

Combining correlated unbiased estimators of the mean of a normal distribution

Timothy Keller¹ and Ingram Olkin²

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 useful in 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; several laboratories might assay a sample of one. 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 5.

It is perhaps surprising that the earliest methods for combining estimators were nonparametric. Fisher (1932) and Tippett(1931) proposed methods for combining p -values obtained from independent studies. Fisher was motivated by agriculture and Tippett by industrial engineering. These methods have been used to combine the results of independent studies in meta-analysis.

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

¹U.S. Department of Agriculture, National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 22030-1504, USA. e-mail: keller-timothy@hotmail.com

²Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, CA 94305-4065, USA. e-mail: iolkin@stat.stanford.edu

Keywords and phrases: meta-analysis, unbiased estimators, correlated estimators.

AMS 2000 subject classifications: 62H12, 62H10.

combined estimator

$$T = w_1 T_1 + w_2 T_2 \quad (1.1)$$

with

$$w_1 = \sigma_1^{-2} / (\sigma_1^{-2} + \sigma_2^{-2}), w_2 = \sigma_2^{-2} / (\sigma_1^{-2} + \sigma_2^{-2}) \quad (1.2)$$

is unbiased and has variance

$$\text{Var}(T) = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \leq \min(\sigma_1^2, \sigma_2^2). \quad (1.3)$$

Consequently, the combined estimator dominates either single estimator in terms of having a smaller variance.

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,

$$T^* = \hat{w}_1 T_1 + \hat{w}_2 T_2. \quad (1.4)$$

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 to study the effect of using estimates of the variances. 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.

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

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

where $\hat{\rho}$ is independent of T_1 and T_2 . Then $T^{(1)}$ is unbiased. The efficiency of $T^{(1)}$ cannot be given in closed form, and Zacks (1966) provides graphs of the efficiency relative to the estimator $T^{(1)}$ with ρ replacing $\hat{\rho}$.

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

$$T^{(2)} = (1 - \hat{b}) T_1 + \hat{b} T_2 \quad (1.6)$$

is unbiased, and $\text{var } T^{(2)} \leq \min(\text{var } T_1, \text{var } T_2)$ provided $\text{Var } \hat{b} \leq b^2$ and $\text{Var}(1 - \hat{b}) \leq (1 - b)^2$. The key point is that in certain BIB designs there is an intra-block and inter-block estimator, and also an estimator \hat{b} .

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

$$T^{(3)} = \hat{\alpha}_1 T_1 + \hat{\alpha}_2 T_2, \quad (1.7)$$

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

$$T^{(4)} = \hat{b}_1 T_1 + \hat{b}_2 T_2, \quad \hat{b}_1 + \hat{b}_2 = 1, \quad (1.8)$$

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 $\mathcal{N}(\theta, \sigma_i^2)$, $i = 1, \dots, k$, is discussed by Norwood and Hinkelmann (1977). Here

$$T^{(5)} = \hat{w}_1 T_1 + \dots + \hat{w}_k T_k \quad (1.9)$$

with $\hat{w}_i = \hat{\sigma}_i^{-2} / \sum_{j=1}^k \hat{\sigma}_j^{-2}$. They show that $\text{var}(T^{(5)}) \leq \min\{\text{var} T_i\}$ if each sample size is greater than 9, or if some sample size is equal to 9, and the others are greater than 17.

For the case $k = 2$ Nair (1980) computes the variance of T^* as an infinite series, 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 $\mathcal{N}(\theta, \Sigma_1)$ and $\mathcal{N}(\theta, \Sigma_2)$. As in the univariate case, there are estimators $\hat{\theta}_1, \hat{\theta}_2$ of the mean vectors and independent covariance matrix estimators S_1, S_2 , each having a Wishart distribution. For the loss function

$$L(\hat{\theta}, \theta, \Sigma_1, \Sigma_2) = (\hat{\theta} - \theta)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\theta} - \theta), \quad (1.10)$$

with Σ_1 and Σ_2 known the estimator

$$\hat{\theta} = (S_1^{-1} + S_2^{-1})^{-1}(\Sigma_1^{-1}\hat{\theta}_1 + \Sigma_2^{-1}\hat{\theta}_2) \quad (1.11)$$

is shown to be best linear unbiased.

The model that we here consider is that there are k normal populations $\mathcal{N}(\theta, \sigma_i^2)$, $i = 1, \dots, k$. This model was considered by Halperin (1961) who provided an extensive analysis in which the estimator of θ is a weighted combination of the individual means, which are permitted to be correlated. For this model Halperin (1961) obtains the same variance as given in (2.8) below. In the present analysis the estimator of θ is a weighted combination of any unbiased estimators, and thereby permits somewhat more flexibility. Our derivation makes use of invariance arguments. In a later paper, Krishnamoorthy and Rohatgi (1990) show that the simple arithmetic mean is dominated by a shrinkage estimator that takes advantage of the variances.

2. The correlated case

As our starting point suppose that the data available are k unbiased estimators T_1, \dots, T_k of θ . However, the vector $T = (T_1, \dots, T_k)$ 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, \dots, T_k) are independent.

When Σ is known, the linear estimator

$$\hat{\theta} = w_1 T_1 + \dots + w_k T_k, \quad w_1 + \dots + w_k = 1, \quad (2.1)$$

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

$$w' = (e'\Sigma^{-1})/(e'\Sigma^{-1}e), \quad (2.2)$$

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

$$Var(\hat{\theta}) = \frac{e'\Sigma^{-1}[\mathcal{E}(T - \theta e)'(T - \theta e)]\Sigma^{-1}e}{(e'\Sigma^{-1}e)^2} = \frac{1}{e'\Sigma^{-1}e}. \quad (2.3)$$

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

$$(w'\Sigma w)(e'\Sigma^{-1}e) \geq (w'e)^2 = 1 \quad (2.4)$$

with equality if and only if (2.2) holds. Also,

$$(e'\Sigma^{-1}e)^{-1} \leq \min\{\sigma_1^2, \dots, \sigma_k^2\}, \quad (2.5)$$

which follows from (2.4) with $w \equiv e_i = (0, \dots, 0, 1, 0, \dots, 0)'$.

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

$$\tilde{\theta} = (e'S^{-1}T)/(e'S^{-1}e). \quad (2.6)$$

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

$$\begin{aligned} Var(\tilde{\theta}) &= \mathcal{E}_S \mathcal{E}_T \frac{e'S^{-1}[(T - \theta e)'(T - \theta e)]S^{-1}e}{(e'S^{-1}e)^2} \\ &= \mathcal{E}_S \frac{e'S^{-1}\Sigma S^{-1}e}{(e'S^{-1}e)^2}. \end{aligned} \quad (2.7)$$

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

$$Var(\tilde{\theta}) = \left(\frac{n-1}{n-k}\right) Var(\hat{\theta}). \quad (2.8)$$

3. Proof of the main result

The Wishart density of S is

$$f(S) = C(k, n) |\Sigma|^{-n/2} |S|^{\frac{(n-k-1)}{2}} \exp\left(-\frac{1}{2} tr \Sigma^{-1}S\right), \quad S > 0, \quad (3.1)$$

where

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

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

$$f(Y) = C(k, n) |Y|^{\frac{(n-k-1)}{2}} \exp\left(-\frac{1}{2} tr Y\right), \quad Y > 0. \quad (3.2)$$

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

$$Var(\tilde{\theta}) = \mathcal{E} \left[\frac{b'Y^{-2}b}{(b'Y^{-1}b)^2} \right]. \quad (3.3)$$

Because the density (3.2) 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 (3.3) in a more convenient form. Let $e_1 = (1, 0, \dots, 0)'$, 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 G . Then, by construction, $Gb = \sqrt{b'b} e_1$. Consequently, with $Z = G'YG$, (3.3) becomes

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

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

$$\text{Var}(\tilde{\theta}) = \mathcal{E} \left[\frac{e_1' Z^{-2} e_1}{(e_1' Z^{-1} e_1)^2} \right] \text{Var}(\hat{\theta}) \tag{3.4}$$

Remark. For any vector a of unit length, and a positive definite matrix B , $a'B^2a \geq (a'Ba)^2$. Hence (3.4) demonstrates that $\text{Var}(\tilde{\theta}) \geq \text{Var}(\hat{\theta})$ under the hypothesis that S and $T = (T_1, \dots, T_k)'$ 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 (3.4). Towards that end, partition the $k \times k$ matrix Z and its inverse as

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

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, 2003.) Employing these relationships, and that $(I - uu')^{-1} = I + \frac{uu'}{1-u'u}$ the expression inside the expectation brackets in (3.4) can be written as:

$$\frac{e_1' Z^{-2} e_1}{(e_1' Z^{-1} e_1)^2} = \frac{\tilde{z}_{11}^2 + \tilde{z}_1' \tilde{z}_1}{\tilde{z}_{11}^2} = 1 + z_{11} u' Z_{22}^{-1} u, \tag{3.5}$$

where $u = Z_{22}^{-1/2} z_1 / \sqrt{z_{11}}$; then (3.4) becomes:

$$\text{Var}(\tilde{\theta}) = [1 + E(z_{11} u' Z_{22}^{-1} u)] \text{Var}(\hat{\theta}). \tag{3.6}$$

The density of Z has the form (3.2), which can be written as

$$f(Z_{22}, z_{11}, u) = C(k, n) |Z_{22}|^{\frac{(n-k)}{2}} \exp\left(-\frac{1}{2} tr Z_{22}\right) z_{11}^{\frac{n}{2}-1} \exp\left(-\frac{1}{2} z_{11}\right) (1 - u'u)^{\frac{(n-k-1)}{2}}. \tag{3.7}$$

Again, using orthogonal invariance, the expectation in (3.6) is

$$\mathcal{E}[z_{11} u' Z_{22}^{-1} u] = C(k, n) I_1 I_2 I_3, \tag{3.8}$$

where

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

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 distribution $\mathcal{W}(I; k-1, n)$. 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

$$I_3 = \left[(n-k) 2^{\frac{n}{2}} \pi^{\frac{k-1}{2}} \Gamma\left(\frac{n-k+1}{2}\right) \right]^{-1}. \tag{3.9}$$

Combining these results we obtain

$$\text{Var}(\tilde{\theta}) = (1 + I_1 I_2 I_3) \text{Var}(\hat{\theta}) = \frac{n-1}{n-k} \text{Var}(\hat{\theta}). \tag{3.10}$$

4. Discussion of efficiency for $k = 2$ and $n = N - 1$

The result that $\text{Var}(\tilde{\theta}) = \frac{n-1}{n-k} \text{Var}(\hat{\theta})$ coincides with what intuition suggests: when $k = 1$, $\text{Var}(\tilde{\theta}) = \text{Var}(\hat{\theta})$; when $k > 1$, $\text{Var}(\tilde{\theta}) > \text{Var}(\hat{\theta})$, and for all k , $\lim_{N \rightarrow \infty} \text{Var}(\tilde{\theta}) = \text{Var}(\hat{\theta})$. 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 $\text{Var}(\tilde{\theta}) \leq \min(\sigma_{11}, \sigma_{22})$ when

$$\frac{1}{N-3} \leq \frac{(\sigma_{11} - \sigma_{12})^2}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}. \tag{4.1}$$

In the special case for which $\text{cov}(T_1, T_2) = 0$, (4.1) is $1/(N-3) \leq \sigma_{11}/\sigma_{22} \leq 1$, which implies that $\text{Var}(\tilde{\theta}) < \min(\sigma_{11}, \sigma_{22})$ 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 $N_1 = N_2 = N$. When T_1 and T_2 are uncorrelated, there are corresponding sample sizes N_1 and N_2 used in estimating the variances. However, when the T 's are correlated, the covariance matrix is estimated from a single sample of size 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 , (4.1) can be written as

$$\frac{1}{N-3} \leq \frac{(\alpha - \rho)^2}{1 - \rho^2}. \tag{4.2}$$

Then it is apparent that if one varies the parameters α and ρ so that $\alpha - \rho \rightarrow 0$, the sample size N necessary for (4.2) to hold increases without bound. But this also is intuitive: $\alpha - \rho \rightarrow 0$ is equivalent to $\hat{\theta} \rightarrow T_1$. Given a rough initial estimate for the parameters α and ρ , one may use (4.2) 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 $\sigma_{11} = \sigma_{22}$, (4.2) becomes

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

This form of equation (4.1) implies that the sample size for (4.1) to hold increases without bound as $\rho \rightarrow 1$. 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.

5. An agricultural application: Forecasting yield

The National Agricultural Statistics Service (NASS), an agency of the USDA, makes monthly pre-harvest yield forecasts for the major US agricultural commodities at several geographic levels. In the final analysis, the official forecast of yield announced to the public is necessarily the result of a mixed process of both objective scientific technique and subjective expert judgment. Nevertheless, subjective expert judgment is most effective when it has an objective estimate of yield with which to commence its operation. Given an historical data series for the most important estimators of yield, and the corresponding published *final* yield, one can estimate the covariance structure and biases for those estimators. These are then the basis for computing a composite estimate of yield. The question of how best to use historical data to estimate the biases in the constituent estimators of yield is important in itself. In order to avoid a long digression, we pick up the discussion of the application at the point where a ‘bias correction’ has already been applied to the historical data; hence only the problem of estimating the covariance matrix remains.

Table 1 presents the predicted yield based on a biological yield model (T_1) and the predicted yield based on a survey of producer expectations (T_2). These data have been masked for security considerations. Make the following assumptions:

- (1) The true yield β_i for year i is the yield published by NASS (Table 2) at the end of the growing season.
- (2) T_1 and T_2 are independent.
- (3) The covariance matrix is essentially constant over time.

Under these assumptions the maximum likelihood estimator for the covariance matrix based on the data in Table 1 is:

Table 1: Predicted yields (weight per area) of commodity Z for state X in month Y .

Year	Survey of biological yield	
1	88.0	87.5
2	82.5	80.0
3	83.0	86.5
4	73.5	79.0
5	79.0	84.5
6	82.0	83.5
7	83.0	79.8
8	80.8	84.0
9	81.0	83.0
10	79.0	79.0
11	64.0	76.0
12	80.5	83.8
13	83.0	87.0
14	81.5	78.5

$$S = \begin{pmatrix} 9.50 & 2.19 \\ 2.19 & 15.30 \end{pmatrix},$$

and the vector of weights for the linear combination of T_1 and T_2 which is the Graybill-Deal estimator of yield is $w' = (0.642, 0.358)$.

A word about the operational implementation of these ideas is in order. It is unreasonable to expect that the assumptions underlying the estimate of the covariance matrix hold for all time; hence, in practice, one envisions that yield data from a 'moving window' of N past years would be used to estimate the vector of coefficients, w , used to compute the composite estimate of yield for the *current* year. This concept has been tested by a cross-validation scheme in which each of $N + 1$ years is sequentially treated as the 'current' year, and the remaining N years are treated as the 'past', where $N + 1$ is the length of the relevant data series which is available; but, for the sake of a simple exposition, the calculations presented in Table 2 are based on all 14 years of data at once, the results of the cross-validation scheme being very similar.

Looking at Table 2, one notes that the root mean square error for the composite estimator was less than that of either constituent estimator of yield, and only slightly larger than the root mean square error for the yield forecast produced by the panel of commodity experts. Given that this panel was privy to a great many sources of information relevant to setting yield, in *addition* to the constituent estimators of yield, this is a remarkable result. One cannot hope to replace expert judgement with statistical methodology; nevertheless, these results demonstrate

Table 2:

Year	Composite Estimate ($\hat{\theta}$)	Panel of Experts	Final Published Yield (θ)
1	87.8	89.5	87.8
2	81.5	82.5	87.3
3	84.2	85.8	85.3
4	75.3	76.3	76.8
5	81.3	83.3	78.3
6	82.5	83.8	89.0
7	81.8	85.0	82.5
8	81.8	81.3	84.0
9	81.7	81.8	82.3
10	79.0	81.0	80.8
11	68.3	67.5	68.3
12	81.6	83.0	83.0
13	84.4	85.0	85.0
14	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

that the techniques of composite estimation can provide a useful starting point for the overall process of setting a yield forecast.

Acknowledgement

The authors are grateful to Brian Cullis for many helpful comments and insights relating to the paper by Halperin (1961), and to the referee for helpful suggestions.

References

- [1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition. John Wiley and Sons, New York. MR1990662
- [2] 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.
- [3] Cohen, A. (1976). Combining estimates of location, *Journal of the American Statistical Association* **71** 172–175. MR426258
- [4] Cohen, A. and Sackrowitz, H. B. (1974). On estimating the common mean of two normal distributions. *Annals of Statistics* **2** 1274–1282. MR365851
- [5] Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.) Oliver and Boyd, London.
- [6] Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators, *Biometrics* **15** 543–550. MR107925
- [7] Green, E. J. and Strawderman, W. E. (1991). A James–Stein type estimator for combining unbiased and possibly biased estimators, *Journal of the American Statistical Association* **86** 1001–1006. MR1146348
- [8] Halperin, M. (1961). Almost linearly-optimum combination of unbiased estimates. *Journal of the American Statistical Association* **56** 36–43. MR141181
- [9] Krishnamoorthy, K. and Rohatgi, V. K. (1990). Unbiased estimation of the common mean of a multivariate normal distribution. *Communications in Statistics – Theory and Methods* **19** (5) 1803–1810. MR1075503
- [10] Kshirsagar, A. (1978). *Multivariate Analysis*, Marcel Dekker, Inc., New York. MR343478
- [11] Kubokawa, T. (1987). Admissible minimax estimation of a common mean of two normal populations. *Annals of Statistics*, **15** 1245–1256. MR902256
- [12] Loh, W.-L. (1991). Estimating the common mean of two multivariate normal distributions, *Annals of Statistics*, **19** 297–313. MR1091852
- [13] 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. MR557567
- [14] Norwood, T. E. and Hinkelmann, K. Jr. (1977). Estimating the common mean of several normal populations, *Annals of Statistics* **5** 1047–1050. MR448679
- [15] Raj, D. (1998). *Sampling Theory*, McGraw-Hill, New York. MR230440

- [16] Seshadri, V. (1963). Constructing uniformly better estimators. *Journal of the American Statistical Association* **58** 172–175. MR145628
- [17] Sobel, M., Uppuluri, R., and Frankowski, K. (1977). *Selected tables in mathematical statistics*, Vol. 4: Dirichlet Distribution – Type 1, American Mathematical Society, Providence, Rhode Island. MR423747
- [18] Tippett, L. H. C. (1931). *The method of statistics*. Williams and Norgate, London.
- [19] 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. MR207100