Impact of Systematic Measurement Error on Inference: Sensitivity Analysis Using Beliefs*

Max Gallop
Duke University
max.gallop@duke.edu

Simon Weschle
Duke University
simon.weschle@duke.edu

June 12, 2012

Abstract

Abstract

*Version 0.1

1 Introduction

2 Theory

Suppose we observe a variable $\mathbf{x}$ with elements x_i, where $i = 1, \ldots, n$. It is believed that this variable has some systematic measurement error for one of the reasons described above. Denote the true but unknown values of $\mathbf{x}$ with $\mathbf{t}$. We suspect that variable $\mathbf{w}$ systematically influences how close the elements x_i are to t_i. The relationship between $\mathbf{w}$, $\mathbf{x}$ and $\mathbf{t}$ is unknowable, but researchers can formulate beliefs about it. In this section, we present a hierarchical model approach that allows researchers to assess what their conclusions about the substantive inference they are interested in should be, give the beliefs they hold about how $\mathbf{w}$ influences $\mathbf{t}$ to produce $\mathbf{x}$. Both $\mathbf{x}$ and $\mathbf{w}$ can be continuous or non-continuous. For each of the four possible combinations a somewhat different approach is necessary. We discuss them in turn.

2.1 Both x and w Continuous

Suppose $x_i \in \mathbb{R}$ is an element of $\mathbf{x}$ and $t_i \in \mathbb{R}$ is an element of $\mathbf{t}$. We specify the following hierarchical model describing the relationship between w_i, t_i, and x_i:

$$\begin{aligned}
t_i &= x_i \cdot m_i \\
m_i &= \alpha_0 + \textstyle\sum_{k=1}^{K} \alpha_k w_i^k \\
\alpha_0 &\overset{\text{iid}}{\sim} F_0 \\
&\vdots \\
\alpha_K &\overset{\text{iid}}{\sim} F_K
\end{aligned} \tag{1}$$

The term m_i is a constant with which the observed value x_i is multiplied to give the true value t_i. If $m_i = 1$ then x_i measures t_i correctly, if $m_i < 1$ then x_i is larger than t_i and if $m_i > 1$ then x_i is smaller than t_i.

The value of m_i depends on w_i and the relationship is modeled through a polynomial of order k where α_0 is the intercept and α_1 is the first-order coefficient of w_i on m_i and so on. The intercept and coefficients are draws from independent and identically distributed random variables with density $F_1, \dots F_k$. It is through these distributions that the researcher quantifies her beliefs how w_i influences the difference between t_i and x_i. In the most simple case they are just constants – but this is only appropriate if the researcher is absolutely sure about the data generating process. More appropriately, they are standard distributions such as the Normal or more complex mixture distributions. We describe prior elicitation in more detail in section 4.

2.2 x Continuous, w Non-Continuous

Now suppose x_i is still continuous but w_i is non-continuous (nominal or ordinal) with L categories. In this case we specify the following hierarchical model describing the relationship between w_i, t_i, and x_i:

$$\begin{aligned} t_i &= x_i \cdot m_i \\ m_i &= \textstyle\sum_{l=1}^{L} \alpha_l \mathbb{I}_{w_i=l} \\ \alpha_1 &\overset{\text{iid}}{\sim} F_1 \\ &\vdots \\ \alpha_L &\overset{\text{iid}}{\sim} F_L \end{aligned} \tag{2}$$

The term m_i, is still a constant with which the observed value x_i is multiplied to give the true value t_i. The value of m_i depends on which of the L categories w_i is. There is an α_l, which is an i.i.d. draw from a distribution F_l, for each of the L categories ($\mathbb{I}$ is an indicator function that take the value of unity if the condition specified is fulfilled and zero otherwise). In essence, the researcher specifies her beliefs about how $\mathbf{w}$ influences the difference between $\mathbf{t}$ and $\mathbf{x}$ for each category separately.

2.3 x Non-Continuous, w Continuous

2.4 Both x and w Non-Continuous

3 Monte Carlo Study

4 Empirical Examples

5 Discussion