J. R. Statist. Soc. A (1999)
162, *Part* 2, *pp.* 227–245

Multivariate shrinkage estimation of small area means and proportions

Nicholas T. Longford

De Montfort University, Leicester, UK

[Received October 1997. Final revision August 1998]

Summary. The familiar (univariate) shrinkage estimator of a small area mean or proportion combines information from the small area and a national survey. We define a multivariate shrinkage estimator which combines information also across subpopulations and outcome variables. The superiority of the multivariate shrinkage over univariate shrinkage, and of the univariate shrinkage over the unbiased (sample) means, is illustrated on examples of estimating the local area rates of economic activity in the subpopulations defined by ethnicity, age and sex. The examples use the sample of anonymized records of individuals from the 1991 UK census. The method requires no distributional assumptions but relies on the appropriateness of the quadratic loss function. The implementation of the method involves minimum outlay of computing. Multivariate shrinkage is particularly effective when the area level means are highly correlated and the sample means of one or a few components have small sampling and between-area variances. Estimation for subpopulations based on small samples can be greatly improved by incorporating information from subpopulations with larger sample sizes.

Keywords: Between-area variation; Economic activity rate; Ethnic minorities; Samples of anonymized records; Shrinkage estimation; UK census

1. Introduction

Local authorities require a variety of statistics about the sociodemographic, economic, housing and other attributes of their residents. Many of these statistics are proportions or percentages, such as the proportion of economically active men or women in a certain age and ethnic group. The national (aggregate) versions of such statistics are also collated by large scale surveys and censuses.

Recently, researchers in local government in the UK have gained access to national samples with local area identifiers, such as the samples of anonymized records (SARs) drawn from the 1991 UK census. The principal motivation for using SARs is to take advantage of the complementary strengths of two sources of information: the local area subsample and the entire (national) sample. A given rate (proportion or percentage), such as the economic activity rate for a local area, is estimated from the local area subsample of the SAR with little precision, although the estimator is (nearly) unbiased. However, the national survey yields a very precise estimator of the *national* rate, but it is a biased estimator of the local area rate because, in general, the local areas have different rates.

Local authorities require the economic activity rates for predicting future income and expenditure as well as demand for various services. The 'exact' rates, based on the 1991 UK census, are available for adult men and women (Local Government Management Board and

Address for correspondence: Nicholas T. Longford, Department of Medical Statistics, De Montfort University, The Gateway, Leicester, LE1 9BH, UK.
E-mail: ntl@dmu.ac.uk

 0964–1998/99/162227

the London Research Centre, 1992), but not for ethnic or age groups. For these subpopulations, and ethnic groups in particular, many area level subsamples are so small that the sample rates, the naïve estimators based solely on the local area data, are clearly unsatisfactory and more efficient estimation is essential. We apply shrinkage estimation to this problem, develop a multivariate extension and outline how shrinkage can be used more generally to borrow strength not only across local areas but also across subpopulations and SARs drawn from different censuses. The theoretical development is for outcomes with an arbitrary and unspecified distribution; the application to SAR data is for dichotomous outcomes.

Shrinkage (composite) estimators combine the local area and the national estimators, with weights determined to optimize (approximately) the precision of the combination. The weights have intuitively appealing properties. For instance, for a given national estimator, the more precise the local area estimator, the more weight it is given. When the underlying ('true') means for the local areas in the nation differ greatly the national estimator is given little weight; when the means are almost identical, the local information is almost ignored and the shrinkage estimator for each local area is close to the national mean. The gain in precision *vis-à-vis* the sample mean is greatest when the underlying local area means are very similar across the nation and the national mean is estimated with much greater precision than the local area mean.

The method is computationally elementary and undemanding. The calculations are not iterative and they can be organized so that only one pass through the survey data is required. We emphasize that the method is not guaranteed to improve the estimate for *every* local area (over the estimate based solely on the local area). The improvement is *in expectation*, i.e. after averaging over the samples that could have been drawn, as well as averaging over the local areas. An assessment of the average improvement, in terms of the reduction in the expected mean-squared error (EMSE), is an integral part of the method.

The multivariate version of the shrinkage estimator can be applied to both a vector of local area means and to a subvector of means by using other variables, subpopulations or surveys as auxiliary information. When the component means are correlated at the local area level, multivariate shrinkage is more efficient than univariate shrinkage applied separately to each component. For instance, an estimation of the rates for an ethnic minority can incorporate information about white subjects. When the underlying rates for minority and white subjects are highly correlated, the estimation for the minority is almost as precise as for the union of minority and white subjects, resulting in a substantial gain in precision over the univariate shrinkage.

The results for the univariate shrinkage estimator are reviewed in Section 2. The multivariate shrinkage estimator is introduced in Section 3. Section 4 applies the method to the 1991 SAR of individuals in Great Britain. Section 5 discusses validation of the method, some extensions and alternatives, and the effect of the uncertainty about the between-area variance. The concluding section outlines how the information in the SAR based on the forthcoming 2001 UK census could be augmented by that in the current SAR.

1.1. Notation

Suppose that the national database, collected using a simple random sampling design, contains the values y_{il} of a variable y for subjects $i = 1, \ldots, n_l$ from local areas $l = 1, \ldots, L$. Let $N = n_1 + \ldots + n_L$ be the total sample size and define $q_l = n_l/N$. The local area sample means are

$$\hat{p}_l = \frac{1}{n_l} \sum_i^{n_l} y_{il},$$

and the national sample mean is

$$\hat{p} = \frac{1}{N} \sum_l \sum_i y_{il} = \sum_l q_l \hat{p}_l.$$

For the means underlying $\hat{p}_l$ and $\hat{p}$ assume the model

$$p_l = p + \delta_l, \tag{1}$$

where δ_l, $l = 1, \ldots, L$, are independent and identically distributed with zero mean and unknown (between-area) variance σ^2. No assumptions are made about the shape or family of the candidate distributions for δ_l.

For a dichotomous variable y it is more traditional to use logit models, such as

$$\text{logit}(p_l) = \text{logit}(p) + \xi_l,$$

with a distributional assumption for $\{\xi_l\}$; see Thomas *et al.* (1994) for an application to health care quality control. In the application to economic activity rates in Section 4, we prefer the identity link because it is better suited to the quadratic loss function. When the proportions p_l attain values within the range 0.20–0.80 the logit, probit and complementary log–log-links are closely approximated by a linear function, so the choice of the link is immaterial. For very small and very large proportions, the quadratic loss function with a non-linear link, such as the logit, imposes a large penalty for estimation error because small differences between observed and fitted proportions translate to large differences in logits. When an estimated proportion is used in a linear formula for the local authority's expenditure or for a similar purpose, this is undesirable; the loss due to an error in estimating a proportion should not depend on the underlying proportion.

2. Univariate shrinkage

The local area estimator $\hat{p}_l$ and the national estimator $\hat{p}$ are combined into the *shrinkage* estimator

$$\tilde{p}_l = (1 - b_l)\hat{p}_l + b_l \hat{p}. \tag{2}$$

The area-specific coefficients b_l are chosen to minimize (approximately) EMSE $= E(\tilde{p}_l - p_l)^2$; the expectation is taken over both sampling (other samples that could have been drawn) and the (population of) local areas:

$$\text{EMSE} = E_s[E_1\{(\tilde{p}_l - p_l)^2 | p_l\}],$$

where E_1 denotes the conditional expectation given the local area and E_s denotes the (unconditional) expectation over the local areas.

EMSE is a quadratic function of b_l; if $q_l < \frac{1}{2}$, its minimum is attained for

$$b_l^* = \frac{v_l(1 - q_l)}{v_l(1 - 2q_l) + \text{var}(\hat{p}) + \sigma^2},$$

and the mean-squared error (MSE) is reduced from the sampling variance of $\hat{p}_l$, $v_l = \text{var}(\hat{p}_l | p_l)$, to

$$\mathrm{EMSE}(b_l^*) = v_l \left\{ 1 - \frac{v_l(1-q_l)^2}{v_l(1-2q_l) + \mathrm{var}(\hat{p}) + \sigma^2} \right\} = v_l\{1 - b_l^*(1-q_l)\}. \tag{3}$$

The shrinkage estimator is

$$\tilde{p}_l = \frac{\{\mathrm{var}(\hat{p}) + \sigma^2 - v_l q_l\}\hat{p}_l + v_l(1-q_l)\hat{p}}{v_l(1-2q_l) + \mathrm{var}(\hat{p}) + \sigma^2}. \tag{4}$$

These results are derived in Section 3 for a more general case.

In applications, the fraction q_l as well as the sampling variance $\mathrm{var}(\hat{p})$ are commonly ignored, yielding the more familiar estimator

$$\tilde{p}_l^{\dagger} = \frac{\sigma^2 \hat{p}_l + v_l \hat{p}}{\sigma^2 + v_l}.$$

This is appropriate in large scale surveys (large N) which cover many local areas (large L) and when the subsample from area l is a small fraction of the survey data ($q_l \ll 1$). The EMSE of $\tilde{p}_l^{\dagger}$ is approximately equal to $1/(1/\sigma^2 + 1/v_l)$; this is smaller than the MSE of the trivial estimator $\hat{p}_l$ by $v_l^2/(v_l + \sigma^2)$, i.e. $1 + v_l/\sigma^2$ times.

Shrinkage estimates display less variation than do the underlying quantities. Researchers are therefore reluctant to use shrinkage to compare local areas. However, shrinkage is optimal not only for means but also for linear combinations of means. For instance, the estimator of the difference of the means of two local areas, $p_1 - p_2$, optimal among all the linear combinations of $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}$, is $\tilde{p}_1 - \tilde{p}_2$. (Of course, the areas 1 and 2 must be identified independently of the realized subsamples.) See Shen and Louis (1998) for a Bayesian approach which attempts to reconcile the goals of efficient estimation of p_l, the ranks of p_l and the (area level) empirical distribution of p_l, $l = 1, \ldots, L$. Our objective is only the first of these goals, motivated in Section 1 by the needs of a typical local authority.

The method based on shrinkage estimation is motivated by Fay and Herriot (1979), who applied a general result due to James and Stein (1961). In the next section, we adapt it for the simultaneous estimation of several means. See Ghosh and Rao (1994) for a comprehensive review of small area methods and Singh *et al.* (1998) for a simulation-based comparison of several methods for small area estimation. Shrinkage estimators are an application of combining information that originates from various sources. In our case, the (*noisy*) trivial estimators $\hat{p}_l$ are 'shrunk' towards the (*stable*) overall mean $\hat{p}$. Longford (1995) contains some applications of shrinkage to estimating large sets (ensembles) of parameters. Efron (1996) has described a general Bayesian framework for pooling information across similar units.

3. Multivariate shrinkage

To estimate the local area means for several subpopulations the (univariate) shrinkage estimator can be applied separately to each subpopulation. Although this is more efficient than using the area level sample means, it fails to take advantage of the similarity of the area level means across the subpopulations. As an example, suppose that the within-area differences of the means of the outcomes for an ethnic group and white residents (who are in a majority) are each equal to C. Then an efficient estimator for the white residents, adjusted by an estimate of C, may be superior to the univariate shrinkage for the ethnic group.

Similarly, when estimating the local area means of a vector of outcomes, univariate shrinkage can be applied to each component of the vector. However, additional gains can be realized by borrowing strength across the components, especially when their area level means are highly correlated.

Let $\hat{\mathbf{p}}_l$ be an unbiased estimator of the vector of area level means $\mathbf{p}_l$, and $\hat{\mathbf{p}}$ and $\mathbf{p}$ be their 'national' counterparts. Let $\mathbf{V}_l$ be the sampling variance matrix of $\hat{\mathbf{p}}_l$ and $\boldsymbol{\Sigma}$ the between-area variance matrix, $\boldsymbol{\Sigma} = \mathrm{var}_s(\mathbf{p}_l)$. Let $\mathbf{Q}_l$ be the diagonal matrix of the fractions: $\mathbf{Q}_l = \mathrm{diag}_k(q_l^{(k)})$, where $q_l^{(k)} = n_l^{(k)}/N^{(k)}$ is the fraction q_l for component k and local area l.

For estimating the linear function of area level means $\mathbf{p}_l^{\mathrm{T}}\mathbf{w}$ for a given vector of constants $\mathbf{w}$, we consider the linear combinations

$$\tilde{p}_l = \hat{\mathbf{p}}_l^{\mathrm{T}}(\mathbf{w} - \mathbf{b}_l) + \hat{\mathbf{p}}^{\mathrm{T}}\mathbf{b}_l \tag{5}$$

and seek the vector $\mathbf{b}_l$ which minimizes

$$\mathrm{EMSE}(\mathbf{b}_l) = E_s[E_1\{(\tilde{p}_l - \mathbf{p}_l^{\mathrm{T}}\mathbf{w})^2|\mathbf{p}_l\}].$$

Simple operations yield the identity

$$\mathrm{EMSE}(\mathbf{b}_l) = (\mathbf{w} - \mathbf{b}_l)^{\mathrm{T}}\mathbf{V}_l(\mathbf{w} - \mathbf{b}_l) + \mathbf{b}_l^{\mathrm{T}}\{\boldsymbol{\Sigma} + \mathrm{var}(\hat{\mathbf{p}})\}\mathbf{b}_l + 2\mathbf{b}_l^{\mathrm{T}}\mathbf{Q}_l\mathbf{V}_l(\mathbf{w} - \mathbf{b}_l).$$

The minimum of this quadratic function of $\mathbf{b}_l$ is found either by vector differentiation or by completing the square; the minimum is attained for

$$\mathbf{b}_l^* = \mathbf{D}_l^{-1}(\mathbf{I} - \mathbf{Q}_l)\mathbf{V}_l\mathbf{w}, \tag{6}$$

where $\mathbf{D}_l = \mathbf{V}_l + \mathrm{var}(\hat{\mathbf{p}}) + \boldsymbol{\Sigma} - \mathbf{Q}_l\mathbf{V}_l - \mathbf{V}_l\mathbf{Q}_l$ and $\mathbf{I}$ is the identity matrix of the appropriate size. The minimum attained is

$$\begin{aligned}\mathrm{EMSE}(\mathbf{b}_l^*) &= \mathbf{w}^{\mathrm{T}}\mathbf{V}_l\mathbf{w} - \mathbf{w}^{\mathrm{T}}\mathbf{V}_l(\mathbf{I} - \mathbf{Q}_l)\mathbf{D}_l^{-1}(\mathbf{I} - \mathbf{Q}_l)\mathbf{V}_l\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{V}_l\mathbf{w} - \mathbf{b}_l^{*\mathrm{T}}\mathbf{D}_l\mathbf{b}_l^*.\end{aligned} \tag{7}$$

By substituting equation (6) in equation (5) and extracting the vector $\mathbf{w}$ we obtain the multivariate shrinkage estimator

$$\tilde{\mathbf{p}}_l = \{\mathbf{I} - \mathbf{V}_l(\mathbf{I} - \mathbf{Q}_l)\mathbf{D}_l^{-1}\}\hat{\mathbf{p}}_l + \mathbf{V}_l(\mathbf{I} - \mathbf{Q}_l)\mathbf{D}_l^{-1}\hat{\mathbf{p}}. \tag{8}$$

For any vector $\mathbf{w}$, $\tilde{\mathbf{p}}_l^{\mathrm{T}}\mathbf{w}$ is optimal for $\mathbf{p}_l^{\mathrm{T}}\mathbf{w}$ among all linear combinations of $\hat{\mathbf{p}}_l$ and $\hat{\mathbf{p}}$. The results quoted in Section 2 are the univariate versions of equations (6)–(8), with $\mathbf{w} = 1$.

In practice, the fractions in $\mathbf{Q}_l$ can often be ignored, yielding a simpler expression for $\mathbf{b}_l^*$ and $\mathrm{EMSE}(\mathbf{b}_l^*)$. When the variance matrices $\boldsymbol{\Sigma}$, $\mathbf{V}_l$ and $\mathrm{var}(\hat{\mathbf{p}})$ are known multivariate shrinkage is at least as efficient as univariate shrinkage applied to each component because EMSE is minimized over a wider class of linear combinations. The two estimators coincide when the matrices $\boldsymbol{\Sigma}$, $\mathbf{V}_l$ and $\mathrm{var}(\hat{\mathbf{p}})$ are diagonal. $\mathbf{V}_l$ and $\mathrm{var}(\hat{\mathbf{p}})$ are diagonal when the components $k = 1, \ldots, K$ refer to non-overlapping subpopulations and the sampling schemes for these subpopulations are independent.

For a vector of variables, each observed on the same sample of subjects, $\mathbf{V}_l$ and $\mathrm{var}(\hat{\mathbf{p}})$ are no longer diagonal and some form of replication is required to separate $\mathbf{V}_l$ and $\boldsymbol{\Sigma}$ from the variance matrix of the outcomes. See Longford (1997) for an example from educational measurement.

3.1. Auxiliary information

The multivariate shrinkage estimator can also be applied when the means of some or only of a single component of $\hat{\mathbf{p}}_l$ are of interest. For estimating the mean of the first component we set $\mathbf{w} = (1, 0, \ldots, 0)$. Then $\tilde{p}_l$ may depend on all the sample means in $\hat{\mathbf{p}}_l$; information is 'borrowed' across the components of $\mathbf{p}_l$ as well as across the local areas.

More generally, consider estimating a linear combination $\mathbf{w}^{\mathrm{T}}\mathbf{p}_l$ based on the vectors of outcomes $\mathbf{y}$ corresponding to $\mathbf{p}_l$, and estimating the same linear combination based on the vectors $\mathbf{y}$ augmented by some auxiliary variables $\mathbf{y}^{(a)}$. Let $\mathbf{p}_l^{(a)}$, $\mathbf{V}_l^{(a)}$, $\mathbf{\Sigma}^{(a)}$, and so on, be the area level mean, its sampling variance matrix, the between-area variance matrix, and so on, for $\mathbf{y}^{(a)}$. Suppose that the estimators $\hat{\mathbf{p}}_l$ and $\hat{\mathbf{p}}_l^{(a)}$ are pairwise independent for each l, and denote by $\mathbf{\Sigma}^{(0,a)} = (\mathbf{\Sigma}_l^{(a,0)})^{\mathrm{T}}$ the between-area covariance matrix of $\mathbf{p}_l$ and $\mathbf{p}_l^{(a)}$. In the following derivation, we use the formula for the inverse of a partitioned (positive definite variance) matrix; in our notation

$$\begin{pmatrix} \mathbf{D}_l & \mathbf{\Sigma}^{(0,a)} \\ \mathbf{\Sigma}^{(a,0)} & \mathbf{D}_l^{(a)} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{F}_l & -\mathbf{F}_l\mathbf{\Sigma}^{(0,a)}\mathbf{D}_l^{(a)} \\ -\mathbf{D}_l^{(a)}\mathbf{\Sigma}^{(a,0)}\mathbf{F}_l & \mathbf{D}_l^{(a)}\mathbf{\Sigma}^{(a,0)}\mathbf{F}_l\mathbf{\Sigma}^{(0,a)}\mathbf{D}_l^{(a)} \end{pmatrix},$$

where $\mathbf{F}_l = (\mathbf{D}_l - \mathbf{\Sigma}^{(0,a)}(\mathbf{D}_l^{(a)})^{-1}\mathbf{\Sigma}^{(a,0)})^{-1}$. The EMSE of the shrinkage estimator of $\mathbf{w}^{\mathrm{T}}\mathbf{p}_l$, using the (auxiliary) information in $\mathbf{y}^{(a)}$, is

$$\mathbf{w}^{\mathrm{T}}\mathbf{V}_l\mathbf{w} - \mathbf{s}_l^{\mathrm{T}}\begin{pmatrix} \mathbf{D}_l & \mathbf{\Sigma}^{(0,a)} \\ \mathbf{\Sigma}^{(a,0)} & \mathbf{D}_l^{(a)} \end{pmatrix}^{-1}\mathbf{s}_l = \mathbf{w}^{\mathrm{T}}\mathbf{V}_l\mathbf{w} - \mathbf{s}_l^{\mathrm{T}}\mathbf{F}_l\mathbf{s}_l, \tag{9}$$

where

$$\mathbf{s}_l = \begin{pmatrix} \mathbf{I} - \mathbf{Q}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^{(a)} - \mathbf{Q}_l^{(a)} \end{pmatrix}\begin{pmatrix} \mathbf{V}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_l^{(a)} \end{pmatrix}\begin{pmatrix} \mathbf{w} \\ \mathbf{0} \end{pmatrix} = (\mathbf{I} - \mathbf{Q}_l)\mathbf{V}_l\mathbf{w}$$

and $\mathbf{0}$ is the matrix of 0s of the appropriate size. Hence the information in $\mathbf{y}^{(a)}$ brings about a reduction in EMSE by $\mathbf{s}_l^{\mathrm{T}}(\mathbf{F}_l - \mathbf{D}_l^{-1})\mathbf{s}_l$; this is non-negative because $\mathbf{F}_l - \mathbf{D}_l^{-1}$ is non-negative definite. Naturally, there is no gain in precision when $\mathbf{\Sigma}^{(a,0)} = \mathbf{0}$. Given a choice of auxiliary information, preference should be given to variables for which the between-area variance matrix $\mathbf{\Sigma}^{(a)}$ is small (has small eigenvalues), $\mathbf{p}_l^{(a)}$ is estimated with high precision ($\mathbf{V}_l^{(a)}$ is small) and the correlations implied by $\mathbf{\Sigma}^{(a,0)}$ are large.

In the examples in Section 4, we apply bivariate shrinkage with a subpopulation (say, white residents) serving as auxiliary information for the subpopulation of interest (say, for ethnic minority residents). In such a case, the naïve estimators $\hat{p}_l$ and $\hat{p}_l^{(a)}$ are independent so long as the subpopulations have no overlap and their sampling is independent. In sampling without replacement from finite subpopulations there is some dependence but it is negligible when the sampling proportion is small. In the context of SARs, this dependence can be ignored. Setting $\mathbf{w} = (1, 0)^{\mathrm{T}}$ in equation (9), the EMSE of the bivariate shrinkage estimator is

$$v_l - \frac{v_l^2(1-q_l)^2}{d_l - \sigma_{a,0}^2/d_l^{(a)}}, \tag{10}$$

where $\sigma_{a,0}$ and $d_l^{(a)}$ are the respective scalar versions of $\mathbf{\Sigma}^{(a,0)}$ and $\mathbf{D}_l^{(a)}$. The EMSE of the univariate shrinkage estimator of p_l is obtained by substituting 0 for $\sigma_{a,0}$ in expression (10). Hence the gain in precision of the bivariate shrinkage over the univariate shrinkage is an increasing function of $\sigma_{a,0}^2/d_l^{(a)}$; there is no gain only when $\sigma_{a,0} = 0$. Given a choice between alternatives for a single auxiliary subpopulation, it is preferable to select it so that the area level means p_l and $p_l^{(a)}$ are highly correlated (the sign of the correlation is immaterial), and $d_l^{(a)}$

is small, i.e., ideally, the within-area means $p_l^{(a)}$ and the national mean $p^{(a)}$ are estimated with high precision and the between-area variance σ_a^2 is small.

3.2. Conditional mean-squared error

Since we are concerned with estimation for a single local area, the MSE, given the mean outcome $\mathbf{p}_l$, would be more appropriate than the EMSE as the standard by which to compare alternative estimators. MSE as well as its minimizer depend on the unknown value of $\mathbf{p}_l$, and so they cannot be used. Nevertheless, it is instructive to explore the performance of the EMSE-optimal shrinkage estimator with respect to the MSE standard. The MSE of the shrinkage estimator $\tilde{p}_l$ is obtained by replacing $\mathbf{\Sigma}$ in $\mathbf{D}_l$ in equation (7) with $(\mathbf{p}_l - \mathbf{p})(\mathbf{p}_l - \mathbf{p})^{\mathrm{T}}$:

$$\mathrm{MSE}(\mathbf{b}_l^*) = \mathbf{w}^{\mathrm{T}}\mathbf{V}_l\mathbf{w} - \mathbf{b}_l^{*\mathrm{T}}\{2\mathbf{\Sigma} + \mathrm{var}(\hat{\mathbf{p}}) + \mathbf{V}_l - \mathbf{Q}_l\mathbf{V}_l - \mathbf{V}_l\mathbf{Q}_l - (\mathbf{p}_l - \mathbf{p})(\mathbf{p}_l - \mathbf{p})^{\mathrm{T}}\}\mathbf{b}_l^*.$$

When the expression inside the braces is positive definite $\mathrm{MSE}(\mathbf{b}_l^*)$ is smaller than the MSE of the naïve estimator $\mathbf{w}^{\mathrm{T}}\hat{\mathbf{p}}_l$, $\mathrm{MSE}(\mathbf{0})$. Thus, the shrinkage estimator *may* be conditionally less efficient than the naïve estimator only when the area level means $\mathbf{p}_l$ are very distant from the national mean $\mathbf{p}$. In the univariate case, the sample rate has smaller MSE than the shrinkage estimator when

$$(p_l - p)^2 > 2\sigma^2 + v_l(1 - 2q_l).$$

When q_l is small and v_l is at least of the same order of magnitude as σ^2 this is likely to be satisfied for all except a few areas with outlying means p_l.

4. Application

In this section, we discuss the estimation of the local area rates of economic activity for two subpopulations of interest: white residents 16–19 years old and minorities in Great Britain. In both cases, univariate shrinkage is much more efficient than the sample rates. For young white men and women, the bivariate shrinkage, with one sex providing auxiliary information for the other, yields little gain over univariate shrinkage. In contrast, for the minorities, the bivariate shrinkage with white subjects providing auxiliary information is very effective for many local areas. Throughout, we use percentages instead of proportions.

A resident is said to be economically active if he or she is either employed or is actively seeking employment (for example, full-time students and housewives are not economically active). The $L = 278$ local areas were defined by the constructors of the SAR to provide as fine geographical detail as possible while ensuring anonymity of all individuals. A local area is roughly equivalent to a large local authority, with an estimated population (in mid-1989) of at least 120000. For instance, areas 1–32 coincide with the London boroughs, with the City of London attached to the City of Westminster in area 1. The areas are geographically contiguous. See Heath (1994) for details.

4.1. Economic activity rates for residents 16–19 years old

To avoid repetition in the illustration of univariate shrinkage, we focus on the results for young men. The effect of univariate shrinkage on the estimated rates for young women is very similar, because the sample sizes and the between-area variances are similar for the two sexes.

The database, a random sample (without replacement) of individuals from the census, contains records of $N = 26963$ white men (ETHGROUP = 1 and SEX = 1 in the SAR) aged

16–19 years, i.e. on average about 97 subjects per area. The numbers of subjects within the areas are in the range 30–359; their median is 90. Four local areas have subsample sizes that are greater than 200: Glasgow City (270), Sheffield (277), Leeds (331) and Birmingham (359). The within-area sample rates $100\hat{p}_l$ are in the range 35.5–79.3%; their sample (observed) standard deviation $100\sqrt{\mathrm{var}_s(\hat{p}_l)}$ is 7.33%. Since even the 'largest' area forms only 1.33% of the subsample, the correlation of $\hat{p}_l$ with $\hat{p}$ can be ignored and all q_l set to 0 in equation (4). Also, in estimating v_l we do not apply the finite sample correction $f = 0.98$ that would reflect the 2% sampling fraction in the SAR.

The (national) sample rate of economic activity among white men aged 16–19 years is $100\hat{p} = 63.2\%$. The estimate of the between-area standard deviation $100\hat{\sigma}$ is 4.65%. Thus, the within-area rates vary much less than what would be judged from the sample rates (their standard deviation is 7.33%). The variation in the sample rates in excess of the variation in the underlying ('true') rates is due to small sample sizes within the local areas. For instance, the estimated sampling variance of the rate for an area with 60 economically active and 40 economically inactive residents in the SAR is 24.0, of the same order of magnitude as the between-area variance ($4.65^2 = 21.6$), but it is a source of variation additional to (and independent of) the area level variation.

The shrinkage estimates of the economic activity rates are graphically summarized in Fig. 1. The sections of the plot correspond to London (L), the other conurbations in England (EC), the rest of England (ER), Wales (W) and Scotland (S). The horizontal axis is the area

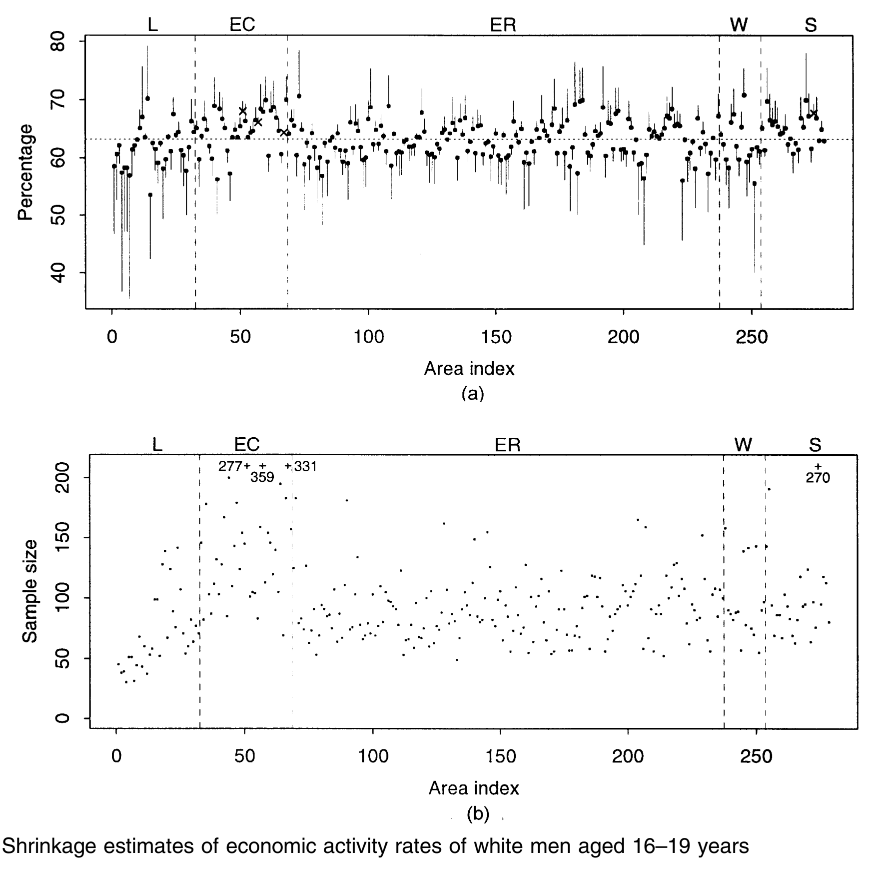


Fig. 1. Shrinkage estimates of economic activity rates of white men aged 16–19 years

order number in the SAR. The areas are in alphabetical order within counties or regions, and the regions are in the following order: Inner London (13 areas), Outer London (19 areas), Greater Manchester (10 areas), Merseyside (six areas), South Yorkshire (four areas), Tyne and Wear (five areas), West Midlands (seven areas) and West Yorkshire (five areas), followed by the remaining counties in England (169 areas), Welsh counties (14 areas) and Scottish regions (25 areas), each in alphabetical order. The areas are listed in Heath (1994).

In Fig. 1(a) each area is represented by a vertical segment connecting the sample and the shrinkage estimates of the economic activity rates (the latter are marked by dots). In Fig. 1(b) the sample sizes are plotted. To improve the resolution of the diagram, the vertical axis is restricted to the maximum of 200. The four areas with subsample sizes greater than 200 are marked by the symbol + with the subsample sizes attached; their shrinkage estimates are marked by the symbol ×. The shrinkage for these areas is only slight. The greatest differences between the sample and shrinkage estimates occur for the areas represented by fewest subjects, several of them in London.

For brevity, we refer to the square root of the estimated EMSE as the standard error. The standard errors of the univariate shrinkage estimators are 1.13–2.15 times smaller than the standard errors of the sample rates. The reduction in the standard error over the sample rates is greatest for the smallest areas; shrinkage reduces the standard error for areas 2 (Camden — 38 subjects), 4 (Hammersmith and Fulham — 30 subjects) and 7 (Kensington and Chelsea — 31 subjects), all three in Inner London, by factors greater than 2.0. For the areas with the largest subsamples the reduction is only slight.

The estimation of the rates for young white men and women can be improved by bivariate shrinkage, so that information is borrowed across the sexes. The national sample rates for young white men and women are $100\hat{\mathbf{p}} = (63.2, 56.3)$ and the moment matching estimate of the between-area variance matrix is

$$100^2 \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 21.6 & 21.0 \\ 21.0 & 24.6 \end{pmatrix};$$

see Appendix A for details of the estimator. The estimated correlation of the local area rates is $21.0/\sqrt{(21.6 \times 24.6)} = 0.91$, implying that the pairs of rates are nearly linearly related. The bivariate shrinkage takes advantage of this association and pulls the pairs of sample rates $\hat{\mathbf{p}}_l$ towards their regression line. This is illustrated in Fig. 2 where the sample rates and the univariate and bivariate shrinkage estimates for young white men and women are plotted. The three pairs of estimates for a local area are represented by three points joined by the full lines. The univariate shrinkage estimates are marked by circles and the bivariate shrinkage estimates by dots. The national sample rates are indicated by the horizontal and vertical broken lines, and the two dotted lines are the regressions of the underlying rates for men on those for women and vice versa. The bivariate shrinkage estimates are pulled, quite radically, towards an average of these two lines. To avoid clutter of the points and lines, the estimates (shrinkage paths) are plotted only for every third local area.

The improvement of the bivariate shrinkage over the univariate shrinkage is quite modest. For instance, the sample rates for local area 3 (Hackney in Inner London, 39 men and 57 women in the SAR) are (59.0%, 42.1%), with estimated standard errors (7.9%, 6.5%). The univariate shrinkage estimates are (62.1%, 51.1%), with standard errors (4.0%, 4.0%), and the bivariate shrinkage estimates are (58.8%, 51.1%), with standard errors (3.5%, 3.7%). Note that the univariate shrinkage towards the national mean for men is completely reversed by the bivariate shrinkage towards the regression line. This is a consequence of the high correlation in $\boldsymbol{\Sigma}$.

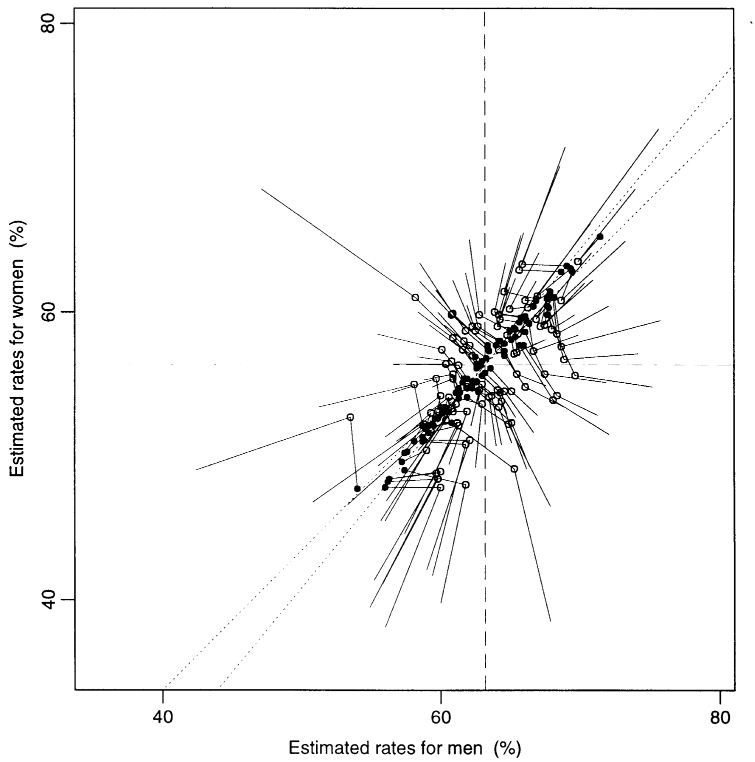


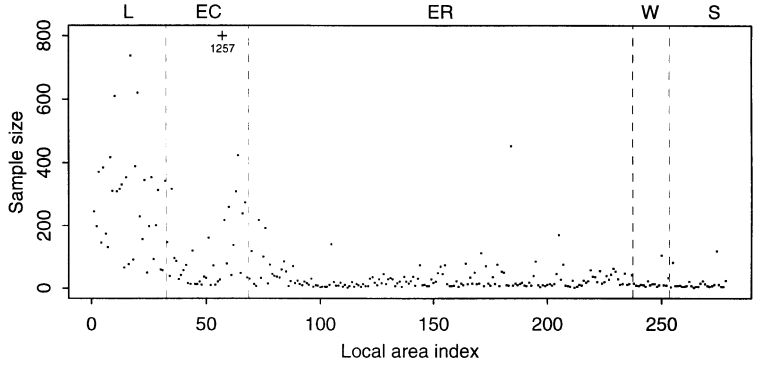
Fig. 2. Paths of shrinkage estimation of economic activity rates for white men and women aged 16–19 years

For the 278 local areas, the bivariate shrinkage estimates for young white men and women have estimated standard errors that are lower than their univariate counterparts by between 0.3% and 0.6%. Bivariate shrinkage is not very effective because the area level subsample sizes for the two sexes are very similar. For local areas with small subsample sizes there is little auxiliary information from the other sex; for local areas with large subsample sizes even the sample rates are fairly precise and there is little scope for improvement over univariate shrinkage.

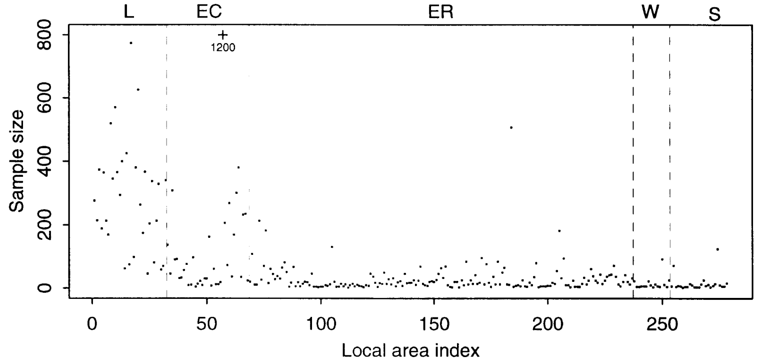
4.2. Economic activity rates of minorities

Economic activity rates for ethnic minorities are of particular interest, but their estimation is most difficult because the minorities are very sparsely and unevenly distributed across the local areas. Here, bivariate (multivariate) shrinkage radically improves the estimation by incorporating information about white residents. Fig. 3 contains plots of the numbers of adult men and women from all ethnic minorities (categories 2–10 of ETHGROUP in the SAR) within the local areas. The vertical axes are restricted to 800, to maintain good resolution of the plots. The SAR database contains records of 19343 men and 19586 women from ethnic minorities. Each local area is represented in the SAR by at least one man and one woman from a minority group. The median area level sample size is 21 for both men and women.

The subsample sizes n_l in many local areas are so small that not only the local area rates p_l but even the corresponding sampling variances $p_l(1-p_l)/n_l$ are estimated very poorly. For



(a)



(b)

Fig. 3. Numbers of (a) adult men and (b) adult women from ethnic minorities in the SAR, by local areas: + with count attached, local area 57, Birmingham

small n_l, the estimators $\hat{p}_l$ and $\hat{p}_l(1-\hat{p}_l)/n_l$ have very large sampling variances. For instance, the latter is 0 when $\hat{p}_l = 0$ or $\hat{p}_l = 1$. This problem can be alleviated by truncating $\hat{p}_l$ at suitably selected values, such as $p_{\min} = 0.25$ and $p_{\max} = 0.75$. If the between-area variance σ^2 is small the sampling variances of the sample rates can be estimated much more efficiently as $\hat{p}(1-\hat{p})/n_l$, i.e. by using the national sample rate $\hat{p}$ in place of $\hat{p}_l$. Also, the denominator $n_l - 1$ is preferred to n_l, unless $n_l = 1$.

A more complex alternative is motivated by shrinkage. The 'pooled' estimator $\hat{p}(1-\hat{p})/(n_l - 1)$ can be combined with the naïve estimator $\hat{p}_l(1-\hat{p}_l)/(n_l - 1)$, with the weights reciprocally proportional to the precisions of these variance estimators. The sampling variance of the latter estimator is given in Appendix A.1. This method is a refinement of the simple approach of using $\hat{p}(1-\hat{p})/(n_l - 1)$ when n_l is small $(n_l < n^*)$ and $\hat{p}_l(1-\hat{p}_l)/(n_l - 1)$ otherwise. Since we expect the area level economic activity rates to vary little, we use the sampling variance estimator $\widehat{\text{var}}(\hat{p}_l) = \hat{p}(1-\hat{p})/(n_l - 0.99)$ for local area subsamples of size $n_l < 50$, and $\hat{p}(1-\hat{p})/(n_l - 1)$ when $n_l \geqslant 50$. In the denominator of the former, we subtract 0.99 instead of 1.0, to avoid problems when $n_l = 1$.

The national sample rates of economic activity of the minorities are 75.5% (standard error 0.65%) for men and 52.2% (standard error 1.15%) for women. The estimated between-area standard deviation is 4.44% for men and 8.48% for women. The higher standard deviation for women reflects the variation in the within-area composition (cultural and religious background, attitudes to family, etc.) of the minorities. Neither the size of the area level standard deviation nor the possible skewness of the area level rates p_l is an obstacle to the application of shrinkage. The standard error of the national rate $100\hat{p}$ as an estimator of an area level rate $100p_l$ is about 8.5%. For $100p_l = 80\%$, the area level sample rate $100\hat{p}_l$ has a standard error greater than 8.5% when $n_l \leqslant 22$. Thus, even the extreme shrinkage given by $b_l = 1$, i.e. using $\hat{p}$, yields an estimator that is superior to the sample rate for at least 50% of the local areas (those with subsample sizes smaller than the median). The shrinkage estimator based on b_l^* is more efficient than $\hat{p}$, although, for the smallest n_l, b_l^* is close to 1. For local areas with large subsamples, more weight is given to the local information ($b_l^* \ll 1$), and the national mean is no longer competitive. However, the sample rate is competitive with the shrinkage estimator only for very large subsample sizes; then the two estimators differ only slightly.

Of course, it may be inappropriate to group all ethnic minorities into a single category. For instance, the rates p_l may be very homogeneous within minority groups. Then the national rate for a group would be an efficient estimator of the area level rate for that group. However, the sample sizes of several minority groups in the SAR are so small that even the between-area variance σ^2 would be estimated with little precision. For most local areas, which have small sample sizes n_l, the shrinkage estimate is effectively equal to the national rate $\hat{p}$. The estimates for the subpopulation of all the minorities would be constructed by combining the group-specific estimates. An outstanding problem arises when the proportions of the various minority groups (the weights) are unknown and must be estimated.

Fig. 4 provides a summary of the univariate shrinkage estimates of the economic activity rates, using the same lay-out as Fig. 1. Most radical shrinkage takes place for the areas with sample sizes that are so small that estimation is effectively based on the national sample rate. The areas in London are notable exceptions because they contain many minority residents. Somewhat less shrinkage takes place for women than for men because the between-area variation for them is greater. The gains in precision are greatest for the local areas with fewest minority residents. For instance, shrinkage pulls the sample rate for local area 52, Gateshead in Tyne and Wear (nine men from minorities in the SAR), from 66.7% to 74.8%, most of the way towards the national sample rate of 75.5%. The estimated standard error is reduced from 15.2% to 4.3%. In contrast, shrinkage alters the estimate for Birmingham by only 0.1%, relying almost exclusively on the within-area information, and so the adjustment for the correlation of the local area and the national sample rates is unimportant. (Minority men from Birmingham are about 6.5% of the national subsample.)

For minority men and women, the estimated variance matrix of the local area (true) rates is

$$\begin{pmatrix} 19.7 & 25.1 \\ 25.1 & 72.0 \end{pmatrix};$$

the correlation of the rates ($25.1/\sqrt{(19.7 \times 72.0)} = 0.67$) indicates that bivariate shrinkage is not much more efficient than univariate shrinkage. For instance, the bivariate shrinkage estimates for Gateshead are 73.9% for men and 48.7% for women (sample size 8). The estimated standard error for men is 4.1% (reduced from 4.3% for univariate shrinkage), and the standard error for women is 7.8%, equal, after rounding, to the standard error for univariate shrinkage (the estimated standard error of the women's sample rate is 17.7%).

The estimation of the rates for minorities can be improved more effectively by incorporating information about white residents. The estimated variance matrix of the local area (true) rates for minority and white men is

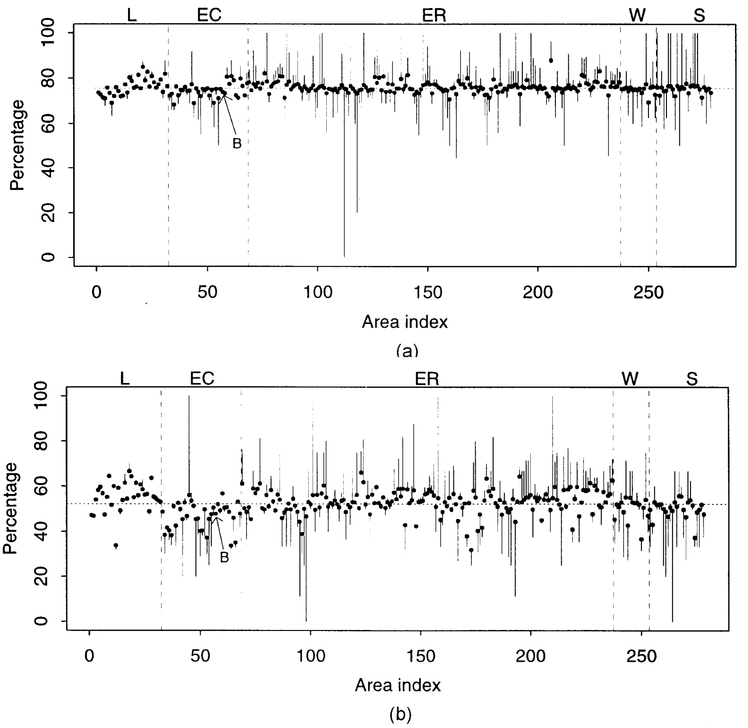


Fig. 4. Sample rates and shrinkage estimates of the local area rates of economic activity for (a) adult men and (b) adult women from all ethnic minorities: B with arrow, estimate for Birmingham, area 57

$$\begin{pmatrix} 19.7 & 15.2 \\ 15.2 & 14.5 \end{pmatrix}.$$

The national sample rates are $\hat{\mathbf{p}} = (75.5\%, 73.3\%)$, with standard errors (0.65%, 0.26%). Since the correlation of the local area rates is high (0.90), the more precisely estimated rates for white men are very informative about the rates for minorities. For instance, the estimated rate for minority men in Gateshead is 71.7%, with an estimated standard error of only 2.3%.

The economic activity rates for women can be estimated similarly, incorporating information about white women. The national sample rates for the minority and white women are (52.2%, 49.7%) with standard errors (1.15%, 0.27%). The estimated between-area variance matrix is

$$\begin{pmatrix} 72.0 & 16.8 \\ 16.8 & 15.3 \end{pmatrix}.$$

The moderate estimated correlation of the rates, 0.51, indicates that bivariate shrinkage is not much more efficient than univariate shrinkage. However, the improvement is still appreciable for the sparsely represented local areas. For instance, the bivariate shrinkage estimate for Gateshead (eight minority women in the SAR) is 48.3% with estimated standard error 6.9%, reduced from 7.8% for the univariate shrinkage.

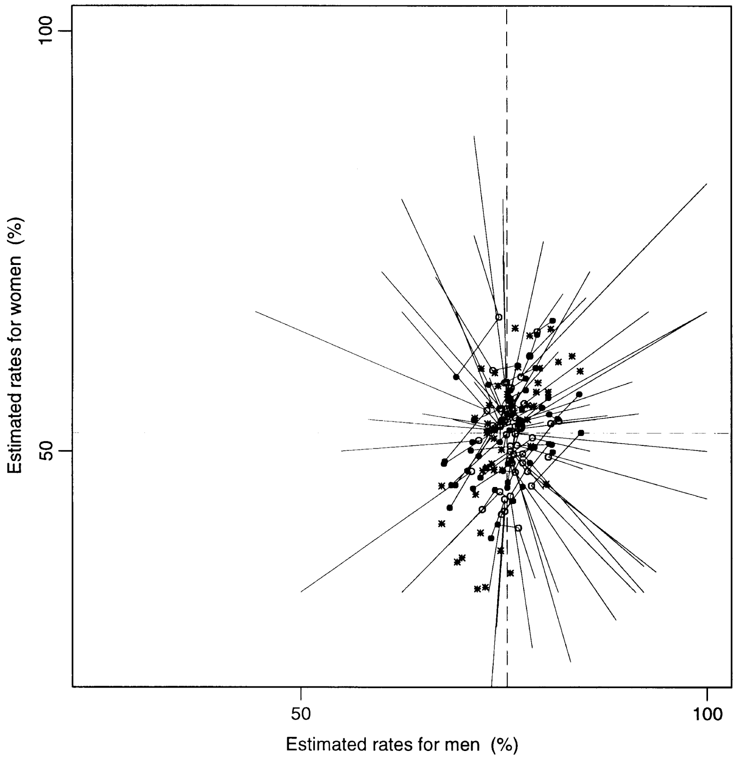


Fig. 5. Paths of shrinkage estimation of economic activity rates for adult men and women from all ethnic minorities; white men and white women respectively were used as auxiliary information in the bivariate shrinkage

The shrinkage paths of the (sample, univariate and bivariate shrinkage) estimates of the local area rates are plotted in Fig. 5. For clarity, the estimates are plotted only for every fourth area; for areas with more than 100 men in the SAR, for which shrinkage yields small adjustments, only the bivariate shrinkage estimates are plotted, using the symbol $*$. Unlike in Fig. 2, the bivariate shrinkage estimates are not aligned along a (regression) line because external information (about white residents) is used.

5. Discussion

In this section, we address two issues that arise in the application of shrinkage: the inflation of EMSE due to uncertainty about the variances σ^2 and v_l, and empirical validation of the method. In Section 5.3, we discuss extensions and an alternative approach (Aitkin, 1996a, b) based on the nonparametric likelihood.

5.1. Estimated σ^2

In both the theoretical development and in the examples, we have ignored the fact that the variances v_l and σ^2, or their matrix counterparts $\mathbf{V}_l$ and $\mathbf{\Sigma}$, are estimated. The inflation of EMSE in equation (3) due to using $\hat{\sigma}^2$ in place of σ^2 can be assessed by using the bootstrap (Efron and Tibshirani, 1994). Its implementation in our setting has the following steps. First, a random sample from the (approximate) sampling distribution of $\hat{\sigma}^2$ is generated. This is done by resampling, with replacement, from the local areas in the SAR, and evaluating the moment matching estimator $\hat{\sigma}^2$. In the second step, the shrinkage estimate $\tilde{p}_l$ is calculated for each value of $\hat{\sigma}^2$. The sample variance of these bootstrap values of $\tilde{p}_l$ is an estimate of the contribution to EMSE due to the sampling variation in $\hat{\sigma}^2$. We applied this procedure with a (bootstrap) sample size of 400.

In the examples discussed in Section 4, EMSE is inflated only slightly. For instance, for area 1 (the City of London and the City of Westminster, 45 men aged 16–19 years), the nominal standard error (ignoring the uncertainty about σ^2) is 3.89%, and it increases to 3.96% when the uncertainty about σ^2 is taken into account by the bootstrap procedure. However, the standard error of the sample rate $100\hat{p}_1$, 7.44%, is still much higher. For areas with greater subsample sizes, the correction of the EMSE is smaller because the shrinkage estimator gives greater weight to the within-area information. For instance, for area 67 (Leeds — 331 subjects), the standard error is increased from 2.282% to 2.284%, which is still smaller than the standard error of the sample rate, 2.63%.

The inflation of EMSE due to the uncertainty about v_l can be assessed similarly. In our examples it is much smaller than the inflation due to the uncertainty about σ^2 because v_l is a very flat function of the underlying rate p_l. Applying the bootstrap for the multivariate shrinkage involves no additional complexity. It yields similar conclusions about the inflation of EMSE, although the reduction in EMSE from the univariate to the bivariate shrinkage is much smaller than from the sample means to the univariate shrinkage.

5.2. Validation

The ideal form of validation would be to compare the alternative estimates with the true values. The area level rates of economic activity, based on the 1991 UK census, are available for all adult men and women, but not for the smaller subpopulations analysed in Section 4. Thus, one way of validating the properties of the shrinkage estimator would be to compare the estimates based on the SAR with the census-based values for all adult men and women. Such comparisons are not very useful because the subsample sizes of the adult men and women, 411 000 and 451 000 respectively, are so large that the adjustments due to shrinkage, $\tilde{p}_l - \hat{p}_l$, are very small. Nevertheless, SAR-based shrinkage estimates are a (small) improvement over the sample rates. The univariate shrinkage estimates for men are closer to the census-based values in 173 local areas (62%) and for women in 163 local areas (59%).

The validation would be more relevant if the area level sample sizes were much smaller, similar to those for the subpopulations of interest. This can be arranged by subsampling from the SAR data. For this, we drew independent 2% random samples of adult men and women from the SAR, yielding two 0.04% random samples from the census. The estimates of the economic activity rates based on these subsamples, comprising 8061 men and 9077 women respectively, are 72.1% and 49.6%. The estimated between-area standard deviations $100\hat{\sigma}$ are 3.48% for women and 4.98% for men. The univariate shrinkage estimates are closer to the census-based rates for 218 (78%) and 209 areas (75%) for men and women respectively. The bivariate shrinkage estimates are closer to the census-based rates for 225 and 215 areas for men and women respectively. The bivariate shrinkage estimates are closer to the census-based values than are the univariate shrinkage estimates in 154 (55%) and 149 areas (54%) for men and women respectively. Other discrepancy measures, such as the mean-squared deviation, are also smaller for the bivariate shrinkage, followed by the univariate shrinkage and the sample means.

The properties of the shrinkage estimator for large and non-symmetrically distributed probabilities p_l can be explored similarly. For illustration, we generated such probabilities from the square of the uniform distribution on (0.7, 1.0). This distribution has the mean $100p = 80\%$ and standard deviation $100\sigma = \surd 80\%$. The dichotomous outcome variable was simulated for the areas with the same area level subsample sizes as for minority women. The shrinkage estimator was closer to the true rate in 194 areas (70%). This percentage, as a random variable, has the approximate standard error $100\surd(0.21/278) = 2.75\%$.

5.3. Extensions and alternatives

The evaluation of the shrinkage estimator consists of two steps: estimation of the between-area variance (matrix) and combining the local area and national estimators. Both steps can be altered. For instance, instead of the independent and identically distributed pattern of deviations from the national rate a spatial correlation pattern of the local area rates can be defined and its parameters estimated. For such a pattern, it is more natural to combine the local area rate with the rates of its neighbouring areas in addition to the national rate. Defining the neighbourhood of a local area may be problematic because the accessibility by the means of transport (in London in particular) is more relevant than geographical proximity. Also, the association with neighbours is likely to be very different in small densely populated urban areas and in large sparsely populated rural areas.

Even if the rates of the local areas vary greatly, this variation may be substantially reduced by regressing on one or a few covariates. If the regression is well determined the resulting (regression) shrinkage estimator is more efficient. However, such an estimator requires the local area means of the covariates, which are themselves estimated. As a consequence, a better fitting model which reduces the between-area variance does not necessarily yield a more efficient regression shrinkage estimator. See Longford (1996) for an example with normally distributed outcomes.

For a regression with dichotomous outcomes, the identity link is inappropriate, although the quadratic loss function for logits (or probits) may be equally inappropriate, depending on the substance of the problem. Our approach can be generalized by replacing the quadratic loss with a more complex function for which EMSE can be expressed analytically. Various fully parametric (generalized linear mixed) models can be applied, but their assumption of a given family of mixture distributions is often too restrictive. Even when there are many local areas, as in our examples, the data contain only modest information about the between-area (mixture) distribution, and so its parsimonious modelling is paramount. In the non-parametric maximum likelihood approach of Aitkin (1996a, b), the mixture distribution is estimated by a distribution with a finite support. The advantages of this approach are that it draws on the powerful theory of generalized linear models, naturally accommodates adjustment for covariates and makes no assumptions about the mixture distribution. A disadvantage of the nonparametric maximum likelihood method is that the estimated mass points and weights do not have a natural interpretation and they cannot be related to the gains in efficiency in as simple a way as the between-area variance (matrix) in shrinkage estimation.

6. Conclusions: using a sample of anonymized records based on the 2001 UK census

The shrinkage estimator derived and illustrated in this paper is an application of the general idea of pooling information across similar units. The similarity is measured by the between-unit variance (matrix) which plays a pivotal role in the estimation. Univariate shrinkage is computationally very simple; multivariate shrinkage provides a further improvement, especially when the rates underlying the additional components are highly correlated and are estimated with precision. The method entails no distributional assumptions. In multivariate shrinkage, the choice of the components (sources of auxiliary information) is up to the analyst. In the examples discussed in Section 4, bivariate shrinkage is superior to univariate shrinkage, but pooling information across more than two components leads to only a slight improvement.

The examples dealt with the estimation of area level economic activity rates for subpopulations of adult residents. Since they refer to 1991, the date of the last UK census, they are of diminishing practical utility. However, they show how the SARs drawn from the 2001 UK census could be used more effectively. Particularly attractive is the idea of incorporating information from the 1991 SAR to improve estimation based on the 2001 SAR. This will be successful if the area level deviations from the general trend over the intervening decade are small. Of course, if the census-based rates for all subpopulations of interest are made available, shrinkage does not have to be applied. Such detailed local area information is unlikely to be published because confidentiality of the census data might be breached.

Acknowledgements

The research described in this paper was suggested to me by Ludi Simpson. He assisted in the retrieval of the data and offered me support and stimulus throughout the research. The Census Microdata Unit at Manchester University enabled me to gain access to the SAR database. Robert Alston, Louise Linsell, Ludi Simpson and two referees provided stimulating comments and suggested improvements on previous versions of the manuscript. Funding by research grant R999B60005 from the National Center for Educational Statistics, USA, is acknowledged.

Appendix A

This appendix derives the moment matching estimator of the between-area variance σ^2 for dichotomous outcomes conditionally independent within local areas. The proof of an identity for shrinkage estimation of binomial variances is outlined in Appendix A.1. All programming was done in S-PLUS (Statistical Sciences, 1993). The code (S-PLUS functions) is available from the author on request.

The unconditional variances of the local area sample rates $\hat{p}_l$, $l = 1, \ldots, L$, are

$$\begin{aligned}
\operatorname{var}(\hat{p}_l) &= \frac{1}{n_l^2} E_s\left\{\operatorname{var}_1\left(\sum_{i=1}^{n_l} y_{il}|\delta_l\right)\right\} + \frac{1}{n_l^2} \operatorname{var}_s\left\{E_1\left(\sum_{i=1}^{n_l} y_{il}|\delta_l\right)\right\} \\
&= \frac{1}{n_l} E_s\{(p+\delta_l)(1-p-\delta_l)\} + \operatorname{var}_s(p+\delta_l) \\
&= \frac{p(1-p)}{n_l} + \frac{n_l - 1}{n_l}\sigma^2 \\
&= v_l + \frac{n_l - 1}{n_l}\sigma^2.
\end{aligned}$$

Weighted versions of this formula apply when the subsampling within local areas is with unequal probabilities. Assuming independence of the $\hat{p}_l$, the variance of the national proportion $\hat{p} = (n_1\hat{p}_1 + \ldots + n_L\hat{p}_L)/N$ is

$$\mathrm{var}(\hat{p}) = \frac{p(1-p)}{N^2} \sum_{l=1}^{L} n_l + \frac{\sigma^2}{N^2} \sum_{l=1}^{L} n_l(n_l - 1)$$
$$= \frac{p(1-p)}{N} + \frac{\sigma^2}{N}(M-1),$$

where $M = (n_1^2 + \ldots + n_L^2)/N$. The between-area variance σ^2 is estimated by matching the between-area sum of squares

$$S_\mathrm{b} = \sum_l n_l (\hat{p}_l - \hat{p})^2 \tag{11}$$

to its expectation. This leads to a simple linear equation; see below. Note that S_b has a continuum of alternative definitions based on different weights (e.g. 1 instead of n_l) and different ways of averaging (e.g. $\tilde{p} = (\hat{p}_1 + \ldots + \hat{p}_L)/L$ instead of $\hat{p}$). The expectation of a summand of S_b in equation (11) is

$$E\{(\hat{p}_l - \hat{p})^2\} = \mathrm{var}(\hat{p}_l - \hat{p})$$
$$= \mathrm{var}(\hat{p}_l) - \frac{2n_l}{N}\,\mathrm{var}(\hat{p}_l) + \mathrm{var}(\hat{p}),$$

and so

$$E(S_\mathrm{b}) = \sum_l \left(1 - \frac{2n_l}{N}\right) n_l\,\mathrm{var}(\hat{p}_l) + N\,\mathrm{var}(\hat{p})$$
$$= (L-1)p(1-p) + \sigma^2(N - M - L + 1).$$

Hence, the solution of the moment matching equation $S_\mathrm{b} = E(S_\mathrm{b}; \sigma^2)$ is

$$\hat{\sigma}^2 = \frac{S_\mathrm{b} - (L-1)\hat{p}(1-\hat{p})}{N - M - L + 1}. \tag{12}$$

Since σ^2 is required only as part of the total variance $\mathrm{var}(\hat{p}) + \sigma^2$, instead of using equation (12) the total can be estimated directly as

$$\widehat{\mathrm{var}}(\hat{p}) + \hat{\sigma}^2 = \frac{S_\mathrm{b}}{N}\,\frac{N + M - 1}{N - M - L + 1} - \frac{\hat{p}(1-\hat{p})}{N}\,\frac{N(L-2) + LM}{N - M - L + 1}. \tag{13}$$

On the one hand, σ^2 is of interest as a measure of between-area variation; on the other hand, equation (13) is useful for discussing the stability of the coefficients $\hat{b}_l^*$.

For two non-overlapping subpopulations,

$$\mathrm{cov}_\mathrm{s}(p_l^{(1)}, p_l^{(2)}) = E_1\{\mathrm{cov}_\mathrm{s}(\hat{p}_l^{(1)}, \hat{p}_l^{(2)})\}.$$

Hence, the between-area covariance of the rates, an off-diagonal element of $\boldsymbol{\Sigma}$, is estimated by the sample covariance of the sample rates $\hat{p}_l^{(1)}$ and $\hat{p}_l^{(2)}$.

A.1. Shrinkage estimation of sampling variances

The shrinkage method can be also applied to improve the estimation of an ensemble of sampling variances $v_l = p_l(1 - p_l)/n_l$. The approach requires an expression for the sampling variance of the sample estimator of $\hat{v}_l$. For $\hat{v}_l = \hat{p}_l(1 - \hat{p}_l)/n_l$ we have the identity

$$\mathrm{var}(\hat{v}_l) = \frac{p_l(1-p_l)(n_l - 1)}{n_l^3}\{(n_l - 1)(1 - 2p_l)^2 + 2p_l(1 - p_l)\}.$$

The proof of this identity is elementary, though tedious. It is based on the identity

$$E\{P(X, k)\} = p^k\, P(n, k),$$

where $P(x, k) = x!/(x-k)!$, for a binomially distributed random variable $X \sim \mathrm{bin}(n, p)$.

References

Aitkin, M. (1996a) Empirical Bayes shrinkage using posterior random effect means from non-parametric maximum likelihood estimation in general random-effect models. In *Statistical Modelling: Proc. 11th Int. Wrkshp Statistical Modelling* (eds A. Forcina, G. M. Marchetti, R. Hatzinger and G. Galmacci), pp. 87–94. New York: Springer.

——— (1996b) A general maximum likelihood analysis of overdispersion in generalized linear models. *Statist. Comput.*, **6**, 251–262.

Efron, B. (1996) Empirical Bayes methods for combining likelihoods. *J. Am. Statist. Ass.*, **91**, 538–565.

Efron, B. and Tibshirani, R. J. (1994) *An Introduction to the Bootstrap*. New York: Chapman and Hall.

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

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

Heath, S. (1994) *A User Guide to the SARs: Samples of Anonymised Records from the 1991 Census of Population of Great Britain*, 2nd edn. Manchester: Census Microdata Unit.

James, W. and Stein, C. (1961) Estimation with quadratic loss. In *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, vol. 1, pp. 361–379. Berkeley: University of California Press.

Local Government Management Board and London Research Centre (1992) *SASPAC User Manual*, part 2. London: Local Government Management Board and London Research Centre.

Longford, N. T. (1995) *Models for Uncertainty in Educational Testing*. New York: Springer.

——— (1996) Small-area estimation using adjustment by covariates. *Qüestió*, **20**, 187–212.

——— (1997) Shrinkage estimation of linear combinations of true scores. *Psychometrika*, **62**, 237–244.

Shen, W. and Louis, T. A. (1998) Triple-goal estimates in two-stage hierarchical models. *J. R. Statist. Soc.* B, **60**, 455–471.

Singh, A. C., Stukel, D. M. and Pfeffermann, D. (1998) Bayesian *versus* frequentist measures of error in small area estimation. *J. R. Statist. Soc.* B, **60**, 377–396.

Statistical Sciences (1993) *S-PLUS Guide to Statistical and Mathematical Analysis, Version 3.2*. Seattle: MathSoft.

Thomas, N. S., Longford, N. T. and Rolph, J. E. (1994) Empirical Bayes methods for estimating hospital specific mortality rates. *Statist. Med.*, **13**, 889–903.