TECHNOMETRICS VOL. 12, NO. 4 NOVEMBER 1970

On Inverse Estimation in Linear Regression

MAX HALPERIN

National Heart and Lung Institute, Bethesda, Maryland

A procedure suggested by Krutchkoff (1967) for inverse estimation in linear regression is compared with the classical procedure from other points of view than that taken by Krutchkoff, i.e. comparative mean square error. In particular, comparisons are made on the basis of "closeness" in estimation (Pitman, 1937), consistency (in a setting where this concept is relevant), and mean square error of the relevant *asymptotic* distributions. It is found that, for large samples, Krutchkoff's estimate is superior in the sense of "closeness" if values of the independent variable are restricted to a certain closed interval around the mean of the independent variates in the experiment and inferior elsewhere. However, the width of this interval varies inversely as the product of the absolute value of the standardized slope (i.e. scaled by the error standard deviation) and the standard deviation of the independent variables used in the experiment. As a practical matter the parameter tends to be large so that the interval where the Krutchkoff estimate is superior will be trivially small. In addition large values of this product parameter imply that the two estimates being compared are virtually indistinguishable. Coupling these latter remarks with the fact that the classical procedure allows an exact confidence interval for the parameter under estimate while the Krutchkoff procedure does not, suggests the classical estimate is to be preferred using the "closeness" criterion. If one uses the criterion of mean square error applied to the relevant asymptotic distribution, one reaches conclusions similar to the above, except that the interval of superiority of the Krutchkoff estimate is no longer trivially small even at best. However, the mean square error criterion fails to take into account the fact that the estimates are correlated and so should be considered an intrinsically less appropriate criterion than closeness. We also find that, in circumstances where the concept is applicable, Krutchkoff's estimate is not consistent whereas the classical estimate is; Krutchkoff's estimate can be trivially modified so that it is consistent but will then tend to never be better in the sense of closeness than the classical estimate.

1. INTRODUCTION AND SUMMARY

The usual assumptions and procedure for inverse estimation in linear regression can be described as follows. One has a sample (y_i, x_i), $i = 1, 2, \cdots, n$, $n \geq 2$, where the x_i are known constants, at least two of which are distinct, and the y_i are independent and $N(\alpha + \beta x_i, \sigma)$. The parameters (α, β) are estimated by least squares as (a, b) and, given a new y, Y say, the corresponding x, X say, is estimated by $\hat{X} = (Y - a)/b$. In a recent paper [2] Krutchkoff has suggested that instead of using the estimated regression of y on x to estimate X, one use the estimated regression of x on y (by least squares) to estimate X as say $\hat{X}_k = c + dY$. In the above:

Received July 1969; revised Nov. 1969

$$a = \bar{y} - b\bar{x}, \qquad c = \bar{x} - d\bar{y},$$
$$b = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{1.1}$$
$$d = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

where

$$\bar{y} = \sum_{i=1}^{n} y_i/n, \qquad \bar{x} = \sum_{i=1}^{n} x_i/n.$$

Kruthchkoff concludes on the basis of a Monte Carlo investigation (in which values of $|b| < .001$ were replaced by $\pm .001$ as appropriate) that the mean square error (MSE) of $\hat{X}_k$ is uniformly less than that of $\hat{X}$. The Monte Carlo work involved 10,000 repetitions and considered both normal and non-normal error distributions. As recently pointed out by E. J. Williams [5], Krutchkoff's results are hardly surprising since the MSE of $\hat{X}$ (untruncated) is infinite; the truncated version of $\hat{X}$ considered by Krutchkoff in his simulation would not be expected to seriously alter any comparison with alternative estimates. As Williams emphasizes, an estimate which is a constant and thus clearly irrelevant to the issue would be superior to $\hat{X}$ in the MSE sense; to make the point another way, a random drawing from *any* distribution with finite variance would provide a better estimate than $\hat{X}$ in the MSE sense. Williams concludes that the MSE is an inappropriate criterion, that the finite variance of $\hat{X}_k$ seems to be its only merit and thus dimisses $\hat{X}_k$ as an estimate of X. One can not quarrel with the conclusion that the MSE is an inappropriate criterion. One might examine further, however, what merits $\hat{X}_k$ has, as compared to $\hat{X}$, according to other intuitively reasonable criteria than MSE.

We pursue this question in the following discussion. It should be pointed out that Krutchkoff is aware that $\hat{X}$ has an infinite MSE [3] but apparently does not think this fact makes his comparisons in [2] of dubious value.

The criteria we consider under the usual regression assumptions, as outlined above are:

1. The relative "closeness" of $\hat{X}$ and $\hat{X}_k$ to X. Here "closeness" is in the sense of Pitman [3]. That is, $\hat{X}$ is a closer estimate than is $\hat{X}_k$ if, for all X,

$$\Pr\{|\hat{X} - X| < |\hat{X}_k - X|\} > \tfrac{1}{2}.$$

2. Consistency. This, of course, is only relevant when Y, a, b are all based on large samples.

3. Mean square error of relevant asymptotic distributions. It should be pointed out here, that the MSE, in general, fails to take into account the correlation between the estimates and thus is intrinsically less meaningful for comparing estimates than the "closeness" concept.

We obtain both large and small sample results on closeness; for simplicity of presentation we indicate below only the large sample results. Somewhat similar results for small samples are given in the body of the paper. We shall also assume, for convenience only, that $\bar{x}$ is fixed for all sample sizes.

To give our results in a compact way we need a modest amount of notation. Thus, let

$$\rho = \beta/\sigma, \quad \delta = \bar{x} - X, \quad \sigma_x^2(n) = n^{-1} \sum_{i=1}^{n} (x_i - \bar{X})^2, \quad \lim_{n\to\infty} \sigma_x^2(n) = \sigma_x^2(\text{finite}).$$

We shall also suppose that our "new Y" is in fact an average of N observations each with variance σ^2 and correspondingly use a notation $\bar{Y}_N$ and denote by $\hat{X}(N)$ the classical estimate of the corresponding X. We also note that it follows easily from (1.1) that the Krutchkoff estimate is given by

$$(1 - R)\bar{x} + R\hat{X}(N) \tag{1.2}$$

where

$$R = db = \frac{nb^2\sigma_x^2(n)}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + nb^2\sigma_x^2(n)} \tag{1.3}$$

and $\hat{y}_i = a + bx_i$. Finally it is useful to define a generalized Krutchkoff estimate by

$$\hat{X}_k(N, M) = (1 - R_M)\bar{x} + R_M\hat{X}(N) \tag{1.4}$$

where

$$R_M = \frac{nb^2\sigma_x^2(n)}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2/M + nb^2\sigma_x^2(n)}. \tag{1.5}$$

Our analysis is in terms of $\hat{X}_k(N, M)$ and $\hat{X}(N)$.

Our asymptotic results are then as follows:

1. For given $\rho(\neq 0)$, δ, N, M, $\rho\sigma_x$

$$\lim_{n\to\infty} \Pr\{|\hat{X}_k(N, M) - X| < |\hat{X}(N) - X|\}$$

$$= \Phi(-|\rho\delta|\sqrt{N}) + \Phi[-|\rho\delta|\sqrt{N}/(1 + 2M\rho^2\sigma_x^2)],$$

where $\Phi(u) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{u} (\exp - t^2/2)\, dt$. It is clear from this result that $\hat{X}_k(N, M)$ will be a closer estimate than $\hat{X}(N)$ for a closed interval on X, depending on N, M, ρ, σ_x and $\bar{x}$. It is also clear that if $|\rho\sigma_x|$ is very large and $|\delta| = r\sigma_x$, then for every fixed r, N, M, the two estimates are indistinguishable in the sense of closeness; in fact, from (1.4) and (1.5), it is apparent that $\hat{X}_k(N, M)$ will tend to be very nearly $\hat{X}(N)$ so that as a practical matter there will be very little to choose between the two estimates. We also observe that if $M = N$ and N is large then, for $|\sqrt{N}\,\rho\sigma_x|$ large and $|\delta| = r\sigma_x$, the estimates will again be indistinguishable. In short, if Y is well determined, the absolute standardized slope ($|\rho|$) is large, or the values of the independent variable are widely dispersed, the estimates are indistinguishable. We note too that if $M \ll N$ and N is large, the interval in which $\hat{X}_k(N, M)$ is superior becomes very small.

2. For given $\rho(\neq 0)$, δ, σ_x^2, $M = N$ and both n and N large, $\hat{X}(N)$ and $\hat{X}_k(N, N)$ both converge in probability to X; however, if M is fixed and N is large $\hat{X}_k(N, M)$ converges in probability to $(1 - R_\infty)\bar{x} + R_\infty X$ where $R_\infty = M\rho^2\sigma_x^2/(1 + M\rho^2\sigma_x^2)$.

3. For given $\rho(\neq 0)$, σ_x^2, M, N, δ, and n large, the MSE of $\hat{X}_k(N, M)$ is less than the MSE of $\hat{X}(N)$ providing $|\delta| < \sqrt{(1/N\rho^2) + 2M\sigma_x^2/N}$. This calculation refers to the asymptotic distributions of $\hat{X}(N)$ and $\hat{X}_k(N, M)$.

The above results are proved in Section 2. We also give in that section some tables which are helpful in assessing whether the Krutchkoff estimate is preferable in practice and an illustrative example. Finally, it should be noted that, with slight modifications due to differences in assumptions, conclusions similar to the above apply to estimation of relative potency in bioassay.

2. Derivations and Discussion

Before proceeding with our analysis it is convenient to reduce $\hat{X}(N)$ and $\hat{X}_k(N, M)$ by a series of transformations. Thus, let

$$\left.\begin{aligned} u &= \frac{(b-\beta)}{\sigma}\sqrt{n}\,\sigma_x(n) \\ v &= \frac{[(\bar{Y}_N - \bar{y}_n) - \beta(X - \bar{x})]}{\sigma}\sqrt{\frac{Nn}{N+n}} \\ \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 &= \sigma^2\chi^2_{n-2} \end{aligned}\right\} \tag{2.1}$$

so that

$$\left.\begin{aligned} \hat{X}(N) &= \frac{\left(v\sqrt{\frac{N+n}{N}} - \rho\delta\sqrt{n}\right)\sigma_x(n)}{u + \rho\sqrt{n}\sigma_x(n)} + \bar{x} \\ R_M &= \frac{[u + \rho\sqrt{n}\,\sigma_x(n)]^2}{(\chi^2_{n-2}/M) + [u + \rho\sqrt{n}\,\sigma_x(n)]^2} \\ \hat{X}_k(N, M) &= \bar{x} + \frac{R_M\left[\left(v\sqrt{\frac{N+n}{N}} - \rho\delta\sqrt{n}\right)\sigma_x(n)\right]}{[u + \rho\sqrt{n}\,\sigma_x(n)]} \end{aligned}\right\} \tag{2.2}$$

We first observe that for n, M, N and $\sigma_x(n)$ fixed it follows from (2.2) that $R_M \to 1$ as $|\rho\sigma_x(n)| \to \infty$ for almost all u and χ^2_{n-2}. Thus, with all other parameters fixed if the regression slope is very large and/or $\sigma_x(n)$ is very large $\hat{X}(N)$ and $\hat{X}_k(N, M)$ are virtually identical with, in fact, $\hat{X}_k(N, M)$ reducing to $\hat{X}(N)$. In other words the better our design for estimation of slope [$\sigma_x(n)$ large] and the more intrinsically precise our calibration ($|\beta|/\sigma$ large) the less difference it makes which of the two estimates of X we use. We would expect these desiderata to obtain in many practical problems and should therefore be inclined to use $\hat{X}(N)$ since, from the standpoint of point estimation, there would be little difference between the estimates while $\hat{X}(N)$ allows an exact interval estimate of X via Fieller's Theorem and $\hat{X}_k(N, M)$ does not.

But now let us consider what can be said about the two estimates when $|\rho\sigma_x(n)|$ is not large; we note that this assumption does not necessarily preclude the possibility that one of $|\rho|$ and $\sigma_x(n)$ is quite large. We also note that this assumption will seldom be relevant in practice.

We first address ourselves to the question of closeness and thus consider the evaluation of

$$P_c(X) = \Pr\{|\hat{X}_k(N, M) - X| < |\hat{X}(N) - X|\}, \tag{2.3}$$

the probability for given X that $\hat{X}_k(N, M)$ deviates less from X than does $\hat{X}(N)$. Clearly, we can write (2.3) as

$$P_c(X) = \Pr\{[\hat{X}_k(N, M) - X]^2 < [\hat{X}(N) - X]^2\}. \tag{2.4}$$

Then, using the definitions (2.2) and letting $T = X(N) - \bar{x}$ we can write (2.4), after some reduction, as

$$P_c(X) = \Pr\left\{T\left(T + \frac{2\delta}{1 + R_M}\right) > 0\right\}. \tag{2.5}$$

We note that reduction of (2.4) to (2.5) involved division by $(1 - R_M)$ so that the following analysis requires $1 - R_M \neq 0$ and hence excludes the case $|\rho\sigma_x(n)| \to \infty$.

We first observe that if $\delta = 0 (X = \bar{x})$, then, from (2.5), $P_c(\bar{x}) = 1$. Hence, either $\hat{X}_k(N, M)$ is a closer estimate than $\hat{X}(N)$ or there are some values of X for which $P_c(X) < \frac{1}{2}$. To proceed further we first suppose that $\delta > 0$. In such case a straightforward decomposition of (2.5) shows that

$$P_c(X) = \Pr\left\{T < -\frac{2\delta}{1 + R_M}\right\} + \Pr\{T > 0\}. \tag{2.6}$$

Then, using the definition of T and observing that u and v are independent $N(0, 1)$ variates, we find that

$$\begin{aligned}\Pr\{T > 0\} &= \Pr\left\{v > \rho\delta\sqrt{\frac{nN}{N + n}}\right\} \Pr\{u > -\rho\sqrt{n}\,\sigma_x(n)\} \\ &\quad + \Pr\left\{v < \rho\delta\sqrt{\frac{nN}{N + n}}\right\} \Pr\{u < -\rho\sqrt{n}\,\sigma_x(n)\} \\ &= \Phi\left(-\rho\delta\sqrt{\frac{nN}{N + n}}\right)\Phi(\rho\sqrt{n}\,\sigma_x(n)) + \Phi\left(\rho\delta\sqrt{\frac{nN}{N + n}}\right)\Phi(-\rho\sqrt{n}\,\sigma_x(n)),\end{aligned} \tag{2.7a}$$

where Φ is the cumulative distribution function for given argument of a $N(0, 1)$ variate.

Similarly, taking account of the fact that R_M is a random variable depending on χ^2_{n-2} as well as u, we find we can write $\Pr\{T < -2\delta/(1 + R_M)\}$ as

$$\int_0^\infty \left\{\int_{-c}^{\infty} \Phi[-g(u)]\varphi(u)\,du + \int_{-\infty}^{-c} \Phi[g(u)]\varphi(u)\,du\right\} p(\chi^2_{n-2})\,d\chi^2_{n-2}, \tag{2.7b}$$

where $c = \rho\sqrt{n}\,\sigma_x(n)$, $\varphi(u)$ is a $N(0, 1)$ density, $p(\chi^2_{n-2})$ is the density of a χ^2_{n-2} variate and

$$g(u) = \left[\frac{2u}{\sigma_x(n)} + (1 - R_M)\rho\sqrt{n}\right] \frac{\sqrt{N}\,\delta}{\sqrt{N + n}\,(1 + R_M)}. \tag{2.8}$$

Now let us also suppose $\rho > 0$ and let $\delta \to \infty$. Then (2.7a) approaches $\Phi(-\rho\sqrt{n}\,\sigma_x(n))$. To evaluate (2.7b) we observe that $\Phi[g(u)]$ will be zero if the

sign of $g(u)$ is negative and will be unity otherwise. Thus, the second integral of (2.7b) is clearly zero as $\delta \to \infty$. In the first integral $\Phi[-g(u)]$ will approach unity as long as $u < (1 - R_M)\rho\sqrt{n}\,\sigma_x(n)/2$. Thus, although an explicit evaluation appears impossible for general values of the parameters, it is clear that the first integral of (2.7b) will (for some θ depending on the particular values of the various parameters, with $0 < \theta < 1$) be of the form $\Phi(-\theta\rho\sqrt{n}\,\sigma_x(n)/2) - \Phi(-\rho\sqrt{n}\,\sigma_x(n))$.

It follows that $P_c(X) \not> \frac{1}{2}$, for all $X \geq \bar{x}$ and $\rho > 0$. An analysis along the lines of the above for the cases $(\rho < 0, \delta > 0)$, $(\rho < 0, \delta < 0)$, $(\rho > 0, \delta < 0)$, shows that $P_c(X)$ depends only on the absolute values of ρ and δ. It follows that $\hat{X}_k(N, M)$ is *not* a closer estimate of X than is $\hat{X}(N)$. However, it is clear that for values of X near $\bar{x}$, $P_c(X) > \frac{1}{2}$. Since one may reasonably argue that inverse estimation should only be used when the X's to be estimated can be expected to be within the range of the x_i used in the calibration experiment, the above remark suggests $\hat{X}_k(N, M)$ may be preferable to $\hat{X}(N)$ for X restricted as indicated above. To decide the merits of this proposition requires a closer analysis of $P_c(X)$. Due to the non-linearity introduced by R_M, if nothing else, this appears most difficult for small samples. One might expect that large sample considerations would be somewhat relevant. Thus, assuming $\lim_{n\to\infty} \sigma_x(n) = \sigma_x$ (finite) and all other parameters fixed, inspection of (2.7a) shows that as $n \to \infty$ (2.7a) approaches $\Phi(-|\rho\delta|\sqrt{N})$. (Note we are now assuming the dependence of $P_c(X)$ on the absolute values of ρ and δ as remarked above.) To evaluate (2.7b) as $n \to \infty$ we observe from (2.2) that R_M approaches $\rho^2\sigma_x^2/(M^{-1} + \rho^2\sigma_x^2)$ for almost all u and χ^2_{n-2}. Utilizing this fact in (2.8) leads us to conclude that $g(u)$ approaches $\sqrt{N}\,|\rho\delta|/(1 + 2M\rho^2\sigma_x^2)$ for almost all u and χ^2_{n-2}. We are thus led to conclude that, as $n \to \infty$, with N, M, ρ, δ, fixed and $\lim_{n\to\infty} \sigma_x(n) = \sigma_x$ (finite), $P_c(X)$ approaches

$$\Phi[-|\rho\delta|\sqrt{N}] + \Phi[-|\rho\delta|\sqrt{N}/(1 + 2M\rho^2\sigma_x^2)] \tag{2.9}$$

We note that if $M = N$ and N is at least $0(n)$ then from (2.2) $R_M \to 1$ as $n \to \infty$ for almost all u and χ^2_{n-2} so that one is again in the situation where one cannot differentiate between $\hat{X}_k(N, M)$ and $\hat{X}(N)$. We observe that (2.9) is monotone decreasing in $|\delta|$ approaching zero for large $|\delta|$, all else being equal. The most interesting case would seem to be $M = N$ in which case putting $|\delta| = r\sigma_x$ and $\Delta = \sqrt{N}\,|\rho|\,\sigma_x$, (2.9) becomes $\Phi(-r\Delta) + \Phi[-r\Delta/(1 + 2\Delta^2)]$. If we set this latter quantity equal to $\frac{1}{2}$, we can use implicit function theory to deduce that the solution of this equation is first decreasing in Δ somewhat past $\Delta = \frac{1}{2}$, with the behavior for larger Δ not very clear. Numerically, two computations appear of interest. One of them is based on the notion that inverse estimation should only be used when X is within the range of x's in the calibration experiment. This suggests taking $r = 2.5$, on the supposition that $5\sigma_x$ would encompass most of the experimental range, and computing $P_c(\bar{x} \pm 2.5\,\sigma_x)$ from (2.9) as a function of Δ. If $P_c(\bar{x} \pm 2.5\,\sigma_x)$ is a decreasing function of Δ, then as long as $P_c(\bar{x} \pm 2.5\,\sigma_x) > .5$ it is reasonable to prefer $\hat{X}_k(N, N)$. The other computation which appears worthwhile is to compute the value of r, as a function of Δ, which is such that (2.9) (with $M = N$) is equal to .5. With this latter information one can make a judgement, for other ranges than $\bar{x} \pm 2.5\,\sigma_x$ as to which

statistic is preferable. Presumably one would estimate Δ from the known N, $\sigma_x^2(n)$ and an estimate of ρ as

$$\frac{\sqrt{n-2}b}{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \tag{2.10}$$

Table I gives the asymptotic value of $P_c(\bar{x} \pm 2.5\ \sigma_x)$ for selected values of $\Delta = \sqrt{N}\,|\rho|\,\sigma_x$. Table II gives, for selected Δ, the value of r, $r(.5)$ say, such that the asymptotic value of $P_c[\bar{x} \pm r(.5)\sigma_x] = .5$. The solution for $r(.5)$ can be carried out by a simple algorithm. Specifically, let $w = r\Delta$; then from

$$\Phi(-w) + \Phi[-w/(1 + 2\Delta^2)] = .5, \qquad w > 0, \tag{2.11}$$

we have

$$\Phi[-w/(1 + 2\Delta^2)] = \Phi(w) - \tfrac{1}{2}$$

and

$$\frac{w}{1 + 2\,\Delta^2} = k\{\Phi(w) - \tfrac{1}{2}\} \tag{2.12}$$

where $k(p)$ is the standardized normal deviate exceeded with probability p. From (2.12)

and

$$\left.\begin{aligned} \Delta &= \sqrt{\frac{w}{k\{\Phi(w) - \tfrac{1}{2}\}} - 1}\Big/\sqrt{2} \\ r &= w/\Delta. \end{aligned}\right\} \tag{2.13}$$

Thus, to compute Table II, we first select $w > 0$, compute $\Phi(w) - \frac{1}{2}$ and $k\{\Phi(w) - \frac{1}{2}\}$; if $w > k\{\Phi(w) - \frac{1}{2}\}$ we compute Δ and r from (2.13); if $w < k\{\Phi(w) - \frac{1}{2}\}$ we choose $w' > w$ and compute $\Phi(w') - \frac{1}{2}$ and $k\{\Phi(w') - \frac{1}{2}\}$ and proceed as above.

TABLE I

$P_c(\bar{x} \pm 2.5\sigma_x)$, the Probability that $|X_k(N, N) - X| < |\hat{X}(N) - X|$, for $|X - \bar{x}| = 2.5\sigma_x$, as a Function of $\Delta = \sqrt{N}\,|\rho|\,\sigma_x$.

Δ	$P_c(\bar{x} \pm 2.5\sigma_x)$	Δ	$P_c(\bar{x} \pm 2.5\sigma_x)$
.01	.980	2.0	.289
.05	.901	2.5	.322
.10	.804	3.0	.347
.15	.714	3.5	.366
.20	.630	4.0	.381
.25	.555	5.0	.403
.29	.502	6.0	.419
.30	.489	7.0	.430
.40	.383	8.0	.438
.50	.308	9.0	.445
1.00	.209	10.0	.451
1.50	.248	15.0	.467

It is evident from Table 1 that although $P_c(\bar{x} \pm 2.5\ \sigma_x)$ is not a decreasing function of Δ, once it gets below .5, it remains below .5. We also see that for $\Delta \leq .29$, $\hat{X}_k(N, N)$ is always to be preferred over $\hat{X}(N)$.

Table II indicates that $r(.5)$ is a decreasing function of Δ; moreover, it is easy to show that $P_c(X)$ as a function of r and Δ is decreasing in r for fixed Δ. Thus, if $[-r(.5)\sigma_x + \bar{x}, \bar{x} + r(.5)\sigma_x]$ covers the range expected for X, one should use $\hat{X}_k(N, N)$.

The above remarks concerning Tables I and II do not take into account the magnitude that Δ is likely to be in practice. First we note that N will almost always be unity so that we may write $\Delta = |\beta|\ \sigma_x/\sigma$. In general $(\sigma_x/\sigma) \gg 1$ and we would ordinarily demand that $|\beta|$ be quite large before we would seriously consider a linear relationship for inverse estimation. With these points in mind our tables (especially Table II) and the monotonicity of

$$\Phi(-r\Delta) + \Phi[-r\Delta/(1 + 2\Delta^2)]$$

in r strongly suggest that in practice $\hat{X}_k(N, N)$ will rarely be preferable to $\hat{X}(N)$. We illustrate this with an example from Bowker and Lieberman [1].

Bowker and Lieberman give an example of a calibration problem in which one has a number of measurements (y) of calcium oxide when large amounts of magnesium are also present and the corresponding (known) amounts (x) of calcium oxide. The data and relevant summary statistics are as follows

i	$x_{n_i}(mg)$	$y_i(mg)$	
1	20.0	19.8	$n = 10,\ \bar{x} = 31.1,$
2	22.5	22.8	$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 427.9$
3	25.0	24.5	$S_{yy} = 438.889,$
4	28.5	27.3	$S_{xy} = 430.69,$
5	31.0	31.0	$b = S_{xy}/S_{xx} = 1.0065,$
6	33.5	35.0	$d = S_{xy}/S_{yy} = .9813$
7	35.5	35.1	$\sigma_x = \sqrt{n^{-1}S_{xx}},$
8	37.0	37.1	$S_{y\cdot x} = .82088.$
9	38.0	38.5	
10	40.0	39.0	

We note, in our example, that the estimates of slope, b and d, are very close and that if, as would usually be the case, $N = 1$, we would estimate Δ, using (2.10), as about 8. For this value of Δ it is clear from Tables I and II, that $\hat{X}_k(1, 1)$ would be better in the sense of closeness only for values of X very near to $\bar{x}$ in units of σ_x. In terms of MSE of the asymptotic distribution, $\hat{X}_k(1, 1)$ would be better than $\hat{X}(1)$ if, approximately

$$\left|\frac{X - \bar{x}}{\sigma_x}\right| < \sqrt{\frac{1}{64} + 2}$$

TABLE II

Solutions of $\Phi(-r\Delta) \pm \Phi[-r\Delta/(1 + 2\Delta^2)] = .5$, *in* r *for Selected Values of* Δ.

Δ	$r(.5)$	Δ	$r(.5)$
.01	67.44	2.500	0.66
.05	13.61	3.005	0.59
.101	6.82	3.501	0.54
.151	4.56	4.000	0.49
.198	3.53	4.501	0.46
.302	2.42	5.000	0.43
.401	1.92	6.003	0.37
.502	1.63	7.002	0.34
1.000	1.08	8.006	0.30
1.500	0.88	9.005	0.28
2.006	0.75	10.006	0.26

or about 1.42. Thus $\hat{X}_k(1, 1)$ appears in a much better light in terms of the MSE criterion than in terms of the closeness criterion. However, as remarked earlier, the closeness measure seems clearly the more meaningful criterion.

We go on now to indicate briefly the line of argument leading to the other conclusions given in the summary.

It follows from (2.2) for N, M, $\bar{x}$ and X fixed, that, as is well known, $\hat{X}(N)$ is asymptotically $N(X, 1/\sqrt{N}\,|\rho|)$ and that $\hat{X}_k(N, M)$ is asymptotically $N[(1 - R_\infty)\bar{x} + R_\infty X, R_\infty/\sqrt{N}\,|\rho|]$. An easy calculation then shows that for the asymptotic distributions:

$$\begin{aligned} \text{MSE}\,[\hat{X}(N)] &= 1/N\rho^2 \\ \text{MSE}\,[\hat{X}_k(N, M)] &= R_\infty^2/N\rho^2 + (1 - R_\infty)^2\,\delta^2. \end{aligned} \tag{2.14}$$

It follows from (2.14) that

$$\text{MSE}\,[\hat{X}_k(N, M)] < \text{MSE}\,[\hat{X}(N)]$$

if

$$|X - \bar{x}| < \sqrt{1/N\rho^2 + 2M\sigma_x^2/N},$$

as indicated in the summary.

We also see from (2.2) that if N is $O(n)$ then as $n \to \infty$, $\hat{X}(N)$ converges, for every fixed v, to X; it follows from this that $\hat{X}(N)$ converges in probability to X. By the same type of argument $\hat{X}_k(N, M)$, for fixed M, converges in probability to $\bar{x} + R_\infty(X - \bar{x})$. Thus, $\hat{X}_k(N, M)$ is not consistent. However, if we also take M to be $O(n)$ as well, then, as $n \to \infty$, R_M converges to unity in probability, so that a modified Krutchkoff estimate converges in probability to X but also (see 2.9) is never a closer estimate than $\hat{X}(N)$. It might be added here that choosing $M = N$ can be made plausible as follows. Take X to have a normal prior with mean $\bar{x}$ and variance $\sigma_x^2(n)$, and suppose that $\bar{Y}_N$ given X has a normal distribution with mean $\alpha + \beta X$ and variance σ^2/N. Suppose also that α and β are

known. It is then easy to see that a Bayes estimate of X is of the form (1.4) with $\hat{X}(N) = (\bar{Y}_N - \alpha)/\beta$ and corresponding to (1.5) with $M = N$,

$$R_N = [\beta^2\sigma_x^2(n)] \Big/ \left[\frac{\sigma^2}{N} + \beta^2\sigma_x^2(n)\right].$$

Thus, the modified Krutchkoff estimate is of the form of the Bayes estimate above with certain parameters replaced by sample estimates.

References

[1] Bowker, A. H. and Lieberman, G. J. (1959). *Engineering Statistics.* Prentice-Hall.
[2] Krutchkoff, R. G. (1967). Classical and Inverse Regression Methods of Calibration. *Technometrics 9,* 425–40.
[3] Krutchkoff, R. G. (1968). Letter to the Editor. *Technometrics 10,* 430–1.
[4] Pitman, E. J. G. (1937). The "Closest" Estimates of Statistical Parameters. *Proc. Cam. Phil. Soc. 33,* 212 *et seq.*
[5] Williams, E. J. (1969). A Note on Regression Methods in Calibration. *Technometrics 11,* 189–92.

References

[2] **Classical and Inverse Regression Methods of Calibration**
R. G. Krutchkoff
Technometrics, Vol. 9, No. 3. (Aug., 1967), pp. 425-439.
Stable URL:
http://links.jstor.org/sici?sici=0040-1706%28196708%299%3A3%3C425%3ACAIRMO%3E2.0.CO%3B2-5

[5] **A Note on Regression Methods in Calibration**
E. J. Williams
Technometrics, Vol. 11, No. 1. (Feb., 1969), pp. 189-192.
Stable URL:
http://links.jstor.org/sici?sici=0040-1706%28196902%2911%3A1%3C189%3AANORMI%3E2.0.CO%3B2-D