A Coverage Approach to Evaluating Mean Square Error

Alan H. Dorfman
Office of Survey Methods Research, U.S. Bureau of Labor Statistics
2 Massachusetts Ave, N.E., Washington, D.C. 20212

Abstract

We propose a new method for evaluating the mean square error (mse) of a possibly biased estimator $\hat{\theta}_1$, or, rather, the *class* of estimators to which it belongs. The method uses confidence intervals c of a corresponding unbiased estimator $\hat{\theta}$ and makes its assessment based on the extent to which c includes $\hat{\theta}_1$. The method does not require an estimate, implicit or explicit, of the bias of $\hat{\theta}_1$, is indifferent to the bias/variance breakdown of $\hat{\theta}_1$'s mse, and does not require surety of a model on which $\hat{\theta}_1$ is based.

1. Introduction. The Problem

Suppose one is presented with the task of choosing between two or more methods of estimating quantities of interest in a given context, for example, synthetic estimators or hierarchical Bayes estimators in a small area estimation context. In other words, suppose the real interest is in comparing *classes* of estimators in a given context. One might wish to do this on the basis of which class tended to have estimators having smaller mean square error for the variety of quantities targeted. But estimating the mean square error of a (possibly) biased estimator is not easy (cf., for example, Rao 2003, Section 4.2.4).

As an example, suppose a regularly administered survey provides hundreds of estimates of wages for different occupations, and for budgetary reasons direct estimates for many of the targeted occupations will not be stable or even necessarily available. Suppose different approaches for indirect estimation have been suggested and one wants to choose between them. One might hope to get the mean square error of all the estimates for each of the indirect approaches and choose the approach which tends to have the smaller mean square errors, occupation by occupation. However, such "point" estimation of mean square error can itself be sufficiently imprecise (due to the difficulties of estimating the latent biases involved) as to leave doubts about whether a satisfactory comparison has been made.

In this paper we present a way of making the comparison that relies on the availability of unbiased estimators (for at least some sizeable subset of the desired estimates). Individual calculations of mean square error are eschewed; the problem of estimating bias or bias squared is circumvented. The basic idea is to evaluate how often the (possibly) biased estimators lie within corresponding confidence intervals generated by the unbiased estimators and their variance estimates. It turns out that such coverage is a good reflection of the overall mean square error of the (class of) biased estimators.

In the next section we lay out the basic evaluation-through-coverage idea, considering two cases: (a) variance of unbiased estimator known, (b) variance of unbiased estimator not known. In Section 3 we state the conditions in which our results may reasonably be applied and note some practical considerations. Section 4 describes an application to an occupational wage survey. Section 5 concludes.

2. The basic coverage idea

Suppose we wish to evaluate the mean square error (mse) of a possibly biased estimator $\hat{\theta}_1$ and we have many repetitions of the process (sample and estimation) yielding realizations of $\hat{\theta}_1$ and likewise of an unbiased estimator $\hat{\theta}$ with its corresponding variance estimate. The method we suggest uses confidence intervals c of the corresponding unbiased estimator $\hat{\theta}$ and determines the extent to which c includes $\hat{\theta}_1$. Roughly speaking, the greater the percent included, the lower the mse of $\hat{\theta}_1$.

This methodology does not require an estimate, implicit or explicit, of the bias of $\hat{\theta}_1$, nor, in fact, of its variance. It is indifferent to the bias/variance breakdown of $\hat{\theta}_1$'s mse. It does not require surety of a model on which $\hat{\theta}_1$ might be based.

To be explicit, assume we have unbiased estimates $\hat{\theta}_g$, corresponding to estimates of interest $\hat{\theta}_{1g}$, for g =1,…,G. Form $(1-\alpha)100\%$ ci's $c_{g\alpha}=\left(\hat{\theta}_g-t_{g\alpha}\sqrt{\hat{v}_g},\hat{\theta}_g+t_{g\alpha}\sqrt{\hat{v}_g}\right)$, where $\hat{v}_g$ is a variance estimate for var($\hat{\theta}_g$), $t_{g\alpha}$ is the $(1-\alpha/2)100\%$ percentile of the t-distribution with degrees of freedom (δ) appropriate to $\hat{v}_g$. We ask what proportion p of $\hat{\theta}_{1g}$ are in $c_{g\alpha}$. This proportion allows us to assess the mean square error (mse) of estimators of type $\hat{\theta}_{1g}$ relative to mse ($\hat{\theta}_g$) = var($\hat{\theta}_g$)
Roughly speaking, the smaller p, the larger mse($\hat{\theta}_{1g}$)/mse($\hat{\theta}_g$). We first consider the preliminary, non-practical case where the variance of $\hat{\theta}$ is known.

Result 1. Suppose $\hat{\theta}\sim N(\theta,v)$ and, independently, $\hat{\theta}_1\sim N(\theta_1,v_1)$. Suppose the target is θ so that $b=\theta_1-\theta$ is the bias of $\hat{\theta}_1$ and $mse\left(\hat{\theta}_1\right)\equiv b^2+v_1=f^2v\equiv f^2\cdot mse\left(\hat{\theta}\right)$.
Consider $C=\left(\hat{\theta}-t\sqrt{v},\hat{\theta}+t\sqrt{v}\right)$, t a positive constant. Let $\Phi(.)$ be the cumulative distribution function for the standard normal distribution.

Then the *inclusion probability* p that $\hat{\theta}_1$ lies in C is $p=\Phi\left(\frac{t\sqrt{v}-b}{\sqrt{v+v_1}}\right)-\Phi\left(\frac{-t\sqrt{v}-b}{\sqrt{v+v_1}}\right)$

$$=\Phi\left(\frac{t\sqrt{v}-\sqrt{f^2v-v_1}}{\sqrt{v+v_1}}\right)-\Phi\left(\frac{-t\sqrt{v}-\sqrt{f^2v-v_1}}{\sqrt{v+v_1}}\right)$$

$$=\Phi\left(\frac{t\sqrt{v}-b}{\sqrt{\left(f^2+1\right)v-b^2}}\right)-\Phi\left(\frac{-t\sqrt{v}-b}{\sqrt{\left(f^2+1\right)v-b^2}}\right).$$

We note:

1. Normality for $\hat{\theta}$ and $\hat{\theta}_1$ is reasonable, given that they will tend to be averages of some sort.
2. The expressions for p are invariant under scale changes, that is, when $\sqrt{v}'=k\sqrt{v}$, $\sqrt{v_1}'=k\sqrt{v_1}$, and $b'=\pm kb$. This fact implies there is no loss of generality in assuming $v=1$ in making the calculations described below (for example, in Figure 1).
3. If we take $t=z_\alpha$, z_α the $(1-\alpha/2)100\%$ percentile of the standard normal distribution, so as to have a $(1-\alpha)100\%$ confidence interval $C_\alpha=\left(\hat{\theta}-z_\alpha\sqrt{v},\hat{\theta}+z_\alpha\sqrt{v}\right)$,

then the probability p that $\hat{\theta}_1$ lies in C_α is $p=\Phi\left(\frac{z_\alpha\sqrt{v}-b}{\sqrt{v+v_1}}\right)-\Phi\left(\frac{-z_\alpha\sqrt{v}-b}{\sqrt{v+v_1}}\right)$.

4. The unknown ratio f of root mean square error (rmse) of $\hat{\theta}_1$ to rmse of $\hat{\theta}$ can arise from an infinity of possible combinations of bias b and variance v_1 of $\hat{\theta}_1$.

Most importantly:

5. For a given ratio of rmse's f, p is virtually *constant* over the range of possible biases b.

For example, for f = 1 and taking $\alpha = 0.05$, we have for b = 0 and v_1 =v, i.e. when $\hat{\theta}_1$ is unbiased, that $p=\Phi\left(\frac{z_\alpha}{\sqrt{2}}\right)-\Phi\left(\frac{-z_\alpha}{\sqrt{2}}\right)=0.8342$. At the other extreme, when the mse of $\hat{\theta}_1$ is simply its bias squared, i.e. for $b=\pm\sqrt{v}$ and v_1 = 0, we find $p=\Phi(z_\alpha-1)-\Phi(-z_\alpha-1)=0.8300$. Indeed, it turns out, by straight calculation, that so long as f = 1, the probability p that $\hat{\theta}_1$ lies within the 95% confidence interval for $\hat{\theta}$, is about 83%.
The bias squared/variance breakdown of mse($\hat{\theta}_1$) has miniscule impact on p. This remarkable result is depicted in Figure 1. Please note the narrow range of the vertical scale.

Similar statements can be made for other values of f. This means that we can work backwards: knowing the percentage of times that $\hat{\theta}_{1g}$'s fall within intervals for $\hat{\theta}_g$'s, we can gauge the mse of $\hat{\theta}_1$ relative to $\hat{\theta}$. Since we can estimate the mse of $\hat{\theta}$ by standard variance estimation techniques, we can get the mean square error of $\hat{\theta}_1$, or rather some (fairly narrow) bounds on it. Furthermore, if we carry out the same procedure for another (possibly) biased estimator $\hat{\theta}_2$, we should be able to gauge the relative sizes of mse($\hat{\theta}_i$), i = 1, 2. This can be useful in comparing competing approaches to estimating θ.

2.1 The case of unknown variance

In practice the variance of $\hat{\theta}$ is unknown. We now consider the unknown variance case.

Result 2. Suppose $\hat{\theta}\sim N(\theta,v)$ and, independently, $\hat{\theta}_1\sim N(\theta_1,v_1)$. Suppose the target is θ and $b=\theta_1-\theta$ is the bias of $\hat{\theta}_1$ and that $mse(\hat{\theta}_1)\equiv b^2+v_1=f^2v\equiv f^2\cdot mse(\hat{\theta})$. Let $\hat{v}$ be an estimate of v. Consider $c=(\hat{\theta}-t\sqrt{\hat{v}},\hat{\theta}+t\sqrt{\hat{v}})$.

With $s=\sqrt{\hat{v}}$, let $F(s)=\Phi\left(\frac{ts-b}{\sqrt{v+v_1}}\right)-\Phi\left(\frac{-ts-b}{\sqrt{v+v_1}}\right)$.

Then the probability p that $\hat{\theta}_1$ lies in c is

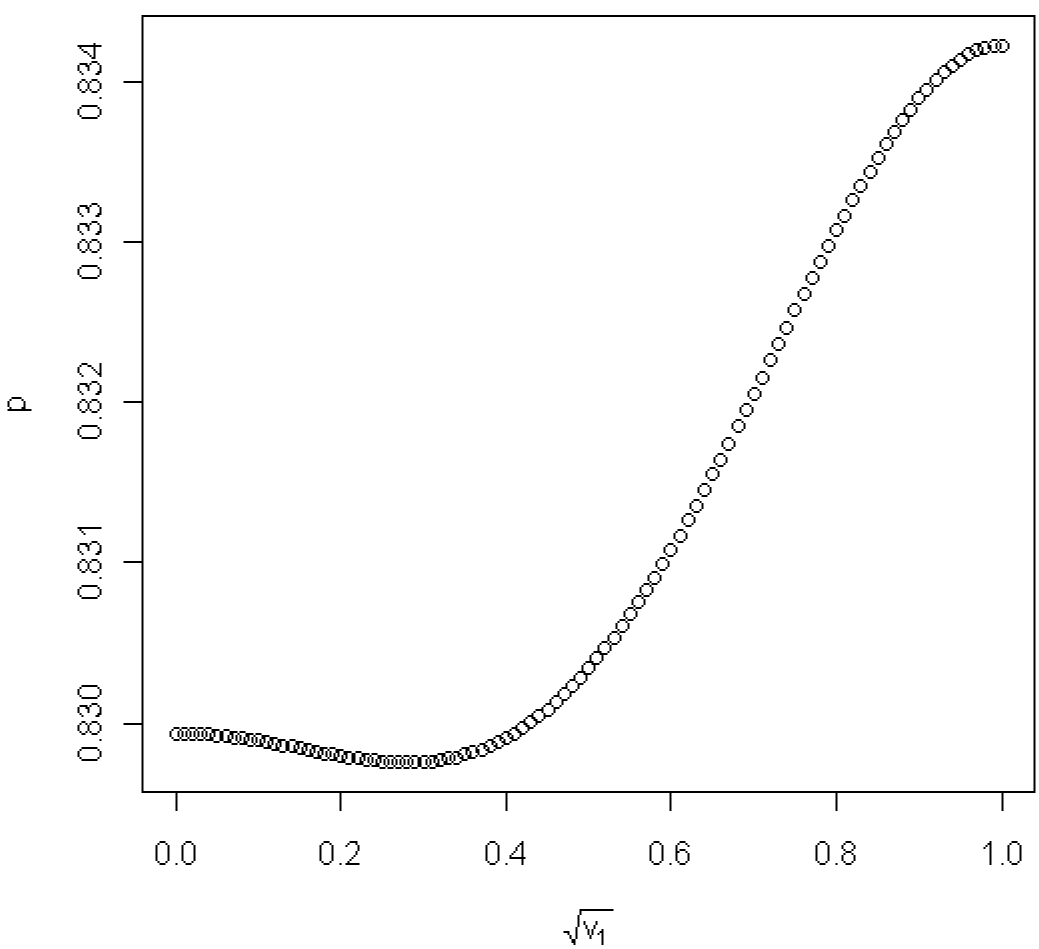
$p=\int_0^\infty F(s)\,f_s(s)\,ds$, where $f_s(s)$ is the density function of $s=\sqrt{\hat{v}}$.

We note:

6. The expression for p is invariant under a constant transformation of bias and standard errors.

7. Taking $t = t_\alpha$, the $(1-\alpha/2)100\%$ percentile of a t-distribution with degrees of freedom δ appropriate to $\hat{v}$, gives the $(1-\alpha)100\%$ confidence interval $c_\alpha = \left(\hat{\theta} - t_\alpha\sqrt{\hat{v}}, \hat{\theta} + t_\alpha\sqrt{\hat{v}}\right)$ and $F(s) = \Phi\left(\frac{t_\alpha s - b}{\sqrt{v + v_1}}\right) - \Phi\left(\frac{-t_\alpha s - b}{\sqrt{v + v_1}}\right)$.

Fig. 1 Coverage of θ_1 by 95% CI of θ, equal mse case



8. If $\delta\hat{v}/v$ is distributed as a chi-square with δ degrees of freedom, then $f_s(s) = 2\chi_\delta\left(\frac{\delta s^2}{v}\right)\frac{\delta s}{v}$, where $\chi_\delta(\bullet)$ is the density function for the chi-square distribution with δ degrees of freedom.

9. The integral can be closely approximated by calculating a finite sum of components on a fine mesh: $p \approx \sum_{j=1}^{N} F(s_j) f_s(s_j)\Delta$, where $s_j = j\ \Delta$, $j = 1,\ldots,N$, with Δ small and $N\Delta$ large enough that further terms do not materially affect the approximation.

And again, most importantly,

10. for given confidence coefficient α, for given degrees of freedom δ, for particular rmse ratio *f*, the probability *p* that $\hat{\theta}_1$ is in the confidence interval c_α is virtually constant.
We show this with some tabulations. It is convenient to refine the notation for the inclusion probability at this point, taking into account the variables it can depend on. We let $p = p(\alpha, \delta, f, b)$, $p_{\min}(\alpha, \delta, f) = \min\{p(\alpha, \delta, f, b); 0 \le b \le f\sqrt{v}\}$ and $p_{\max}(\alpha, \delta, f) = \max\{p(\alpha, \delta, f, b); 0 \le b \le f\sqrt{v}\}$.
For fixed α, δ, f, $p_{\max}(\alpha, \delta, f)$ and $p_{\min}(\alpha, \delta, f)$ are the outer bounds on *p* over the possible bias/variance breakdowns of $\text{mse}(\hat{\theta}_1)$; the min, for example, may occur at 0 bias, 0 variance, or someplace in between.

Tables 1 – 2 give values of p_{min}, p_{max} respectively (as percents) for alpha = .05, corresponding to 95% confidence intervals for theta.hat, using a grid of values of *b* over which to estimate these quantities. One thing to note is that in both tables changes in *p* are very gradual as the degrees of freedom change. Similar tables can be constructed for say 90% and 99% confidence intervals.

Table 1. Minimal *p* for α = .05

f / δ	0.2	0.4	0.6	0.8	0.9	1	1.1	1.2	1.4	1.6	1.8	2
3	94.8	94.0	92.8	91.1	90.0	88.9	87.6	86.2	83.0	79.4	75.5	71.1
4	94.7	93.9	92.4	90.4	89.1	87.8	86.2	84.6	80.8	76.4	71.6	66.4
5	94.7	93.7	92.1	89.9	88.5	87.0	85.3	83.4	79.2	74.3	69.0	63.2
6	94.7	93.6	91.9	89.5	88.1	86.4	84.6	82.6	78.0	72.8	67.1	60.9
7	94.7	93.6	91.8	89.2	87.7	86.0	84.1	82.0	77.2	71.7	65.7	59.2
8	94.6	93.5	91.7	89.0	87.4	85.6	83.7	81.5	76.5	70.8	64.6	57.9
9	94.6	93.5	91.6	88.9	87.2	85.4	83.3	81.1	75.9	70.1	63.7	56.9
10	94.6	93.5	91.5	88.7	87.0	85.2	83.1	80.7	75.5	69.5	63.0	56.0
11	94.6	93.4	91.4	88.6	86.9	85.0	82.8	80.5	75.1	69.1	62.4	55.4
12	94.6	93.4	91.4	88.5	86.8	84.8	82.6	80.2	74.8	68.7	61.9	54.8
13	94.6	93.4	91.3	88.4	86.7	84.7	82.5	80.1	74.6	68.3	61.5	54.3
14	94.6	93.4	91.3	88.3	86.6	84.6	82.3	79.9	74.3	68.0	61.1	53.9
15	94.6	93.4	91.2	88.3	86.5	84.5	82.2	79.7	74.1	67.8	60.8	53.5
16	94.6	93.3	91.2	88.2	86.4	84.4	82.1	79.6	73.9	67.5	60.5	53.2
17	94.6	93.3	91.2	88.2	86.3	84.3	82.0	79.5	73.8	67.3	60.3	52.9
18	94.6	93.3	91.2	88.1	86.3	84.2	81.9	79.4	73.6	67.2	60.1	52.7
19	94.6	93.3	91.1	88.1	86.2	84.2	81.9	79.3	73.5	67.0	59.9	52.4
20	94.6	93.3	91.1	88.1	86.2	84.1	81.8	79.2	73.4	66.9	59.7	52.2
30	94.6	93.2	91.0	87.8	85.9	83.7	81.3	78.7	72.7	65.9	58.6	51.0
40	94.6	93.2	90.9	87.7	85.7	83.6	81.1	78.4	72.3	65.5	58.0	50.3
50	94.6	93.2	90.9	87.6	85.7	83.4	81.0	78.2	72.1	65.2	57.7	49.9
60	94.6	93.2	90.9	87.6	85.6	83.4	80.9	78.1	71.9	65.0	57.5	49.7
70	94.6	93.2	90.8	87.6	85.6	83.3	80.8	78.0	71.8	64.9	57.3	49.5
80	94.5	93.2	90.8	87.5	85.5	83.3	80.8	78.0	71.8	64.8	57.2	49.4
90	94.5	93.2	90.8	87.5	85.5	83.2	80.7	77.9	71.7	64.7	57.1	49.3
100	94.5	93.2	90.8	87.5	85.5	83.2	80.7	77.9	71.6	64.6	57.0	49.2

Table 2. Maximum *p* for α = .05

f / δ	0.2	0.4	0.6	0.8	0.9	1	1.1	1.2	1.4	1.6	1.8	2
3	94.8	94.0	92.8	91.1	90.1	89.0	87.8	86.6	83.9	81.0	78.0	75.0
4	94.7	93.9	92.4	90.4	89.2	87.9	86.5	85.0	81.8	78.5	75.1	71.8
5	94.7	93.7	92.2	89.9	88.6	87.1	85.6	83.9	80.5	76.9	73.3	69.8
6	94.7	93.7	92.0	89.6	88.1	86.6	84.9	83.2	79.5	75.8	72.0	68.4
7	94.7	93.6	91.8	89.3	87.8	86.2	84.4	82.6	78.8	75.0	71.1	67.5
8	94.6	93.5	91.7	89.1	87.5	85.8	84.1	82.2	78.3	74.4	70.5	66.7
9	94.6	93.5	91.6	88.9	87.3	85.6	83.8	81.9	77.9	73.9	70.0	66.2
10	94.6	93.5	91.5	88.8	87.1	85.4	83.5	81.6	77.6	73.5	69.5	65.7
11	94.6	93.4	91.5	88.7	87.0	85.2	83.3	81.4	77.3	73.2	69.2	65.4
12	94.6	93.4	91.4	88.6	86.9	85.1	83.2	81.2	77.1	72.9	68.9	65.1
13	94.6	93.4	91.4	88.5	86.8	84.9	83.0	81.0	76.9	72.7	68.7	64.8
14	94.6	93.4	91.3	88.4	86.7	84.8	82.9	80.9	76.7	72.5	68.5	64.6
15	94.6	93.4	91.3	88.4	86.6	84.7	82.8	80.7	76.6	72.4	68.3	64.4
16	94.6	93.4	91.3	88.3	86.5	84.7	82.7	80.6	76.4	72.2	68.1	64.3
17	94.6	93.3	91.2	88.2	86.5	84.6	82.6	80.5	76.3	72.1	68.0	64.1
18	94.6	93.3	91.2	88.2	86.4	84.5	82.5	80.5	76.2	72.0	67.9	64.0
19	94.6	93.3	91.2	88.2	86.4	84.5	82.5	80.4	76.1	71.9	67.8	63.9
20	94.6	93.3	91.2	88.1	86.3	84.4	82.4	80.3	76.1	71.8	67.7	63.8
30	94.6	93.3	91.1	87.9	86.1	84.1	82.0	79.9	75.5	71.2	67.1	63.2
40	94.6	93.2	91.0	87.8	85.9	83.9	81.8	79.7	75.3	70.9	66.8	62.9
50	94.6	93.2	90.9	87.7	85.8	83.8	81.7	79.6	75.1	70.8	66.6	62.7
60	94.6	93.2	90.9	87.7	85.8	83.8	81.6	79.5	75.0	70.7	66.5	62.5
70	94.6	93.2	90.9	87.6	85.7	83.7	81.6	79.4	75.0	70.6	66.4	62.5
80	94.6	93.2	90.9	87.6	85.7	83.7	81.6	79.4	74.9	70.5	66.3	62.4
90	94.6	93.2	90.9	87.6	85.7	83.6	81.5	79.3	74.9	70.5	66.3	62.3
100	94.5	93.2	90.9	87.6	85.7	83.6	81.5	79.3	74.8	70.4	66.2	62.3

Figures 2 and 3 give graphical representation of $p_{\max}(\alpha,\delta,f)$ and $p_{\min}(\alpha,\delta,f)$ for selected values of α and δ. In each of these figures the lower curve represents the smallest value of *f* compatible with the level of coverage p, and the upper curve the largest value. The shape of the area between these lines varies in detail when we change α and δ but in general is a cornucopia opening to the upper left. Thus, we can note that it is easier to precisely assess small *f* than large *f* from their corresponding compatible inclusion probabilities.

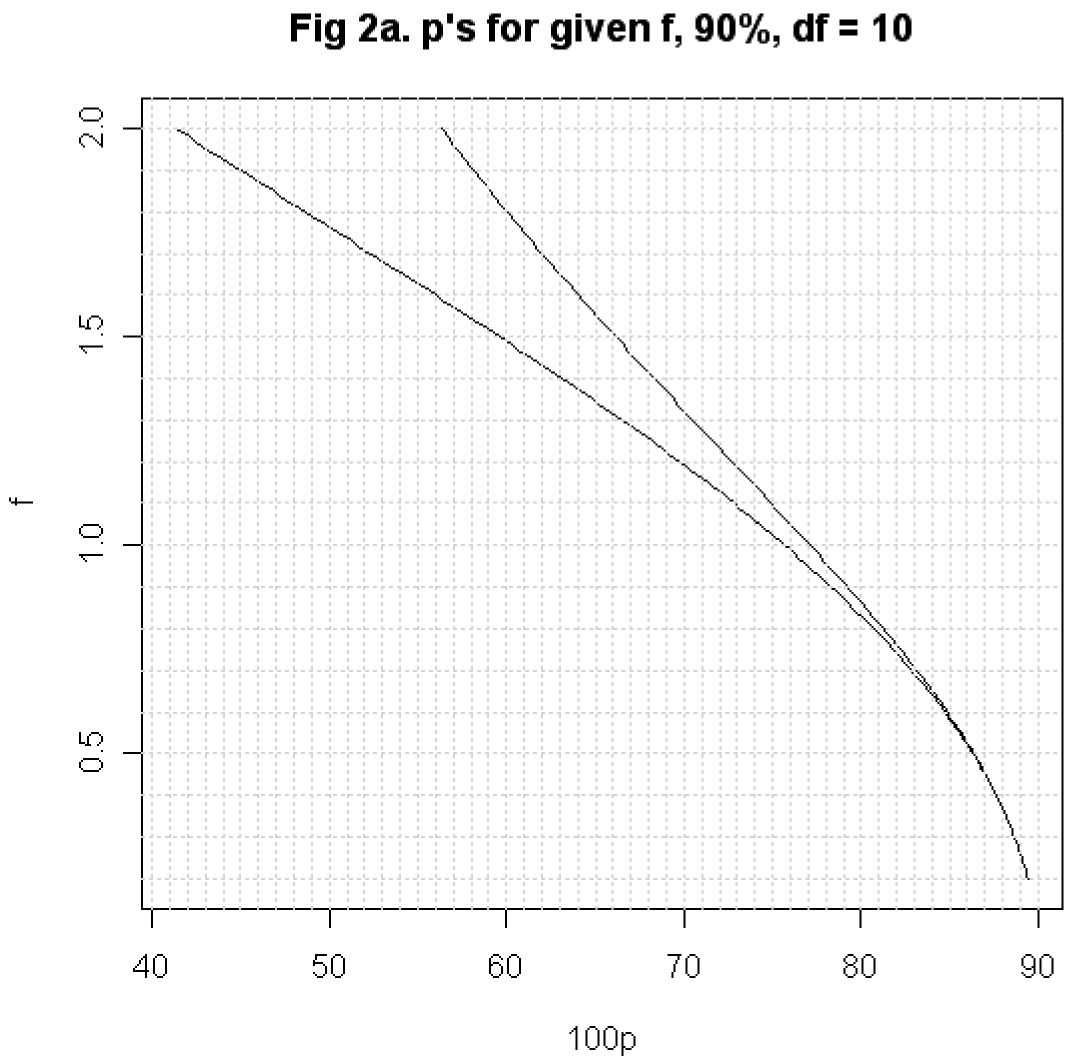
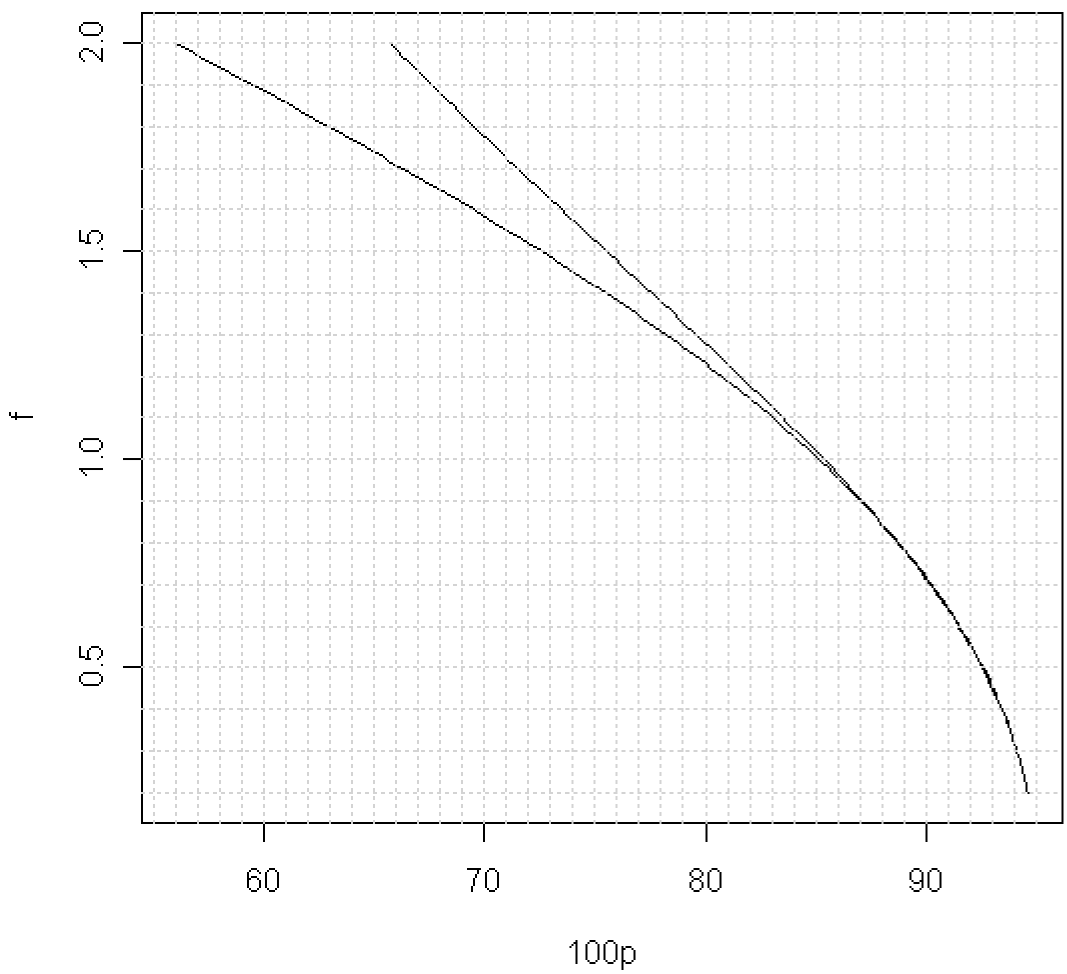


Fig 2a. p's for given f, 90%, df = 10
2.0
1.5
1.0
0.5
f
40
50
60
70
80
90
100p

Fig 2b. p's for given f, 95%, df = 10



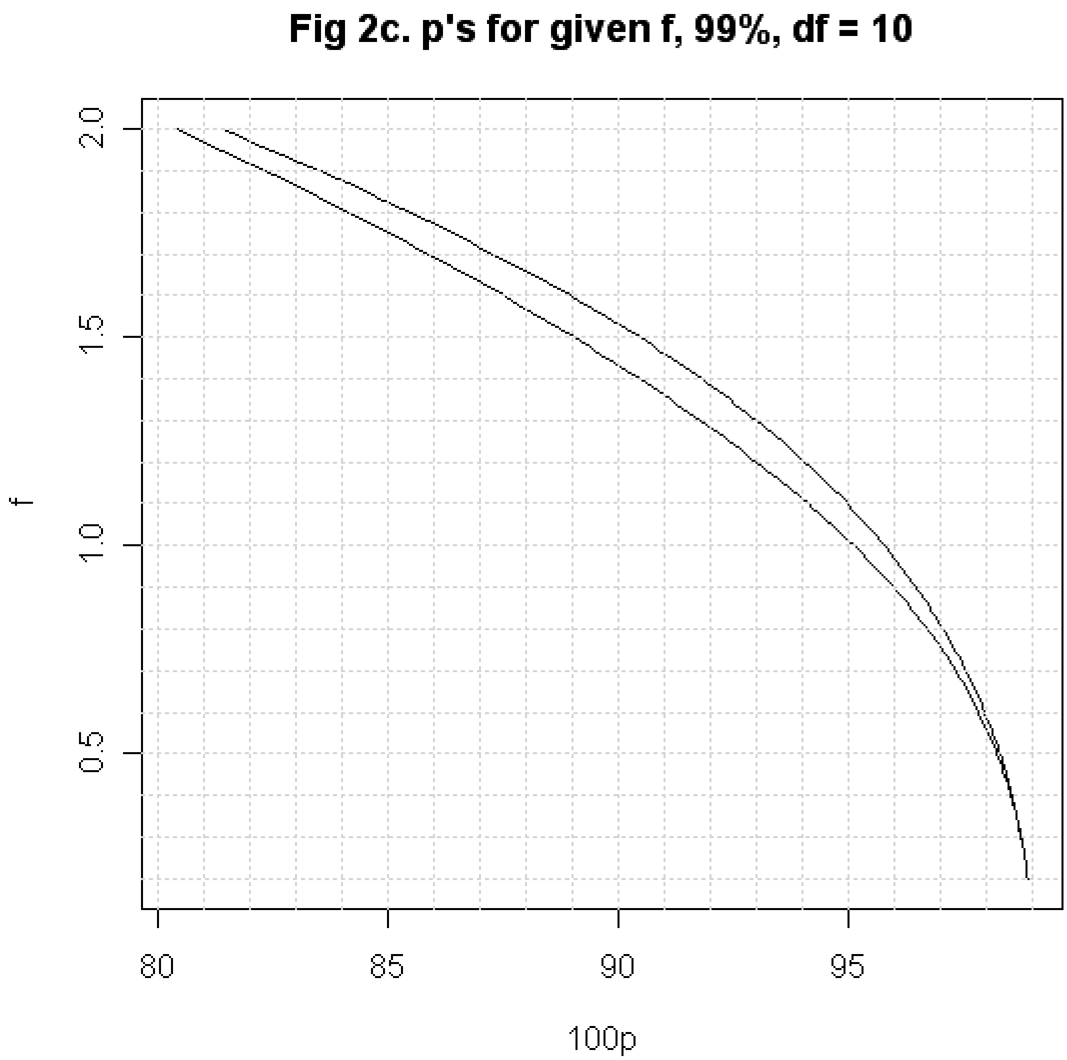


Fig 2c. p's for given f, 99%, df = 10
f
2.0
1.5
1.0
0.5
80
85
90
95
100p

Fig 3a. p's for given f, 90%, df = 20

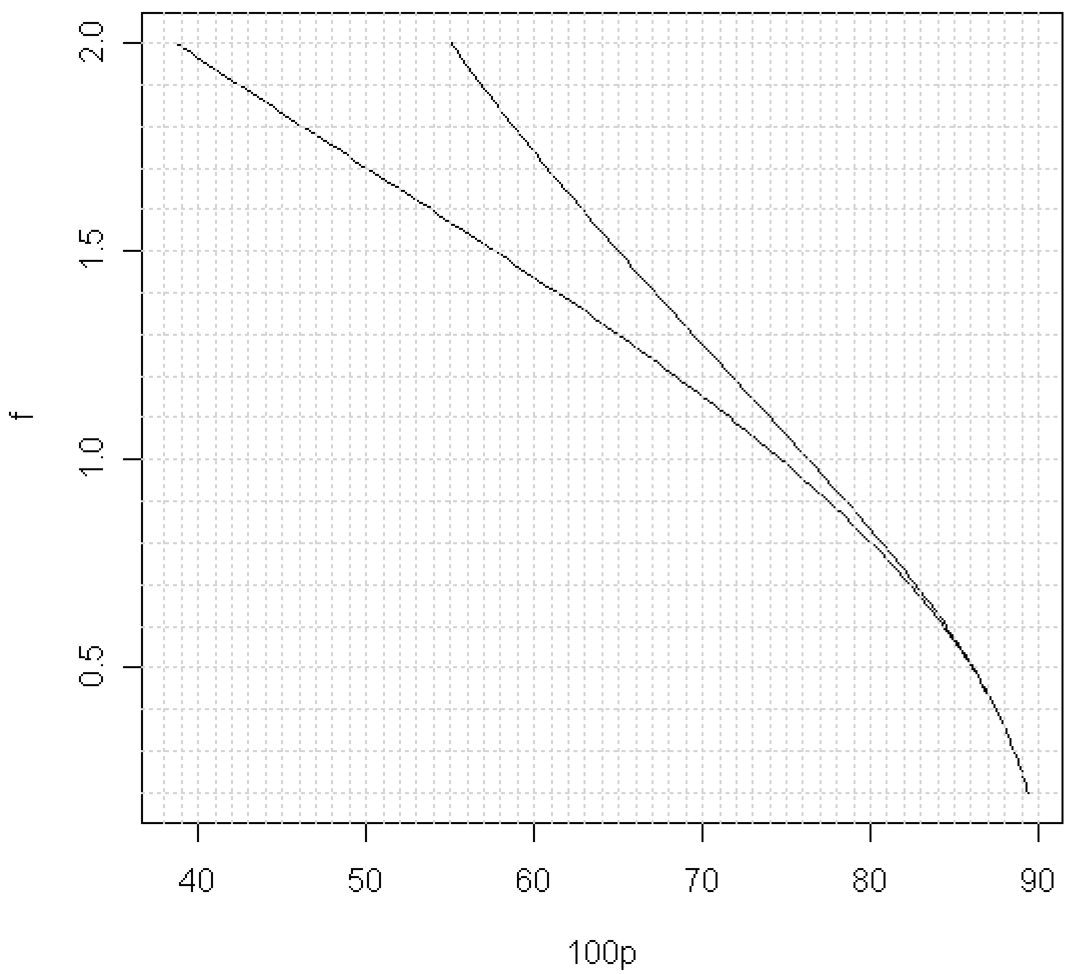


Fig 3b. p's for given f, 95%, df = 20

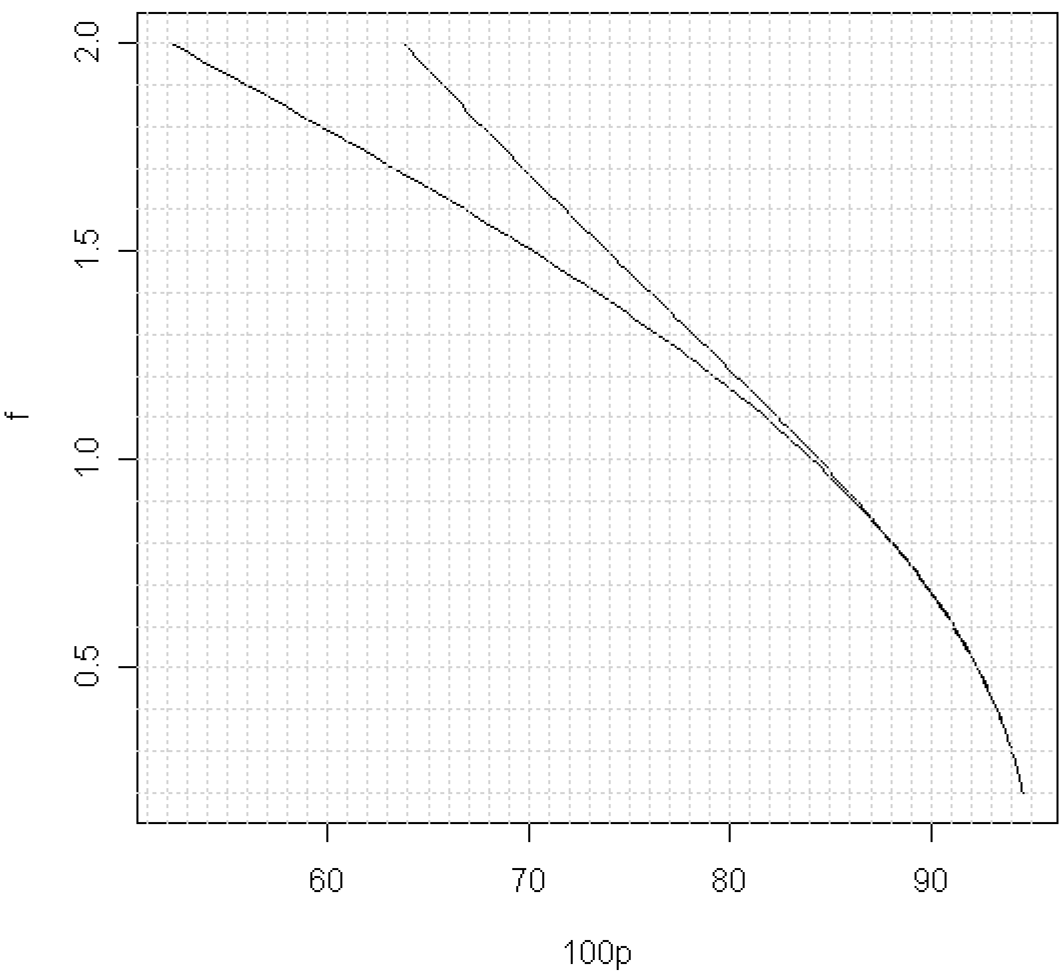
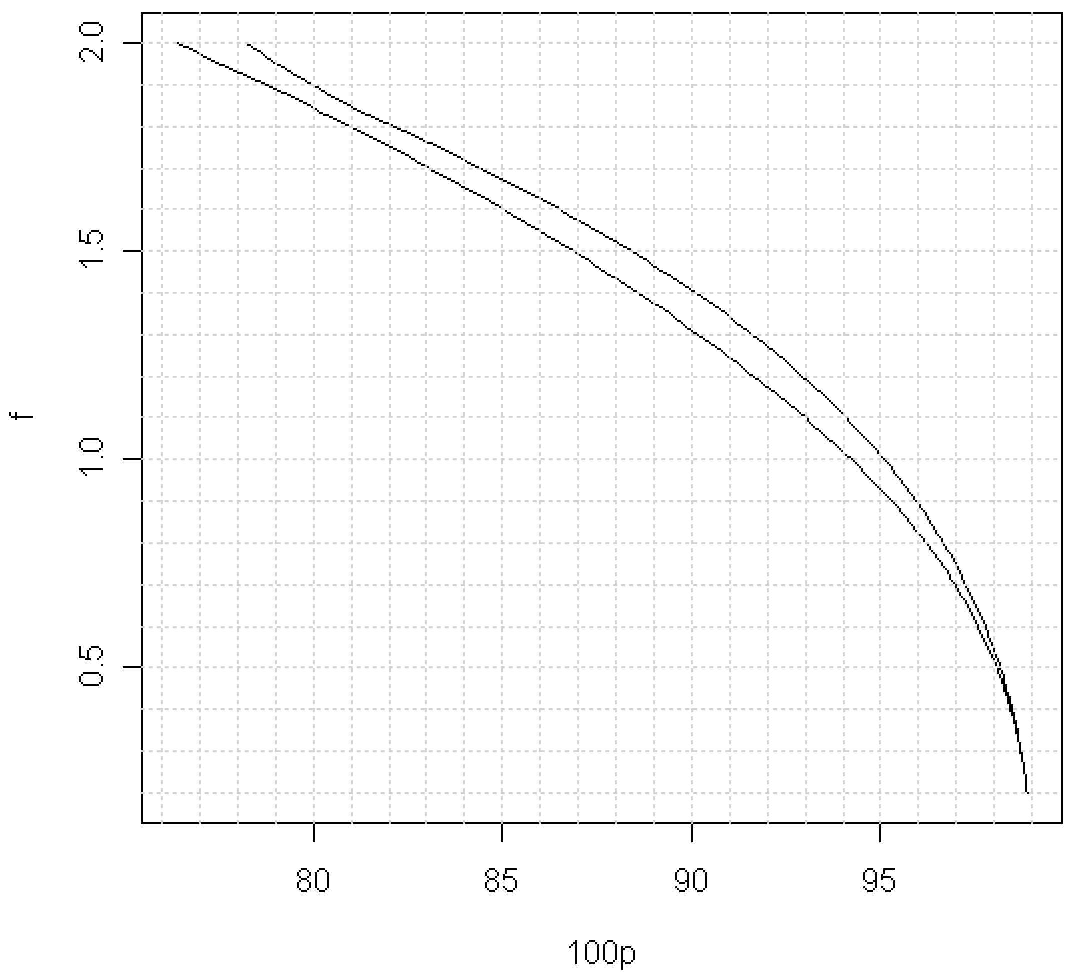


Fig 3c. p's for given f, 99%, df = 20



What level coverage intervals might be best to use? A 95% confidence intervals is probably best, for a combination of reasons:
(a) for f near 1, it has the narrowest band, and the estimators we are interested in comparing are most likely to have f in this region
(b) although it will be harder to be precise when $f > 1$, this may not be of too much concern, since estimators with a really large f are not likely to be estimators we'd want to use anyway. It won't matter that much what the exact value of f is for f known to be large already.

3. Practical considerations

Suppose we are interested in comparing two or more *methods* of estimation, each giving rise to families of estimates, say $\left\{\hat{\theta}_{1g}\right\}_{g\in G}$, $\left\{\hat{\theta}_{2g}\right\}_{g\in G}$. Each g might correspond to a different target, for example, wage of a particular occupation. It is of course possible that rmse will vary with g because of the population structure for different targets, even where the amount and type of data available for the estimates is the same. Thus in comparing the methods we must rely on overall relative rmse's f_1, f_2, which we can hope will reflect the average of f's we would get, if we could repeatedly estimate each target many times.

We assume we can get $\{\hat{\theta}_g\}_{g\in G}$, the corresponding family of unbiased estimators, and their respective confidence intervals $\{c_{g\alpha}\}_{g\in G}$. Then we estimate the inclusion probabilities p_1 and p_2 by the fraction of times $\hat{\theta}_{1g} \in c_{g\alpha}$ and $\hat{\theta}_{2g} \in c_{g\alpha}$ respectively.

Since the relationship between p's and f's depends on the degrees of freedom associated with the variance estimate of $\hat{\theta}_g$, we will want to take our overall estimates over groups within which the amount of data is about the same for the constituent estimators. We will also need to be able to at least roughly assess the degrees of freedom.

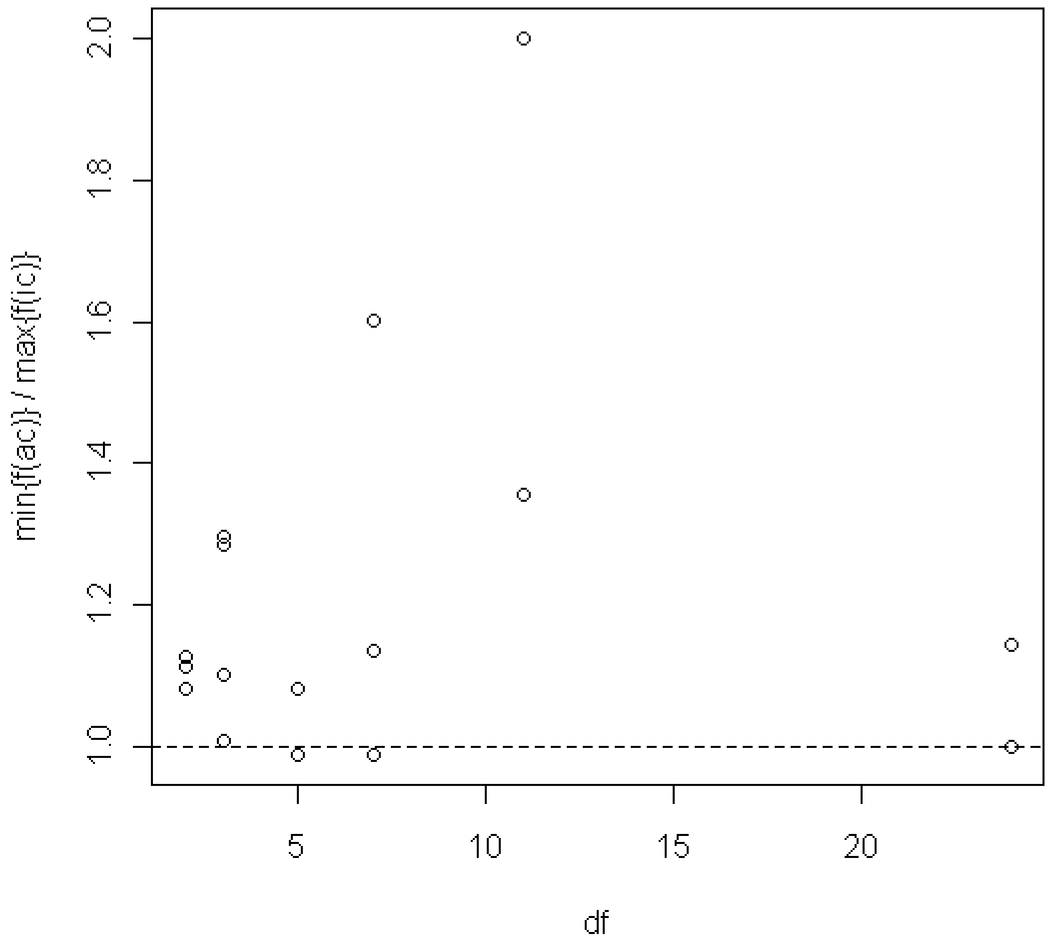
We need furthermore to take into account two assumptions that entered into the above *Results*:
(1) independence of the estimators $\hat{\theta}$ and $\hat{\theta}_1$ and
(2) normality of these estimators.

To achieve independence, we can divide the sample into mutually equal-sized, exclusive, uncorrelated parts, calculating $\hat{\theta}$ on one part and $\hat{\theta}_1$ on the other. In practice this may complicate calculation of variances and degrees of freedom, and reduce the chances of normality. For this reason, auto-calibration is a good idea: f should equal 1, if the estimators on the two partial samples are both the unbiased estimator. That is, we can assess the mean square error of $\hat{\theta}_1 = \hat{\theta}$ calculated from one sub-sample by the amount of coverage of confidence intervals for $\hat{\theta}$ derived from the other sub-sample. If this process yields coverage less (more) than that corresponding to f = 1 for the nominal degrees of freedom, we determine what degrees of freedom is needed to match the coverage we would get with f = 1. We then assume this degrees of freedom in assessing the biased estimators.

4. An Application

Two approaches (labeled "ac" and "ic") to estimating average annual wage for an occupation, depending on different assumptions and methodology, were applied to data from a occupation wage survey, for which a conventional design-unbiased estimator was available for many occupations. These occupations were divided into sub-groups that were fairly homogeneous with respect to degrees of freedom, and the coverage method applied to each such group.
Results are depicted in Figures 4a and 4b. Each point gives the value of f_1/f_2 = mse($\hat{\theta}_{1g}$)/mse($\hat{\theta}_{2g}$) for a different sub-group. A ratio of *one* would have meant they were on a par. The smaller this ratio, the better $\hat{\theta}_{1g}$. In the first figure, there seems to be a high consistency of results, suggesting that the *ac* estimator is worse than the *ic* estimator. (Note that the *miniumum* value of f(ac) is being compared to the *maximum* of f(ic), which favors *ac*-.) There seems little question that, by our test, *ic* has smaller mean square error, overall.

Fig 4a. Ratio of min{f(ac)} to max{f(ic)} within groups using confidence intervals from first sub-sample



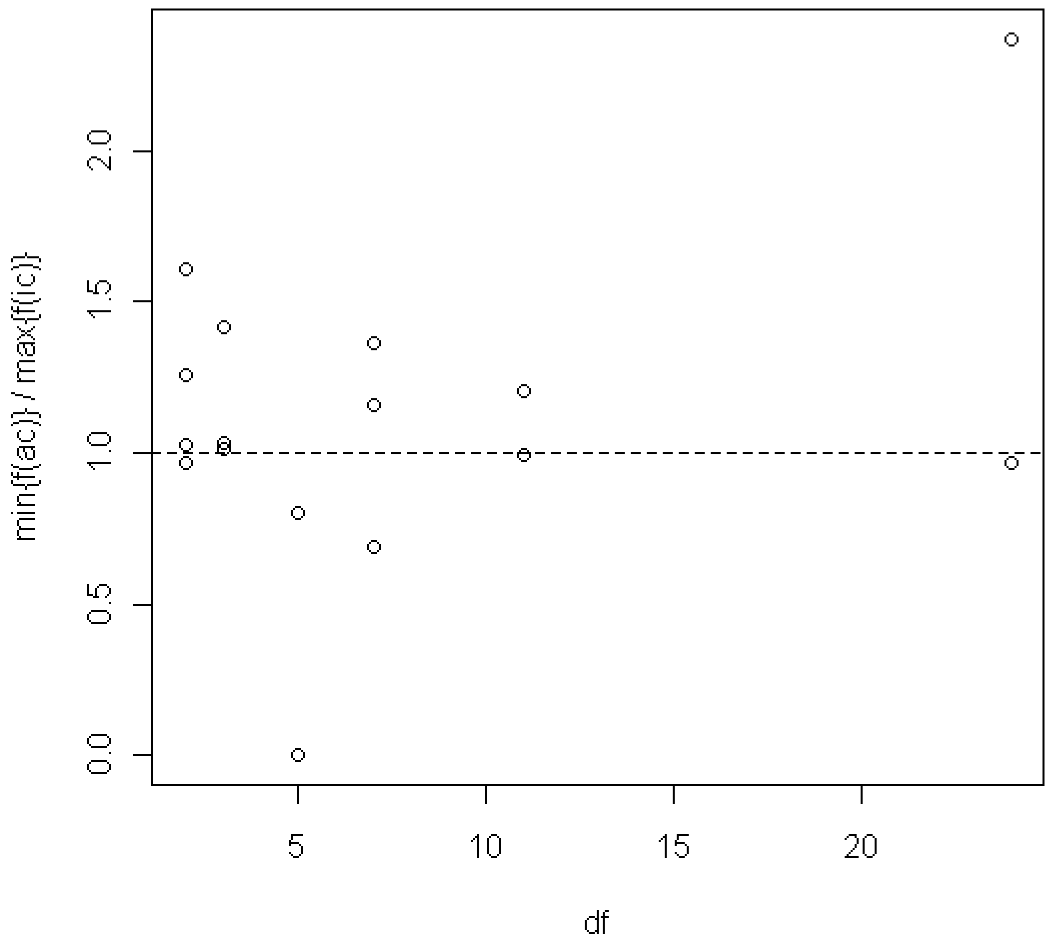
The procedure could equally well be done reversing the roles of the partial samples. The second figure gives the results which are somewhat less clear cut. There are some anomalies, but *ic* still seems favored overall.

5. Conclusion

This paper introduces a new way of evaluating classes of (possibly biased) estimators with respect to their mean square error. This "coverage approach" rests on the availability of an unbiased estimator and on meeting certain assumptions, but is simple to implement and in particular does not require tenuous estimates of bias or mean square error for individual estimators. The assumptions are normality of the estimators involved, independence of the unbiased estimator from the tested estimator, and assessment of degrees of freedom. Independence can be achieved by appropriate choice of data sets. The procedure seems fairly robust to some inaccuracy in assessing degrees of freedom. It is likely vulnerable to sharp deviations from normality. In practice, using the process above called "auto-calibration" may be a sufficient remedy for any lurking violations. Some further exploration of the non-normal case would be helpful.

It would be desirable to explore the application of the new technique in a variety of circumstances. One inviting venue is small area estimation (Rao, 2003). The coverage approach also might provide a new tool for assessing the effectiveness and relative effectiveness of different forms of generalized variance estimation (Wolter 1985).

Fig 4b. Ratio of min{f(ac)} to max{f(ic)} within groups using confidence intervals from second sub-sample



Acknowledgements

The views expressed are the author's and do not reflect Bureau of Labor Statistics policy.

References

Rao, J.N.K. (2003) *Small Area Estimation*, Wiley, Hoboken
Wolter, K. (1985) *Introduction to Variance Estimation*, Springer-Verlag, New York