▩ 연구논문

A Note on the Small-Sample Calibration

Beong-Soo So
Dept. of Statistics, Ewha Womans University

Abstract

We consider the linear calibration model: $y_i = \alpha + \beta x_i + \sigma\varepsilon_i$, $i=1, \cdots, n$, $y = \alpha + \beta x + \sigma\varepsilon$ where $(y_1, \cdots, y_n, y)$ stands for an observation vector, $\{x_i\}$ fixed design vector, (α, β) vector of regression parameters, x unknown *true* value of interest and $\{\varepsilon_i\}$, ε are mutually uncorrelated measurement errors with zero mean and unit variance but otherwise *unknown* distributions. On the basis of simple *small-sample low-noise* approximation, we introduce a new method of comparing the mean squared errors of the various competing estimators of the true value x for *finite* sample size n. Then we show that a class of estimators including the classical and the inverse estimators are consistent and first-order efficient within the class of all regular consistent estimators irrespective of type of measurement errors.

1. Introduction

We consider the following linear calibration model :

$$y_i = \alpha + \beta(x_i - \bar{x}) + \sigma\varepsilon_i, \ i=1, \cdots, n \qquad (1.1)$$
$$y = \alpha + \beta(x - \bar{x}) + \sigma\varepsilon$$

where $\{y_i\}$, y represent observations of the response variable, $\{x_i\}$ known values of the variable of interest with $\bar{x} = \sum_{i=1}^{n} x_i / n$, (α, β) vector of regression parameters, x unknown *true* value of interest and $\{\varepsilon_i\}$, ε are mutually uncorrelated measurement errors with zero mean and unit variance but otherwise *unknown* distributions. In the typical calibration problem, we are mainly interested in the prediction of *new* x value of the variable of interest on the basis of both the *current* y value and the previous *test* data (y_i, x_i), $i=1, \cdots, n$ with exactly known x_i values.

In the literature two different methods of estimation were proposed in order to make an inference about x. As is well known, the classical least squares method minimizes the y residuals and yields as estimates of the parameters (α, β) :

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta} = S_{xy}/S_{xx}$$

where $\bar{x}=\sum_{i=1}^{n} x_i/n,\ \bar{y}=\sum_{i=1}^{n} y_i/n$ and $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}),\ S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Then the *classical* estimator $\hat{x}$ of x is defined by :

$$\hat{x} = \bar{x} + (y - \hat{\alpha})/\hat{\beta} = \bar{x} + (S_{xx}/S_{xy})(y - \bar{y}). \qquad (1.2)$$

On the othe hand several authors including Krutchkoff (1967), Hoadley (1972), Hunter and Lamboy (1981) among others advocated to use the inverse regression procedure of minimizing x residuals as an alternative to (1.2) which yields the so-called *inverse* estimator $\tilde{x}$ of x :

$$\tilde{x} = \bar{x} + (S_{xy}/S_{yy})(y - \bar{y}). \qquad (1.3)$$

Historically, the discussion in the literature on how to make valid inference about x has been characterized by the disagreement and confusion. Well-known controversy between the two estimators was motivated by the considerations of *consistency* and *mean squared errors* of the estimators. Specifically, under the usual normal error model, the classical estimator is supported by the maximum likelihood (ML) approach and thus is typically consistent and asymptotically efficient as $n \to \infty$. But it has *infinite* mean squared error (MSE) for finite sample size n as is noted by Williams (1969). While the inverse estimator has a Bayesian justification as is shown by Hoadley (1972), Hunter and Lamboy (1981) and it has typically smaller MSE than the classical estimator in the region of interest. But, as noted by Berkson (1967) and Shukla (1972), it is known to be generally biased and is not consistent as $n \to \infty$.

But we note that most of previous works in this area were based on the *large-sample* approximation to the biases and MSEs of the competing estimators. Thus they are mainly of only academic interest because, for practical purposes, calibration errors are often small and thus sample size needs not be very large but is typically small in practice as is emphasized by most of practioners of the real calibration work. See the comment by Rosenblatt and Spiegelman in the discussion of Hunter and Lamboy (1981).

In this paper, motivated by the above consideration, we reconsider the practical calibration problem from the *small-sample low-noise* viewpoint and try to throw the new light on the issues such as *consistency* and *efficiency* of the various competing estimators including the classical and the inverse estimators.

Specifically, on the basis of simple approximate expression for the asymptotic mean squared errors (AMSE) of the arbitrary regular consistent estimators, we will show that both the classical and the inverse estimators together with other compound estimators have justifications in their own right as low-noise consistent and first-order efficient estimators without any reference to the specific distributional assumptions such as normality.

This paper is organized as follows. In Section 2 we first introduce the new definitions of the concepts of small-sample low-noise consistency and efficiency of the estimators of x. Then we derive an important lower bound for the AMSE of the arbitrary regular consistent estimators which depends only on the second-order moments of the measurement errors but is independent of the type of the errors.

In Section 3 we show that both the classical and the inverse estimators are consistent and first-order efficient irrespective of the type of the measurement errors.

Finally in Section 4 we give some examples which illustrate the relevance of the small-sample low-noise approach and also discuss the possible extensions to non-linear and multivariate calibration problems.

2. Main Results

In this paper we always assume that the sample size n is *fixed* finite number and also use following vector notations. Let $Y=(y_1, \cdots, y_n, y)$ be the $(n+1)$-dimensional random vector of observations and let $\mu=(\mu_1, \cdots, \mu_n, \mu^*)$ be the mean vector $E(Y)$ of the observation vector Y defined by $\mu_i=\alpha+\beta x_i$ and $\mu^*=\alpha+\beta x$, $i=1, \cdots, n$. We note that the mean vector μ depends on the parameter vector $\theta=(\alpha, \beta, x)$ and thus we will denote it by $\mu(\theta)$ or by $\mu(\alpha, \beta, x)$ showing the explicit dependence on the relevant parameters.

We first introduce the following definition of regular-consistent estimator of x :

Definition 1. An estimator $h(Y)$ of x is called regular-consistent if $h(\cdot)$ is a continuously differentiable function of Y and satisfies the condition :

$$h(\mu(\theta))=h(\mu(\alpha, \beta, x)) = x \quad \text{holds for all} \quad \theta=(\alpha, \beta, x). \tag{2.1}$$

Remark 1. Note that our definition of consistency (2.1) is completely different from the usual large-sample definition of consistency which is used in most of previous works as in Berkson (1967) and Shukla (1972) because we do not consider the behaviour of the estimator as the sample size n gets large but study the performance of the estimator as the variance σ^2 of the measurement error gets small for finite sample size n.

Remark 2. For regular estimators, we note that condition (2.1) is equivalent to the more familiar concept of *asymptotic unbiasedness* as $\sigma \to 0$ which is defined by :

$$\lim_{\sigma \to 0} E[h(Y)] = x \quad \text{for all} \quad \theta = (\alpha, \beta, x). \tag{2.2}$$

Similarly we note that the condition (2.1) is equivalent to that of the *consistency in probability* as $\sigma \to 0$:

$$\lim_{\sigma \to 0} h(Y) = x \quad \text{for all} \quad \theta = (\alpha, \beta, x). \tag{2.3}$$

Remark 3. Any reasonable estimator of x must be consistent in the sense of (2.1) because when there is no measurement error it seems perfectly reasonable to require that we should be able to recover the true value x exactly no matter where it is. In fact every estimator considered in the literature satisfies this requirement including the classical and inverse estimators.

In order to compare the performances of various regular consistent estimators, we introduce the definition of the AMSE (Asymptotic Mean Squared Error) of the estimator as follows.

Definition 2. AMSE (Asymptotic Mean Squared Error) of the regular consistent estimator $h(Y)$ of x is the quantity defined by :

$$AMSE[h(Y)] = \sigma^2 \lim_{\sigma \to 0}(E[h(Y)-x]^2/\sigma^2). \tag{2.4}$$

Now we are ready to establish the fundamental lower bound for the AMSEs of the regular consistent estimators.

Theorem 1. If $h(Y)$ is a regular consistent estimator of x, then we have the inequality :

$$AMSE[h(Y)] \geq \frac{\sigma^2}{\beta^2}[1+1/n+(x-\bar{x})^2/S_{xx}] \text{ for all } (\alpha, \beta, x), \beta \neq 0 \tag{2.5}$$

where $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Proof. By the regular consistency of the estimator $h(\cdot)$, we have the identity :

$$h(\mu(\alpha, \beta, x)) = x \quad \text{for all } (\alpha, \beta, x).$$

Differentiating above identity partially with respect to α, β, x respectively in sequence, we get the following series of identities :

$$\sum_{i=1}^{n} \partial h / \partial y_i + \partial h / \partial y = 0 \tag{2.6}$$

$$\sum_{i=1}^{n} (\partial h / \partial y_i) x_i + (\partial h / \partial y) x = 0 \tag{2.7}$$

$$(\partial h / \partial y) \beta = 1. \tag{2.8}$$

Multiplying (2.6) by k_1 and (2.7) by k_2 respectively and subtracting them from (2.8), we have the identity :

$$\sum_{i=1}^{n} \partial h / \partial y_i (-k_1 - k_2 x_i) + \partial h / \partial y (\beta - k_2 x) = 1. \tag{2.9}$$

Now by the Cauchy-Schwartz inequality, we have the inequality :

$$(\partial h / \partial y)^2 + \sum_{i=1}^{n} (\partial h / \partial y_i)^2 \geq [(\beta - k_2 x)^2 + \sum_i (k_1 + k_2 x_i)^2]^{-1}. \tag{2.10}$$

Here, by the simple projection method, we obtain the identity :

$$inf_{(k_1, k_2) \in R^2} [(\beta - k_2 x)^2 + \sum_{i=1}^{n} (k_1 + k_2 x_i)^2] = \beta^2 [1 + 1/n + (x - \bar{x})^2 / S_{xx}]^{-1}. \tag{2.11}$$

Now we note that the regular consistency of the estimator $h(\cdot)$ implies

$$\begin{aligned} h(Y) - x &= h(Y) - h(\mu) \\ &= \sum_{i=1}^{n} (\partial h / \partial y_i) \sigma \varepsilon_i + (\partial h / \partial y) \sigma \varepsilon + o(\sigma) \end{aligned}$$

and

$$AMSE[h(Y)] = \sigma^2[\sum_{i=1}^{n}(\partial h/\partial y_i)^2 + (\partial h/\partial y)^2] \quad (2.12)$$

immediately. Therefore, taking supremum of the left hand side of (2.10) with respect to k_1, k_2 and using the identities (2.11) and (2.12), we get the required inequality (2.5) immediately.

Remark 4. Note that the lower bound (2.5) does not depend on the *type* of the distribution of measurement errors as long as they are mutually uncorrelated with zero mean and constant variance. Thus lower bound (2.5) is in fact *semiparametric* information bound which is valid for arbitrary type of error distributions.

In the next section we will consider the problem of identifying a class of regular consistent estimators which attain the lower bound (2.5) for all x.

3. Efficient Estimators

We first define efficiency of the regular consistent estimator as follows.

Definition 3. A regular consistent estimator $h(Y)$ of x is called *efficient* if we have the identity :

$$AMSE[h(Y)] = \frac{\sigma^2}{\beta^2}[1 + 1/n + (x-\bar{x})^2/S_{xx}] \text{ for all } (\alpha, \beta, x)\ \beta \neq 0. \quad (3.1)$$

Now we are ready to establish the important optimality results of the classical and inverse estimators.

Theorem 2. If $\beta \neq 0$, then both the classical estimator $\hat{x}$ defined by (1.2) and the inverse estimator $\check{x}$ defined by (1.3) are regular consistent and efficient.

Proof. By the direct substitution, we can check the regular consistency of the two estimators immediately. As for the proof of the efficiency, note that direct Taylor expansions yield the identities :

$$\begin{aligned}\hat{x} - x &= (S_{xx}/S_{xy})(y - \bar{y}) \\ &= (\sigma/\beta)[(\varepsilon - \bar{\varepsilon}) - (S_{x\varepsilon}/S_{xx})(x - \bar{x})] + o(\sigma) \quad (3.2)\end{aligned}$$

and

$$\hat{x}-x = (S_{xy}/S_{yy})(y-\bar{y})$$
$$= (\sigma/\beta)[(\varepsilon-\bar{\varepsilon})-(S_{x\varepsilon}/S_{xx})(x-\bar{x})] + o(\sigma) \qquad (3.3)$$

where $\bar{\varepsilon}=\sum_{i=1}^{n}\varepsilon_i/n$, $S_{x\varepsilon}=\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})$.

Because measurement errors are mutually uncorrelated with zero mean and unit variance, (3.2) and (3.3) imply immediately the efficiency of the estimators.

Remark 5. Since the estimators $\hat{x}$ and $\tilde{x}$ are both efficient estimators, we can construct a new *compound* estimator x^* of x which is defined by :

$$x^* = \phi^{-1}[p\phi(\hat{x}) + (1-p)\phi(\tilde{x})] \qquad (3.4)$$

where $\phi(\cdot)$ is an arbitrary monotonically increasing function of x and $0\le p\le 1$. Then we can easily show that the compound estimator x^* is also regular consistent and efficient estimator. Two important special cases are the estimators defined by :

$$x_1^* = \frac{\hat{x}+\tilde{x}}{2} \quad \text{and} \quad x_2^* = \sqrt{\hat{x}\tilde{x}}\ .$$

Remark 6. In order to further discriminate the so-called second-order efficient estimator among the various first-order efficient estimators, we have to take into account more terms in the Taylor expansion and should also assume the knowledge of the third and forth-order moments of the measurement errors which is typically not available in the small-sample experiment. Therefore this topic will not be considered in this paper.

4. Examples and Discussions

As a simple practical example of typical small-sample calibration experiment, we first consider the calibration experiment of measuring the moment of inertia of the product discussed in 22.2 of Taguchi (1987).

Example 1. In this calibration experiment, test sample size is $n=12$ and the estimated variance is $s^2=0.013$ and $\hat{\alpha}=0$ and $\hat{\beta}=142.35$. We note that the classical estimator $\hat{x}$ gives 0.0590 ± 0.0018 as an estimate of the moment of inertia x of the product while the inverse estimator $\tilde{x}$ gives 0.05898 ± 0.0018 which is very close to the value $\hat{x}=0.0590$ ± 0.0018. This result seems to strongly support the conclusion of our study about the practical equivalence of the two estimators in the small-sample low-noise set-up. See 22.2 of Taguchi (1987) for more details of the calculation.

Next example provides simple numerical comparison between the approximate results of our work and the exact results from the small-sample simulation study for the computation of the MSEs of the classical and inverse estimators as is done by Krutchkoff (1967).

Example 2. Krutchfield (1967) conducted intensive simulation study for the comparison of MSEs of classical and inverse methods of calibration. Typical sample sizes were $n=4$, 6 and 8 and the standard deviation was $\sigma=0.1$ with $\alpha=0$, $\beta=0.5$. On close examination, the simulated values of the MSEs of the inverse estimators are generally in good agreement with the values given by the approximate formula (3.1) for AMSEs derived in Section 3. This results seems to justify the validaty of the small-sample low-noise approximation developed in this paper. Following table provides typical comparison between our results and those from the simulation study. See the Table 5 of Krutchfield (1967) for more details.

Simulated MSEs and AMSE for n=4, design 2(x=0), 2(x=1)

x	MSE classical	MSE inverse	AMSE
0	.075	.060	.060
.2	.061	.050	.053
.4	.058	.048	.050
.6	.059	.049	.050
.8	.063	.052	.053
1.0	.073	.059	.060

Finally we discuss the problems of extending our results to non-linear and multivariate calibration models.

Remark 7. If we consider general *non-linear* calibration model :

$$y_i = f(x_i; \beta) + \varepsilon_i$$
$$y = f(x; \beta) + \varepsilon \quad i=1, \cdots, n$$

where $f(\cdot;\beta)$ represents an arbitrary momotonic non-linear function of x and β denotes vector of regression parameters. We can establish similar optimality results for the non-linear least squares estimators of x within the small-sample low-noise framework.

Remark 8. Instead of the univariate calibration model (1.1), we can consider *multivariate* calibration model where each of the observations in (Y_i, X_i) $i=1, \cdots, n$, (Y, X) are m-dimensional random vectors. Then we can generalize our results to this case without difficulty by considering appropriate small-sample low-noise approximation.

References

[1] Berkson, J. (1967), "Estimation of a Linear Function for a Calibration line," *Technometrics*, Vol. 11, pp. 649-660.

[2] Hoadley, A. B. (1972), "A Baysian Look at Inverse Linear Regression," *Journal of American Statistical Association*, Vol. 65, pp. 356-369.

[3] Hunter, H. G. and Lomboy, W. F. (1981), "A Baysian Analysis of the Linear Calibration Problem," *Technometrics*, Vol. 23, pp. 323-350.

[4] Krutchkoff, R. G. (1967), "Classical and Inverse Regression Methods of Calibration," *Technometrics*, Vol. 9, pp.425-439.

[5] Shukla, G. K. (1972), "On the Problem of Calibration," *Technometrics*, Vol. 14, pp. 547-553.

[6] Taguchi, G. (1987), *System of Experimental Design*, Kraus International Publication.

[7] Williams, E. J. (1969), "A Note on Regression Methods in Calibration," *Technometrics*, Vol. 11, pp. 189-192.