Applied Mathematical Sciences, Vol. 4, 2010, no. 55, 2733 - 2742

Bias Terms of Measurement Error in Treatment by expected Estimating Equations

Manoochehr Babanezhad and Farhad Yaghmaei

Golestan University, Department of Statistics
Faculty of Sciences, Gorgan, Golestan, Iran
m.babanezhad@gu.ac.ir, mbaba22@yahoo.com

Abstract

Inference for the effect of treatment on an outcome often suffers from the unavailability of valid measurements of treatment. Due to this, systematic errors in measurements of treatment occur frequently and are inevitable in practice. Measurement error in treatment thus forms a common source of bias in treatment effect estimates on outcome. This study provides strategy for correcting bias by expected estimating equations in regression models. We proceed this in both continuous and categorical treatment with measurement error assumptions.

Keywords: Estimating equations, Bias, Measurement error, Misclassification

1 Introduction

The so-called ‘measurement error problem’ or ‘errors-in-variables problem’ arises in situations where the treatment X is difficult to measure or cannot be accurately measured for all study subjects. For instance, in environmental problems individual levels of pollution and radiation are difficult to measure, or in agricultural studies the amount of the fertilizer actually absorbed by the plant is a quantity which cannot be accurately measured [1, 2, 3]. Then, random or systematic errors occur in measurements of treatment X. In this paper, we provide strategy for correcting bias by expected estimating equations in regression models. We refer to error in a continuous treatment X as plain measurement error. When X is discrete (categorical), which is often the case in agriculture applications, then misclassification is the term to use. The paper is organized as follows. Section 2 presents the structure of systematic error in measurements of continuous or discrete (categorical) treatment under regression models. In section 3, we extract the bias terms through expected estimating equations and conclusion in section 4.

2 Models for measurement error

The structure of measurement error or misclassification in treatments under regression models initially requires specifying models for the error process. Such models quantify the relationship between the true treatment X and the observed treatment W. In the literature on the measurement error problem [1, 2, 4, 5], two general types of measurement error model are commonly considered for continuous X. The standard error model is typically the 'Classical measurement error model' in which, $W = X + U$, where U, the measurement error term, is assumed to be independent of X; that is, $U \perp\!\!\!\perp X$. In practice, it is common to assume that the measurement error U has mean 0 so that $\mathrm{E}(U|X) = \mathrm{E}(U) = 0$. This implies that $\mathrm{E}(W|X) = X$, suggesting that W is an unbiased measure of X. In fact, virtually all developments on measurement error assume the error to be normally distributed with mean zero and could be homoscedastic or heteroscedastic [6]. The 'Berkson error model' is an alternative to the classical measurement error model. It is based on the assumption that the measurement error is independent of the observed treatment W, in the sense that, $X = W + U$, where U is independent of W; that is, $U \perp\!\!\!\perp W$ [2, 6].

In case where $X$is discrete, the measurement error model can be defined in terms of conditional misclassification probabilities [5]. Misclassification error basically differs from measurement error because the observed treatment variable W cannot be expressed as a sum of the true treatment variable X with an error variable. Rather, one must characterize the measurement error in terms of misclassification probabilities. These probabilities are often expressed as the probability of the observed treatment W given the true treatment X; that is, $P(W|X)$. For dichotomous variables, it is conventional to express these through the sensitivity, $\pi_{1|1} = P(W = 1|X = 1)$, and the specificity, $\pi_{0|0} = P(W = 0|X = 0)$.

3 Impact of measurement error and correcting bias

In the measurement error analysis, measurement error can be either differential or non-differential. Non-differential measurement error in the presence of an error-free covariate Z occurs when Y is independent of W, given X and Z; that is, $Y \perp\!\!\!\perp W|X, Z$ so that $f_{Y|W,X,Z} = f_{Y|X,Z}$. If this assumption fails, then the error is said to be differential. This assumption is useful, because it greatly simplifies the link between the association of Y and W and the association of Y and X.

3.1 Linear regression models

We begin with an illustration of the effects of measurement error for the case of homoscedastic ordinary linear regression where treatment X is continuous,

$$\mathrm{E}(Y_i|X_i, Z_i) = \beta_0^* + \beta_1^* X_i + \beta_2^* Z_i, \tag{1}$$

where Z_i is an error-free covariate, X_i is an error-prone treatment with mean μ_x and variance σ_x^2, and $\beta^* = (\beta_0^*, \beta_1^*, \beta^*)$ is an unknown finite-dimensional parameter, with β_1^* encoding the conditional association between X_i and Y_i (given Z_i). Under the classical measurement error model, $W_i = X_i + U_i$, we observe a variable W_i instead of X_i, where the measurement error U_i is independent of (X_i, Z_i) with mean zero $\mu_u = 0$. Here, we will additionally assume U_i to be normally distributed with constant variance σ_u^2 which is assumed known or can be estimated from supplementary data [2]. Suppose that the primary interest of the study lies in the conditional association β_1^* of X and Y. When the investigator is unaware of the measurement error or chooses to ignore it, he/she may simply regress Y on (W, Z), and would then not obtain a consistent estimate of β^*, but instead obtain an estimate of the naive parameter $\theta^* = (\theta_0^*, \theta_1^*, \theta_2^*)$ indexing the following naive regression model,

$$\mathrm{E}(Y_i|W_i, Z_i) = \theta_0^* + \theta_1^* W_i + \theta_2^* Z_i. \tag{2}$$

The latter is implied by model (1) by the fact that the relationship between Y and (W, Z) is greatly simplified when the measurement error is non-differential,

$$\begin{aligned} \mathrm{E}(Y|W, Z) &= \mathrm{E}\{\mathrm{E}(Y|W, Z, X)|W, Z\} \\ &= \mathrm{E}\{\mathrm{E}(Y|X, Z)|W, Z\} = \beta_0^* + \beta_1^* \mathrm{E}(X|W, Z) + \beta_2^* Z. \end{aligned} \tag{3}$$

The latter implies that the regression of Y on (W, Z) is equal to the regression of Y on $\{\mathrm{E}(X|W, Z), Z\}$. In regression models, one often solves the estimating equation corresponding to that regression model to obtain regression coefficient estimates. The estimating equation is called unbiased if it has expectation zero when evaluated at the true parameter values. If it is unbiased, its solution is a consistent and asymptotically normal estimator for the considered parameters. In the absence of measurement error a consistent and asymptotically normal estimator for β^* under model (1) can be obtained by solving,

$$0 = \sum_{i=1}^{n} U_i(Y_i, X_i, Z_i; \beta^*) = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_i \\ Z_i \end{pmatrix} (Y_i - \beta_0^* - \beta_1^* X_i - \beta_2^* Z_i) .$$

By replacing W instead of X, $U(Y_i, W_i, Z_i; \beta^*)$ may not generally be unbiased. Under the naive model (2), the limiting parameter θ^* is obtained by solving the following expected estimating equation

$$0 = \mathrm{E}\{U_i(\theta^*)\} = \mathrm{E}\left\{ \begin{pmatrix} 1 \\ W_i \\ Z_i \end{pmatrix} (Y_i - \theta_0^* - \theta_1^* W_i - \theta_2^* Z_i) \right\} . \qquad (4)$$

Comparing this with the true model (1) then yields a bias formula. Note that, we use $U(\theta^*)$ to denote an estimating function, and U for random measurement error.

For simplicity, we first start with models in which there is no error-free covariate. Consider a linear model (1) and assume the classical error model holds. As stated, with the classical error model, measurement error U is independent of X with mean 0 and variance σ_u^2. It follows $\mathrm{E}(W|X) = X$; $Var(W|X) = \sigma_u^2$; $\mathrm{E}(W) = \mu_x$; $Var(W) = \sigma_x^2 + \sigma_u^2$; and $Cov(W, X) = Cov(X + U, X) = \sigma_x^2$. Under regression model (2) with no covariate Z, the naive coefficient estimators can be obtained by solving,

$$0 = \mathrm{E}\{U(\theta^*)\} = \mathrm{E}\left\{ \begin{pmatrix} 1 \\ W \end{pmatrix} (Y - \theta_0^* - \theta_1^* W) \right\} = \mathrm{E}\left[\mathrm{E}\left\{ \begin{pmatrix} 1 \\ W \end{pmatrix} (Y - \theta_0^* - \theta_1^* W) | X, W \right\} \right] .$$

This yields

$$\theta_1^* = \frac{Cov(W, X)}{Var(W)} \beta_1^* = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \beta_1^* .$$

Let

$$\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \qquad (5)$$

and $\theta_0^* = \beta_0^* + (1-\lambda)\beta_1^*\mu_w$. As a result, ordinary least squares regression yields biased estimators of the regression slopes of error-prone treatments. In particular, because $\lambda < 1$, the least squares regression coefficient θ_1^* is biased towards zero. This bias does not vanish with increasing sample size. In the measurement error literature, the attenuation factor λ is called the 'reliability ratio' (variance of true treatment variable divided by variance of measured treatment variable, possibly given the error-free covariate); it expresses the degree of attenuation. It suggests that measurement error bias increases with decreasing treatment variance. If there is information on the magnitude of the error variance and the distribution of X, then the above results allow in principle to correct for measurement error in estimating regression slopes, at least for reasonably simple forms of measurement error. An asymptotically unbiased estimator of β_1^* is then given as follows,

$$\hat{\beta}_1 = \frac{\hat{\theta}_1}{\lambda} \tag{6}$$

where $\hat{\theta}_1$ is the ordinary least squares estimate of θ_1^*. The resulting estimator (6) is sometimes called the regression coefficient corrected for attenuation. Further, $\mathrm{E}(\hat{\beta}_1) = \beta_1^*$ and $Var(\hat{\beta}_1) = Var(\hat{\theta}_1)/\lambda^2$. Because $\lambda < 1$, it is clear that $Var(\hat{\beta}_1) > Var(\hat{\theta}_1)$. This implies correcting for bias entails that the corrected estimator will be more variable than the biased estimator and then have wider confidence intervals. This generally means that the price for reduced bias is increased variance. It follows from the normality assumption and the classical error model that

$$\begin{aligned}
Var(Y|W) &= Var(\beta_0^* + \beta_1^* X + \epsilon|W) \\
&= \beta_1^{*2} Var(X|W) + \sigma_\epsilon^2 \\
&= \beta_1^{*2} \frac{\sigma_u^2\sigma_x^2}{\sigma_x^2 + \sigma_u^2} + \sigma_\epsilon^2,
\end{aligned}$$

Consider now model (1) with error-free covariate Z and the classical error model. It follows from the classical error model that $E(W|Z) = E(X|Z)$, $\mathrm{E}(XW|Z) = \mathrm{E}\{X(X+U)|Z\} = \mathrm{E}(X^2|Z)$, and $Cov(W,X|Z) = Cov(X + U, X|Z) = \sigma_{x|z}^2$. Here we suppose that X and Z are linearly related, $E(X|Z) = \eta_0^* + \eta_1^* Z$.

Under the non-differential measurement error assumption, by solving,

$$0 = \mathrm{E}\{U(\theta^*)\} = \mathrm{E}[\mathrm{E}\{U(\theta^*)|X, W, Z\}]$$

$$= \mathrm{E}\left\{\begin{pmatrix} 1 \\ W \\ Z \end{pmatrix} \mathrm{E}(Y|X,Z) - \theta_0^* - \theta_1^* W - \theta_2^* Z\right\}.$$

we can obtain

$$\theta_1^* = \lambda_z \beta_1^*, \tag{7}$$

where

$$\lambda_z = \frac{\sigma^2_{x|z}}{\sigma^2_{x|z} + \sigma^2_u}, \tag{8}$$

and $\sigma^2_{x|z} = Var(X|Z)$. A similar derivation shows that the coefficient of Z is also biased, unless Z is independent of X; $\theta_2^* = \beta_2^* + (1-\lambda_z)\beta_1^*\eta_1^*$. The bias of this regression slope is thus $(1-\lambda_z)\beta_1^*\eta_1^*$. Note that hypothesis tests for the regression slope β_1^* are valid in the presence of random, non-differential measurement error (in the sense of preserving the nominal Type I error rate) because, there is no bias (see expression 7) under the null hypothesis that $\beta_1^* = 0$. Tests may, however, be less powerful than in the absence of measurement error. It follows that, when $E(X|W,Z) = \alpha_0^* + \alpha_1^* W + \alpha_2^* Z$,

$$\mathrm{E}(Y|W,Z) = \beta_0^* + \beta_1^*\alpha_0^* + \beta_1^*\alpha_1^* W + (\beta_2^* + \beta_1^*\alpha_2^*)Z. \tag{9}$$

This implies that the naive model test that none of the predictors are useful for explaining variation in Y is valid in the sense of having the desired Type I error rate. Specifically, examination of (2) and (9) shows that $\theta_2^* = 0$ is equivalent to $\beta_2^* = 0$, only if $\beta_1^*\alpha_2^* = 0$. It follows that the naive test of $H_0 : \beta_2^* = 0$ is valid only if X is unrelated to Y conditional on Z ($\beta_1^* = 0$) or if Z is unrelated to X ($\alpha_2^* = 0$). The naive tests that are valid, that is, those that maintain the Type I error rate, will still suffer reduced power relative to the test based on the true data.

Moreover, when the error in X follows the Berkson error model, $Cov(W,X) = Cov(W, W+U) = Var(W)$. This implies that $\lambda = 1$. That is, the naive estimator of slope is an unbiased estimator of β^*, but there is an increase in the residual variance because under model (1),

$$Var(Y|W) = Var(\beta_0^* + \beta_1^* X + \epsilon|W) = \beta_1^{*2}\sigma_u^2 + \sigma_\epsilon^2.$$

3.2 Nonlinear regression models

Regression coefficients in generalized linear models, including models of particular interest such as logistic regression (or probit regression), are affected by measurement error in much the same manner as are linear model regression coefficients [2, 5, 7]. We now consider the nonlinear model with logistic link without covariates Z,

$$\text{logit} P(Y = 1|X) = \beta_0^* + \beta_1^* X. \tag{10}$$

Assume that the error in X is non-differential and follows the classical measurement error model. Then the observed data model implied by,

$$P(Y = 1|W) = \int_x P(Y = 1|W, X) f_{X|W}(x|w)dx = \int_x P(Y = 1|X) f_{X|W}(x|w)dx.$$

This integral is not easy to handle, and to the best of our knowledge there is no closed form solution for the bias expressions. Now consider the above model with the probit link,

$$\Phi^{-1}\{P(Y = 1|X)\} = \beta_0^* + \beta_1^* X.$$

Under the assumption of normality, we can evaluate the latter integral by the probit link. As stated, $X \sim N(\mu_x, \sigma_x^2)$ and $U \sim N(0, \sigma_u^2)$ then

$$\text{E}(X|W) = \mu_x + \rho_{xw}\frac{\sigma_x}{\sigma_w}(W - \mu_w)$$

$$Var(X|W) = (1 - \rho_{xw}^2)\sigma_x^2 = \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_u^2}\right)^{-1}.$$

Then the latter integral can be written as

$$\begin{aligned} P(Y_i = 1|W_i) &= \int_{-\infty}^{\infty} \Phi(\beta_0^* + \beta_1^* x_i) f(x_i; \mu_i^*, \sigma^{*2})dx_i \\ &= \Phi\left\{\frac{\beta_0^* + \beta_1^*\mu_i^*}{\sqrt{1 + \beta_1^{*2}\sigma^{*2}}}\right\} \end{aligned}$$

where $\sigma^{*2} = (\frac{1}{\sigma_x^2} + \frac{1}{\sigma_u^2})^{-1}$ and $\mu_i^* = \text{E}(X_i|W_i)$ for $i = 1, ..., n$ [8]. A direct comparison of the latter with the naive model, $P(Y_i = 1|W_i) = \Phi(\theta_0^* + \theta_1^* W_i)$ yields

$$\theta_1^* = \frac{\lambda\beta_1^*}{\sqrt{1 + \sigma_u^2\lambda\beta_1^{*2}}}$$

and $\theta_0^* = \frac{\beta_0^* + (1-\lambda)\mu_w\beta_1^*}{\sqrt{1+\sigma_u^2\lambda\beta_1^{*2}}}$. The close relationship between the logit and probit form, namely $G(t) = (1 + \exp(-t))^{-1} \approx \Phi(t/h)$ with $h = 1.70$, allows us to obtain an asymptotic bias formula for the logistic regression coefficients in model (10).

As explained in Section 2, the situation in which a discrete variable is measured with error, is referred to as misclassification. In the case of a dichotomous treatment X, the probability $\pi_{1|1,z} = P(W = 1|X = 1, Z)$, for instance expresses how likely it is for someone who is truly exposed with covariate level Z to be classified as exposed. Likewise, $\pi_{0|0,z} = P(W = 0|X = 0, Z)$ expresses how likely it is for someone who is truly unexposed with covariate level Z, to be classified as unexposed. The extent to which $\pi_{1|1,z}$ and $\pi_{0|0,z}$ are less than 1 reflects the severity of the degree of misclassification, with 1 indicating no misclassification error. Consider now a linear regression model (1) with no error-free covariate Z for a continuous outcome Y given a dichotomous treatment variable X which is subject to misclassification. Assuming non-differential measurement error

$$\begin{aligned}
\mathrm{E}(Y|W) &= \mathrm{E}\{\mathrm{E}(Y|X)|W\} \\
&= \beta_0^* + \beta_1^* P(X=1|W) \\
&= \beta_0^* + \beta_1^* \{WP(X=1|W=1) + (1-W)P(X=1|W=0)\} \\
&= \beta_0^* + \beta_1^* \frac{\pi_{0|1}\mu_x}{1-\mu_w} + \beta_1^* \left\{ \frac{\pi_{1|1}\mu_x}{\mu_w} + \frac{\pi_{0|0}(1-\mu_x)}{1-\mu_w} - 1 \right\} W
\end{aligned}$$

where $\pi_{w|x} = P(W = w|X = x)$ for $w = 0, 1$ and $x = 0, 1$. A direct comparison of the latter with the naive model, $\mathrm{E}(Y|W) = \theta_0^* + \theta_1^* W$, shows that

$$\theta_1^* = \frac{\mu_x(1-\mu_x)}{\mu_w(1-\mu_w)}(\pi_{1|1} + \pi_{0|0} - 1)\beta_1^* \tag{11}$$

and $\theta_0^* = \beta_0^* + \frac{(1-\pi_{1|1})\mu_x}{1-\mu_w}\beta_1^*$. Let $\kappa = \frac{\mu_x(1-\mu_x)}{\mu_w(1-\mu_w)}(\pi_{1|1} + \pi_{0|0} - 1)$. Substituting $\mu_w = 1 - \pi_{0|0} + (\pi_{1|1} + \pi_{0|0} - 1)\mu_x$ into (11) yields

$$\kappa = \frac{\mu_x(1-\mu_x)}{\{1-\pi_{0|0} + (\pi_{1|1}+\pi_{0|0}-1)\mu_x\}\{\pi_{0|0} - (\pi_{1|1}+\pi_{0|0}-1)\mu_x\}}(\pi_{1|1} + \pi_{0|0} - 1).$$

More generally when there is an error-free covariate Z in model (1) with a dichotomous treatment variable X, misclassification error may be expressed in terms of

$$P(W = 1|X, Z = z) = 1 - \pi_{0|0,z} + (\pi_{1|1,z} + \pi_{0|0,z} - 1)X,$$

where $\pi_{w|x,z}$ is related with the error-free covariate Z for $w = 0, 1$ and $x = 0, 1$. Babanezhad et al. [3] investigate the asymptotic bias of the ordinary least squares estimate of the regression coefficient in terms of reclassification probabilities when they are related to the error-free covariate Z. We now investigate the bias of the ordinary least squares estimate of the regression coefficients β^* when $\pi_{w|x,z} = \pi_{w|x}$ for $w = 0, 1$ and $x = 0, 1$. We can obtain by solving the expected estimating equation under the non-differential measurement error, that

$$\theta_1^* = \kappa_1 \beta_1^* \tag{12}$$

$\theta_0^* = \beta_0^* + \mu_x\beta_1^* - \mu_w\theta_1^* + \mu_z(\theta_2^* - \beta_2^*)$
$\theta_2^* = \beta_2^* + \beta_1^*\rho\frac{\sqrt{\mu_x(1-\mu_x)}}{\sigma_z}\left\{1 - (\pi_{1|1} + \pi_{0|0} - 1)\frac{\theta_1^*}{\beta_1^*}\right\}$
where $\rho = \rho_{xz}$ and the coefficient
$\kappa_1 = (\pi_{1|1} + \pi_{0|0} - 1)\left\{\frac{\mu_x(1-\mu_x)(1-\rho^2)}{\mu_w(1-\mu_w)-\mu_x(1-\mu_x)(\pi_{1|1}+\pi_{0|0}-1)^2\rho^2}\right\}$ is attenuation factor.

4 Conclusion

In this paper, we have investigated the impact of treatment measurement and misclassification error on the asymptotic bias of different regression models. Our interest in this stems from the fact that when no adjustments are made for measurement errors, the bias of effect estimates can grow unexpectedly large and lead to a loss of efficiency. It is revealed that in the data analysis that the biased estimate of fertilizer is obtained.

References

[1] W.A. Fuller, *Measurement Error Models* . New York, John Wiley, 1987.

[2] R.J. Carroll, D. Ruppert, L.A. Stefanski and C.M. Crainiceanu, *Measurement Error in Nonlinear Models, Second Edition.* CRC Press, 2006.

[3] M. Babanezhad, S. Vansteelandt and E. Goetghebeur, Comparison of causal effect estimators under exposure misclassification, *Journal of Statistical Planning and Inference*, 140 (2010), 1306–1319.

[4] L.A. Stefanski and J.S. Buzas, Instrumental variables estimation in binary regression measurement error models. *Journal of The American Statistical Association*, 90 (1995), 541–550.

[5] P. Gustafson, *Measurement Error and Misclassification in Statistics and Epidemiology Impacts and Bayesian Adjustments.* Press/CRC, 2003.

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

[7] S. Greenland and P. Gustafson, Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. *American Journal of Epidemiology*, 164 (2006), 63–68.

[8] I.M. Heid, H. Küchenhoff, J. Wellmann, M. Gerken and et al., On the potential of measurement error to induce differential bias on odds ratio estimates: an example from radon epidemiology. *Statistics in Medicine*, 21 (2002), 3261–3278.

Received: March, 2010