Ann. Inst. Statist. Math.
Vol. 42, No. 1, 51-64 (1990)

ESTIMATION OF TWO NORMAL MEANS WHICH MAY BE COMMON*

J. H. VENTER AND S. J. STEEL

Department of Statistics and Operations Research, Potchefstroom University, 2520 Potchefstroom, South Africa

(Received September 26, 1988; revised January 30, 1989)

Abstract. Consider the problem of estimating the mean of a normal population when independent samples from this as well as a second normal population are available. Pre-test estimators which combine the two sample means if a test of the hypothesis of equal population means accepts but otherwise use only the first sample mean, are compared to limited translation estimators which are derived in the spirit of Bickel (1984, *Ann. Statist.*, **12**, 864–879) (we also cover the cases of unknown variances). Our conclusion is that if the accuracy with which the second population mean can be estimated is of the same or better order of magnitude as the accuracy with which the first can be estimated, then the limited translation estimators largely dominate the pre-test estimators in terms of mean square error loss.

Key words and phrases: Common mean, Graybill-Deal estimator, pre-test estimator, limited translation estimator, problem P.

1. Introduction

Suppose X_i is the sample mean and S_i^2 the unbiased sample variance of an independent sample of size n_i from an $N(\mu_i, \sigma_i^2)$-distributed population, $i = 1, 2$. In this paper we are interested in estimation of μ_1 and μ_2. The usual estimators are X_1 and X_2, respectively, but if it is known that μ_1 and μ_2 are nearly equal, then it is possible to improve on these estimators by taking suitable combinations of X_1 and X_2. For definiteness we treat estimation of μ_1 only, but by interchanging symbols estimation of μ_2 can be treated analogously.

If the assumption can be made that $\mu_1 = \mu_2$, then the Graybill-Deal estimator (1959)

*This research was supported by grants from the FRD of the CSIR of South Africa.

$$\hat{\mu} = (n_1 S_2^2 X_1 + n_2 S_1^2 X_2)/(n_1 S_2^2 + n_2 S_1^2) \tag{1.1}$$

can be used to estimate the common mean $\mu_1 = \mu_2$, but if this assumption is not true, then the mean squared error (MSE) of $\hat{\mu}$ considered as an estimator of μ_1 can be quite large (in fact, it is unbounded as $|\mu_1 - \mu_2| \to \infty$). Hence the experimenter who is uncertain about the assumption that $\mu_1 = \mu_2$ needs to use the Graybill-Deal estimator with caution. Ohtani (1987) proposes to do this by means of a pre-test (PT) estimator. With W^* a test statistic of the hypothesis that $\mu_1 = \mu_2$ which is rejected when W^* is large, Ohtani's PT estimator of μ_1 is

$$\hat{\mu}_1 = \hat{\mu}\boldsymbol{I}(W^* \leq c) + X_1\boldsymbol{I}(W^* > c)\,, \tag{1.2}$$

with $\boldsymbol{I}(A)$ the indicator function of the event A and c a constant. Different choices of W^* can be made. Ohtani uses a modified Wald test statistic. The choice of c relates to the size of the pre-test employed, and Ohtani tentatively proposes a size of 25% but says that "... it is a remaining problem to seek an optimal size of the pre-test".

Ohtani (1987) gives further references to relevant papers, but does not refer to the significant work of Bickel (1984). For the case σ_1^2 and σ_2^2 known, Bickel's work can be used to derive a nearly optimal (in a sense to be made clear below) estimator for μ_1 and by extension of the arguments the cases σ_1^2, σ_2^2 equal but unknown and σ_1^2, σ_2^2 arbitrary unknown, can also be treated.

In Section 2 the case is considered where σ_1^2 and σ_2^2 are known, and the problem of estimating μ_1 is placed within the framework of Bickel's theory. As a result a limited translation (LT) estimator of μ_1 can be proposed that has low MSE when $\mu_1 = \mu_2$, while its MSE never exceeds the minimax value by more than a prescribed factor. It also transpires that the proposed estimator achieves its maximum MSE when $|\mu_1 - \mu_2| \to \infty$, i.e., when the prior belief that the two means are equal is completely mistaken. Comparison of MSE's shows that the LT estimator is superior to the PT estimator in the situations outlined below.

Section 3 contains a discussion of the case where $\sigma_1^2 = \sigma_2^2$, but are otherwise unknown. It is found that the LT estimator of Section 2, modified to incorporate the usual pooled unbiased estimator of variance, is still a nearly optimal solution to the problem of estimating μ_1. In Section 4 the general case where σ_1^2, σ_2^2 are arbitrary unknown variances is considered. Once more the LT estimator performs better in terms of MSE than the PT estimators.

2. Known variances

To begin with, assume σ_1^2 and σ_2^2 are known. Since X_1 and X_2 are sufficient for μ_1 and μ_2, we base an estimator for μ_1 on X_1 and X_2 only. Translation equivariance considerations suggest that attention may be restricted to estimators of the form $X_1 + h(X_2 - X_1)$ for a suitable function h, and our concern is to choose h in an optimal way. The minimax estimator of μ_1 is X_1 and its MSE is σ_1^2/n_1. Following Bickel (1984), we define an optimal choice of h as follows: take $\varepsilon > 0$ and restrict the choice of h to functions such that the MSE of $X_1 + h(X_2 - X_1)$ is at most $(1+\varepsilon)\sigma_1^2/n_1$, and among these find the one having the smallest MSE when $\mu_1 = \mu_2$. More formally put

$$R(h, \Delta) = E[X_1 + h(X_2 - X_1) - \mu_1]^2/(\sigma_1^2/n_1)\,, \tag{2.1}$$

and call this the relative MSE (RMSE) of $X_1 + h(X_2 - X_1)$. As will be seen below it depends on μ_1 and μ_2 only through

$$\Delta = (\mu_2 - \mu_1)/\upsilon \quad \text{with} \quad \upsilon^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2\,. \tag{2.2}$$

Then the optimal h (for given $\varepsilon > 0$) minimizes $R(h, 0)$ among all h satisfying $R(h, \Delta) \le 1 + \varepsilon$ for all Δ.

This problem can be reduced to Bickel's problem P as follows. Put

$$Y = (X_2 - X_1)/\upsilon \quad \text{and} \quad \tau = (\sigma_2^2/n_2)/(\sigma_1^2/n_1)\,. \tag{2.3}$$

Then the conditional distribution of $X_1 - \mu_1$ given Y is

$$N(-(\sigma_1^2/n_1\upsilon)(Y - \Delta), (\sigma_1^2/n_1)\tau/(1+\tau))\,. \tag{2.4}$$

Taking conditional expectation given Y in (2.1), we get

$$\begin{aligned} R(h, \Delta) &= \tau/(1+\tau) + E[(\sigma_1^2/n_1\upsilon)(Y - \Delta) - h(\upsilon Y)]^2/(\sigma_1^2/n_1) \\ &= \{\tau + E[Y - \Delta - \tilde{h}(Y)]^2\}/(1+\tau)\,, \end{aligned} \tag{2.5}$$

where

$$\tilde{h}(y) = (n_1\upsilon/\sigma_1^2)h(\upsilon y)\,. \tag{2.6}$$

Now the restriction $R(h, \Delta) \le 1 + \varepsilon$ is seen to be equivalent to

$$E[Y - \Delta - \tilde{h}(Y)]^2 \le 1 + q^2 \quad \text{with} \quad q^2 = \varepsilon(1+\tau)\,, \tag{2.7}$$

and the problem of finding an optimal h is reduced to finding $\tilde{h}$ which minimizes the left-hand side of (2.7) when $\Delta = 0$, subject to (2.7) holding for all Δ, and transforming this $\tilde{h}$ by (2.6) to h. This is just problem P of Bickel (1984). He points out that the exactly optimal solution is complicated, but argues from various points of view that a nearly optimal solution is provided by the function

$$\tilde{h}_q(y) = y\boldsymbol{I}(|y| \le q) + q\boldsymbol{I}(y > q) - q\boldsymbol{I}(y < -q) \tag{2.8}$$

introduced by Efron and Morris (1971). The corresponding nearly optimal estimator of μ_1 is

$$X_1 + h_q(X_2 - X_1) = X_1 + (\sigma_1^2/n_1 \upsilon)\tilde{h}_q((X_2 - X_1)/\upsilon)\,, \tag{2.9}$$

and its RMSE is given by (2.5) as

$$R(h_q, \Delta) = \{\tau + E[Y - \Delta - \tilde{h}_q(Y)]^2\}/(1 + \tau)\,. \tag{2.10}$$

The estimator (2.9) entails a limited translation away from the maximum likelihood estimator, X_1, and hence it will be called the limited translation estimator of μ_1.

Introducing the function

$$g(s, a) = \int_a^\infty (z - s)^2 \phi(z) dz = (a - 2s)\phi(a) + (1 + s^2)\Phi(-a)\,, \tag{2.11}$$

with ϕ and Φ the $N(0, 1)$-density and distribution functions, respectively, we can express the expectation in (2.10) as

$$\begin{aligned} E[Y - \Delta - \tilde{h}_q(Y)]^2 \\ = \Delta^2[\Phi(q - \Delta) - \Phi(-q - \Delta)] + g(q, q - \Delta) + g(q, q + \Delta)\,. \end{aligned} \tag{2.12}$$

It is easily shown that this is symmetric in Δ, has a minimum at $\Delta = 0$ and increases monotonically to a maximum of $1 + q^2$ as $|\Delta| \to \infty$. Hence $R(h_q, \Delta)$ behaves similarly, its maximum being $(\tau + 1 + q^2)/(1 + \tau) = 1 + \varepsilon$.

The obvious PT estimator in this case is

$$\begin{aligned} \frac{n_1\sigma_2^2 X_1 + n_2\sigma_1^2 X_2}{n_1\sigma_2^2 + n_2\sigma_1^2}\boldsymbol{I}(|(X_2 - X_1)/\upsilon| \le c) + X_1\boldsymbol{I}(|(X_2 - X_1)/\upsilon| > c) \\ = X_1 + (\sigma_1^2/n_1\upsilon)\tilde{h}_c((X_2 - X_1)/\upsilon) = X_1 + h_c(X_2 - X_1)\,, \end{aligned} \tag{2.13}$$

with c the critical value and

$$\tilde{h}_c(y) = y\mathbf{I}(|y| \le c)\ . \tag{2.14}$$

By noting the similarity between (2.9) and (2.13), it follows that the RMSE of (2.13) is

$$R(h_c, \Delta) = \{\tau + E[Y - \Delta - \tilde{h}_c(Y)]^2\}/(1+\tau)\ , \tag{2.15}$$

where

$$\begin{aligned} E[Y-\Delta-\tilde{h}_c(Y)]^2 \\ = \Delta^2[\Phi(c-\Delta) - \Phi(-c-\Delta)] + g(0, c-\Delta) + g(0, c+\Delta)\ . \end{aligned} \tag{2.16}$$

For any $c \ge 0$ this function achieves a maximum at a finite Δ, and we can select $c = c(q)$ such that this maximum is $1 + q^2$, with q as in (2.7). Then, (2.13) also has a maximum RMSE of $1 + \varepsilon$ and is comparable to (2.9) in this respect.

In order to comment on the interpretation of the parameter τ, note that since σ_1^2/n_1 and σ_2^2/n_2 may be thought of as measures of the accuracy with which X_1 and X_2 estimate μ_1 and μ_2, respectively, it follows by (2.3) that a small value of τ means that X_2 estimates μ_2 more accurately than X_1 estimates μ_1. Hence, if μ_1 and μ_2 are close together and τ is small, then there is a good opportunity to reduce risk substantially by incorporating X_2 effectively in the estimator. If, however, the value of τ is large, X_2 estimates μ_2 poorly compared to the accuracy with which X_1 estimates μ_1. In such a case it is impossible to improve appreciably on X_1 as an estimator of μ_1 by using the information supplied by X_2.

In Fig. 1(a) the RMSE of the PT estimator (2.13) and that of the comparable LT estimator (2.9) are shown as functions of Δ for the case $\tau = 0.2$ and $\varepsilon = 0.25$ so that the RMSE of both estimators are restricted below 1.25. It is evident that LT is superior to PT for small values of Δ which corresponds to μ_1 and μ_2 being close together. For large Δ the situation is reversed. Since these estimators are particularly appropriate when the experimenter believes that μ_1 and μ_2 are close together, it seems desirable that an estimator should minimize risk in this part of the parameter space at the cost of higher risk in the less likely part where μ_1 and μ_2 are far apart. It is in this sense that LT is preferable to PT. Similar conclusions are evident in Figs. 1(b) and 1(c) which show the cases $\tau = 1.0$ and $\tau = 5.0$, respectively, although the superiority of LT becomes less pronounced as τ increases. Figure 1(d) represents an extreme case, viz. $\tau = 50.0$. Clearly, LT and PT are now equivalent for small values of Δ, while PT is still superior to LT for large Δ. Hence one would prefer PT to LT under such circumstances. Also shown in Fig. 1(d) is the constant RMSE of the minimax estimator, X_1. It is evident that X_1, which nearly

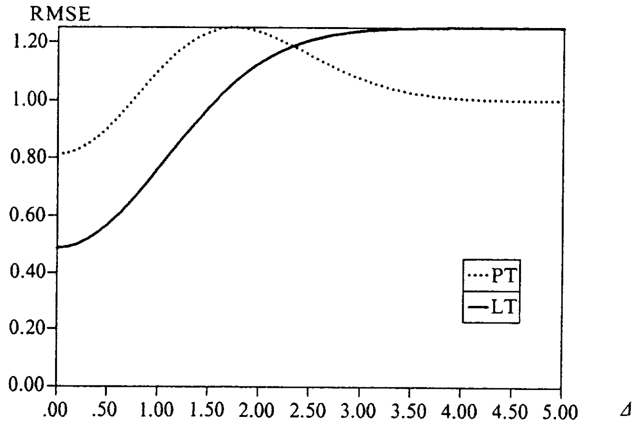

Fig. 1(a). RMSE with max 1.25 and $\tau = 0.2$.

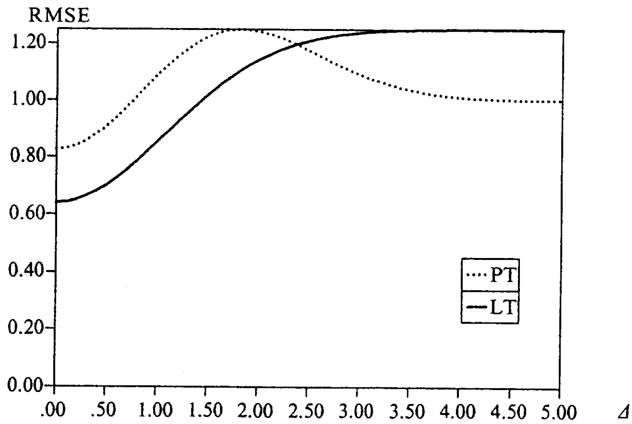

Fig. 1(b). RMSE with max 1.25 and $\tau = 1.0$.

dominates both PT and LT, is preferable to any of these alternative estimators. Very little can apparently be gained in terms of reduced risk at small values of Δ when the value of τ becomes large.

Comparing Figs. 1(a), 1(b), 1(c) and 1(d), we see that the RMSE of PT changes comparatively little with varying τ, whereas the RMSE of LT decreases substantially when τ decreases. Since a small value of τ represents a good opportunity to improve upon X_1 in terms of reduced risk at small Δ, it is clear that LT is much more successful than PT in effectively utilizing the information supplied by X_2. Figures 2(a), 2(b) and 2(c) illustrate these remarks further. These figures show the RMSE's of LT and PT at $\Delta = 0$ (i.e., $\mu_1 = \mu_2$) as functions of τ for $\varepsilon = 0.1$, 0.25 and 0.5. Also shown is the function $\tau/(1+\tau)$ which is the RMSE at $\Delta = 0$ of $(n_1\sigma_2^2 X_1 + n_2\sigma_1^2 X_2)/(n_1\sigma_2^2 + n_2\sigma_1^2)$. This is the optimal estimator for this case and both

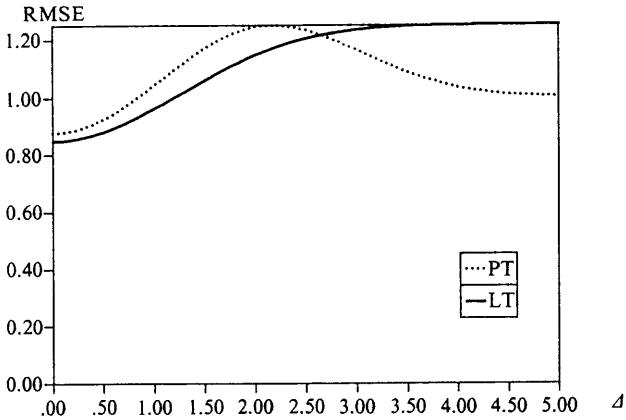

Fig. 1(c). RMSE with max 1.25 and $\tau = 5.0$.

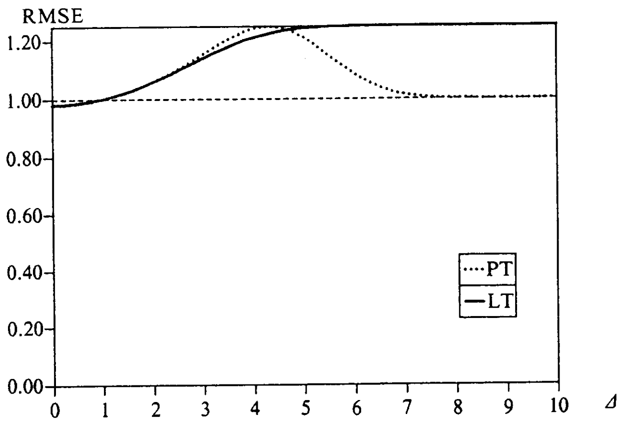

Fig. 1(d). RMSE with max 1.25 and $\tau = 50$.

LT and PT reduces to this when $\varepsilon = \infty$. For small ε the RMSE of LT clearly approximates this limiting RMSE much closer than does the RMSE of PT.

3. Equal unknown variances

Now we turn to the case where σ_1^2 and σ_2^2 are assumed to have the common but unknown value σ^2. Put

$$(3.1) \qquad S^2 = [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/m, \qquad m = n_1 + n_2 - 2 .$$

Since X_1, X_2 and S are sufficient for μ_1, μ_2 and σ, we base our estimator on them. Again translation and scale equivariance considerations suggest that

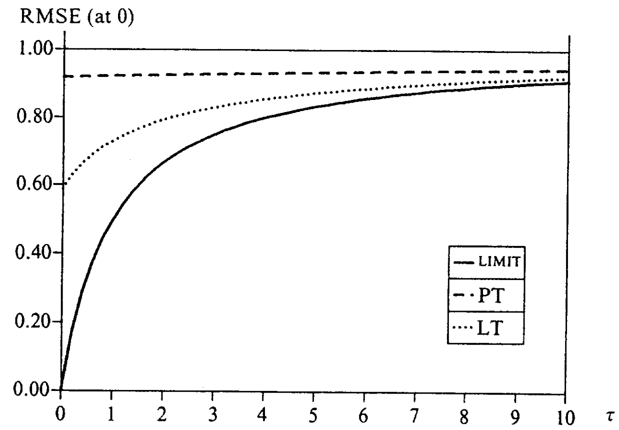

Fig. 2(a). RMSE at 0 with max 1.1.

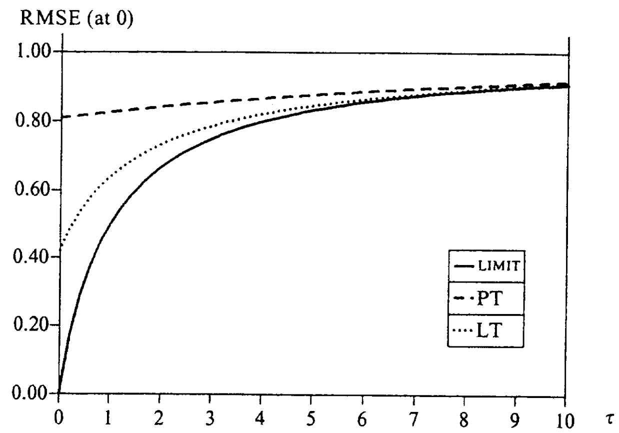

Fig. 2(b). RMSE at 0 with max 1.25.

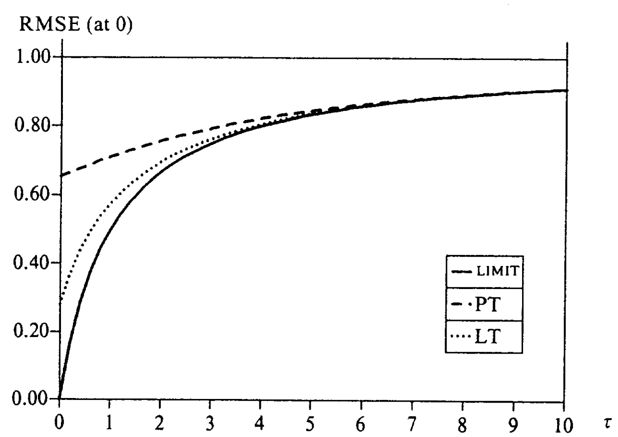

Fig. 2(c). RMSE at 0 with max 1.5.

attention be restricted to estimators of the form

$$X_1 + Sh((X_2 - X_1)/S) \tag{3.2}$$

for suitable h. In this case

$$\tau = n_1/n_2, \quad \upsilon = \sigma u \quad \text{with} \quad u = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,, \tag{3.3}$$

and arguing as in the previous case the RMSE of (3.2) can be expressed as

$$R(h, \Delta) = \{\tau + E[Y - \Delta - T\tilde{h}(Y/T)]^2\}/(1 + \tau)\,, \tag{3.4}$$

where

$$T = S/\sigma \quad \text{and} \quad \tilde{h}(y) = n_1 u h(uy)\,. \tag{3.5}$$

Now note that Y is $N(\Delta, 1)$-distributed and mT^2 is χ_m^2-distributed, independent of Y. The problem of finding an optimal h therefore reduces to the Studentized-version P′ of problem P: among $\tilde{h}$ such that

$$E[Y - \Delta - T\tilde{h}(Y/T)]^2 \le 1 + q^2 \quad \text{for all } \Delta, \quad \text{with} \quad q^2 = \varepsilon(1 + \tau)\,, \tag{3.6}$$

find that choice which minimizes the left-hand side of (3.6) when $\Delta = 0$. Most likely problem P′ is even more difficult than P and an exactly optimal solution is not known. A good solution can be found as follows. From (2.8), we get $T\tilde{h}_q(Y/T) = \tilde{h}_{Tq}(Y)$ so that

$$\begin{aligned} E[Y - \Delta - T\tilde{h}_q(Y/T)]^2 &= E\{E[Y - \Delta - \tilde{h}_{Tq}(Y)]^2 \mid T\} \\ &\le E\{1 + (Tq)^2\} \\ &= 1 + q^2\,. \end{aligned}$$

Hence $\tilde{h}_q$ still satisfies (3.6) and the corresponding "limited translation" estimator of μ_1 is

$$X_1 + (S/n_1 u)\tilde{h}_q((X_2 - X_1)/Su) \tag{3.7}$$

with RMSE

$$\{\tau + E[Y - \Delta - T\tilde{h}_q(Y/T)]^2\}/(1 + \tau)\,. \tag{3.8}$$

It is possible to express (3.8) in terms of the function g of (2.11). By conditioning on T it is found that

$$E[Y-\Delta-T\tilde{h}_q(Y/T)]^2 = E\{\Delta^2[\Phi(Tq-\Delta)-\Phi(-Tq-\Delta)] + g(Tq,Tq-\Delta)+g(Tq,Tq+\Delta)\}\,. \tag{3.9}$$

As in the case where σ_1^2 and σ_2^2 are known, it is possible to show that (3.8) is symmetric in Δ, has a minimum at $\Delta = 0$ and increases monotonically to a maximum of $1+\varepsilon$ as $|\Delta| \to \infty$. We computed the RMSE in (3.8) numerically using subroutine DQDAGI of IMSL. Figure 3 compares this RMSE (with $m = 5$, $\tau = 1.0$ and $\varepsilon = 0.25$) with the corresponding RMSE in (2.10) for the case σ_1^2 and σ_2^2 known. It is clear that even for such a small value of m these two RMSE's differ very little. We found the same conclusion for other values of τ and ε. The cost in terms of increased RMSE of not knowing σ_1^2 and σ_2^2 appears to be quite small, and we conclude that (3.7) is still nearly optimal if m is not too small.

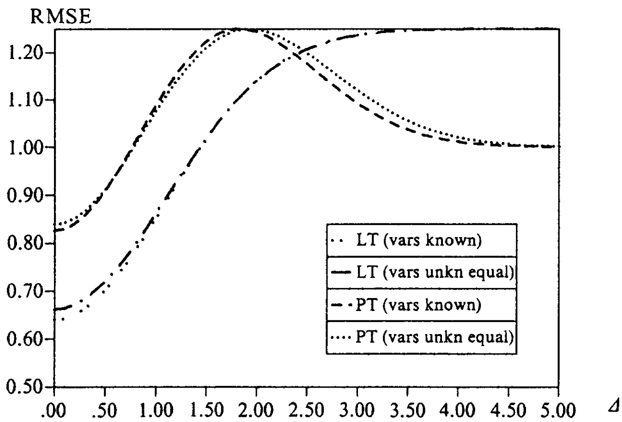

Fig. 3. Comparison of RMSE for cases unknown equal and known variances ($m = 5$, $\tau = 1.0$, max RMSE = 1.25).

The natural PT estimator for this case is given by

$$\frac{n_1X_1+n_2X_2}{n_1+n_2}\boldsymbol{I}\left(\left|\frac{X_2-X_1}{Su}\right|\le c\right)+X_1\boldsymbol{I}\left(\left|\frac{X_2-X_1}{Su}\right|>c\right) = X_1+\frac{S}{n_1u}\tilde{h}_c\left(\frac{X_2-X_1}{Su}\right), \tag{3.10}$$

where c once again denotes a critical value and $\tilde{h}_c$ is given by (2.14). As before the RMSE of (3.10) is given by

(3.11) $$\{\tau + E[Y - \Delta - T\tilde{h}_c(Y/T)]^2\}/(1 + \tau)\,,$$

where

(3.12) $$\begin{aligned} &E[Y - \Delta - T\tilde{h}_c(Y/T)]^2 \\ &\quad = E\{\Delta^2[\Phi(Tc - \Delta) - \Phi(-Tc - \Delta)] \\ &\qquad + g(0, Tc - \Delta) + g(0, Tc + \Delta)\}\,. \end{aligned}$$

Again we can select $c = c(q, m)$ such that the maximum value of (3.12) is $1 + q^2$. It turns out that m has very little influence on $c(q, m)$. For example, with $\tau = 1.0$ and $\varepsilon = 0.25$ we have $q = 0.7071$ and $c(q, 5) = 1.265$ while $c(q, \infty) = 1.277$. In Fig. 3 we also compare the RMSE's of the PT estimators for these cases and it is again clear that the cost of not knowing σ_1^2 and σ_2^2 is small when judged from a risk point of view. Consequently, the conclusions of Section 2 regarding comparison of the PT and LT estimators remain valid.

4. Arbitrary unknown variances

If σ_1^2 and σ_2^2 are arbitrary and unknown, equivariance considerations do not suggest a specific form for an estimator of μ_1. We therefore define the LT estimator by replacing σ_1^2 and σ_2^2 by their respective unbiased estimators, S_1^2 and S_2^2, in (2.9), i.e.,

(4.1) $$X_1 + (S_1^2/n_1 V)\tilde{h}_Q((X_2 - X_1)/V)\,,$$

where

$$V^2 = S_1^2/n_1 + S_2^2/n_2 \qquad \text{and} \qquad Q^2 = \varepsilon(1 + (S_2^2/n_2)/(S_1^2/n_1))$$

are now random variables, while $\tilde{h}_Q$ is still defined by (2.8) with Q replacing q. The RMSE of (4.1) is given by

(4.2) $$\{\tau + E[Y - \Delta - (T_1^2/U^2)\tilde{h}_{QU}(Y)]^2\}/(1 + \tau)\,,$$

where

$$T_i^2 = S_i^2/\sigma_i^2, \quad i = 1, 2 \qquad \text{and} \qquad U^2 = V^2/v^2 = (T_1^2 + \tau T_2^2)/(1 + \tau)\,.$$

Then introducing the function

(4.3) $$\tilde{g}(r, s, a, b) = \int_a^b (rz - s)^2 \phi(z)\,dz$$

$$= r\phi(b)(2s - br) - r\phi(a)(2s - ar) + (r^2 + s^2)[\Phi(b) - \Phi(a)] ,$$

the expectation in (4.2) can be written as

$$(4.4) \qquad E[\tilde{g}(1 - T_1^2/U^2, \quad \Delta T_1^2/U^2, \quad -\Delta - QU, \quad -\Delta + QU) + g(QT_1^2/U, \quad QU - \Delta) + g(QT_1^2/U, \quad QU + \Delta)] .$$

Since U and also $Q = \sqrt{\varepsilon(1 + \tau T_2^2/T_1^2)}$ are functions of T_1 and T_2, (4.4) can be evaluated numerically as a bivariate integral with respect to the densities of T_1 and T_2 (e.g., using subroutine DQAND of IMSL). It is straightforward to show that (4.4) is symmetric in Δ. In addition, when $\Delta \to \infty$, the first and third terms in (4.4) approach zero, while the second term becomes $1 + E(QT_1^2/U)^2 = 1 + q^2$ after simplification. Hence the RMSE (4.2) approaches $1 + \varepsilon$ as $\Delta \to \infty$. In addition, it is easily verified that

$$E[Y - \Delta - t\tilde{h}_q(Y)]^2 \leq 1 + t^2q^2$$

for all $t, q \geq 0$ and with Y still $N(\Delta, 1)$-distributed. Utilizing this result, it follows from (4.2) by conditioning on T_1 and T_2 that the RMSE of (4.1) never exceeds $1 + \varepsilon$, i.e., the estimator (4.1) still satisfies the required RMSE restriction.

In Fig. 4 we compare the RMSE's of the two LT estimators (4.1) and (2.9) for the case $n_1 = n_2 = 5$, $\tau = 1.0$ and $\varepsilon = 0.25$. The two RMSE's are again quite close, even for such small sample sizes. Similar results were found for other choices of sample sizes and parameters. This again suggests

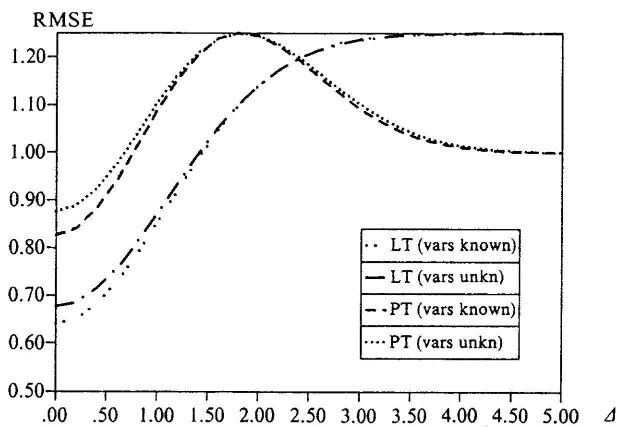

Fig. 4. Comparison of RMSE for cases of unknown and known variances ($n_1 = 5$, $n_2 = 5$, $\tau = 1.0$, max RMSE = 1.25).

that the cost of not knowing the variances is relatively small, and we think that the LT estimator (4.1) is still fairly close to being optimal.

A PT estimator for this situation may be based on the Wald test statistic $W = (X_2 - X_1)/V$ which leads to the estimator

$$X_1 + (S_1^2/n_1 V)\tilde{h}_c((X_2 - X_1)/V) . \tag{4.5}$$

The RMSE of this estimator is given by the expression in (4.2) with the expectation replaced by

$$E[\tilde{g}(1 - T_1^2/U^2, \quad \Delta T_1^2/U^2, \quad -\Delta - cU, \quad -\Delta + cU) + g(0, \quad cU - \Delta) + g(0, \quad cU + \Delta)] . \tag{4.6}$$

As in the previous cases we can now try to choose c such that the maximal RMSE of (4.5) is $1 + \varepsilon$. Unfortunately, it turns out that this choice now depends on τ which is unknown in the present case. If e.g., $n_1 = n_2 = 5$, $\varepsilon = 0.25$, then we get $c = 1.228$ for $\tau = 1.0$ while $c = 1.022$ for $\tau = 0.2$. If we should use $c = 1.228$ as a possible compromise value, calculations show that the maximal RMSE varies between 1.468 at $\tau = 0$ and 1.000 as $\tau \to \infty$. In contrast with the LT estimator, it thus becomes clear that c cannot be chosen to control the maximal RMSE of the PT estimator exactly at a prescribed level in the case of arbitrary unknown variances.

In Fig. 4 we also compare the RMSE of (4.5) for the case $n_1 = n_2 = 5$, $\varepsilon = 0.25$ assuming $\tau = 1.0$ so that $c = 1.228$ with the RMSE of the corresponding known variances PT estimator (2.13) having $c = 1.277$. It is clear that if the difficulty of unknown τ could be overcome, the resulting PT estimator could be expected to perform much like the corresponding known variances PT estimator, and the previous conclusions regarding the relative merits of the PT and LT estimators would remain valid.

Ohtani (1987) bases his PT estimator on the modified Wald statistic $W^* = W^2 A^2$ where $A^2 = 1 - 2\{(S_1^2/n_1)^2/(n_1 - 1) + (S_2^2/n_2)^2/(n_2 - 1)\}/V^4$. For $n_1, n_2 \geq 3$, we have $A^2 \geq 0$ and then we can write the corresponding PT estimator as

$$X_1 + (S_1^2/n_1 V)\tilde{h}_{c/A}((X_2 - X_1)/V) , \tag{4.7}$$

and the RMSE of this PT estimator can be obtained as before. It turns out that there is very little difference between the RMSE's of (4.5) and (4.7) if their c's are chosen to give them the same maximum $1 + \varepsilon$. Moreover, the same difficulty appears again: the choice of c depends on the unknown τ. Thus the conclusions of our comparisons between the RMSE's of the LT and PT estimators remain true for this version of the PT estimator. The motivation for the above modification W^* of W is to obtain a test whose size is closer to the nominal value. Evidently, this has little bearing on the estimation problem when estimators are judged according to their RMSE's.

In summary, the LT estimators are easy to apply in all three cases considered and they perform substantially better than the PT estimators in situations where it is worthwhile to use an estimator which combines the data from the two samples.

Acknowledgement

The authors would like to thank the referee for suggestions that improved this paper.

REFERENCES

Bickel, P. J. (1984). Parametric robustness: small biases can be worthwhile, *Ann. Statist.*, **12**, 864–879.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators—Part I: the Bayes case, *J. Amer. Statist. Assoc.*, **66**, 807–815.

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

Ohtani, K. (1987). On the use of the Graybill-Deal estimator when two normal population means may not be common, *Comm. Statist. B—Simulation Comput.*, **16**, 855–870.