

# Predictive validity under partial observability

## Erratum

Eduardo Alarcón-Bustamante, Ernesto San Martín, and Jorge González

February 12, 2021

In [Alarcón-Bustamante, San Martín, and González \(2020\)](#) it is shown that

$$D_{0x} = \max_{x \in X} \left\{ \frac{2\mathbb{E}(Y|X = x, Z = 1) - y_0 - y_1}{2\mathbb{P}(Z = 0|X = x)} \frac{d\mathbb{P}(Z = 0|X)}{dX} \Big|_{X=x} - \frac{\mathbb{P}(Z = 1|X = x)}{\mathbb{P}(Z = 0|X = x)} M.E_{Z=1}^{X=x} \right\}.$$

However, this expression is incorrect because it depends on an assumption incorrectly wrote, namely

$$0 \leq \frac{d\mathbb{E}(Y|X)}{dX} \Big|_{X=x}.$$

The assumption should be focused on the non-observed group, more specifically

$$0 \leq \frac{d\mathbb{E}(Y|X, Z = 0)}{dX} \Big|_{X=x}; \quad (1)$$

that is, we assume that the selection test is such that higher scores would translate to higher values of the outcome. Under this assumption, it is clear that  $D_{0x} = 0$ , and therefore identification bounds for the marginal effect are given by:

$$\mathbb{P}(Z = 1|X = x)M.E_{Z=1}^{X=x} + [y_0 - \mathbb{E}(Y|X = x, Z = 1)] \frac{d\mathbb{P}(Z = 0|X)}{dX} \Big|_{X=x} \leq M.E^{X=x} \leq$$

$$D_{1x}\mathbb{P}(Z = 0|X = x) + \mathbb{P}(Z = 1|X = x)M.E_{Z=1}^{X=x} + [y_1 - \mathbb{E}(Y|X = x, Z = 1)] \frac{d\mathbb{P}(Z = 0|X)}{dX} \Big|_{X=x} \quad (2)$$

where  $D_{1x} = \max_{x \in X} \{M.E_{Z=1}^{X=x}\}$ . This means, by considering that the selection process correctly selects, the marginal effect in the non-observed group can not be higher than the maximum observed marginal effect. Under the identification restriction (1), the new plots for the Marginal effect of both Mathematics test score and Language and Communication one are shown in Figures (1) and (2), respectively.

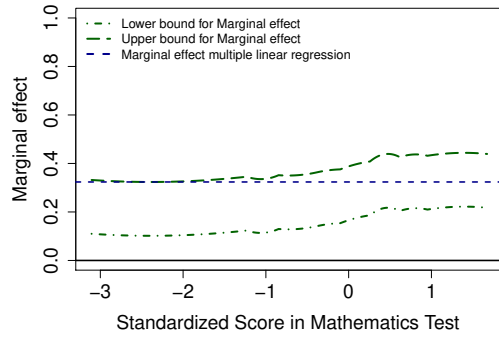


Figure (1): Identification bounds for marginal effect in Mathematics test.

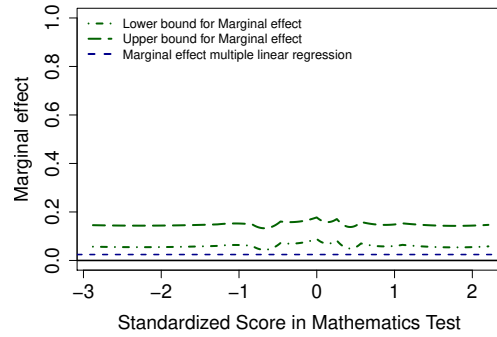


Figure (2): Identification bounds for marginal effect in Language and Communication test.

The blue-dashed lines show us the marginal effect of the multiple linear regression, the currently used model in Chile to measure the Marginal Effect of selection tests. This model is estimated under the ignorability assumption, i.e., by using the available information only.

Identification bounds capture all the plausible solutions for marginal effects when it is assumed that the selection process correctly selects. Under this irrefutable and a non-testable assumption, we are analysing what can be concluded when we combine it with the data. Thus, if a policymaker is ready to believe in (1), it can be concluded that the marginal effect computed under ignorability is coherent with it, as well as the case in the Mathematics test. In contrast, the multiple linear regression model does not allow to conclude the marginal effect of the Language and Communication test over the GPA because its marginal effect is out of the bounds. Hence, it is a not a plausible solution under the identification restriction (1). Moreover, considering that the marginal effect in this test is lower than the lower bound, we can conclude that the current methodological strategy used in Chile for learning about the marginal effect of the Language and Communication test is a non-pertinent solution.

Finally, if we assume that the system select well in the evaluated undergraduate program, and a linear regression model is used to extract conclusions about the marginal effect of these selection tests, it is better to use the Mathematics test instead the Language one.

## References

- Alarcón-Bustamante, E., San Martín, E., & González, J. (2020). Predictive validity under partial observability. In M. Wiberg, D. Molenaar, J. González, U. Böckenholt, & J.-S. Kim (Eds.), *Quantitative psychology* (pp. 135–145). Cham: Springer International Publishing.

Quantitative Psychology

84th Annual Meeting of the Psychometric Society,  
Santiago, Chile, 2019

Wiberg, M.; Molenaar, D.; Gonzalez, J.; Böckenholt, U.;  
Kim, J.-S. (Eds.)

2020, IX, 397 p. 68 illus., 32 illus. in color., Hardcover  
ISBN: 978-3-030-43468-7