COMBINING CORRELATED UNBIASED ESTIMATORS OF THE MEAN OF A NORMAL DISTRIBUTION

by

T. Keller
National Agricultural Statistics Service
U. S. Department of Agriculture

Ingram Olkin
Department of Statistics
Stanford University

Technical Report No. 2002-5
March 2002

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu

Combining Correlated Unbiased Estimators of the Mean of a Normal distribution

T. Keller and I. Olkin

National Agricultural Statistics Service and Stanford University

Abstract

There are many applications in which one seeks to combine multiple estimators of the same parameter. If the constituent estimators are unbiased, then the fixed linear combination which is minimum variance unbiased is well-known, and may be written in terms of the covariance matrix of the constituent estimators. In general, the covariance matrix is unknown, and one computes a composite estimate of the unknown parameter with the covariance matrix replaced by its maximum likelihood estimator. The efficiency of this composite estimator relative to the constituent estimators has been investigated in the special case for which the constituent estimators are uncorrelated. For the general case in which the estimators are normally distributed and correlated, we give an explicit expression relating the variance of the composite estimator computed using the covariance matrix, and the variance of the composite estimator computed using the maximum likelihood estimate of the covariance matrix. This result suggests that the latter composite estimator may be usefulin applications in which only a moderate sample size is available. Details of one such application are presented: combining estimates of agricultural yield obtained from multiple surveys into a single yield prediction.

1. Introduction

The need to combine estimators from different sources arises in many fields of application. In agriculture estimates may come from different experimental stations; in the medical sciences there may be multi-sites or multiple studies; sample surveys may contain subsurveys at different locations. Often

making a prediction requires the combination of estimators. The present analysis was motivated by a model to predict agricultural yield. However, the model is generic, and occurs in a variety of contexts. The specifics of the application are discussed in Section 4.

It is perhaps surprising that the earliest methods for combining estimators were nonparametric. Fisher (1930) and Tippett(1931) proposed methods for combining p -values obtained from independent studies. Fisher was motivated by agriculture and Tippett by industrial engineering.

The parametric problem was first posed by Cochran (1937), who was also motivated by an agricultural problem. For simplicity suppose that we have two estimators T_1 and T_2 of θ from a $\mathcal{N}(\theta, \sigma_1^2)$ and $\mathcal{N}(\theta, \sigma_2^2)$ population, respectively. The combined estimator

$$T = w_1 T_1 + w_2 T_2 \tag{1.1}$$

with $w_1 = \sigma_1^{-2}/(\sigma_1^{-2} + \sigma_2^{-2}), w_2 = \sigma_2^{-2}/(\sigma_1^{-2} + \sigma_2^{-2})$ is unbiased and has variance

$$V(T) = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \leq min(\sigma_1^2, \sigma_2^2). \tag{1.2}$$

Consequently, the combined estimator dominates either single estimator.

In practice the variances are unknown, and estimates $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ independent of T_1, T_2, are substituted in w_1 and w_2, that is,

$$T^* = \hat{w}_1 T_1 + \hat{w}_2 T_2. \tag{1.3}$$

Of course, now the variance of T^* is no longer minimum variance, but it is unbiased.

Cochran's paper was the genesis for a sequence of papers. We briefly describe these in chronological order. Graybill and Deal (1959) started with the Cochran model and assumed that the estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are independent and that each arises from a sample of size larger than 9. Under this

condition, they show that T^* is uniformly better than either T_1 or T_2, where better means smaller variance.

Seshadri (1974), motivated by balanced incomplete block (BIB) design considerations, assumes that there is an unbiased estimator $\hat{b}$ of the ratio $b = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$, independent of T_1 and T_2 . Then the estimator

$$T^{(1)} = (1 - \hat{b})T_1 + \hat{b}T_2 \tag{1.4}$$

is unbiased, and var $T^{(1)} \leq$ min $(varT_1, varT_2)$ provided Var $\hat{b} \leq b^2$ and Var $(1 - \hat{b}) \leq (1 - b)^2$. The key point is that in certain BIB designs there is an intra-block and inter-block estimator, and also an estimator $\hat{b}$.

Zacks (1966) starts with the assumption that the ratio $\rho = \sigma_2^2/\sigma_1^2$ is unknown, and creates an estimator

$$T^{(2)} = (\hat{\rho}T_1 + T_2)/(\hat{\rho} + 1), \tag{1.5}$$

where $\hat{\rho}$ is independent of T_1 and T_2. Then $T^{(2)}$ is unbiased. The efficiency of $T^{(2)}$ cannot be given in closed form, and Zacks provides graphs of the efficiency relative to the estimator $T^{(2)}$ with ρ replacing $\hat{\rho}$.

When the sample sizes of the two samples are equal to n, Cohen and Sackrowitz (1974) discuss estimators of the form

$$T^{(3)} = \hat{\alpha}_1 T_1 + \hat{\alpha}_2 T_2, \tag{1.6}$$

where α_i are functions of sample variances and are chosen with respect to a squared error loss function normalized by σ_1^2. They determine the sample size n for which $T^{(3)}$ is superior to either T_1 or T_2.

Because the estimators T_i of the mean and s_i^2 of the variances are location and scale estimators, Cohen (1974) considers a location-scale family as a more general construct than the normal family. Again, the combined estimator is

$$T^{(4)} = \hat{b}_1 T_1 + \hat{b}_2 T_1, \quad \hat{b}_1 + \hat{b}_2 = 1, \tag{1.7}$$

where now $\hat{b}_2 = c\hat{\sigma}_1^2/(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$, c is a suitably chosen constant, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are appropriately chosen estimators.

The extension from combining two estimators to combining k estimators from k normal populations $N(\theta, \sigma_i^2)$, $i = 1, \ldots, k$, is discussed by Norwood and Hinkelmann (1977). Here

$$T^{(5)} = \hat{w}_1 T_1 + \ldots + \hat{w}_k T_k \tag{1.8}$$

with $\hat{w}_i = \hat{\sigma}_i^{-2} / \sum_1^k \hat{\sigma}_j^{-2}$. They show that Var $(T^{(5)}) \leq \min \{Var \quad T_i\}$ if each sample size is greater than 9, or if some sample size is equal to 9, and the others need be greater than 17.

For the case $k = 2$ Nair (1980) computes the variance of T^* as an infinite series, and as a function of two parameters, σ_1^2 and $\alpha = n_1\sigma_1^2/n_2\sigma_2^2$. Of course, it is symmetric and can be restated as a function of σ_2^2 and $1/\alpha$.

Following the formulation of Cohen and Sackrowitz (1974), Kubokawa (1987) provides a family of minimax estimators under normalized quadratic loss functions. Green and Strawderman (1991) also consider quadratic loss and provide a James-Stein shrinkage estimator. The use of a quadratic loss function is extended to the multivariate case by Loh (1991), where now we have normal populations $N(\theta, \mathbf{\Sigma_1})$ and $N(\theta, \mathbf{\Sigma_2})$. As in the univariate case, there are mean estimators $\hat{\theta}_1, \hat{\theta}_2$ and independent covariance matrix estimators S_1, S_2, each having a Wishart distribution, the loss function is

$$L(\hat{\theta}, \theta, \Sigma_1, \Sigma_2) = (\hat{\theta} - \theta)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\theta} - \theta) \tag{1.9}$$

The estimator

$$\hat{\theta} = (S_1^{-1} + S_2^{-1})^{-1}(\Sigma_1^{-1}\hat{\theta}_1 + \Sigma_2^{-1}\hat{\theta}_2) \tag{1.10}$$

is shown to be best linear unbiased.

The model that we here consider is that there are k normal populations $\mathcal{N}(\theta, \sigma_i^2), i = 1, \ldots, k$. The data available are k unbiased estimators $T_1, \ldots, T_k$ of θ. However, the vector $T = (T_1, \ldots, T_k)$ has covariance matrix Σ, for

which there is a sample covariance matrix S having a Wishart distribution $\mathcal{W}(\Sigma, k, n)$. Furthermore, S and $(T_1, \ldots T_k)$ are independent.

When Σ is known, the linear estimator

$$\hat{\theta} = w_1 T_1 + \ldots + w_k T_k, \quad w_1 + \ldots + w_k = 1, \tag{1.11}$$

with $w_i, i = 1, \ldots, k$, fixed is unbiased. Let $w = (w_1, \ldots, w_k)'$ and $e = (1, \ldots, 1)'$. For the choice

$$w = (e'\Sigma^{-1})/(e'\Sigma^{-1}e), \tag{1.12}$$

$\hat{\theta}$ is also minimum variance. Furthermore,

$$Var(\hat{\theta}) = \frac{e'\Sigma^{-1}[\mathcal{E}(T - \theta e)'(T - \theta e)]\Sigma^{-1}e}{(e'\Sigma^{-1}e)^2} = \frac{1}{e'\Sigma^{-1}e}. \tag{1.13}$$

That Var $(\hat{\theta})$ is minimum variance follows from the Cauchy-Schwartz inequality:

$$(w'\Sigma w)(e'\Sigma^{-1}e) \geq (w'e) = 1 \tag{1.14}$$

with equality if and only if (1.12) holds. Also,

$$(e'\Sigma^{-1}e)^{-1} \leq \min\{\sigma_1^2, \ldots, \sigma_k^2\}, \tag{1.15}$$

which follows from (1.14) with $w = e_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$.

When Σ is unknown it is estimated by S, and we consider the estimator

$$\tilde{\theta} = (e'S^{-1}T)/(e'S^{-1}e). \tag{1.16}$$

The estimator $\tilde{\theta}$ is unbiased and has variance

(1.17)

$$\begin{aligned} Var(\tilde{\theta}) &= \mathcal{E}_S \mathcal{E}_T \frac{e'S^{-1}[(T-\theta e)'(T-\theta e)]S^{-1}e}{(e'S^{-1}e)^2} \\ &= \mathcal{E}_S \frac{e'S^{-1}\Sigma S^{-1}e}{(e'S^{-1}e)^2}. \end{aligned}$$

In the next section we provide a proof of the basic result:

$$Var(\tilde{\theta}) = \left(\frac{n-1}{n-k}\right) Var(\hat{\theta}). \tag{1.18}$$

2. Proof of the Theorem

The Wishart density of S is

$$f(S) = C(k,n) \mid \Sigma \mid^{-n/2} \mid S \mid^{\frac{(n-k-1)}{2}} \quad exp(-\frac{1}{2}\, tr\, \Sigma^{-1}S), S > 0, \tag{2.1}$$

where

$$C(k,n) = \{2^{\frac{nk}{2}} \pi^{\frac{k(k-1)}{4}} \prod_{i=1}^{k} \Gamma(\frac{n-i+1}{2})\}^{-1}$$

and $\Sigma > 0$ (that is, Σ is positive definite).

Let $Y = \Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}}$, so that the density of Y is

$$f(Y) = C(k,n) \mid Y \mid^{\frac{(n-k-1)}{2}} exp(-\frac{1}{2}\, tr\, Y), Y > 0. \tag{2.2}$$

With $b = \Sigma^{-\frac{1}{2}} e$

$$Var(\tilde{\theta}) = \mathcal{E}\left[\frac{b'Y^{-2}b}{(b'Y^{-1}b)^2}\right]. \tag{2.3}$$

Because the density (2.2) is orthogonally invariant, that is, $\mathcal{L}(G'YG) = \mathcal{L}(Y)$ for any orthogonal matrix G, a judicious choice of G allows one to put (2.3) in a more convenient form. Let $e_1 = (1, 0, \ldots, 0)$, and choose G so that the first row of G is $b'/\sqrt{b'b}$ and the remaining $k-1$ rows of G complete an orthonormal basis for $\mathbb{G}$. Then, by construction, $Gb = \sqrt{b'b}e_1$. Consequently, with $Z = G'YG$, (2.3) becomes

$$Var(\tilde{\theta}) = \mathcal{E}[\frac{e_1'Z^{-2}e_1}{(e_1'Z^{-1}e_1)^2}]\ \frac{1}{b'b}.$$

Note that $b'b = e'\Sigma^{-1}e$, and recall that $Var(\hat{\theta}) = e'\Sigma^{-1}e$, so that

$$Var(\tilde{\theta}) = \mathcal{E}[\frac{e_1'Z^{-2}e_1}{(e_1'Z^{-1}e_1)^2}]\ Var(\hat{\theta}) \tag{2.4}$$

Remark: For any vector a of unit length, and a positive definite matrix $B, a'B^2a \geq (a'Ba)^2$. Hence (2.4) demonstrates that $Var(\tilde{\theta}) \geq Var(\hat{\theta})$ under the hypothesis that S and $T = (T_1, \ldots, T_k)'$ are independent, but with no distributional assumptions on S or T.

Now the task of proving the theorem is reduced to computing the expectation on the right side of equation (2.4). Towards that end, partition the $k \times k$ matrix Z and its inverse as

$$\mathrm{Z} = \begin{pmatrix} z_{11} & z_1' \\ z_1 & Z_{22} \end{pmatrix}, \quad \mathrm{Z}^{-1} = \begin{pmatrix} \tilde{z}_{11} & \tilde{z}_1' \\ \tilde{z}_1 & \tilde{Z}_{22} \end{pmatrix},$$

where Z_{22} and $\tilde{Z}_{22}$ are both $(k-1) \times (k-1)$.

In what follows we make use of well-known relationships between the blocks of Z and Z^{-1}. (See, for instance, Anderson, 1984.) Employing these relationships, and that $(I - uu')^{-1} = I + \frac{uu'}{1-u'u}$ the expression inside the expectation brackets in (2.4) can be written as:

$$\frac{e_1'Z^{-2}e_1}{(e_1'Z^{-1}e_1)^2} = \frac{\tilde{z}_{11}^2 + \tilde{z}_1'\tilde{z}_1}{\tilde{z}_{11}^2} = 1 + z_{11}\ u'Z_{22}^{-1}u, \tag{2.5}$$

where $u = Z_{22}^{-1/2}z_1/\sqrt{z_{11}}$ and (2.4) becomes:

(2.6) $$Var(\tilde{\theta}) = [1 + E(z_{11}\, u' Z_{22}^{-1} u)] Var(\hat{\theta}).$$

The density of Z has the form (2.2), which can be written as

(2.7)
$$f(Z_{22}, z_{11}, u) = C(k,n) \mid Z_{22} \mid^{\frac{(n-k)}{2}} exp(-\frac{1}{2} tr Z_{22}) z_{11}^{\frac{n}{2}-1} exp(-\frac{1}{2} z_{11})(1-u'u)^{\frac{(n-k-1)}{2}}.$$

Again, using orthogonal invariance, the expectation in (2.6) is

(2.8) $$\mathcal{E}[z_{11} u' Z_{22}^{-1} u] = C(k,n) I_1 I_2 I_3,$$

where

$$\begin{aligned}
I_1 &= \int_0^\infty z_{11}^{\frac{n+2}{2}-1} exp(-\frac{1}{2} z_{11}) dz_{11} = \Gamma(\frac{n+2}{2}) 2^{\frac{n+2}{2}}, \\
I_2 &= \int_{u'u<1} u'u(1-u'u)^{\frac{n-k-1}{2}} du = (k-1)\pi^{\frac{k-1}{2}} \Gamma(\frac{n-k+1}{2}) / 2\Gamma(\frac{n+2}{2}), \\
I_3 &= \int_{Z_{22}>0} (e_1' Z_{22}^{-1} e) \mid Z_{22} \mid^{\frac{n-k}{2}} exp(-\frac{1}{2} tr Z_{22}) dZ_{22}.
\end{aligned}$$

The integral I_2 can be evaluated using polar coordinates; it is also a Dirichlet Integral of Type-I, (see Sobel, Uppuluri and Frankowski, 1977). To simplify notation in I_3 let $Q = Z_{22}$, so that Q is a $(k-1) \times (k-1)$ matrix having a Wishart distributon $\mathcal{W}(I, k-1, n)$. Then $I_3 = \mathcal{E}(Q^{-1})_{11} / C(k-1, n)$. But this expectation is known (see e.g. Kshirsagar, 1978, p. 72) so that

$$I_3 = [(n-k) 2^{\frac{n}{2}}\, \pi^{\frac{k-1}{2}}\, \Gamma(\frac{n-k+1}{2})]^{-1}$$

Combining these results we obtain

$$\begin{aligned}
Var(\tilde{\theta}) &= (1 + I_1 I_2 I_3) Var(\hat{\theta}) \\
&= \frac{n-1}{n-k} Var(\hat{\theta})
\end{aligned}$$

3. The Relative Efficiency of $\tilde{\theta}$ for $k = 2$ and $n = N - 1$

The result that Var $(\tilde{\theta}) = \frac{n-1}{n-k}$ Var $(\hat{\theta})$ coincides withwhat intuition suggests: when $k = 1$, Var $(\tilde{\theta})$ = Var $(\hat{\theta})$; when $k > 1$, Var $(\tilde{\theta})$ > Var $(\hat{\theta})$, and for all k, $lim_N \to \infty$ Var $(\tilde{\theta})$ = Var $(\hat{\theta})$. But the result gives more precise information that helps one to assess the efficiency of the Graybill-Deal estimator for a given sample size.

Consider the case $k = 2$, $N = n-1$. If, without loss of generality, we take $\sigma_{11} = \min \{\sigma_{11}, \sigma_{22}\}$, then Var $(\tilde{\theta}) \leq \min (\sigma_{11}, \sigma_{22})$ when

$$\frac{1}{N-3} \leq \frac{(\sigma_{11} - \sigma_{12})^2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}. \tag{3.1}$$

In the special case for which cov $(T_1, T_2) = 0$, (3.1) is $1/(N-3) \leq \sigma_{11}/\sigma_{22} \leq 1$, which implies that Var $(\tilde{\theta})$ < min $(\sigma_{11}, \sigma_{22})$ for all $N \geq 5$. Note that this does not contradict the previously quoted result of Graybill and Deal (1959): their hypothesis allows N_1 and N_2, the sample sizes for the respective constituent estimators, to be unequal; whereas the current theorem was derived under the assumption that $N_1 = N_2 = N$.

Writing $\sigma_{11} = \alpha^2\sigma_{22}, 0 \leq \alpha \leq 1$, and denoting the correlation between T_1 and T_2 by ρ, (3.1) can be written as

$$\frac{1}{N-3} \leq \frac{(\alpha - \rho)^2}{1 - \rho^2}. \tag{3.2}$$

Then it is apparent that if one varies the parameters α and ρ so that $\alpha - \rho \to 0$, the sample size N necessary for (3.2) to hold increases without bound. But this also is intuitive: $\alpha - \rho \to 0$ is equivalent to $\hat{\theta} \to T_1$. Given a rough initial estimate for the parameters α and ρ, one may use (3.2) to obtain some idea whether the Graybill-Deal estimator dominates the better of the two constituent estimators for a given sample size.

Taking the special case $\sigma_{11} = \sigma_{22}$, (3.2) becomes

$$\frac{1}{N-3} \leq \frac{1-\rho}{1+\rho}.$$

This form of equation (3.1) implies that the sample size for (3.1) to hold increases without bound as $\rho \to 1$. Once again, this is intuitive: to say ρ is close to 1 means the estimator T_2 provides essentially the same information about θ as the estimator T_1, and hence the composite estimator cannot be expected to provide much more information than that provided by T_1 alone.

4. An Agricultural Application: Predicting Yield

The National Agricultural Statistics Service (NASS), an agency of the USDA, makes predictions of yield (defined as production per unit area) for the major US agricultural commodities, both at the level of the primary region of production for the commodity, and also at the level of the individual states that comprise that region. A diverse body of information is employed in making a yield prediction, and the process of incorporating that information into a single estimate of yield is a complex task that relies on the experience and judgment of commodity experts, as well as statistical methodology. The goal of research into composite estimation at NASS is not to replace that experience and judgment, but to apply appropriate statistical methodology to obtain a consistent means of combining key estimators of yield into a single yield prediction, and thereby provide a useful starting point for the process of making a monthly yield prediction.

There are two major types of surveys that form the basis for estimates of final yield. One is a survey that makes various measurements and counts for plants from randomly located plots in randomly selected fields having the commodity in question. The same plots are visited several times throughout the growing season, and the particular counts and measurements which are made depend on the point in the growing season at which the data is collected and the development of the crop, as well as the particular commodity in question. These survey results are used to produce a yield prediction using biological models of yield. The other type of survey, directly asks randomly selected producers of the commodity for their best estimate of the yield they anticipate for their own fields. This survey is also conducted at intervals throughout the growing season. These survey data are the basis for more than one estimate of yield, the estimates depending on how the data are summarized.

At some point after the crop has been harvested, the actual yield is known with some degree of accuracy. Although yield certainly varies widely from year to year, it is reasonable to assume that over some 'moving window' of

time the covariance structure and the biases for a set of these key estimates of yield (for a given combination of commodity, geographic region, and month of the prediction) are relatively stable, and hence the covariance structure and biases can be estimated from historical data.

There are (at least) two possible ways to deal with the biases in the estimators. If one arbitrarily enforces the constraint that $w_1 + \ldots + w_k = 1$, then among all such linear combinations, $w_1T_1 + \ldots + w_kT_k$, of the constituent estimators $T_1, \ldots, T_k$, the linear combination with the smallest mean square error is:

$$\hat{\theta} = \frac{e'M^{-1}T}{e'M^{-1}e}, \tag{4.1}$$

where $T = (T_1, \ldots, T_k)'$ and M is the mean square error matrix. Then, in analogy with the development in the case for which $T_1, \ldots, T_k$ were unbiased, one actually computes

$$\tilde{\theta} = \frac{e'\hat{M}^{-1}T}{e'\hat{M}^{-1}e}, \tag{4.2}$$

where $\hat{M}$ is an estimate of M based on the historical data. The properties of the estimator in (4.2) are not well understood. The condition that $w_1 + \ldots + w_k = 1$ has a very natural statistical interpretation in the unbiased setting: the condition insures that the composite estimate is also unbiased. In the more general setting for which the constituent indications may be biased, the condition that $w_1 + \ldots + w_k = 1$ has no such natural statistical interpretation, and seems to be merely a mathematical convenience. In particular, the composite estimator in (4.2) is, in general, biased. An alternative procedure is to use the historical data to estimate the biases of the constituent estimators, and then use these bias estimates to calculate 'corrected' estimators which are unbiased, and proceed as discussed in the first section of the paper. There are several statistical issues that arise in considering this latter procedure that need to be addressed. However, in the majority of applications to yield prediction that we have examined thus far,

the latter procedure has resulted in a somewhat smaller mean square error of prediction than that resulting from the former procedure.

Security issues preclude a presentation of an actual NASS data set. Nevertheless, in order to give the reader some idea how the methodology works in practice, a masked data set has been prepared. In Table 1, the columns headed 'farmer reported yield' and 'biological yield model' correspond to 14 years of yield predictions based on surveys of the anticipated yield reported by producers, and of the crop itself, respectively. These data have already been corrected for bias, using bias estimates based on historic data. To mimic the operation of a moving window of time, each of the 14 estimates in the column labeled 'composite estimate' is calculated using the sample covariance matrix based on the other 13 years. The column headed 'panel of experts' gives the yield prediction produced by a group of commodity experts, who are privy to a wide variety of information. That information includes not only the two constituent estimators of yield listed in the table, but other yield estimators, the raw survey data, yield estimates from previous months, data on weather, growing conditions, and cultural practices, reports on genetic improvements, information on economic inputs to the crop, prices, and so forth. The column headed 'true yield' is the last value of yield for the combination of year, geographic area and commodity in question which has been published by the USDA.

Note that the root mean square error for the composite estimator was less than that for either of the individual estimators of yield, and only slightly more than the root mean square error for the yield estimate produced by the panel of commodity experts. This latter observation is particularly remarkable in view of the bulk of additional information relevant to predicting yield that is available to the panel. The results displayed in Table 1 generalize to all the combinations of crop, month, and geographic region which we have tried so far: the composite estimator, on average, does about as well, or a little better, as a predictor of yield than the estimate of yield produced by the panel of experts. In some instances, the composite estimator does quite a bit better than any of the competing estimators, and seldom is it the case that the mean square error for the composite estimator is very much larger than the mean square error for the prediction based on the estimate of yield produced by the expert panel. Thus there is evidence that a significant reduction in the mean square error in predicting yield can be achieved by computing a Graybill-Deal type composite estimator as an initial step in the

overall yield prediction process.

Table 1

Predicted yields (weight per area) of commodity Z for state X in month Y.

year	farmer reported yield	biological yield model	composite estimate	panel of exports	true yield
1	88.0	87.5	87.8	89.5	87.8
2	82.5	80.0	81.5	82.5	87.3
3	83.0	86.5	84.2	85.8	85.3
4	73.5	79.0	75.3	76.3	76.8
5	79.0	84.5	81.3	83.3	78.3
6	82.0	83.5	82.5	83.8	89.0
7	83.0	79.8	81.8	85.0	82.5
8	80.8	84.0	81.8	81.3	84.0
9	81.0	83.0	81.7	81.8	82.3
10	79.0	79.0	79.0	81.0	80.8
11	64.0	76.0	68.3	67.5	68.3
12	80.5	83.8	81.6	83.0	83.0
13	83.0	87.0	84.4	85.0	85.0
14	81.5	78.5	80.4	82.0	81.8

Root Mean Square Error:

Farmer Reported Yield	3.06
Biological Yield Model	3.92
Composite Estimator	2.68
Panel of Experts	2.58

References

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis*, John Wiley Sons, New York.

Cochran, W.G. (1937) Problems arising in the analysis of a series of similar experiments, Supplement to the *Journal of the Royal Statistical Society* **4** 102-118.

Cohen, A. (1976) Combining estimates of location, *Journal of the American Statistical Association* **71** 172-175.

Cohen, A. and Sackrowitz, H.B. (1974) On estimating the common mean of two normal distributions. *Annals of Statistics* **2** 1274-1282.

Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.) Oliver and Boyd; London.

Graybill, F.A. and Deal, R.B. (1959) Combining unbiased estimators, *Biometrics* **15** 543-550.

Green, E.J. and Strawderman, W.E. (1991) A James-Stein type estimator for combining unbiased and possibly biased estimators, *Jornal of the American Statistical Association* **86** 1001-1006.

Kshirsagar, A. (1978) *Multivariate Analysis*, Marcel Dekker, Inc.

Kubokawa, T. (1987) Admissible minimax estimation of a common mean of two normal populations. *Annals of Statistics*, **15** 1245-1256.

Loh, W-L. (1991) Estimating the common mean of two multivariate normal distributions, *Annals of Statistics*, **19** 297-313.

Nair, K.A. (1980) Variance and distribution of the Graybill-Deal estimator of the common mean of two normal populations, *Annals of Statistics* **8** 212-216.

Norwood, T.E. and Hinkelmann, K. Jr. (1977) Estimating the common mean of several normal populations, *Annals of Statistics* **5** 1047-1050.

Raj, D. (1998) *Sampling Theory*, McGraw-Hill, 16-17.

Seshadri, V. (1963) Constructing uniformly better estimators. *Journal of the American Statistical Association* **58** 172-175.

Sobel, M., Uppuluri, R., and Frankowski, K. (1977) Selected tables in mathematical statistics, Vol. 4: Dirichlet Distribution - Type 1, *American Mathematical Society.*

Tippett, L.H.C. (1931). *The method of statistics* Williams and Norgate: London.

Zacks, S. (1966) Unbiased estimation of the common mean of two normal distributions based on small samples of equal size, *Journal of the American Statistical Association* **61** 467-476