual method provides an unbiased measure of it. Normal probability plots of the data reveal no obvious violations to the assumptions of normality of the three observed variables. We further assume that $\sigma_{xe} = \sigma_{xu} = \sigma_{eu} = 0$ so that all remaining model parameters may be estimated.

Using the method of IV estimation discussed in Subsection 3.3 and the maximum likelihood estimates of the covariances, we have $\widehat{\beta}_1 = 376.26/440.91 = 0.853$ with estimated standard error 0.228. Computations of the standard error are done similarly to that given by Carter (1981). To illustrate that an attenuation towards zero has been corrected, if we run the regression of the measurements of machine B on the measurements of the manual method and ignore the measurement error, the naive ordinary least squares estimate of the slope parameter is

Table 1.
Three Measures of Blood Glucose.

Patient	Manual	Machine A	Machine B
1	99	100	94
2	118	118	111
3	94	92	90
4	98	102	96
5	71	70	67
6	96	96	92
7	133	132	125
8	86	88	86
9	137	132	127
10	99	100	96
11	153	150	140
12	116	116	112
13	74	80	78
14	108	108	102
15	88	90	85
16	117	116	110

$\widehat{\beta}_{1,OLS} = 0.847$. To address the over-all main goal of the study that investigates a comparison of the less expensive machines, one would perform a formal analysis to determine, for example, if the error variance of machine A is less than the error variance of machine B. This can be done by regressing the IV on x_i, as opposed to regressing x_i on the IV, as was demonstrated in Subsection 3.3, and then estimating the variance of the error term of that linear regression. When all parameters are estimated by making use of the IV defined above as the identifying information, it can be shown that the estimates of these variances are 0.27 for machine A and 1.13 for machine B. One, of course, would conduct a formal test of hypotheses on these variances prior to stating the formal conclusion that machine A should be used to measure blood glucose over machine B.

4 Nonlinear ME Models

In this section we discuss ME models that are nonlinear in either $\underline{x}_i$ or $\underline{\beta}$. Fuller (1987) says that it is conventional to consider ME models to be nonlinear only when the β's enter the mean function in a nonlinear manner or when the mean function is nonlinear in the explanatory variables measured with error. Let the model be defined by

$$y_i = h(\underline{x}_i; \underline{\beta}) + q_i, \tag{15}$$

$$Y_i = y_i + w_i, \tag{16}$$

$$\underline{X}_i = \underline{x}_i + \underline{u}_i, \tag{17}$$

where $i = 1, 2, \ldots, n$ and $h(.)$ is a real valued continuous, nonlinear function. We may combine Equations (15) and (16) into one equation giving $Y_i = h(\underline{x}_i; \underline{\beta}) + e_i$, where e_i is the sum of both measurement error, w_i, and random equation error, q_i, if they exist, and $\underline{\varepsilon}_i' = (e_i, \underline{u}_i')$ are independent random vectors having mean $\underline{0}$ and covariance matrix $\Sigma_{\varepsilon\varepsilon}$. Often it is assumed that the errors are normally distributed. We will primarily be interested in estimating the unknown vector $\underline{\beta}$. Fuller (1987, p. 226) provides a formal definition of a nonlinear ME model stating the model defined by (15), (16) and (17) is nonlinear if $h(\underline{x}; \underline{\beta})$ is nonlinear in $\underline{x}$ when $\underline{\beta}$ is fixed or

if $h(\underline{x};\ \underline{\beta})$ is nonlinear in $\underline{\beta}$ when $\underline{x}$ is fixed. We will once again focus mainly on the structural case where the $\underline{x}_i$ are assumed to be random with mean $E(\underline{x}_i) = \underline{\mu}_x$ and variance–covariance matrix Σ_{xx}. If $\underline{x}_i$ are assumed to be independent of the errors, then we have

$$\begin{bmatrix} \underline{x}_i \\ e_i \\ \underline{u}_i \end{bmatrix} \sim F\left(\begin{bmatrix} \underline{\mu}_x \\ 0 \\ \underline{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \underline{0} & \underline{0} \\ \underline{0} & \sigma_{ee} & \Sigma_{eu} \\ \underline{0} & \Sigma'_{eu} & \Sigma_{uu} \end{bmatrix} \right), \tag{18}$$

where F might be any distribution assumed for the model. In the next subsection, we discuss the case where F is a multivariate normal distribution.

Not much research has been done on fully general forms of the nonlinear ME model and thus presents an area open for further research. Results on specific forms and under certain assumptions of nonlinear ME models may be found, for example, in Carroll *et al.* (1995).

4.1 Normal Theory Models: Non-Identifiability and Bias

This class of ME models are those whose error distributions are independent of their mean function. Griliches & Ringstad (1970, p. 370) showed that the bias in the classical least squares estimators is exacerbated when the regression function is nonlinear. Their results are for a specific nonlinear model, but they state that their work "can be viewed, however, as an approximation to the estimation of more general nonlinear models". They assumed additive measurement error, $X = x + u$, with u and x normally distributed, $E(u) = 0$, $E(xu) = 0$, and parameterized the model such that $\sigma_{XX} = 1$. The authors denoted σ_{uu} by λ. Their model was $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$, nonlinear in the random variable x, with $E(e) = E(xe) = E(x^2 e) = 0$, and all variables were univariate. They showed that the naive ordinary least squares estimate of β_1, $\widehat{\beta}_1$, is biased towards zero by a factor of $(1 - \lambda)$, where in their notation, λ was the fraction of error variance in the total variance in the observed variable, i.e. σ_{uu}. The problem became even more serious for the nonlinear terms, in that the naive ordinary least squares estimate of β_2, $\widehat{\beta}_2$, was biased towards zero by the square of the bias factor of the linear term. As in the linear ME model, the naive estimator would result from least squares estimation using the observed $\underline{X}$ in place of the latent, unobserved $\underline{x}$.

In the general model, the naive approach is to fit the model $Y_i = h(\underline{X}_i;\ \underline{\beta}) + e_i$ by classical least squares or by a semi-parametric or non-parametric fitting method. Gleser (1990) improves upon this naive estimator in a paper from the proceedings of the 1989 AMS-IMS-SIAM Joint Summer Research Conference. He works under the assumptions that $E(e) = E(u) = 0$ and that e, u, and x are mutually statistically independent. In fitting the model, he takes independent and identically distributed (i.i.d.) observations, (Y_i, X_i), on (Y, X) and therefore (x_i, e_i, u_i) are i.i.d. each with the same distribution as (x, e, u). In his words, the naive approach "in general, leads to inconsistent estimators with a high degree of asymptotic bias" (Gleser 1990, p. 99). He points out that consistent estimators have been obtained in special cases, but the general problem of finding consistent and efficient estimators is still unsolved.

4.2 Normal Theory Models: Additional Information Required

Known parameters

Gleser (1990) discusses solutions to the problem of parameter estimation where identifying parameters are assumed to be known, a case not realistic in application, and also the case where they are estimated from the data. His solution is to replace $\underline{x}$ with the best linear predictor, $\widehat{\underline{x}}$, of

$\underline{x}$ given $\underline{X}$. This amounts to the regression calibration method. He points out that when $\underline{\mu}_x$, Σ_{xx}, and Σ_{uu} are known, $\underline{x}$ and $\underline{u}$ are independent, and $\underline{X}$ is normally distributed

$$E(\underline{x} \mid \underline{X}) = \underline{X}\Lambda + \underline{\mu}_x(I_k - \Lambda), \tag{19}$$

where $\Lambda = (\Sigma_{xx} + \Sigma_{uu})^{-1}\Sigma_{xx}$ is the reliability matrix, I is the $k \times k$ identity matrix, and k is the dimension of $\underline{X}$. Of course, in practice, some of these quantities may not be known. The equality in (19) does not hold when $\underline{X}$ is not normally distributed, but the right-hand side of Equation (19) is still the best linear (in $\underline{X}$) predictor of $\underline{x}$ under mean square error (Gleser, 1990). So when these parameters are known, Gleser suggests fitting the model $Y = h(\underline{X}\Lambda + \underline{\mu}_x(I_k - \Lambda); \underline{\beta}) + e = h(E(\underline{x} \mid \underline{X}); \underline{\beta}) + e$.

An iterative estimation procedure for coefficients of nonlinear "functional" relations was proposed by Wolter & Fuller (1982) and assumes a known covariance matrix of $(e_i, \underline{u}_i')$, i.e. Σ_{eu}, or minimally its order is known. Their estimator is a modification of the maximum likelihood estimator for the nonlinear model with normal measurement error. Assuming $(e_i, \underline{u}_i')$ are independent normal $(\underline{0}, \Sigma_{eu})$ random variables, the maximum likelihood estimators are those values of $\underline{x}$ and $\underline{\beta}$ that minimize the sum of squares

$$\sum_{i=1}^{n} q(\underline{\beta}, \underline{x}_i; Y_i, \underline{X}_i) = \sum_{i=1}^{n} \{Y_i - h(\underline{x}_i; \underline{\beta}), \underline{X}_i - \underline{x}_i\}\Sigma_{eu}^{-1}\{Y_i - h(\underline{x}_i; \underline{\beta}), \underline{X}_i - \underline{x}_i\}'.$$

Under their nonlinear model it is not possible to derive an explicit expression for the MLE of $\underline{\beta}$ but Wolter and Fuller do develop two estimators under slightly different assumptions of the order of Σ_{eu} and provide theorems for the limiting distributions of these iteratively defined estimators.

Repeated measures

In Gleser's solution mentioned above, when $\underline{\mu}_x$ and/or Λ are unknown, consistent estimators $\widehat{\underline{\mu}}$ and $\widehat{\Lambda}$ can be used as substitutes. Such consistent estimators can be obtained from the data, $\underline{X}_1, \ldots, \underline{X}_n$, and information from prior calibration studies on $(\underline{x}, \underline{X})$ (Gleser, 1990). Using repeated measures, an estimate of Σ_{uu} can be obtained as previously mentioned in the sections on repeated measures under the linear ME models, and $\widehat{\Sigma}_{uu}$ can then be used to estimate Λ by $\widehat{\Lambda} = S_{XX}^{-1}(S_{XX} - \widehat{\Sigma}_{uu})$. In such cases, the model $Y = h(\widehat{\underline{x}}; \underline{\beta}) + e$, where $\widehat{\underline{x}} = \underline{X}\widehat{\Lambda} + \widehat{\underline{\mu}}_x(I_k - \widehat{\Lambda})$, should be fitted to estimate $\underline{\beta}$. Gleser (1990) also points out that a better substitution for $h(\underline{x}; \underline{\beta})$ may be obtained from a Taylor series expansion of $h(\underline{x}; \underline{\beta})$ about $\widehat{\underline{x}}$. He proved consistency and asymptotic distributional properties for his estimator.

Instrumental variables

In addition to the model defined by (15), (16) and (17), assume that there are observations available on an instrumental variable, $\underline{W}_i$. In nonlinear models, IV's must still satisfy the same requirements as those stated in Subsection 3.3 for the simple linear ME model. That is, IV's must be correlated with $\underline{x}_i$, independent of the measurement error in $\underline{x}_i$, i.e. independent of $\underline{u}_i$, and must also be surrogate, i.e. $\underline{W}_i$ must be independent of Y_i given $\underline{x}_i$ (which is equivalent to saying independent of model error, q_i).

Not much research has been done on general forms of structural nonlinear ME models that use IV's as the identifying information. In fact, in a recent paper, Carroll *et al.* (2004, p. 736) mention that many of the results in IV estimation methods "are applicable either only for special parametric models or for general parametric models that rely on small-error approximations known to fail for some nonlinear and nonparamteric models".

Buzas (1997) does assume a structural model in using a regression calibration approach with IV's to construct unbiased score equations in nonlinear measurement error models. His method, however, requires the use of an unbiased estimate of the true, unknown variable.

Carroll *et al.* (2004) use Bayesian Markov Chain Monte Carlo to arrive at asymptotically efficient estimators for nonlinear and non-parametric regression models. Their approach uses a Bayesian penalized spline, or P-spline, fitting method similar to that of Berry *et al.* (2002). Details of P-spline fitting are not discussed here, but can be found in Eilers & Marx (1996). In their solution, Carroll *et al.* (2004) present a method for estimating the measurement error variance that makes use of IV's that are linearly related to the unknown, true values. This estimator for the measurement error variance makes use of a root-n-consistent estimator of the variance of the unobserved true measurements that, in turn, is based on the covariance between the IV and the error contaminated observations. They also provide theorems that state identification results for nonlinear and non-parametric models under relatively weak conditions, in the presence of such IV's. They present a polynomial regression case which, in their words, "illustrate(s) how difficult the estimation problem of IV for nonlinear models can be" (Carroll *et al.*, 2004, p. 737).

References

Berry, S.A., Carroll, R.J. & Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *J. Am. Stat. Assoc.*, **97**(457), 160–169.

Buzas, J.S. (1997). Instrumental variable estimation in nonlinear measurement error models. *Commun. Stat.: Theor. Methodol.*, **26**(12), 2861–2877.

Campbell, A. (2002). The potential role of aluminium in Alzheimer's disease. *Nephrol. Dial. Transplant.*, **17**, 17–20.

Carroll, R.J., Ruppert, D., Crainiceanu, C.M., Tosteson, T.D. & Karagas, M.R. (2004). Nonlinear and nonparametric regression and instrumental variables. *J. Am. Stat. Assoc.*, **99**(467), 736–750.

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

Carroll, R.J. & Stefanski, L.A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *J. Am. Stat. Assoc.*, **85**(411), 652–663.

Carter, R.L. (1981). Restricted maximum likelihood estimation of bias and reliability in the comparison of several measuring methods. *Biometrics*, **37**(4), 733–741.

Carter, R.L. & Fuller, W.A. (1980). Instrumental variable estimation of the simple errors-in-variables model. *J. Am. Stat. Assoc.*, **75**(371), 687–692.

Cook, J.R. & Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Am. Stat. Assoc.*, **89**(428), 1314–1328.

Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Stat. Sci.*, **11**(2), 89–121.

Fuller, W.A. (1980). Properties of some estimators for the errors in variables model. *Ann. Stat.*, **8**(2), 407–422.

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

Gleser, L.J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Contemporary Mathematics Volume 112: Statistical Analysis of Measurement Error Models and Applications*, Eds. Brown P. J. and Fuller, W.A., pp. 99–114. Providence, RI, USA: Amer. Math. Soc.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.*, **29**(4), 722–729.

Griliches, Z. & Ringstad, V. (1970). Error-in-the-variables bias in nonlinear contexts. *Econometrica*, **38**(2), 368–370.

Kendall, M. & Stuart, A. (1979). *The Advanced Theory of Statistics. Vol. 2: Inference and Relationship*, London: Griffin.

Klepper, S. & Leamer, E.E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, **52**(1), 163–183.

Neyman, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**(1), 1–32.

Pierce, D.A., Stram, D.O. & Vaeth, M. (1990). Allowing for random errors in radiation dose estimates for the atomic bomb survivor data. *Radiat. Res.*, **123**(3), 275–284.

Pierce, D.A., Stram, D.O., Vaeth, M. & Shafer, D. (1992). The errors in variables problem: Considerations provided by the radiation dose-response analyses of the A-bomb survivor data. *J. Am. Stat. Assoc.*, **87**(418), 351–359.
Prentice, R.L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69**(2), 331–342.
Stefanski, L.A. (2000). Measurement error models. *J. Am. Stat. Assoc.*, **95**(452), 1353–1358.
Wolter, K.M. & Fuller, W.A. (1982). Estimation of nonlinear errors in variables models. *Ann. Stat.*, **10**(2), 539–548.

Résumé

Cet article donne une introduction et une vue d'ensemble au modèle souvent sous-utilisé d'erreur de mesure. Le but est de fournir un résumé simple des problèmes qui surgissent de l'erreur de mesure et des solutions qui ont été proposées. Nous commençons par décrire comment les modèles d'erreur de mesure se produisent dans des situations réelles. Alors nous continuons de définir le modèle d'erreur de mesure, présentant au commencement la forme multivariable du modèle, et puis, commençant par la forme la plus simple du modèle discutez complètement ses dispositifs et solutions à l'en raison présenté par problèmes de l'erreur de mesure. Nous discutons des formes dimensionnelles et plus avançées plus élevées du modèle et donnons une brève illustration numérique.

[*Received January 2006, accepted April 2007*]