

Bayesian and Non-Bayesian Sensitivity Analysis for Correcting Measurement Error

Manoochehr Babanezhad¹

Abstract

Measurements of the exposure are often contaminated by the systematic error. The systematic error then causes inconsistent exposure effect estimates and leads to erroneous conclusions to various degrees in statistical analysis. While some methods develop to correct for measurement error, techniques for correcting this still are controversial. In this paper, we compare Bayesian and Non-Bayesian methods to correct measurement error. This is performed both through theoretically and simulations. We investigate the efficiency of these methods by estimating the effect of hypothetical exposure which is subject to measurement error by sensitivity analysis.

Mathematics Subject Classification: 62C10

Keywords: Bayesian, Simulation extrapolation, Measurement error, Error Structure

1 Introduction

In most statistical analysis the measurements of exposure are frequently

¹ Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran, e-mail: m.babanezhad@gu.ac.ir.

subject to error [5, 7]. That is, the true values of exposure cannot be precisely measured and we observe some contaminated measurements instead. The systematic error is here the difference between the value of true exposure, which is unobserved, and the value, in which are observed, of measured. As is discussed by many authors [3, 5, 7], measurement error reduces power of exposure effect estimate on an outcome of interest. Ignoring this issue may lead to serious bias estimate. Despite the rather elaborate work on correcting the measurement error in exposure over the past decades, techniques for correcting this still are controversial.

In non-Bayesian context the methods of simulation extrapolation (SIMEX) and instrumental variable (IV) to surmount measurement error are carried out during [2, 3]. SIMEX method originally proposed by Cook and Stefanski, which is well suited to estimate and reduce the bias due to additive measurement error. While the literature on measurement error relies on the availability of an auxiliary dataset, we establish the method of instrumental variable (IV) for correcting measurement error [1]. To implement the instrumental variables (IVs) remedy for the measurement error, one must have a variable an IV, that is correlated with the true exposure and not the measurement error [1, 6]. Furthermore IV must be correlated with outcome only through its correlation with exposure. The aim of this paper is to compare Bayesian and (two) non-Bayesian methods for measurement error when exposure is subject to complex error model. Further, a sensitivity analysis for the performance of considered methods with different error variability was performed. Moreover, in this paper, we deal the correction of measurement error when the structure of the considered error model is different with classic and Berkson error models.

2 Complex Error Structure

Suppose that a model explores the relationship between a outcome Y and exposure X by a conditionally mean function,

$$Y_i = E(Y_i|X_i) + \epsilon_i \quad (1)$$

where e.g, $E(Y_i|X_i) = \beta_0^* + \beta_x^*X_i$ in linear, and $E(Y_i|X_i) = \text{expit}(\beta_0^* + \beta_x^*X_i)$ in nonlinear case [1, 6].

Complex error structure is deviation from the classical additive error [2]. This

type of error allows correlation between errors in linear regression model with the error in the exposure. That is

$$W = \gamma_0 + \gamma_1 X + U \quad (2)$$

where W is observed or error-prone exposure, and U refers to the measurement error component independent of X ($U \perp\!\!\!\perp X$). This non-classical model assumes the true exposure is biased, that is, $E(W|X) \neq X$. In addition, it can be assumed the measurement error is non-differential. The assumption of non-differential measurement error refers to the fact that the distribution of W depends only on the actual exposure X and not on the response variable or other variables in the model. That is, the conditional distribution of $f(y|x, w)$ is identical to the conditional distribution of $f(y|x)$. Under the assumption that (X, U, ϵ) are jointly normal distributed, we will obtain $\beta_0 = \beta_0^* + \beta_x^* \mu_x - \beta_1(\gamma_0 + \gamma_1 \mu_x)$ and

$$\beta_w = \frac{\beta_x^* \gamma_1 \sigma_x^2 + \rho \sqrt{\sigma_x^2 \sigma_u^2}}{\gamma_1 \sigma_x^2 + \sigma_u^2} \quad (3)$$

where $E(Y_i|W_i) = \beta_0 + \beta_w W_i$.

3 Correcting Measurement Error Method

Over the past decades, a number of statistical techniques have been proposed for correcting the impact of measurement error in an exposure X . The choice of methods depends on the information distribution of the variables, the magnitude of the error variance, and the type of error model. They indeed differ according to the assumptions about the distribution of the unobserved exposure X , the availability of additional data about the unobserved exposure X and the theoretical background of the approach, which may be parametric or nonparametric. In this section, we investigate two methods of non-Bayesian so called simulation extrapolation (SIMEX) and instrumental variable (IV). SIMEX is a useful tool for correcting measurement error in a very broad range of settings. This is the only method that provides a visual display of the impact of measurement error on regression parameter estimations. To use the methods of IV, we start by making natural assumptions regarding the conditional densities of all the variables of the model.

Assumptions. (i) $f(y|x, w, z) = f(y|x)$ and (ii) $f(w|x, z) = f(w|x)$, where Z is an instrumental variable. To use the Bayesian paradigm, the posterior density of the unknown quantities is given by

$$f(x, \theta|w, \gamma) \propto f(x, w|\gamma, \theta)f(\theta) \quad (4)$$

where $f(x, \theta|w, \gamma)$ is the joint posterior distribution of the true exposures X and the observed exposure W , θ refers to the unknown parameter, and $f(\theta)$ is prior distribution of θ .

4 Sensitivity Analysis

To compare of the considered methods, a sensitivity analysis of the measurement error variability with assuming $U \sim N(\mu_u = 0, \sigma_u^2 = 0.25)$ in the model (2) is carried out. All simulation results are based on 1000 replications. The sensitivity analysis was carried out under the assumption of normality and nondifferential measurement error. We have generated $Z \sim N(0, 1)$ as continuous IV, $W|Z \sim N(0.5 + 0.25Z, \sigma_{w|z}^2)$, and $X|Z \sim N(0.5 + 0.25Z, \sigma_{x|z}^2)$. For the Bayesian method (4), we consider the prior density of $\theta \sim N(1, 4)$ and conjugate posterior density. The results were summarized in Table I. Posterior means and 95% equal-tailed credible intervals of the estimation obtained for the models under the naive Bayesian analysis and the proposed Bayesian method to correct for measurement error are depicted in Table 1. This table indicates that under the two types of analysis, the point and credible intervals estimates of the estimation under both cases are not similar, suggesting that a slight adjustment is manifested for the level of measurement error exhibited in the sensitivity analysis. Figure 1 shows how variability exits on the true and observed exposure. In addition, it shows the the behaviour of Simulation extrapolation method.

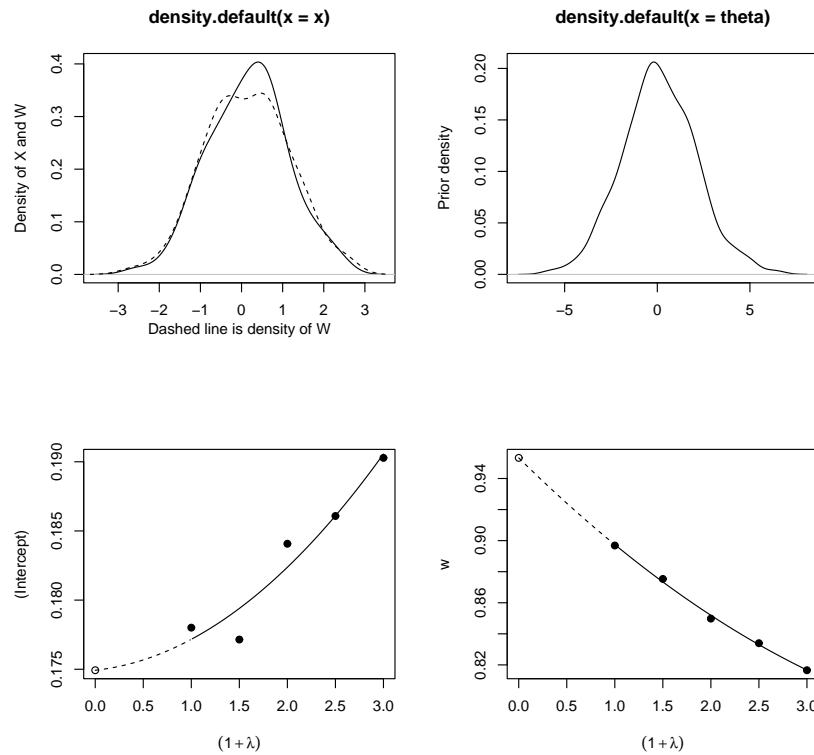


Figure 1: True and error-prone, and prior distribution. The behavior of simulation extrapolation method

5 Conclusion

In this paper we compare three methods of correcting complex systematic measurement error. In sensitivity analysis shows Bayesian method performs well under with small variance of U and outperform SEMIX and IV techniques. While there is misspecified prior density and posterior density, sensitivity analysis shows IV and SEMIX are doing well and outperform Bayesian method. Moreover, the analysis shows also in case where the variability of U is somewhat large by postulating correct prior density and posterior density three considered methods are doing approximately the same (unreported). This might be because the structure of measurement error is complex. We have also carried out with the additive classic and Berkson error models. This shows three considered methods are doing approximately the same (unreported) whether misspecified prior and posterior density. The choice of the SIMEX approach al-

Table 1: Sensitivity analysis where U is normally distributed in two cases with $\mu_u = 0$; $\sigma_u^2 = 0.5$ and $\sigma_u^2 = 0.25$

Methods	Coefficient	95 % C.I.	P.value
Bayesian	0.17	(0.03, 0.30)	0.04
SIMEX	0.33	(0.12, 0.65)	0.02
IV	0.45	(0.10, 0.85)	0.02
Bayesian	-0.11	(-0.23, 0.04)	0.01
SIMEX	-0.33	(-0.58, 0.31)	0.04
IV	-0.29	(-0.14, 0.72)	0.07

lows us to consider linear or nonlinear regression models, because this approach does not depend on the functional form of the measurement error model. In general, SIMEX [3, 4], and IV can reduce the bias except for the structure of complex error in simple linear regression. IV tends more relatively stable on estimation with the increasing of measurement error compared with SIMEX.

References

- [1] M. Babanezhad and F. Yaghmaei, Bias terms of measurement error in treatment by estimating equations, *Applied Mathematical Sciences*, **55**, (2010), 2733-2742.
- [2] R.J. Carroll, D. Ruppert, L.A. Stefanski and C.M. Crainiceanu, *Measurement Error in Nonlinear Model*, Second Edition, CRC Press, 2006.
- [3] J. Cook and L. Stefanski, A simulation extrapolation method for parametric measurement error models, *Journal of the American Statistical Association*, **89**, (1994), 1314-1328.
- [4] I.M. Heid, H. Küchenhoff, J. Wellmann, M. Gerken, L. Kreienbrock 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.

- [5] G.E. Hernandez, P. Gustafson¹ and I. Burstyn, Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study, *BMC Medical Research Methodology*, **11**(67), (2011), 1-11.
- [6] Y. Hu and S.M. Schennach, Instrumental variable treatment of nonclassical measurement error models, *Econometrica*, **76**, (2008), 195-216.
- [7] J. Liu, P. Gustafson, N. Cherry and I. Burstyn, Bayesian analysis of a matched case-control study with expert prior information on both the misclassification of exposure and the exposure-disease association, *Statistics in Medicine*, **28**(27), (2009), 14-19.