A cautionary tale on instrument vector calibration for the treatment of unit nonresponse in surveys

Éric Lesage[*], David Haziza [†] and Xavier D'Haultfœuille [‡]

July 1, 2017

Supplementary material

Performance of the variance estimator

1 Setup

We conducted a simulation study in order to assess the properties of variance estimator $\widehat{V}_1$, given in equation (25) of the main article, in terms of relative bias. We generated a finite population of size $N = 50,\ 000$, each consisting of a variable of interest Y_k, a set of calibration variables $X_k^{(\Gamma_1,\Gamma_2)}$, an instrument Z_k and an unobserved variable U_k. Then we let $\boldsymbol{X}_k = \left(1, X_k^{(\Gamma_1,\Gamma_2)}\right)^\top$ and $\boldsymbol{Z}_k = (1, Z_k)^\top$. The variables Z_k and U_k were first generated from a uniform distribution $\left(-\sqrt{3}, \sqrt{3}\right)$ so that $\mathbb{E}(Z_k) = \mathbb{E}(U_k) = 0$ and $\mathbb{V}(Z) = \mathbb{V}(U_k) = 1$. Then, given Z_k, Y_k was generated according to two models:
(i) a linear regression model:

$$Y_{1,k} = 10 + 5Z_k + \varepsilon^y_{1,k}, \tag{1}$$

where the errors $\varepsilon^y_{1,k}$ were generated from a normal distribution with mean equal to 0 and variance equal to 4. The resulting coefficient of determination was equal to 85%;
(ii) an exponential model:

$$Y_{2,k} = \exp(2.5\ Z_k) + \varepsilon^y_{2,k}, \tag{2}$$

where the errors $\varepsilon^y_{2,k}$ were generated from a normal distribution with mean equal to 0 and variance equal to 4.

[*]INSEE, Paris, France
[†]Department of mathematics and statistics, Université de Montréal, Montreal, Canada
[‡]CREST-ENSAE, Paris, France

Finally, given the values of Z_k and U_k, the $X_k^{(\Gamma_1,\Gamma_2)}$-values were generated according to the linear regression model

$$X_k^{(\Gamma_1,\Gamma_2)} = \Gamma_1 Z_k + \Gamma_2 U_k + \sigma_{(\Gamma_1,\Gamma_2)} \varepsilon_k^{(\Gamma_1,\Gamma_2)},$$

where $\sigma^2_{(\Gamma_1,\Gamma_2)} = 1 - \Gamma_1^2 - \Gamma_2^2$ and the errors $\varepsilon^{(\Gamma_1,\Gamma_2)}$ were normally distributed with mean equal to 0 and variance equal to 1. We used the following values for Γ_1 and Γ_2: $\Gamma_1 \in \{0.2, 0.4, 0.6\}$ and $\Gamma_2 \in \{0, 0.1, 0.3, 0.5\}$.

From the population, we selected $K = 10\ 000$ samples, of size $n = 1,500$ according to simple random sampling without replacement.

In each population, units were assigned a response probability p_k according to

$$p_k = \frac{1}{2 + 0.35\ Z_k} + 0.1\ U_k.$$

This led to an overall response rate of around 50%. Finally, the response indicators R_k were generated independently from a Bernoulli distribution with parameter p_k, $k \in U$.

We computed the instrumental calibration estimator $\widehat{t}_C(\Gamma_1, \Gamma_2)$, and its variance estimator $\widehat{V}_C(\Gamma_1, \Gamma_2)$ based on linear weighting for different values of Γ_1 and Γ_2. The weights were computed so that the calibration constraints

$$\sum_{k \in s} d_k R_k F(\widehat{\boldsymbol{\lambda}}^\top \mathbf{Z}_k) \boldsymbol{X}_k = \sum_{k \in U} \boldsymbol{X}_k$$

were satisfied.

We computed the Monte Carlo percent relative bias of $\widehat{V}_1$ given by

$$RB_{MC}\left(\widehat{V}_1\right) = 100 \times \frac{\left\{\mathbb{E}_{MC}\left(\widehat{V}_1\right) - \mathbb{V}_{MC}(\widehat{t}_c)\right\}}{\mathbb{V}_{MC}(\widehat{t}_c)}.$$

2 Results

The results are presented in Table 1 and Table 2. The results in both tables suggest that the proposed variance estimator performs very well in terms of relative bias in all the scenarios with an absolute relative bias less than 4.0%.

Γ_1 \ Γ_2	0	0.1	0.3	0.5
0.6	-0.3	-1.7	-1.1	-1.5
0.4	-3.2	0.4	-2.6	-0.2
0.2	2.2	3.9	4.0	2.6

Table 1: Monte Carlo percent relative bias of the variance estimators, $\widehat{V}_C(\Gamma_1, \Gamma_2)$, of the instrumental calibration estimator for different pairs (Γ_1, Γ_2) corresponding to population generated according to (1.)

Γ_1 \ Γ_2	0	0.1	0.3	0.5
	0	0.1	0.3	0.5
0.6	-1.0	-1.2	-1.1	-0.6
0.4	-2.5	0.5	-1.6	-0.0
0.2	2.4	3.4	3.6	3.2

Table 2: Monte Carlo percent relative bias of the variance estimators, $\widehat{V}_C(\Gamma_1, \Gamma_2)$, of the instrumental calibration estimator for different pairs (Γ_1, Γ_2) corresponding to population generated according to (2).