A PRACTICAL USE FOR INSTRUMENTAL-VARIABLE CALIBRATION

Phillip S. Kott
National Agricultural Statistics Service, Fairfax, VA 20230, USA

KEY WORDS: Remainder weights; Model unbiased; Randomization consistent; population mean; Remainder mean

1. Introduction

Recently, Estavao and Särndal (2000) introduced a "functional form" calibration estimator for $T = \sum_U y_k$, where U is a population of N elements. Their estimator can be expressed as

$$t_{CALF} = \sum_{k \in S} w_k y_k, \qquad (1)$$

where S is the sample,

$$w_k = a_k + (\sum_{i \in U} \mathbf{x}_i - \sum_{i \in S} a_i \mathbf{x}_i)(\sum_{i \in S} q_i \mathbf{z}_i' \mathbf{x}_i)^{-1} q_k \mathbf{z}_k', \qquad (2)$$

$a_k = 1/\pi_k \geq 1$ is the original sampling weight for element k, $\mathbf{x}_k$ is a row vector of J auxiliary variables associated with k, q_k is an arbitrary constant, and $\mathbf{z}_k$ is a row vector of J instrumental variables, some of which may also be components of $\mathbf{x}_k$. This assumes that $\sum_S q_i \mathbf{z}_i' \mathbf{x}_i$ is invertible. The w_k are called "calibration weights" in part because they satisfy the calibration equation, $\sum_U \mathbf{x}_k = \sum_S w_k \mathbf{x}_k$.

It is easy to show that t_{CALF} is an unbiased estimator for T under the model $y_k = \mathbf{x}_k \beta + \epsilon_k$, where $E(\epsilon_k | \mathbf{x}_k) = 0$. Moreover, t_{CALF} is randomization consistent under mild conditions, which we assume here to hold. Finally, under those same conditions and some equally mild restrictions on the variance structure of the ϵ_k, the anticipated variance (model expected randomization mean squared error) of t_{CALF} is asymptotically invariant to the choice of q_k and $\mathbf{z}_k$.

This estimator is an interesting, if not new, generalization of the standard GREG made popular by Särndal et al. (1992). An earlier version of t_{CALF} can be found in Brewer et al. (1988), although not in calibration form. In practice, it is not obvious why one would contemplate using a vector for $\mathbf{z}_k$ other than $\mathbf{x}_k$ itself, the usual GREG formulation. As for q_k, it is frequently set equal to a_k. Brewer (1994), however, has argued that setting $q_k = a_k - 1$, the *remainder weight*, more often returns a set of calibration weights where $w_k \geq 1$ for all elements in the sample. Many find this a desirable property since then each sample element can be thought of as at least representing itself.

In this note, we will consider the scenario where $\mathbf{x}_k = (1, x_k)$, and the x_k vary within the population and the sample (so matrices are invertible when need be). Using the remainder weights for the q_k helps assure that all $w_k \geq 1$ (for those k in the sample). The addition of a well chosen $\mathbf{z}_k$, introduced in Section 2, makes that property even more likely. Section 3 discusses the more modest goal of finding a complete set of positive calibration weights. Section 4 contains a modest empirical investigation. The discussion in Section 5 concludes that a particular choice of the $\mathbf{z}_k$ will produce a set of calibration weights with all $w_k \geq 1$ if any such set exists except in a degenerative circumstance.

2. The Instrument and its Calibration Weights

Let $\mathbf{z}_k = (1, z_k)$, where $z_k = 1$ when $x_k \geq A$, and $z_k = -1$ otherwise. *A* can be anywhere within the range of the x_k. As suggested in the introduction, we also let $q_k = a_k - 1$, the remainder weight.

Let S_1 be that part of the sample for which $z_k = 1$, and S_2 be the complement of S_1 within the sample. Let m be the remainder-weighted mean of the x_k in S, m_1 be the remainder-weighted mean of the x_k in S_1, and m_2 be the remainder-weighted mean of the x_k in S_2. Let $\hat{R}_1$ be the sum of the remainder weights in S_1, $\hat{R}_2$ be the sum of the remainder weights in S_2, and $\hat{R} = \hat{R}_1 + \hat{R}_2$. Finally, let M_R be the mean value of all x_k in $R = U - S$; that is, those elements in the universe but not in the sample. We will let R also stand for the size of R. Not surprisingly, it is estimated by $\hat{R}$. Under many designs, the two are identical.

Inspecting equation (2) one can see that the calibration weights are invariant to transformations of $\mathbf{z}_k$ or $\mathbf{x}_k$ (e.g., $\mathbf{z}_k$ can be replaced by $\mathbf{z}_k \mathbf{H}$ where $\mathbf{H}$ is any nonsingular J xJ matrix without it affecting the result). Consequently, we can replace each x_k in $\mathbf{x}_k$ by $x_k - m$. In the matrix $\mathbf{z}_k$, we can replace the 1 by $1/\hat{R}$, each z_k in S_1 by $1/\hat{R}_1$, and each z_k in S_2 by $-1/\hat{R}_2$. As a result of these substitutions, the 2 x 2 matrix $\sum_S q_i \mathbf{z}_i' \mathbf{x}_i$ becomes diagonal. Its upper left hand corner contains a 1, and its lower right the value $m_1 - m_2$.

A little manipulation reveals

$$w_k = a_k + (N - \sum_{i \in S} a_i)(a_k - 1)/\hat{R} + (m_1 - m_2)^{-1} [\sum_{i \in U} (x_i - m) - \sum_{i \in S} a_i (x_i - m)](a_k - 1) c_k, \qquad (3)$$

where $c_k = 1/\hat{R}_1$ when $k \in S_1$, and $c_k = -1/\hat{R}_2$ otherwise. Observe that for sample designs where $\sum_S a_i = N$, equation (3) has a much simpler form:

$$w_k = a_k + (m_1 - m_2)^{-1} [\sum_{i\in U} x_i - \sum_{i\in S} a_i x_i](a_k - 1)c_k,$$

Continuing from equation (3),

$$w_k = a_k + (R - \sum_{i\in S} [a_i - 1])(a_k - 1)/\hat{R} + (m_1 - m_2)^{-1} [\sum_{i\in R} (x_i - m) - \sum_{i\in S} (a_i - 1)(x_i - m)](a_k - 1)c_k$$

$$= 1 + (a_k - 1) + (R - \hat{R})(a_k - 1)/\hat{R} + (m_1 - m_2)^{-1} R(M_R - m)c_k$$

$$= 1 + (a_k - 1)\{(R/\hat{R}) + (m_1 - m_2)^{-1} R(M_R - m)c_k\}.$$

$$= 1 + (a_k - 1)R (m_1 - m_2)^{-1} \{(m_1 - m_2)/\hat{R} + (M_R - m)c_k\}$$

$$= 1 + (a_k - 1)(R/\hat{R}_1)(m_1 - m_2)^{-1} (M_R - m_2) \quad (4.1)$$ when $k \in S_1$

$$= 1 + (a_k - 1)(R/\hat{R}_2)(m_1 - m_2)^{-1} (m_1 - M_R) \quad (4.2)$$ when $k \in S_2$.

This last step uses the equality $\hat{R}_1 m_1 + \hat{R}_2 m_2 = \hat{R} m$.

It is easy to see that w_k is equations (4.1) and (4.2) will be 1 or greater as long as $m_2 \le M_R \le m_1$. Now, m_1 is a randomization consistent estimator of the mean of the x_k values in R that are greater than or equal to A, while m_2 is a randomization consistent estimator of the mean of the x_k values in R that are less then A.

In principle, A can be anywhere within the range of the x_k in U. In practice, it makes sense to put it somewhere in the "center" of the distribution. Although the population median seems a reasonable choice, the population mean proved more effective in the modest empirical example to be discussed in Section 4. In Section 5, we see that setting $A = M_R$ will find calibration weights with all $w_k \ge 1$ if such a set exists. Whatever the choice of A, it needs to be made before one looks at the sample. Otherwise, t_{CALF} might not really be randomization consistent.

3. Sample Weights Versus Remainder Weights

One can think of the conventional ratio estimator as having the same form of t_{CALF} in equation (1) with $\mathbf{x}_k = x_k$, $\mathbf{z}_k = 1$, and $q_k = a_k$, the original sample weight of element k. Brewer (1979) proposed a variant with $q_k = a_k - 1$, what we have called the "remainder weight" because $\sum_S (a_k - 1)y_k$ estimates $\sum_R y_k$. Each of the calibration weights under the conventional ratio formulation must be positive as long as all $x_i \ge 0$ and one sample element has a positive x-value, since $w_k = [\sum_U x_i / \sum_S a_i x_i]a_k$. Brewer's approach assures more. No calibration weight will be less than 1, since $w_k = 1 + \{\sum_R x_i / \sum_S [a_i - 1]x_i\}[a_k - 1]$, and $a_k \ge 1$. Note that $[a_i - 1]x_i$ must be positive for at least one sample element for Brewer's w_k to be defined. That is to say, at least one noncertainty sample element must have a positive x-value.

A similar thing happens in our scenario. Defining z_k as in the previous section but letting $q_k = a_k$, the interested reader can derive these calibration-weight formulae:

$$w_k = a_k (N/\hat{N}_1^*)(m_1^* - m_2^*)^{-1} (M - m_2^*) \quad (5.1)$$ when $k \in S_1$

$$= a_k (N/\hat{N}_2^*)(m_1^* - m_2^*)^{-1} (m_1^* - M) \quad (5.2)$$ when $k \in S_2$,

where $\hat{N}_1$ ($\hat{N}_2$) is the sum of the w_k in S_1 (S_2), m_1^* (m_2^*) is the sample-weighted mean of the x_k in S_1 (S_2), and M is the mean of the x_k in U. Under this regime, all the calibration weights are positive when $m_1^* < M < m_2^*$. This is no guarantee, however, that each weight is at least 1.

4. A Modest Empirical Investigation

In this section, we investigate self-weighted samples of size 16 drawn from a very large population, U. The population is so large that the differences between equations (4) and (5) (virtually) vanish. The x-values are generated by a chi-squared distribution with 1 degree of freedom. The population mean of the x_k in U is assumed to be 1, the mean of the chi-squared distribution. Likewise, its median is assumed to be 0.455.

Although we are rarely interesting in samples of size 16 in practice, this example has instructive value. Moreover, it is not that uncommon to use a separate regression estimator where there are a few of 16 elements per stratum.

Table 1 displays the results of a simulation comparing calibration weights computed using equation (5) with conventional simple regression weights:

$$w_k' = a_k[1 + 16(1 - m^*) (x_k - m^*) / \sum_{i\in S} (x_i - m^*)^2], \quad (6)$$

where m^* is the sample-weighted mean of the x_i. In the first 1000 simulations, A is set equal to the population mean, 1. It is not possible to compute equation (5) in five simulations, because all sample x-values are less

than 1. Consequently, S_1 is empty, and m_1* does not exist. In the other 995 simulations, the calibration weights are all positive. By contrast, equation (6) produces a nonpositive weight in 6.7% of the simulations. (We are focusing on nonpositive weights here because N is assumed to be so large that virtually any positive weight will be greater than 1.)

Note that is conceivable for the largest x-value in a sample to be exactly 1, rendering equation (5) computable and a calibration weight equal to zero. That did not happen in any of the 1,000 simulation.

In the next set of simulations, equation (5) is calculated using the population median as A. The calibration weights can always be calculated in each of the1000 simulations, but some weights are nonpositive in 3.6% of them. Although this is better than using the simple regression weights, it is not as good as setting A equal to the population mean had been.

Table 1 also displays results from simulations using the sample mean and then the sample median as A. Using the sample median produces nonpositive weights in *more* simulations than the conventional regression method. Using the sample mean is much better, but not as good as using the population mean.

According to Table 1, setting A equal to the population mean (A = 1) is clearly the best thing to do. Increasing the sample size to 25 has little qualitative effect on the results, except that complete sets of positive calibration weights become more common. On the one hand, using equation (5) with A = 1 produces a positive calibration weight for every sample element in all 1,000 simulation (not displayed). On the other, equation (6) with 25 replacing 16 returns at least one nonpositive calibration weight in only 2.3% of the simulations. A small fraction, but not zero.

5. Discussion

The population mean works well as A in our modest simulations because equation (5) will always return nonnegative weights as long as there is a single sample element with an x-value greater than or equal to the population mean and a single same element with an x-value below the population mean.

When N is not nearly infinity, equation (4) can be different from equation (5). The former was constructed to assure that no calibration weight would be less than 1. If we set $A = M_R$ (the mean x-value among population elements not is the sample), then that will always be the case as long as there is a single sample element with an x-value greater than or equal to M_R and single sample element with an x-value below M_R. This is why equation (4) with $A = M_R$ is usually preferable to equation (5) with A = M. Nevertheless, under certain unusual conditions, it is possible that equation (5) will return all positive weights, while equation (4) will not be computable. This can happen when there is no sample element with an x-value greater than or equal to M_R but there is one with an x-value greater than $M < M_R$.

Suppose $x_k \le x_{smax} < M_R$ for all elements k in the sample, so that equation (4) with $A = M_R$ is not computable. It is easy to see that no set of calibration weights satisfying $w_k \ge 1$ exists. Suppose one did. Then $\sum_S (w_k - 1) = R$, and $\sum_S (w_k - 1)x_k = \sum_R x_k = RM_R$. But

$$\sum_S (w_k - 1)x_k / \sum_S (w_k - 1) \le \sum_S (w_k - 1)x_{smax} / \sum_S (w_k - 1) = x_{smax} < M_R.$$

A contradiction. An analogous argument applies when $x_k \ge x_{smin} > M_R$ for all elements in the sample. There remains the possibility that the minimum value of x_k in the sample is M_R, so that equation (4) is not computable (because m_2 is undefined) but a set of calibration weights all greater than or equal 1 exists (with $w_i = 1$ whenever $x_i \ne M_R$).

It is easy to see that unless the minimum value of x_k in the sample is M_R, equation (4) with $A = M_R$ fails to produce a set of calibration weights that are all 1 or greater only when no such set exists. Moreover, when the minimum value of x_k in the sample is M_R, we can redefine z_k to be 1 when $x_k > M_R$ and -1 otherwise. Thus, except in the degenerate case where every $x_k \in S$ equals M_R, equation (4) (or something like it) with $A = M_R$ fails to produce a set of calibration weights that are all 1 or greater only when no such set exists. A similar series of arguments can be made about equation (5) with A = M and nonnegative calibration weights.

Table 1. Fractions of 1,000 Simulations With at Least One Nonpositive Weight

Choice for A	Using Equation (5)	Using Equation (6)[a]
Population Mean	0.005[b]	0.067
Population Median	0.036	0.065
Sample Mean	0.025	0.060
Sample Median	0.087	0.055

[a] These values vary because they are based on different simulations

[b] Calibration weights could not be calculated at all in five simulations.

6. References

Brewer, K.R.W. (1979). A Class of Robust Sample Designs for Large-Scale Surveys. *Journal of the American Statistical Association*, 83, 128-132.

Brewer, K.R.W. (1994). Survey Sampling Inference: Some Past Perspectives and Present Prospects. *Pakistan Journal of Statistics*, 10(1)A, 213-233.

Brewer, K.R.W., Hanif, M., and Tam, S.M. (1988). How Nearly Can Model-Based and Prediction and Design-Based Estimation Be Reconciled? *Journal of the American Statistical Association*, 83, 128-132.

Estevao, V.M. and Särndal, C-E. (2000). A Functional Form Approach to Calibration. *Journal of Official Statistics*, 16, 379-400.

Särndal, C-E, Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.