Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(1) 2012 31-41

Allocating a Limited Budget to Small Areas

Nicholas T. Longford
SNTL and UPF, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain
Received 19 April 2011; Revised 18 August 2011; Accepted 18 August 2011

SUMMARY

We consider the general problem of allocating funds from a fixed budget to the districts of a country according to the values of a district-level indicator. The obvious approach might seem to be to efficiently estimate the relevant district-level quantities, and then to apply the allocation scheme that would be optimal if these quantities were equal to the estimates. We show that such a two-stage strategy is suboptimal. By a simulation approach, we find allocation schemes that are superior to this strategy. We offer no single universal solution, but motivate the results by intuition. We discuss the implications of our finding on the separation of the remits of a statistical agency and its client.

Keywords : Composite estimation, Optimal allocation, Shrinkage, Small-area estimation.

1. INTRODUCTION

For a modern information-based government, small-area statistics plays an important role in the monitoring of various economic, social and demographic indicators of the local administrative units (districts) of the country. It is essential for devising effective measures and incentives to reduce iniquities and other undesirable differences among the country's regions or districts. Production of estimates for the districts, which entails the conduct of national surveys, construction of databases and statistical inference, is usually the task of a statistical agency, such as the Office for National Statistics in the UK, and taking action based on the estimates, after synthesis with other information, is the responsibility of a government department or agency. This paper is concerned with the separation of these two activities when estimation is associated with appreciable uncertainty. Its conclusion, based on a simulation study of allocating a limited budget to the country's districts, is that the activities of the statistical agency (estimation) and of the government (allocation of funds) have to be integrated. The method of allocation that would be optimal if the relevant district-level (population) quantities were established with precision is suboptimal when efficient (small-area) estimates are used in their stead. We find and motivate some inefficient estimators that result in a better allocation of funds.

We consider the following problem. A (national) government department wishes to allocate a fixed budget B to its D districts according to district-level quantities θ_d, $d = 1, ..., D$. The sizes of the relevant populations of the districts are N_d. For instance, θ_d may be the unemployment rate and N_d the size of the labour force in district d. The department sets a threshold T and establishes a unit value U, and would in ideal circumstances award to district d a grant of size $G_d = (\theta_d - T)N_dU$ if θ_d exceeds T, and no grant otherwise; $G_d = 0$ when $\theta_d < T$. In the context of the unemployment problem, the amount G_d can be interpreted as awarding amount U for every unemployed above the 'tolerated' level of unemployment T. Two problems are encountered. The quantities θ_d are not known and have to be estimated, and the fixed budget B may be insufficient, $B < G_1 + \cdots + G_D$.

E-mail address : NTL@sntl.co.uk

The former problem is addressed by estimating the quantities θ_d, from a survey or by combining information from several sources. The latter problem can be addressed by raising the threshold T, reducing the amount U, or by setting a *shortfall* S_d for each district and awarding only $A_d = G_d - S_d$. The award A_d is always nonnegative, so $0 \le S_d \le G_d$. In the setting we consider, a government statistical agency supplies a set of (nearly) efficient estimates $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_D)$ to the department, which applies an algorithm for allocating the budget, regarding the estimates $\hat{\theta}$ as if they were the population quantities $\theta = (\theta_1, \ldots, \theta_D)$.

We demonstrate that such a two-stage strategy is suboptimal, even when both stages, estimation of θ and the algorithm for allocating B based on θ are optimal. We compare allocations by their total squared shortfall, $S = S_1^2 + \ldots + S_D^2$, and prefer allocations with smaller S. The shortfall is a natural quantity for assessing the failure of the allocation scheme in a district, and adding up the squares amounts to penalising a large shortfall more harshly then several smaller shortfalls in total. Some alternatives to S are outlined in Section 2.1. When S_d are random quantities, based on estimators $\hat{\theta}_d$, we consider the expected squared shortfall E(S) instead.

In the next section, we describe solutions for the two stages and show, by example, that their combination yields a disappointing result — a less efficient estimator of θ leads to a superior allocation scheme. In Section 3, we find allocation schemes that are better than the default by using thresholds different from T. Section 4 addresses design of a survey on which allocation would be based. The discussion in the concluding Section 5 summarises the findings and connects them to the government statistician's remit. We relate the conclusions to those of Shen and Louis (1998) who demonstrated, in the context of small-area estimation, that the property of efficiency is not maintained by nonlinear transformations; for example, the ranks of efficient estimators of $\theta_1, \ldots, \theta_D$ are not efficient estimators of the ranks of these quantities. Different estimators should be used for ranking, for estimating (or graphically representing) the distribution of the district-level quantities and for making inferences about the extremes and other summaries of θ_d. See Longford (2005a and b, Part II) for related discussion.

2. SMALL-AREA ESTIMATION AND ALLOCATION PROBLEMS

The problem of estimating $\theta = (\theta_1, \ldots, \theta_D)$ for a partition (division) of a country, or of a similar domain, is commonly referred to as small-area estimation (Platek *et al*. 1987; Ghosh and Rao 1994; Rao 2003; and Longford 1999 and 2005b, Part II). Its particular challenge arises when the subsamples for one or several districts (subdomains) are too small and direct estimation, based solely on the values of the relevant variable on the observed units in the district concerned, is unsatisfactory. Following Fay and Herriot (1979) and the development of empirical Bayes and multilevel methods (Robbins 1955; Efron and Morris 1972; Longford 1993; and Goldstein 2002), a concensus has been formed that the direct estimator of θ_d can be improved by borrowing strength across the districts, that is, by exploiting their similarity. This is in accord with the theory of James and Stein (1961), although they do not construct a solution, and the established small-area estimators do not satisfy their theoretical standard of admissibility.

Most approaches to small-area estimation can be described as combining alternative estimators, the unbiased but inefficient direct estimator and a synthetic estimator that entails a model and its estimator for an average district. When no information other than the values of the relevant variable is available, this amounts to setting

$$\tilde{\theta}_d = (1 + b_d)\hat{\theta}_d + b_d\hat{\theta}, \qquad (1)$$

where $\hat{\theta}$ is an estimator of the national population quantity θ and b_d a suitably defined coefficient. For normally distributed outcomes, the empirical Bayes approach yields

$$b_d = \frac{1}{1 + n_d\hat{\omega}}, \qquad (2)$$

where ω is the ratio of the between-and within-district variances, $\omega = \sigma_B^2/\sigma^2$, $\hat{\omega}$ its estimate and n_d the subsample size (or its equivalent) for district d.

For binary data, the rates (probabilities) θ can be estimated from the fit of a logistic regression model with random coefficients. An alternative, related to moment matching, is described and applied in Longford (1999 and 2004). The between-district variance σ_B^2 is estimated as the part of the sample variance of the direct estimators that is in excess of what would be expected if the rates θ_d were identical:

$$\hat{\sigma}_B^2 = \frac{1}{N}\left\{\sum_{d=1}^{D} N_d\left(\hat{\theta}_d - \hat{\theta}\right)^2 - \sum_{d=1}^{D}\left(1-\frac{N_d}{N}\right)\hat{\theta}_d\left(1-\hat{\theta}_d\right)\right\}. \quad (3)$$

As most of the district-level rates are estimated by $\hat{\theta}_d$ with large sampling variances, the sampling variance of $\hat{\sigma}_B^2$ is reduced when each term $\hat{\theta}_d\,(1 - \hat{\theta}_d)$ in (3) is replaced by their approximate average $\hat{\theta}\,(1 - \hat{\theta})$. This yields the estimator

$$\hat{\sigma}_B^2 = \frac{1}{N}\sum_{d=1}^{D} N_d\left(\hat{\theta}_d - \hat{\theta}\right)^2 - \frac{D-1}{N}\hat{\theta}\,(1-\hat{\theta})\,.$$

The direct estimator $\hat{\theta}_d$ is then combined with the national estimator $\hat{\theta}$, as in (1), with coefficient b_d set so as to minimise the mean squared error (MSE) of the combination. The ideal coefficient b_d depends on the squared deviation $(\theta_d - \theta)^2$, which is unknown and could be estimated only with low precision. It is therefore replaced by its district-level expectation σ_B^2, which is estimated with much greater precision when each of many districts has a substantial represention in the data source (survey). The same coefficient b_d is obtained as in (2), although the variance ratio estimator $\hat{\omega}$ may be different. The uncertainty about θ is easy to take care of, although its impact on estimating θ_d is usually trivial. Addressing the uncertainty about σ_B^2 presents a much sterner challenge. See Rao (2003) and Longford (2005b, Part II, and 2007) for alternative approaches.

2.1 Allocation of a Limited Budget

The optimal allocation of a limited budget is found by minimising the total of the squared shortfalls, $S = S_1^2 + ... + S_D^2$. Objective functions other than S can be adopted. These may incorporate exceptional arrangements for some districts, contributions (losses) according to different formulas for groups of districts (e.g., regions), a steeper loss function, such as $|S_d|^3$, or a loss function with change points, reflecting the critical nature of estimation errors that exceed a certain level. In fact, the squared shortfall S_d^2 as a (loss) function of A_d has a change point at G_d; for $A_d > G_d$, there is no loss, and for $A_d < G_d$ the loss is quadratic, equal to $(G_d - A_d)^2$.

When the ideal amounts G_d are known the minimum of S is easy to find. The solution is trivial when $B \geq G_+ = G_1 + ... + G_D$; each district is awarded the full grant G_d and the distribution of the remainder of the budget is immaterial. With an insufficient budget, when $B < G_+$, the method of Lagrange multipliers, or substitution of $S_1 = G_+ - B - S_2 - ... - S_D$ in the expression for S, yields the condition

$$\frac{1}{2}\frac{\partial S}{\partial S_d} = S_d - (G_+ - B - S_2 - ... - S_D) = 0$$

for any district $d \neq 1$ with a positive award A_d. Hence, the districts that receive awards should share the shortfall equally. If this average shortfall is greater than G_d for a district, then this district is reassigned to those receiving no award, and the shortfall is distributed equally among the rest. If such a reassignment takes place the average shortfall is increased and further districts may have to be reassigned to receive no award. A reassignment should be carried out for one district at a time and repeated as many times as necessary.

2.2 Simulated Version of the Country

In Section 3, we explore a range of estimators and adjustments of the threshold in pursuit of an optimal allocation of the budget B. They are based on a simulated set of quantities $\theta = (\theta_1, ..., \theta_D)$ held fixed in replications. The quantities θ (district level rates of unemployment) are plotted in Fig. 1 against the population sizes of the districts. The population sizes N_d are themselves simulated, from a scaled beta distribution with an admixture of normal, and the maximum of the sample is multiplied by the factor 1.75 to represent the country's capital. There are $D = 100$ districts.

The rates θ_d are generated from another scaled beta distribution and are linearly combined with the population sizes N_d so that the correlation of $\{\theta_d\}$ and $\{N_d\}$ would be negative for the less populous and positive for the more populous districts. The population sizes are held fixed throughout, and all the results are reported for this vector θ until Section 4.1, where the results for other vectors θ are summarised. The country's population (labour force) is $N = N_1 + ... + N_D$ = 25 million. The national rate $\theta = (N_1\theta_1 + ... + N_D\theta_D)/D$ is equal to 9.26%, and the mean of the district-level rates θ_d is $\bar{\theta} = (\theta_1 + ... + \theta_D)/D = 9.37\%$.

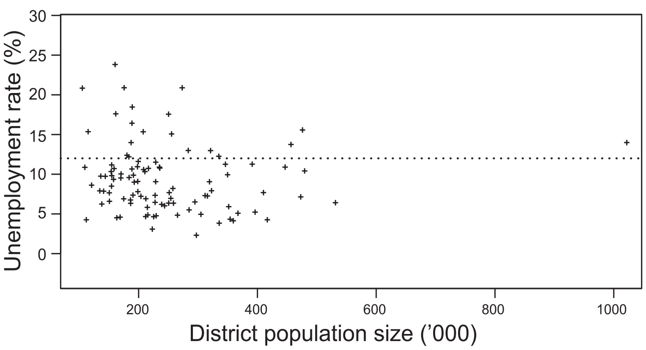

Fig. 1. The population sizes and unemployment rates of the districts. Computer generated values.

The government department has budget B = 4 million units, and each unemployed in excess of the rate of T =12%, marked in Fig. 1 by horizontal dots, is associated with the ideal grant of U = 35 units. The threshold level of T =12% is exceeded by 20 districts with total population 5.53 million (22.9%). The ideal awards G_d for these districts add up to 6.554 million units, accounting for 187260 excess unemployed in total. That is, there is a total shortfall of 2.554 million. If sufficient resources were available and θ were known, the capital would receive a grant of 0.711 million. Only one district would receive more, 0.850 million; its population is 273400 and unemployment rate is 20.9%, the third highest. In this setting, the optimal allocation within the budget is positive only for thirteen districts, with total squared shortfall of $S_* = 0.3704 \times 10^{12}$. Henceforth we regard 10^{12} as the unit for S and drop this factor when reporting other values of S or $E(S)$. The shortfall for each of these thirteen districts is 0.155 million; their total squared shortfall amounts to 0.3123, 84% of the total for the twenty districts that deserve a grant.

Suppose a survey with sample size 9000 is the sole source of information about θ. The survey has a stratified sampling design with simple random sampling within the districts, with subsample sizes n_d proportional to the subpopulation sizes N_d. The subsample sizes of the four least populous districts are 38, 39, 40 and 41; the capital is represented in the survey by 368 subjects. We emphasise that these sample sizes are far too small for reliable direct estimation. By way of illustration, suppose the district with sample size n_d = 40 has unemployment rate 10% (θ_d = 0.10). The standard error of the direct estimator is $100\sqrt{0.1 \times 0.9/40}$ = 4.74%, so sample rates of 5% (implying no problem with unemployment) and 15% (acute problem) are quite plausible. Thus, indirect estimation is indispensable in this context.

Trivial solutions include the same allocation of 40000 units to each district and allocation proportional to the population size of the district. They are associated with respective quadratic shortfalls of 2.811 and 2.673. We will derive allocations that are far superior, but these trivial allocations can nevertheless be used as points of reference, together with the ideal quadratic shortfall of 0.370. For example, we may assess the value of an allocation scheme that results in a given value of E(S) by the relative effectiveness

$$E = \{2.673 - \mathrm{E}(S)\}/2.303,$$

so that E = 1 corresponds to the ideal allocation and E = 0 to a totally ineffective use of the survey. We use the allocation proportional to size as the reference, because equal allocation would be considered as inequitable even in the absence of any information about the unemployment in the districts.

3. EXPECTED SHORTFALL WITH COMPOSITE ESTIMATION

We consider the direct and composite estimators of θ_d, $\hat{\theta}_d$ and $\tilde{\theta}_d$, d = 1, ..., D, and use them to allocate the total budget naively, treating them as if they were the underlying quantities θ_d. In the simulations we conduct, we replicate the processes of sampling and estimation of $\hat{\theta}_d$, followed by evaluation of $\tilde{\theta}_d$ for each district d, and calculate the squared shortfall S for each set of D estimates. Each simulation comprises 5000 replications. To avoid any confusion, we refer to a set of D estimators of θ_d used in an allocation scheme as a single estimator.

With the direct estimators $\hat{\theta}_d$, E(S) is equal to S_{dir} = 1.311, so that E_{dir} = 0.576. The difference $S_{\mathrm{dir}} - S_*$ = 0.941, or the S_{dir}/S_* = 3.54-fold increase, can be interpreted as the price of incomplete information. Given a survey of respectable size, it is quite steep, although a comparison on the square-root scale, as $\sqrt{S_{\mathrm{dir}}}/\sqrt{S_*}$ = 1.88, may reflect this more appropriately. It is therefore natural to look for estimators more efficient than $\hat{\theta}_d$. However, with the composite estimator we fare no better. Using it instead of $\hat{\theta}_d$ results in E(S) equal to S_{cmp} = 1.388, exceeding S_{dir} by 0.077.

Shen and Louis (1998) showed that using $b_d = 1/(1 + n_d \hat{\omega})$ in (1) amounts to too much shrinkage for representing the district-level variation of θ_d by their estimates. This problem is resolved with the coefficients $b_d^* = 1 - \sqrt{1 - b_d}$; the D estimates $\tilde{\theta}_d = (1 - b_d^*)\hat{\theta}_d + b_d^*\hat{\theta}$ are dispersed as much as the estimated between-district variance $\hat{\sigma}_B^2$. This motivates our exploration of the shrinkage estimators (1) with the coefficients, or estimates of,

$$b_d^* = 1 - (1 - b_d)^q$$

for $q > 0$. These sets of coefficients generate a continuum of estimators between the direct and composite estimators. The exponent $q > 1$ corresponds to shrinkage greater than in the established composite estimators. We summarise the efficiency of the sets of D estimators based on an exponent q by the average of the MSEs, $\mathrm{MSE}^{(q)} = \frac{1}{D}\sum_d \mathrm{MSE}\left(\tilde{\theta}_d^{(q)}; \theta_d\right)$, or its square root, estimated empirically. That is, let $\hat{\theta}_{d,i}$, $i = 1, ..., R$, be the values of an estimator $\hat{\theta}_d$ of θ_d for a given district d, obtained in the $R = 5000$ replications. Then its MSE is estimated by

$$\widehat{\mathrm{MSE}}\left(\hat{\theta}_d; \theta_d\right) = \frac{1}{R}\sum_{i=1}^{R}\left(\hat{\theta}_{d,i} - \theta_d\right)^2.$$

Given the large number of replications, the sampling variation of this MSE estimator can be ignored.

The left-hand panel of Fig. 2 summarises the values of E(S), denoted by S_q, based on these estimators as a function of q. It confirms that efficient estimation of θ_d, with $q = 1$, is detrimental to effective allocation of the budget B. Optimal allocation is attained for $q = 0.5$, but S_q is a flat function of q in the vicinity of 0.5, so any choice in the interval $q \in (0.4, 0.55)$ would be suitable. The allocation based on the direct estimators $\hat{\theta}_d$ is superior to any allocation based on shrinkage with exponent $q > 0.8$, and S_q rises steeply for such q.

For completeness, the right-hand panel displays $\sqrt{\mathrm{MSE}^{(q)}}$ as a function of q, confirming that the shrinkage estimators are more efficient (on average) than the direct estimators ($q = 0$). In fact, $\mathrm{MSE}^{(q)}$ attains its minimum for $q = 0.90$, just short of $q = 1.0$ which is optimal according to a different criterion, averaging over the (estimated) distribution of the deviations $\theta_d - \theta$.

Fig. 2 conveys a discomforting message. Efficient estimators do not yield an optimal decision (allocation). The cause of this seeming contradiction is the essentially small-sample nature of the estimators of θ_d and their nonlinear involvement in the allocation formula. An alternative class of shrinkage estimators is based on the coefficients $b_d^\dagger = f b_d$ for a positive constant f. It yields results very similar to those for b_d^*.

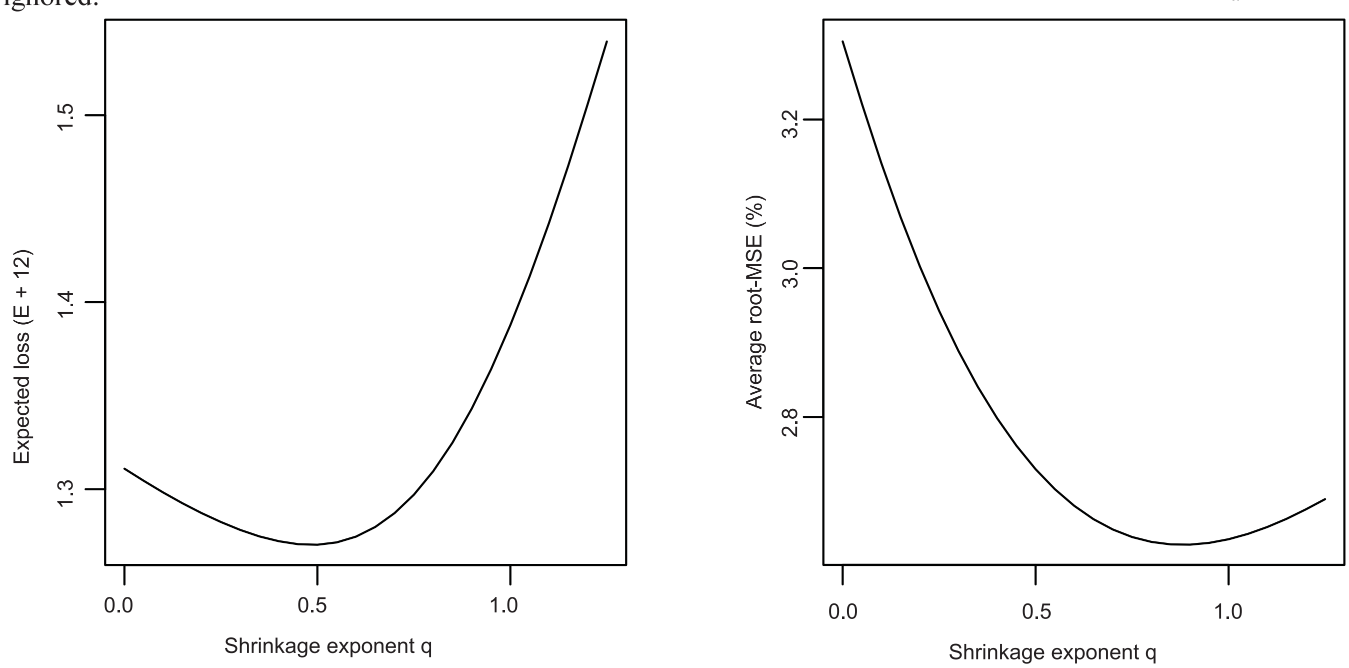

Fig. 2. The expected squared shortfall S_q and the average root-MSE, $\sqrt{\mathrm{MSE}^{(q)}}$, as functions of the shrinkage exponent q.

3.1 Altered Threshold

The naive allocation scheme explored thus far may be suboptimal because of the asymmetry in how the shortfall is assessed. Failure to allocate funds to a deserving district, for which $\hat{\theta}_d < T < \theta_d$, amounts to a relatively greater squared shortfall than allocating funds to a district that does not deserve them, for which $\hat{\theta}_d > T > \theta_d$, and similarly for $\tilde{\theta}_d$. This suggests that erring on the side of greater allocation may have less severe consequences. Of course, such an allocation has to be adjusted to satisfy the budget constraint $A_1 + ... + A_D \leq B$, so the result may be that some districts get smaller awards. A simple way of arranging a greater allocation is by using, in conjuction with $\hat{\theta}_d$ or $\tilde{\theta}_d$, a threshold T' lower than T. The assessment of such an allocation by the squared shortfall S is with respect to the original threshold T, even when the awards A_d are set with reference to a treshold $T' < T$.

The expected values and the standard deviations of the squared shortfall with the threshold set at $0.08 < T' < 0.13$ are plotted in Fig. 3 for the sets of direct and composite estimators and for the (compromise) shrinkage estimators based on the exponent $q = 0.5$. The minima of E(S) for the three (sets of) estimators are attained at around 11.3% for the composite, 11.0% for the direct, and 11.5% for the compromise shrinkage estimator based on $q = 0.5$. The minimum E(S), 1.260, is incurred for the compromise shrinkage, 0.035 lower than the minimum for the composite estimator, which is 0.014 lower than the minimum for the direct estimator. The gains over the allocation scheme based on $T = 12\%$ are substantial for the composite estimator and very small for the direct estimator.

The right-hand panel compares the (empirical) sampling variation of the total shortfalls. It highlights another problem with the allocation based on the composite estimator — its uneven performance in replications with settings $T' \geq 11\%$. In some replications, the total squared shortfall S is quite small, but in others it is substantial. The two other estimators perform much more evenly. This is an unexpected result because the values of the D estimates $\tilde{\theta}_d$ are dispersed less (around $\hat{\theta}$ or θ) than the direct estimates $\hat{\theta}_d$ for every realisation of the survey.

3.2 District-specific Thresholds

If there were no shortfall and the value of θ_d were known for a district d, then the allocation of G_d to d would incur no contribution to the squared shortfall S. This suggests that S may be reduced by setting the threshold T' more flexibly as a function of the estimated MSE. For a composite estimator, the MSE depends on $(\theta_d - \theta)^2$, so it is estimated with poor precision. It is more practical to estimate the averaged MSE, denoted by eMSE:

$$\text{eMSE}\left(\tilde{\theta}_d ; \theta_d\right) = \text{E}_{\mathcal{D}}\left\{\text{MSE}\left(\tilde{\theta}_d ; \theta_d , n_d\right)\right\},$$

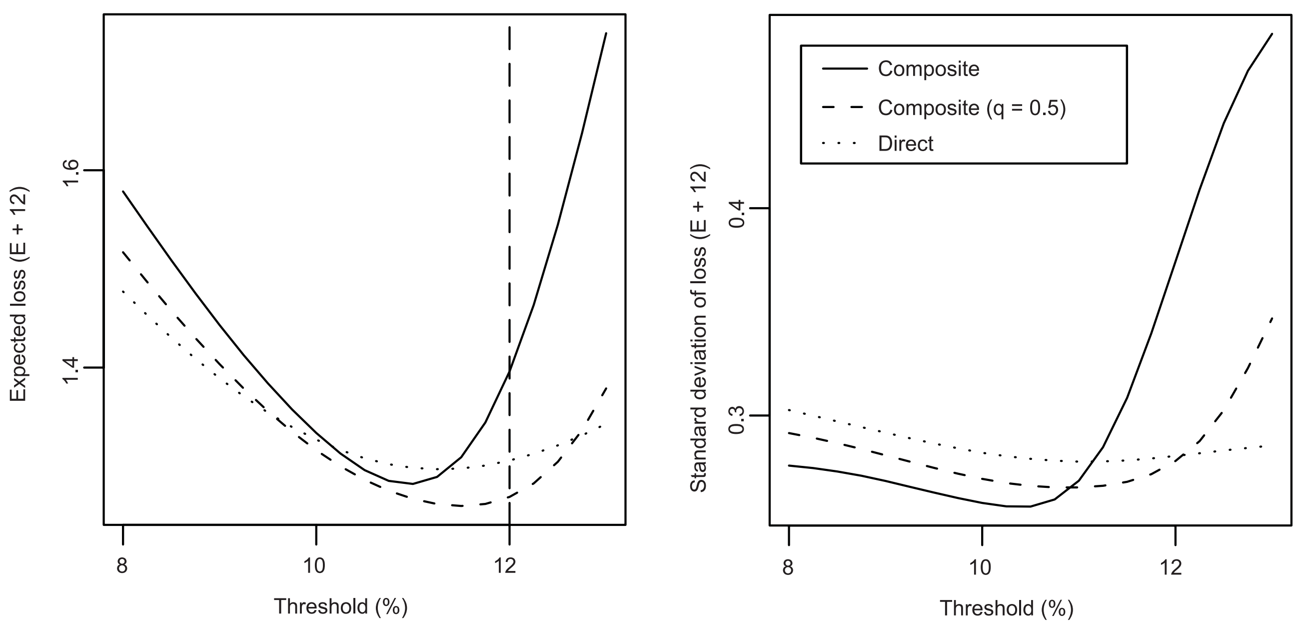

Fig. 3. The expected squared shortfall $\text{E}\{S(T')\}$ and the standard deviation $\sqrt{\text{var}\{S(T')\}}$, as functions of the altered threshold T'.

where the expectation is over the districts, but with the sample size n_d held fixed. It amounts to replacing $(\theta_d - \theta)^2$ in the expression for MSE($\tilde{\theta}_d$; θ_d) by σ_B^2, the variance of the (population) rates θ_d, $d = 1, ..., D$, in agreement with the standard empirical Bayes method. See Longford (2007) for details.

Thus, for a range of positive coefficients c, we set the thresholds $T'_d = T - c\hat{s}_d$, where $\hat{s}_d$ is the estimate of the root-MSE of the estimator ($\hat{\theta}_d$ or $\tilde{\theta}_d$) used in place of θ_d. Fig. 4 displays the results for $c \in (-0.5, 1.0)$. The compromise shrinkage estimator yields the superior allocation, with squared shortfall 1.264 for $c = 0.20$. The minimum squared shortfalls for the direct and composite estimators are 1.310 and 1.267, attained with the respective coefficients $c = 0.05$ and $c = 0.40$. For the direct estimator, E(S) depends on c very weakly. For the composite estimator, E(S) decreases steeply for $c \in (0, 0.3)$. The compromise estimator is superior to the direct throughout $c \in (0, 1)$ and is nearly constant for $c \in (0.1, 0.3)$.

The sampling variation, indicated by the standard deviations of the squared shortfall in the right-hand panel of Fig. 4, decreases substantially with c for the composite estimator and depends on it very weakly for the direct estimator. The sampling variation for the compromise shrinkage estimator is smaller than for the direct estimator for all values $c > 0$. The sets of squared shortfalls in the replications are highly correlated.

In summary, setting the thresholds flexibly can reduce E(S), but only by as much as reducing the threshold uniformly. These two ways of altering the threshold could be combined, by setting $T_d^{\dagger} = T' - c\hat{s}_d$, but we believe that such a setting would be unstable and highly contingent on the values of θ.

3.3 Altering the Focus of Shrinkage

The shrinkage estimator in (1) can be interpreted as pulling the direct estimator toward $\hat{\theta}$ as a particular *focus*. Altering this focus, and setting it to a constant of our choice, defines another continuum of estimators. This has an interpretation similar to altering the threshold, but doing so flexibly, taking into account the sampling variation of the direct estimator in a different way than in the previous section, namely, as a nonlinear function of $1 + n_d\hat{\omega}$.

The results for these estimators are summarised by Fig. 5 drawn using the same layout as Fig. 4. A surprising outcome is that the optimal focus is greater than the threshold of 12% for all three estimators. The smallest E(S), equal to 1.2526, is attained for the shrinkage exponent $q = 0.75$. For $q = 0.5$, E(S) depends on the focus of shrinkage T' very weakly, and is smaller than for $q = 1.0$ throughout the range $T' \in (0.09, 0.16)$. In contrast, E(S) for $q = 1.0$ depends on T' much more; it attains its minimum of 1.2585 at around 13%. Note that for small q the values b_d^* are small, and so the shrinkage does not have a strong impact even when the

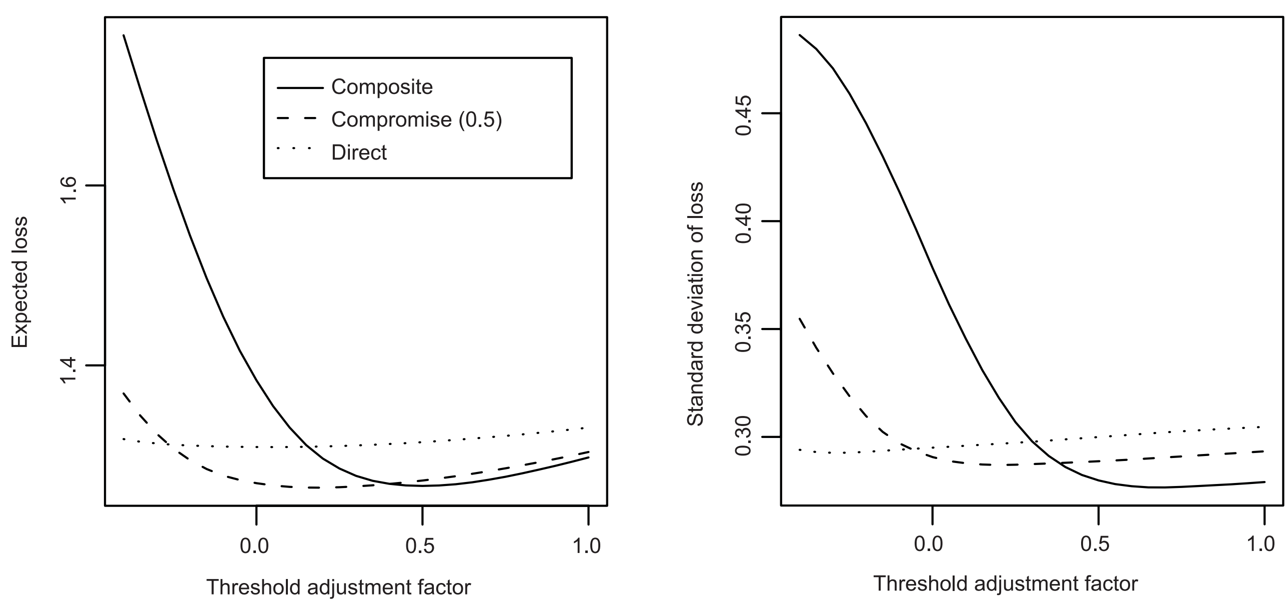

Fig. 4. The expected squared shortfall $E\{S(T')\}$ and the standard deviation $\sqrt{\mathrm{var}\{S(T')\}}$, as functions of the threshold adjustment c.

focus is far away from $\hat{\theta}$. In the limit, as $q \to 0$, the focus is irrelevant for the direct estimator.

The right-hand panel of Fig. 5 summarises the uncertainty associated with a particular realisation of the survey. It gives further support for the choice of a focus in excess of the threshold 12%, and therefore way beyond the focus of the composite estimator ($\hat{\theta}$, a random variable with expectation 9.3% and standard deviation 0.3%).

4. SAMPLING DESIGN

The analysis in the previous section is concerned principally with effective allocation of funds, as assessed by E(S). Estimation is secondary, and emphasis on its efficiency (small eMSE) would detract from this task. In this section, we assume that a single budget is available for the survey and allocation of funds, and seek the optimal sample size for the survey. A survey with a smaller sample size is cheaper and leaves more funds for allocation, but the squared shortfall is likely to be large because of estimation errors. With a larger sample size the estimation errors are reduced but fewer funds are available for allocation; shortfall is likely to be smaller, although if too much is spent on the survey, shortfall may arise due to insufficient funds for allocation. In brief, the expected (squared) shortfall is a U-shaped function of the survey costs.

If we are committed to the class of stratified sampling designs with simple random sampling within districts, with subsample sizes proportional to the district sizes, only one degree of freedom can be manipulated in the design — the overall sample size. We may search for the smallest sample size n for which a pre-set squared shortfall S_* is not exceeded, or seek the sample size for which S_* is minimised. Table 1 gives the minimum squared shortfall and the corresponding optimal focus for several shrinkage estimators and a range of sample sizes. For comparison, E(S) with the direct estimator is added in the right-most column. The sample size for a set value of S_* can reliably be approximated by linear interpolation.

The table shows that the minima for each sample size depend on the shrinkage exponent only slightly, but are attained with very different foci. It may seem counterintuitive that the optimal focus with q =0.5 increases with sample size so radically. However, for large sample sizes the shrinkage coefficients b_d^* are smaller, so the amount of shrinkage, although much greater than if the focus were set to the national rate $\hat{\theta}$, or to the threshold T, remains moderate even for the least populous districts. In fact, the dependence of E(S) on the focus is very weak for q =0.5. We emphasise that the shrinkage estimators with focus set to $\hat{\theta}$ lead to much less efficient allocation. For example, for sample

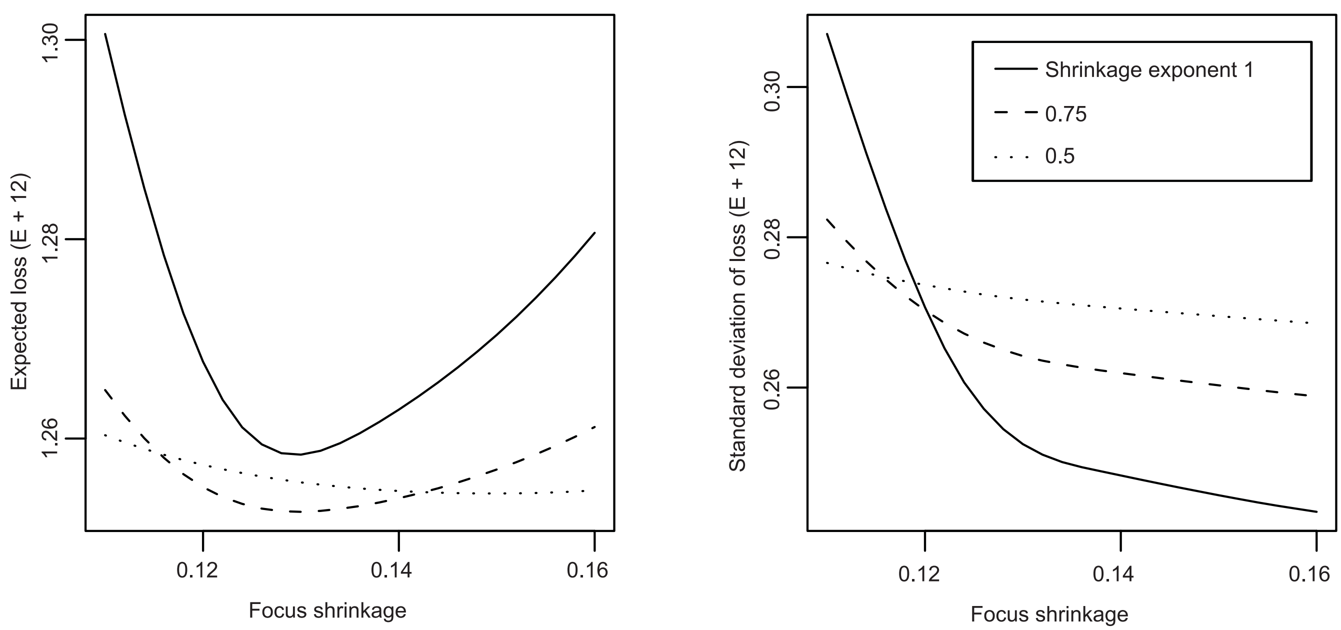

Fig. 5. The expectation and standard deviation of the squared shortfall as functions of the focus of shrinkage.

Table 1. The minimum expected squared shortfall for several sample sizes and shrinkage exponents. The columns headed 'Shortfall' give the minimum E(S) with the corresponding standard deviations in parentheses underneath.

Sample size	Shrinkage exponent						
	$q = 1$		$q = 0.75$		$q = 0.50$		$q = 0$
	Shortfall	Optimal focus	Shortfall	Optimal focus	Shortfall	Optimal focus	Direct estimation
9000	1.2584 (0.2525)	0.125	1.2526 (0.2642)	0.130	1.2546 (0.2699)	0.145	1.3112 (0.2917)
12000	1.1305 (0.2392)	0.125	1.1280 (0.2380)	0.130	1.1156 (0.2408)	0.155	1.1670 (0.2585)
15000	1.0306 (0.2160)	0.125	1.0317 (0.2172)	0.130	1.0189 (0.2191)	0.150	1.0602 (0.2311)
20000	0.9156 (0.1863)	0.120	0.9127 (0.1895)	0.130	0.9111 (0.1901)	0.160	0.9369 (0.2011)
25000	0.8313 (0.1650)	0.115	0.8263 (0.1674)	0.135	0.8243 (0.1679)	0.160	0.8509 (0.1776)
40000	0.6931 (0.1247)	0.110	0.6918 (0.1235)	0.140	0.6899 (0.1248)	0.180	0.7013 (0.1315)

size 25000, E(S) with the composite estimator ($q = 1$) is 0.8792, with standard deviation 0.1885, greater by 0.048 (5.5%) than with the optimal focus $T' = 11.5\%$. With a smaller exponent ($q = 0.5$), we gain both stability and a lower minimum.

If the funds for allocation and for the conduct of the study originate from the same source the problem of how to split the funds for the two activities has to be addressed. Suppose an increase of the sample size by 1000 is associated with the additional outlay of 50000 units. So, after a survey with sample size of 15000, the funds for allocation would be reduced to 3.7 million, and with 25000 to 3.2 million.

Table 2 displays the minimum squared shortfalls for the same sample sizes as in Table 1. The row for $n = 9000$ is copied to make the comparisons easier.

The table indicates that the optimal sample size is somewhere in the range 20000 – 25000. For example, for $q = 0.5$, the optimal sample size is close to 21000 (E(S) equal to 1.0707). The exact determination is not useful, because the minimum squared shortfall is a flat function of the sample size in this range, and the optimum is bound to depend also on the values of θ_d, on which our analysis is conditioned. The allocation based on direct estimation is consistently less optimal than the allocations based on shrinkage estimators with a selected focus, although the differences decrease with sample size. For shrinkage estimators, the dependence of E(S) on the focus becomes weaker with the sample size, and for $n = 40000$ the difference between the optimal focus and focus $\hat{\theta}$ is of the order $O(0.01)$.

4.1 Sensitivity Analysis

While illustrating that efficient estimation of θ is not in accord with effective allocation of funds is relatively easy, finding an (almost) optimal allocation is much harder because we have to rely almost exclusively on a search based on simulations. We have to respond to the concern that relatively small changes in the settings of the simulations may result in substantial changes in the relative sizes of E(S). In particular, we have to explore the results for different vectors of rates θ. The smallest E(S) is bound to depend on θ a great deal, but it suffices if there is a setting (exponent q and focus θ^*) for which E(S) is close to its minimum for all plausible values of θ (distributions of the district-level rates).

One set of simulations, with 5000 replications, takes about 60 seconds of CPU time, so extensive exploration of scenarios is feasible, even though the results are difficult to summarise in a compact fashion. We have found that the empirical Bayes

Table 2. The minimum expected squared shortfall for several sample sizes and shrinkage exponents, with the funds for allocation adjusted for survey costs (50000 units for each 1000 subjects above the sample size 9000).

Sample size (Total alloc.)	Shrinkage exponent						
	$q = 1$		$q = 0.75$		$q = 0.50$		$q = 0$
	Shortfall	Optimal focus	Shortfall	Optimal focus	Shortfall	Optimal focus	Direct estimation
9000 (4.00×10^{12})	1.2584 (0.2525)	0.125	1.2526 (0.2642)	0.130	1.2546 (0.2699)	0.145	1.3112 (0.2917)
12000 (3.85×10^{12})	1.1692 (0.2378)	0.125	1.1679 (0.2370)	0.130	1.1558 (0.2405)	0.150	1.2046 (0.2553)
15000 (3.70×10^{12})	1.1133 (0.2143)	0.125	1.1155 (0.2160)	0.125	1.1036 (0.2182)	0.155	1.1467 (0.2330)
20000 (3.45×10^{12})	1.0767 (0.1870)	0.110	1.0822 (0.1871)	0.125	1.0710 (0.1892)	0.160	1.1039 (0.1990)
25000 (3.20×10^{12})	1.0880 (0.1648)	0.105	1.0938 (0.1655)	0.125	1.0835 (0.1721)	0.160	1.1087 (0.1717)
40000 (2.45×10^{12})	1.2927 (0.1221)	0.105	1.2958 (0.1203)	0.145	1.2883 (0.1222)	0.180	1.3047 (0.1272)

shrinkage is not optimal in any setting, and the shrinkage exponent should be set between 0.5 and 0.75. The threshold T is always a better focus of shrinkage than the national rate θ (or $\bar{\theta}$, the average of the district-level rates).

5. DISCUSSION

We have presented a simulation-based method for improving the naive allocation of a fixed amount of funding to the districts of a country, according to a policy described by an allocation formula. We assumed a threshold for making grants that is attained by a minority of the districts. In the first step, we reduced the threshold for making awards. In the second, we set district-specific thresholds. In a transparent system, the first step would attract little controversy, but the second may appear as objectionable to districts for which more relevant information is available. This can be interpreted as a conflict between the appearance of fairness and effectiveness of the allocation procedure. However, interpreting district-specific thresholds as unfair to some districts would be appropriate only if the district-level quantities θ were determined with precision.

By way of an illustration, consider one district with little information (large MSE of $\hat{\theta}_d$) and one with a lot of information. Suppose their values of θ_d coincide and are smaller than the threshold. The 'small' district will have greater probabilities $P(\hat{\theta}_d > T)$ and $P(\tilde{\theta}_d > T)$, and so its expected award will be greater than for the 'large' district. Although the rare awards to the large district (in replications) will tend to be greater, the two factors are unlikely to cancel out.

An element of unfairness and inefficiency of the allocation scheme with district-specific thresholds is due to biased estimation of MSE($\tilde{\theta}_d$; θ_d), arising from the use of eMSE instead of MSE (Longford 2005b, Chapter 8; Longford 2007). In any case, the improvement made in the second step over the first is small and it may be expedient to forego it.

The conclusions of our study imply that the established organisation of statistical work by one party (the statistical office) producing estimates of relevant quantities, with indication of their precision, and another applying their algorithm for decision making, is ineffective when the latter algorithm uses nonlinear transformations of the estimates. The source of the problem is the fragile nature of statistical efficiency in sub-asymptotic samples: if an estimator $\hat{\xi}$ is efficient for ξ, a nonlinear transformation $g(\hat{\xi})$ need not be efficient for $g(\xi)$. Reporting estimated standard errors or other measures of uncertainty about the estimated parameters is not satisfactory unless the uncertainty can be incorporated in the client's algorithm. In brief, the combination of the two stages, estimation of θ and allocation based on $\hat{\theta}$, is not optimal when the two stages are optimal, but are performed separately, the second treating $\hat{\theta}$ as if it were θ; the two stages have to be integrated.

The EM algorithm (Dempster *et al.* 1977), applied with θ regarded as the missing information, provides an example of such an integration in the context of estimation. It is not applicable in the allocation problem directly because of the discontinuities involved, because there is no short list of (linear) sufficient statistics for the second stage, and a criterion other than maximum likelihood is used. However, the theory of the EM algorithm provides an explanation for the 'paradoxes' observed.

Finding analytical support for the empirical conclusions of this paper is an outstanding challenge. A simpler problem, of deciding between two groups as to which one has a greater expectation, when losses (negative utilities) associated with the two kinds of bad decisions are specified, is solved in Longford (2012). The problem addressed in this paper is much more complex, because it involves several groups (districts) and their utilities are linked by the limited budget.

ACKNOWLEDGEMENTS

Part of the research for this paper was conducted during the author's visiting appointment in the Department of Economics and Business, University Pompeu Fabra, Barcelona, Spain. Support by the Grants SAB2004–0190 and SEJ2006–13537, is acknowledged.

REFERENCES

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood for incomplete data via the EM algorithm. *J. Roy. Statist. Soc.,* Ser. **B39**, 1–38.

Efron, B. and Morris, C.N. (1972). Limiting the risk of Bayes and empirical Bayes estimators —Part II. The empirical Bayes case. *J. Amer. Statist. Assoc.*, **67**, 130–139.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269–277.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Stat. Sci.*, **9**, 55–93.

Goldstein, H. (2002). *Multilevel Statistical Models*. 3rd ed. Edward Arnold, London.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In : J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* **1**, pp. 361–379. University of California Press, Berkeley, CA.

Longford, N.T. (1993). *Random Coefficient Models*. Oxford University Press, Oxford, UK.

Longford, N.T. (1999). Multivariate shrinkage estimation of small-area means and proportions. *J. Roy. Statist. Soc.,* Ser. **A162**, 227–245.

Longford, N.T. (2004). Missing data and small-area estimation in the UK Labour Force Survey. *J. Roy. Statist. Soc.,* Ser. **A167**, 341–373.

Longford, N.T. (2005a). On selection and composition in small area and mapping problems. *Statist. Methods Medical Res.*, **14**, 1–14.

Longford, N.T. (2005b). *Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician*. Springer-Verlag, New York.

Longford, N.T. (2007). On standard errors of model-based small-area estimators. *Survey Methodology*, **33**, 69–79.

Longford, N.T. (2012). Comparing normal random samples, with uncertainty about the priors and utilities. *Scandinavian J. Stat.*, **39**; to appear (DOI: 10.1111/j.1467-9469.2011.00743.x).

Platek, R., Rao, J N.K., Särndal, C.E., Singh, M.P. (1987). *Small Area Statistics. An International Symposium*. Wiley and Sons, New York.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley and Sons, New York.

Robbins, H.E. (1955). An empirical Bayes approach to statistics. In : J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, pp. 157–163. University of California Press, Berkeley, CA.

Shen, W., and Louis, T.A. (1998). Triple-goal estimates in two-stage hierarchical models. *J. Roy. Statist. Soc.,* Ser. **B60**, 455–471.