Social Statistics Research Centre
University of Southampton

Minimum Variance Stratification of a Finite Population

Dan Hedlin

Abstract

This paper considers the combined problem of allocation and stratification in order to minimise the variance of the expansion estimator of a total, taking into account that the population is finite. The proof of necessary minimum variance conditions utilises the Kuhn-Tucker Theorem. Stratified simple random sampling with non-negligible sampling fractions is an important design in sample surveys. We go beyond limiting assumptions that have often been used in the past, such as that the stratification equals the study variable or that the sampling fractions are small. We discuss what difference the sampling fractions will make for stratification. In particular, in many surveys the sampling fraction equals one for some strata. The main theorem of this paper is applied to two populations with different characteristics, one of them being a business population and the other one a small population of 284 Swedish municipalities. We study empirically the sensitivity of deviations from the optimal solution.

SSRC Methodology Working Paper M03/07

Minimum Variance Stratification of a Finite Population

DAN HEDLIN[1]

ABSTRACT

This paper considers the combined problem of allocation and stratification in order to minimise the variance of the expansion estimator of a total, taking into account that the population is finite. The proof of necessary minimum variance conditions utilises the Kuhn-Tucker Theorem. Stratified simple random sampling with non-negligible sampling fractions is an important design in sample surveys. We go beyond limiting assumptions that have often been used in the past, such as that the stratification equals the study variable or that the sampling fractions are small. We discuss what difference the sampling fractions will make for stratification. In particular, in many surveys the sampling fraction equals one for some strata. The main theorem of this paper is applied to two populations with different characteristics, one of them being a business population and the other one a small population of 284 Swedish municipalities. We study empirically the sensitivity of deviations from the optimal solution.

KEY WORDS: Optimal stratification; certainty stratum; take-all stratum; self-representing stratum; skewed population; business sample survey.

[1] Department of Social Statistics; University of Southampton;Southampton, SO17 1BJ; United Kingdom; E-mail: deh@sosci.soton.ac.uk

1. INTRODUCTION

It is essential in surveys to minimise the sample size because of costs involved. In official statistics it is also required to keep the response burden down. Stratification is a widely used sample survey technique that serves many purposes, one of them being to improve precision or to reduce the sample size. The sampling frame is divided into strata, and independent samples are drawn from each stratum without replacement. For example, the most widely used design in business surveys is stratified simple random sampling, where the population is divided into, for example, subpopulations according to industry. Each subpopulation is stratified by size, say by employment. We focus on size stratification and we use the term population with the meaning subpopulation in the sense just described. For highly skewed populations with a small number of extremely influential units, the size stratum with the largest units is typically a *certainty stratum* (also called self-representing, complete enumeration or take-all stratum) where all units are selected for observation. Other strata in the population are *genuine sampling strata*. This type of design is particularly common in business surveys and other establishment surveys. In practice, the stratum boundaries are often determined by univariate stratification with one continuous stratification variable, where the objective function is usually the estimator variance of one important study variable. Practitioners often use the cum$\sqrt{f}$ rule (Dalenius and Hodges 1959), which assumes that the sampling fractions are negligible. As noted above, this is not a suitable assumption for highly skewed populations. Further, the Dalenius-Hodges rule assumes that the stratification variable is the same as the study variable, which is either unrealistic or, if the two variables are indeed similar, makes stratification almost superfluous as such a powerful auxiliary variable could be used in estimation instead.

Several issues have to be addressed when designing a stratified sample (c.f. Särndal, Swensson and Wretman 1992, p. 101):

Construction of Strata: (A1) Which stratification variable(s) is (are) to be used? (A2) How many strata should there be? (A3) How should strata be demarcated? Choice of Sampling and Estimation Methods: (B1) Sampling design for each stratum. (B2) An estimator for each stratum. (B3) The sample size for each stratum.

This paper focuses on questions A3 and B3 jointly. As set answers to the other questions we assume that (A1) there is a frame with known values of a given stratification variable for every unit; (A2) the number of strata, H, is predetermined; (B1) a simple random sample is drawn from each stratum; (B2) the expansion estimator is used for each stratum. As for B3, we fix the overall sample size to be a predetermined number, n, but the allocation of this sample to strata is determined as part of the optimisation problem.

First we put the problem into its context. For a more comprehensive overview of stratification, see Sigman and Monsour (1995). The population $U = \{1,2,\ldots N\}$ with study variable $\mathbf{y} = (y_1, y_2, \ldots y_N)'$ is stratified and a sample is taken in order to estimate the population total $t = y_1 + y_2 + \ldots + y_N$. Consider the expansion estimator of the total of **y**:

$$\hat{t}_y = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k=1}^{n_h} y_k \ , \qquad (1)$$

where N_h and n_h are the number of frame units in stratum h and the sample size in stratum h, respectively. The problem considered here is to find the univariate stratification that minimises the variance of $\hat{t}_y$,

$$Var(\hat{t}_y) = \sum_{h=1}^{H} N_h^2 \frac{S_{yh}^2}{n_h}\left(1 - \frac{n_h}{N_h}\right), \qquad (2)$$

where $S_{yh}^2 = \sum_{k=1}^{N_h} (y_k - \bar{y}_h)^2 / (n_h - 1)$ is the study variable variance in stratum h, with $\bar{y}_h$ being the mean of the y_k in stratum h. The quantities N_h and S_{yh}^2 are functions of the stratum boundaries. The objective function, $Var(\hat{t}_y)$, is here regarded as a function of the stratum boundaries and the stratum sample sizes. We minimise it under the constraints that the sample sizes add up to n and that each stratum sample size is no greater than the stratum population size.

Dalenius (1950) minimises

$$v(\hat{t}_x) = \sum_{h=1}^{H} N_h^2 \frac{S_{xh}^2}{n_h}, \qquad (3)$$

where S_{xh}^2 is the stratification variable variance in stratum h. Unlike (2), Dalenius presupposes that $n_h = nN_h S_{xh} \Big/ \sum_{h=1}^{H} N_h S_{xh}$ is Neyman allocation under the assumption that $S_{xh} \approx S_{yh}$. Dalenius derives the following equations as a necessary condition for stratum boundaries $b_1 < b_2 < \ldots < b_{H-1}$ minimising (3):

$$\frac{S_{xh}^2 + (b_h - \bar{x}_h)^2}{S_{xh}} = \frac{S_{x,h+1}^2 + (b_h - \bar{x}_{h+1})^2}{S_{x,h+1}}, \; h = 1, 2, \ldots H-1, \qquad (4)$$

where $\bar{x}_h$ is the mean of the stratification variable in stratum h. Schneeberger (1985) points out that a solution to (4) is not necessarily a local or global minimum to (3). There may be, for example, two solutions, one minimum and one maximum. The function $v(\hat{t}_x)$ approximates (2) under the following assumption and approximation.

Assumption A1.a. The values of the study variable equal those of the stratification variable.

Approximation 1. The finite population correction in (2) is ignored.

In this paper we do not use Approximation 1 in any theorem. It is intriguing that when Approximation 1 is dropped, the optimal conditions remain similar to (4) but finite population corrections will emerge, as shown in Theorem 1 below. Thus, this problem is in a sense parallel to many other problems in survey sampling: you obtain formulae for finite populations by inserting finite population corrections at appropriate places in the corresponding formulae for infinite populations. Papers dealing with optimal stratification that use Approximation 1 include Dalenius (1950), Ekman (1959), Dalenius and Hodges (1959), Sethi (1963), Serfling (1968) and Mehta *et al.* (1996). Like Dalenius (1950), we use the following approximation.

Approximation 2. The finite population is approximated with a continuous distribution.

Several authors have addressed the problem of finding the point where the far tail of a skewed distribution should be cut off to form a certainty stratum. Dalenius (1952), Glasser (1962) and Hidiroglou (1986) have solved the two-stratum special case, with one certainty stratum and one genuine sampling stratum. These results are not easily generalised to more than two strata. Although Glasser derives an exact result, as opposed to Dalenius who uses Approximation 2, they arrive at essentially the same condition for the stratum boundary b_1:

$$(b_1 - \bar{x}_1)^2 = \frac{N_1 S_1^2}{n_1}, \qquad (5)$$

where unity as subscript refers to stratum 1, which is the genuine sampling stratum. We generalise this result to an arbitrary but predetermined number of strata.

Thus, this paper generalises the previous results in two ways. Further, we solve the *combined* problem of finding the optimal allocation and optimal stratification when there are several genuine sampling strata and one certainty stratum. Although this can be solved in two steps (first by finding an optimal allocation and then by finding the optimal stratification given the allocation), it is still of theoretical interest that the two steps can be solved simultaneously. Condition (5) turns out to be a special case of our results, whereas (4) does not.

An algorithm given by Lavallée and Hidiroglou (1988) and Hidiroglou and Srinath (1993) minimises the sample size under a precision constraint rather than the other way round. Detlefsen and Veum (1991), Sweet and Sigman (1995), and Slanta and Krenzke (1996) discuss convergence problems of the algorithm and how to implement it. Unlike this algorithm, we do not have any predetermined allocation scheme and there will be no convergence problems, unless a very large proportion of the units in the frame have the same value of the stratification variable.

Baxi (1995) proposes an algorithm for an approximately optimal stratification where one unit is sampled in each stratum. The finite population correction is not ignored.

We obtain further results under an assumption less restrictive than Assumption A1.a:

Assumption A1.b. A stochastic relationship between a superpopulation study variable *Y* and a stratification variable *X* holds. We shall show results under *Model 1a*: $Y = a + bX + e_x$, where a and b are constants, and the e_x are uncorrelated errors with zero mean and variance $s^2 x^g$, for some constants s^2 and g. These results can be extended to *Model 1:* $Y = y(X) + e_x$, where $y(\cdot)$ is a known function and e_x has a general variance structure. The current paper is the first one to obtain the minimum variance of the expansion estimator under Model 1a without relying on Approximation 1.

The optimal number of strata is not discussed here. See Serfling (1968) and Singh (1971), both of which draw on Approximation 1. Discussions of other designs and estimators than stratified simple random sampling and the expansion estimator include Wright (1983) who uses the auxiliary $\mathbf{x} = (x_1, x_2, \ldots x_N)'$ in both the design and estimation stage, the latter with a GREG estimator under a special case of Model 1. Addressing both A3 and B3 Wright finds the allocation and stratification that minimise the anticipated variance (the variance under both the model and the design). The method of Wright is also described in Särndal, Swensson, Wretman (1992, sec. 12.4). Pandher (1996) uses a GREG too, but with only two strata. Another model-based approach is Unnithan and Nair (1995).

Sections 2 and 3 state conditions for stratum boundaries minimising the variance. Applications are presented in section 4. Concluding remarks are given in Section 5.

2. A SOLUTION UNDER ASSUMPTION A1.A

We disregard nonsampling errors, that is nonresponse, measurement and coverage errors and assume that every population unit corresponds to exactly one frame unit. The strata are determined by stratum boundary points $b_1 < b_2 < \ldots < b_{H-1}$ with strata defined as $A_1 = \{u : x_u \le b_1\}$, $A_h = \{u : b_{h-1} < x_u \le b_h\}$, $h = 2,3,\ldots,H-1$, and $A_H = \{u : b_{H-1} < x_u\}$, where **x** is the stratification variable. Set $b_0 = x_1$ and $b_H = x_N$. We seek values of $(\mathbf{n},\mathbf{b}) = (n_1, n_2, \ldots, n_H, b_1, b_2, \ldots, b_{H-1})'$ that minimise (2) under the following constraints:

$$\begin{cases} g_h(\mathbf{n},\mathbf{b}) \equiv n_h - N_h \le 0 \ , \quad h = 1, 2\ldots H \\ g_{H+1}(\mathbf{n},\mathbf{b}) \equiv \sum_{h=1}^{H} n_h - n \le 0 \end{cases} \tag{6}$$

Note that these constraints allow any stratum to be a certainty stratum. As a useful special case the constraints will be further restricted:

$$\begin{cases} g_h(\mathbf{n},\mathbf{b}) \equiv n_h - N_h < 0 \ , \quad h = 1, 2\ldots H-1 \\ g_H(\mathbf{n},\mathbf{b}) \equiv n_H - N_H = 0 \\ g_{H+1}(\mathbf{n},\mathbf{b}) \equiv \sum_{h=1}^{H} n_h - n = 0 \end{cases} \tag{7}$$

We give now a framework that will allow us to apply optimisation theory for continuous functions. The framework can either be seen as a superpopulation model or simply as an approximation approach. In this section we adopt the latter viewpoint, which was introduced above as Approximation 2. Let x_1 and x_N be a priori known lower and upper bounds for the values of X with density $f_X(x)$. We will need three properties of the strata: probability, mean and variance. Let P_h denote the probability that X falls in stratum h:

$$P_h = \int_{b_{h-1}}^{b_h} f_X(x)\,dx \qquad (8)$$

The conditional mean and variance of X given $X \in (b_{h-1}, b_h)$ are:

$$m_{xh} = \int_{b_{h-1}}^{b_h} x f_X(x)\,dx \qquad (9)$$

$$s_{xh}^2 = \int_{b_{h-1}}^{b_h} (x - m_{xh})^2 f_X(x)\,dx \qquad (10)$$

Under the approximation approach, the integer N_h and the finite population mean $\bar{x}_h$ and variance S_{xh}^2 (which equals S_{yh}^2 under Assumption A1.a) are assumed approximately equal to NP_h, m_{xh} and s_{xh}^2, respectively. We will denote NP_h by $N_h(\mathbf{b})$ or just N_h. Thus N_h is regarded as a continuous function of the stratum boundaries. We also treat $n_1, n_2, \ldots n_H$ as continuous variables. The function (2) is then approximated by

$$f(\mathbf{n},\mathbf{b}) = \sum_{h=1}^{H} N_h^2(\mathbf{b}) \frac{s_{xh}^2(\mathbf{b})}{n_h}\left(1 - \frac{n_h}{N_h(\mathbf{b})}\right) \qquad (11)$$

For notational simplicity, we will in the sequel drop the argument **b** in the functions $N_h(\mathbf{b})$ and other functions of the stratum boundaries.

Lemma 1 gives an optimum under constraints (6), whereas Theorem 1 gives an optimum under the more restricted constraints (7). The proofs are in Appendix A and B.

Lemma 1. Suppose $f_X(x) > 0$ on (x_1, x_N). If a stratification and allocation have a local minimum of (11) under constraints (6) with at least two genuine sampling strata, then (12) and (13) are satisfied:

$$\frac{n_h}{n_j} = \frac{N_h s_{xh}}{N_j s_{xj}} \quad \forall h \text{ and } j \text{ where } n_h < N_h \text{ and } n_j < N_j \tag{12}$$

$$(b_h - m_{xh})^2 \left(\frac{N_h}{n_h} - 1 \right) + \frac{N_h}{n_h} s_{xh}^2 -$$

$$(b_h - m_{x,h+1})^2 \left(\frac{N_{h+1}}{n_{h+1}} - 1 \right) - \frac{N_{h+1}}{n_{h+1}} s_{x,h+1}^2 + (l_{h+1} - l_h) = 0, \tag{13}$$

$h = 1, 2 \ldots H-1$, for some non-negative real numbers l_h and l_{h+1}. The nature of l_h and l_{h+1} is discussed in Appendix A.

Theorem 1. Suppose strata 1, 2, ... H–1 are predetermined to be genuine sampling strata and stratum H is predetermined to be a certainty stratum. Then, if $f_X(x) > 0$ on (x_1, x_N), a necessary condition for a local minimum of (11) with respect to stratum sample sizes and stratum boundaries under constraints (7) is the system of equations (14), (15) and (16) below.

Conditions for stratum sample sizes:

$$n_h = (n - N_H) N_h s_{xh} \left(\sum_{h=1}^{H-1} N_h s_{xh} \right)^{-1}, \quad h = 1, 2, 3 \ldots H-1. \tag{14}$$

Conditions for the boundaries $b_1, b_2, \ldots b_{H-2}$ of the genuine sampling strata:

$$\frac{(b_h - m_{xh})^2\left(1-\frac{n_h}{N_h}\right)+s_{xh}^2}{s_{xh}} = \frac{(b_h - m_{x,h+1})^2\left(1-\frac{n_{h+1}}{N_{h+1}}\right)+s_{x,h+1}^2}{s_{x,h+1}}, \quad (15)$$

$h = 1, 2, 3 \ldots H–2$.

Condition for the boundary b_{H-1} of the certainty stratum:

$$(b_{H-1} - m_{x,H-1})^2 = \frac{N_{H-1}}{n_{H-1}} s_{x,H-1}^2. \quad (16)$$

Remarks:

1. This paper does not attempt to provide any sufficient condition for a local minimum.

2. Equation (14) is Neyman allocation when stratum H is a certainty stratum.

3. Finite population correction factors of the type $1-n/N$ are often seen in survey sampling theory. Interestingly, this problem is no exception: the proper finite population result (15) is obtained by inserting finite population corrections at appropriate places in the corresponding formula valid for an infinite population, (4).

4. Equation (16) with $H = 2$ is equivalent to (5).

5. When applying Theorem 1 in a practical situation, the unknown superpopulation parameters m_{xh} and s_{xh}^2 must be estimated or guessed by the corresponding parameters of the finite population and the values of n_h and N_h have to be rounded to nearest integer.

2.1. The special condition for certainty strata

What is the difference between (15) and (16) in Theorem 1? It may be expressed this way. Suppose you stratify by using a condition fairly close to (15), like the cum$\sqrt{f}$ rule, using this rule for all strata. Then you allocate the sample and end up with $n_H = N_H$, what have you done? This approach corresponds to $h = H–1$ and $l_{H-1} = l_H = 0$ in (13) in Lemma 1, as shown in Appendix A. Compare this with an approach where strata 1, 2, ... H–1 are predetermined genuine sampling strata and stratum H may or may not be a certainty stratum. Then, in (13) with $h = H–1$, we have $l_{H-1} = 0$ and $l_H \geq 0$. Thus the absence of l_H in the first approach tend to make either stratum H too narrow or at least one of the other strata too wide.

3. A SOLUTION UNDER ASSUMPTION A1.B

Theorem 1 is now generalised to Model 1a under Assumption A1.b. Under this superpopulation model we have

$$s_y^2 = \int_{x_1}^{x_N} \int_{-\infty}^{\infty} \left[bx + s^2 x^g - bm_x\right]^2 f_{e|x}\left(e|X\right) f_X\left(x\right) de\, dx,$$

where s_y^2 is the variance of Y. We shall use similar notation for all moments of Y and X. Calculating the integral term by term, we obtain

$$s_y^2 = b^2 s_x^2 + \overline{s_e^2},$$

where $\overline{s_e^2}$ is the mean of the conditional variances of e_x, given X:

$\overline{s_e^2} = s^2 \int_{x_1}^{x_N} x^g f_X\left(x\right) dx$. Using the anticipated variance as the measure of effectiveness, the objective function to be minimised is

$$E_M Var\left(\hat{t}_y\right) = \sum_{h=1}^{H} N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left[b^2 s_{xh}^2 + \overline{s_{eh}^2}\right], \qquad (17)$$

where E_M denotes expectation under the model, *Var* is as previously the variance over all possible samples, and $\overline{s}_{eh}^{\,2} = \frac{s^2}{P_h} \int_{b_{h-1}}^{b_h} x^g f_X(x)\,dx$. We state Theorem 2 without proof, as it is a straightforward extension of that of Theorem 1.

Theorem 2. Suppose strata 1, 2, ..., H–1 are predetermined genuine sampling strata and stratum H is a predetermined certainty stratum. Suppose further that Model 1a holds and that $f_X(x) > 0$, $x \in (x_1, x_N)$. Then, a necessary condition for a local minimum of (17) with respect to stratum sample sizes and stratum boundaries under constraints (7) is the system of equations (18), (19) and (20).

Condition for stratum sample sizes:

$$n_h = (n - N_H) N_h \sqrt{s_{xh}^{\,2} + \overline{s}_{eh}^{\,2} b^{-2}} \left(\sum_{h=1}^{H-1} N_h \sqrt{s_{xh}^{\,2} + \overline{s}_{eh}^{\,2} b^{-2}} \right)^{-1}, \qquad (18)$$

Conditions for the boundaries $b_1, b_2, \ldots b_{H-2}$ of the genuine sampling strata:

$$\frac{\left[(b_h - m_{xh})^2 + s^2 b_h^g b^{-2}\right](1 - n_h/N_h) + s_{xh}^{\,2} + \overline{s}_{eh}^{\,2} b^{-2}}{\left(s_{xh}^{\,2} + \overline{s}_{eh}^{\,2} b^{-2}\right)^{-0.5}} = \qquad (19)$$

$$\frac{\left[(b_h - m_{x,h+1})^2 + s^2 b_h^g b^{-2}\right](1 - n_{h+1}/N_{h+1}) + s_{x,h+1}^{\,2} + \overline{s}_{e,h+1}^{\,2} b^{-2}}{\left(s_{x,h+1}^{\,2} + \overline{s}_{e,h+1}^{\,2} b^{-2}\right)^{-0.5}},$$

$h = 1, 2 \ldots H - 2$,

Condition for the boundary b_{H-1} of the certainty stratum:

$$\left(b_{H-1} - m_{x,H-1}\right)^2 + s^{2} b^{g}_{H-1} b^{-2} = \frac{N_{H-1}}{n_{H-1}} \left(s^{2}_{x,H-1} + \overline{s}^{2}_{e,H-1} b^{-2}\right). \qquad (20)$$

Remarks:

1. Equation (18) is Neyman allocation under Model 1a. It is a special case of the optimal allocation scheme shown by Serfling (1968) and Singh (1971) who minimises the variance under Model 1 and Approximations 1 and 2.

2. If $1 - n_h / N_h = 1 - n_{h+1} / N_{h+1} = 1$, (19) is a special case of a condition given by Dalenius and Gurney (1951). They, too, use Model 1 and Approximations 1 and 2.

3.1. Do we need assumption A1.b?

Now we consider heuristically the difference between the conditions (18) – (20) and the parallel conditions (14) – (16) in Theorem 1. To make the comparison more transparent we shall only consider the homoscedastic special case of Model 1 with $g = 0$, which makes $\overline{s}^{2}_{eh} = s^{2}, \ \forall h$. Then the difference between the conditions is additive constants involving $s^{2} b^{-2}$ which are grossed by factors containing N_h. If $N_h s^{2} b^{-2}$ is negligible compared to s^{2}_{xh}, h = 1, 2... H–1, and probably therefore also negligible to $\left(b_h - m_{xh}\right)^2$, the optimal stratification could be done according to Theorem 1, without having to rely on Model 1. There is a close relationship between $s^{2} b^{-2}$ and r_{xy} (now suppressing subscript *h)*. We have $s_{xy} = b s^{2}_{x}$ and $s^{2}_{y} = b^{2} s^{2}_{x} + s^{2}$ under Model 1a. It is easily shown that $s^{2} b^{-2} / s^{2}_{x} = \left(1 - r^{2}_{xy}\right) / r^{2}_{xy}$. Hence a stratification satisfying the conditions in Theorem 2 is not close to a stratification done according to Theorem 1, unless r_{xy} is high. In case of heteroscedasticity, the stratifications can be quite different even if the correlation is high.

4. APPLICATIONS

In this section we give some numerical illustrations of the results obtained in section 2. Applications under Assumption A1.a are of interest, although they may be unrealistic, because a comparison of methods using this assumption provides a more critical test of their performances than Assumption A1.b. Further, as Theorem 1 was derived under Approximation 2, it is interesting to see if there exists a stratification with even lower variance than one given by this theorem. We apply the results to two populations.

The annual census of Swedish manufacturing industry collects data on sales, cost of materials, energy used in the production process, etc, for all businesses above a certain employment level. The census together with derived variables, such as value added, is frequently used as a sampling frame for other surveys. We applied our results to the 1989 frame with value added as stratification variable. The frame here referred to as the *value added population*, contains 7326 units and its skewness is 12.4 (which could be compared with skewness 2.0 of an exponential distribution). The population was divided into $H = 4$ strata. The stratum comprising units with the largest values of the stratification variable was a certainty stratum, the other strata were genuine sampling strata. The sample size was set to 400.

The dataset MU284 contains data on Swedish municipalities. It can be found in Särndal, Swensson, Wretman (1992; Appendix B) and in StatLib (http://lib.stat.cmu.edu) submitted to StatLib by Esbjörn Ohlsson. There are 284 municipalities in Sweden, which makes this dataset a small one. We consider the variable P75, which is the 1975 population (in thousands). Since the P75 variable has only 68 distinct values, it is not a variable you would think of as well approximable by a continuous distribution. For MU284, the sample size was set to 80.

4.1. Performance measure

We searched for the stratification with the smallest estimator variance (2), which we refer to as the *best possible stratification*. We let the maximum x-value of each stratum be the stratum boundary. Clearly, as we now consider a specific situation, with specified values of **x**, sample size n and number of strata H, there exists a best possible stratification (a global minimum). $Var(\hat{t}_y)$ was computed for a large number of combinations of the stratum sizes N_1, N_2 and N_3. We do not, however, give a full account of the search method here. We denote the estimator variance for a specific stratification by $Var(\hat{t}_y;\mathbf{N})$, where $\mathbf{N} = (N_1,\ldots,N_H)$, $H = 3$. The *variance ratio* is defined as the ratio of the estimator variance obtained by a particular stratification and the estimator variance using the best possible stratification.

The best possible stratification of the value added population is shown in Table 1. Even with stratum 4 removed, the remaining population is highly skewed, the skewness being 3.5. The minimum coefficient of variation of this population, $(t_y)^{-1}\sqrt{V(\hat{t}_y;\mathbf{N})}$, constructing 4 strata of any kind and sampling 400 units, is 1.688 %.

4.2. On the equations (4) and (15)

Recall that the Dalenius equations (4) are derived under Approximation 1 and that condition (15) is derived for predetermined genuine sampling strata only. For these reasons, the size of certainty stratum units is held fixed to its best possible size and the analyses in this subsection are confined to genuine sampling strata. The finite population factors in (15) moderate the impact of $\left(y_h - m_h\right)^2$ and $\left(y_h - m_{h+1}\right)^2$, and if they increase from stratum 1 to stratum *H*, which is likely if the population is highly skewed, the effect of them is stronger on the right hand side of each equation. Consequently, (15) tends to produce strata less unequal in size than strata given by (4).

Usually, when (4) or (15) are applied to a finite population an exact solution does not exist. The stratum boundaries b_1 and b_2 in Table 1 is a solution to (4) or (15) in the sense that they minimise the sum of the absolute differences between the right hand and left hand side of each equation. The stratifications are different; however, the difference in variance ratio is not large. For MU284 the stratifications by (4) and (15) are identical, see Table 2.

Table 1. Stratifications for the value added population with four methods sorted by ascending variance ratio

Stratum	Best possible		(15)		SA		(4)	
	N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h
1	5225	74	5096	67	5086	66	5400	85
2	1433	66	1555	72	1572	74	1320	67
3	482	74	489	75	482	74	420	62
4	186	186	*186*	*186*	186	186	*186*	*186*
Variance ratio	1.000		1.001		1.002		1.004	

NOTE: Italicised numbers are fixed to the best possible ones.

Table 2. Stratifications for the MU284 population with four methods sorted by ascending variance ratio

Stratum	Best possible		(15)		(4)		SA	
	N_h	n_h	N_h	n_h	N_h	n_h	N_h	n_h
1	111	12	101	10	101	10	92	9
2	73	10	80	11	80	11	92	14
3	51	9	54	10	54	10	53	10
4	49	49	*49*	*49*	*49*	*49*	47	47
Variance ratio	1.000		1.019		1.019		1.038	

NOTE: Italicised numbers are fixed to the best possible ones.

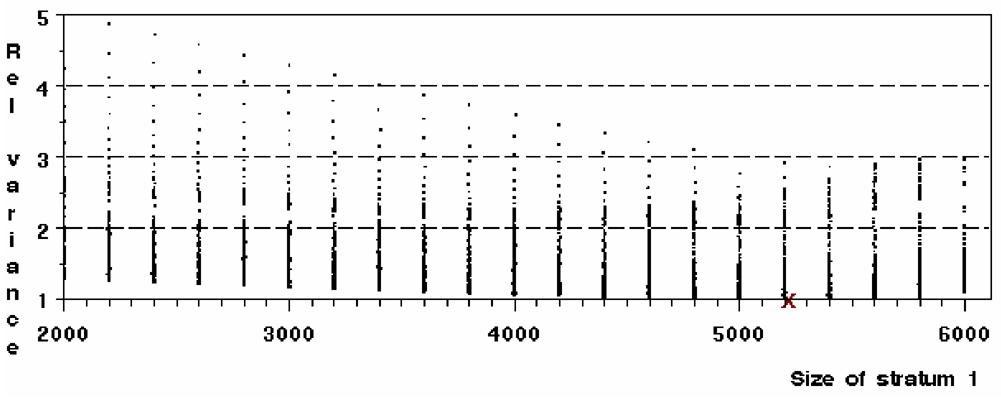
4.3. The certainty stratum

To apply Theorem 1 we need to solve (15) and (16) simultaneously. The results of the previous subsection indicate that the Dalenius equations (4) are satisfactory as an approximate solution to (15). To solve (4) we used the approximate method proposed by Ekman (1959), which has been shown to give excellent results (Cochran 1961; Hess, Sethi and Balakrishnan 1966; Murthy 1967). To solve (16) the algorithm went through all possible values of the size of the certainty stratum from $N_H = 0$ to $N_H = n$-15, and for each value determined the other stratum boundaries by a fast numerical algorithm for the Ekman rule (Hedlin, 2000). This procedure is in Tables 1 and 2 referred to as the stratification algorithm, SA. Note that SA in Tables 1 and 2 solves a bigger problem than (4) and (15) and still is competitive.

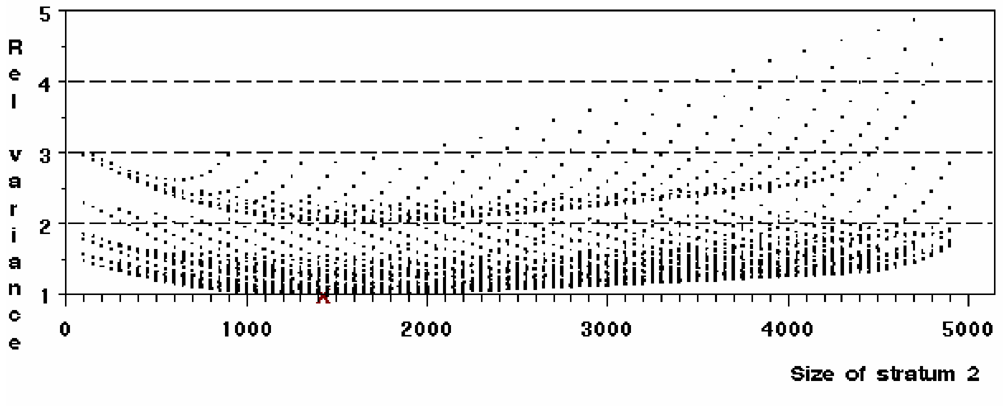
4.4. Flatness of the objective function

Under variation of the three stratification parameters, N_1, N_2, and N_3, the estimator variance $Var(\hat{t}_y;\mathbf{N})$ forms a response surface in a four-dimensional space. Let P_j be the response surface projected on the two-dimensional space $(N_j, Var(\hat{t}_y;\mathbf{N}))$ for j = 1, 2, 3 and 4. Figure 1 displays P_j for j = 1, 2, 3 and 4, with the variance ratio along the y-axis. The most striking feature of the plots in Figure 1 is the flatness of the estimator variance surface. Plot (d), for example, shows that if the size of the certainty stratum is within $(120, 230)$ it is possible to hit the minimum variance if the other strata are chosen optimally. The interval $(120, 230)$ must be considered very wide as the certainty stratum with a total sample size of 400 cannot contain more than 400 units. If the certainty stratum is chosen within this interval and the three genuine sampling strata are determined by the Ekman rule, the worst variance ratio is 1.05 (achieved for N_4 = 120). This repeated for the interval $(140, 230)$ gives 1.02 as the worst variance ratio (achieved for N_4 = 230). A large certainty stratum combined with a small size of stratum 3 yields variance ratios that are unacceptable.

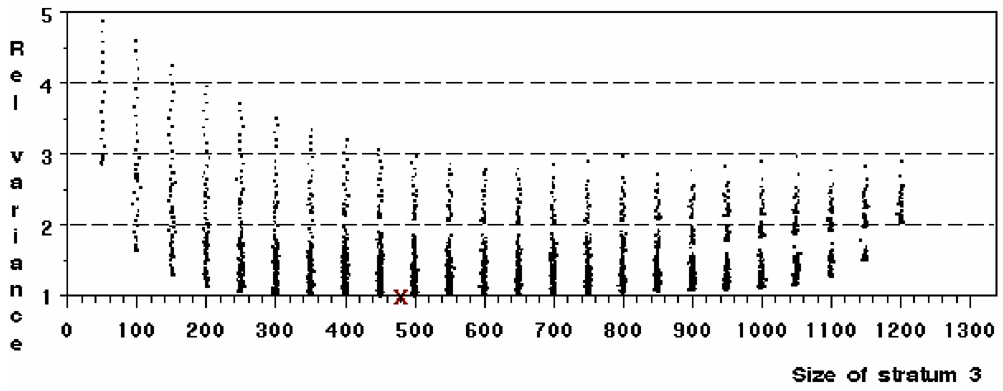
(a)



(b)



(c)



(d)

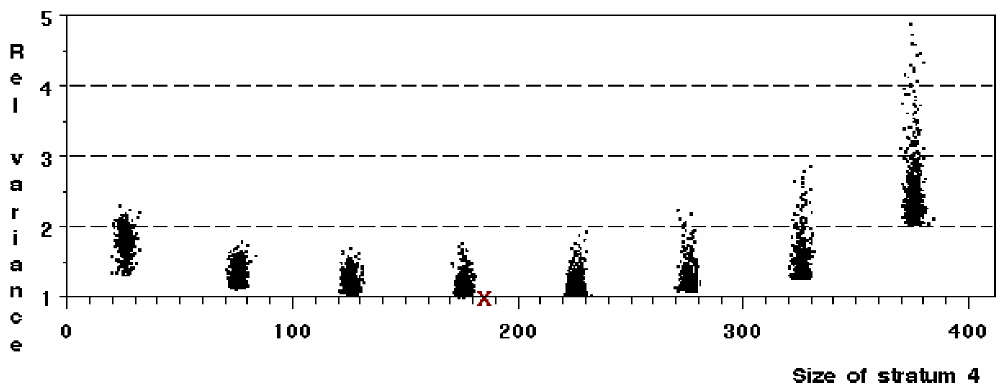


Figure 1. The variance ratio for a large number of stratifications of the value added population. Plot (a) shows the variance ratio for stratum 1 with N_1 along the x-axis. For each N_1 there are a number of choices of N_2, N_3 and N_4. The variance ratio of each combination is represented by a point in the scatter plot. For enhanced visibility, the points are randomly moved horizontally by addition of a small

normally distributed quantity. The best possible size of stratum 1 is marked on the x-axis. Plot (b), (c) and (d) are analogous, with N_2, N_3 and N_4, respectively, instead of N_1. All combinations of N_1, N_2, N_3 and N_4 are shown in each plot.

5. CONCLUDING REMARKS

We have derived necessary conditions for the combined problem of allocation and stratification in order to minimise the variance of the expansion estimator. In doing so, we have relied on the approximation of the finite population with a continuous distribution. An application to a population with 284 units and only 68 distinct values of the stratification variable does not indicate that this approximation is sensitive. This is further supported by the fact that Glasser (1962) obtains the same result as a special case of the main theorem of this paper without this approximation.

If a stratum is predetermined as a certainty stratum, the condition for its minimum variance size is substantially different from those of genuine sampling strata.

As for genuine sampling strata, the finite population correction can give a stratification that is far from what you would get with a conventional method such as the Dalenius-Hodges rule, which is derived for an infinite population. However, the deviation from the optimum that the Dalenius-Hodges rule necessarily gives should not often be of great practical importance. This is due to the empirical fact that in most practical applications the estimator variance surface is flat around the best possible stratum boundaries for genuine sampling strata. Surprisingly, one application indicates that this may be true for the certainty stratum as well.

APPENDIX A. PROOF OF LEMMA 1

To prepare the proof we give the partial derivatives of the function $f(\mathbf{n},\mathbf{b})$, see (11), and the constraints (6) and (7). As $f(x)$ is assumed continuous, $P_h = N_h / N$, m_h and s_h^2, see (8) – (10), are continuous and differentiable functions of b_{h-1} and b_h on (b_0, b_H). This makes (11) and the constraints differentiable functions. From (8) we see that, for h = 1, 2, …, H,

$$\frac{\partial g_h}{\partial n_j} = \begin{cases} 1 \text{ if } h = j \\ 0 \text{ otherwise} \end{cases}$$

$$\frac{\partial g_h}{\partial b_j} = \begin{cases} Nf(b_h), \text{ if } j = h-1 \\ -Nf(b_h), \text{ if } j = h \\ 0 \text{ otherwise} \end{cases}$$

whereas the derivative of g_{H+1} is always one for any of the components of **n**, and always zero for the components of **b**. Rewriting (11) to

$$f(\mathbf{n},\mathbf{b}) = \sum_{h=1}^{H} \left(\frac{N_h}{n_h} - 1 \right) N_h s_h^2 \qquad \text{(A.1)}$$

we see that

$$\frac{\partial f}{\partial n_h} = -\frac{N_h^2 s_h^2}{n_h^2}, \quad h = 1, 2 \ldots H \qquad \text{(A.2)}$$

and

$$\frac{\partial f}{\partial b_h} = Nf(b_h)$$

$$\times \left[(b_h - m_h)^2 \left(\frac{N_h}{n_h} - 1 \right) + \frac{N_h s_h^2}{n_h} - (b_h - m_{h+1})^2 \left(\frac{N_{h+1}}{n_{h+1}} - 1 \right) + \frac{N_{h+1} s_{h+1}^2}{n_{h+1}} \right], \qquad \text{(A.3)}$$

$h = 1, 2 \ldots H - 1.$

To prove Lemma 1, we first show that the gradients of the constraints are linearly independent in all feasible points, that is, all points $(\mathbf{n},\mathbf{b})$ satisfying (6). If they were not, there would exist non-zero scalars $\alpha_1, \alpha_2, \ldots, \alpha_{H+1}$ that would satisfy

$$
\begin{array}{llll}
 & a_1 & (1, 0, 0, ,..., 0, & -Nf(b_1), 0, 0, ..., 0)' \\
+ & a_2 & (0, 1, 0, ,..., 0, & Nf(b_1), -Nf(b_2), 0, ..., 0)' \\
 & \vdots & & \vdots \\
+ & a_H & (0, 0, 0, ,..., 1, & 0, 0, 0, ..., Nf(b_{H-1}))' \\
+ & a_{H+1} & & 0, 0, 0, ..., 0)'
\end{array}
$$

$= \mathbf{0}$.

Then $a_h = a_{H+1}$ and $Nf(b_h)(a_{h+1} - a_h) = 0$, $h = 1, 2, \ldots, H\text{-}1$, and $Nf(b_{H-1})a_H = 0$. Under the presumption that $f(x) > 0$, all $\alpha_h = 0$, $h = 1, 2, \ldots, H$, and we must have $\alpha_{H+1} = 0$. Hence all scalars $\alpha_1, \alpha_2, \ldots, \alpha_{H+1}$ are zero and the gradients are linearly independent in all feasible points.

This is a requirement for the Kuhn-Tucker Theorem (e.g. Luenberger, 1973). By this theorem, if $(\mathbf{n}^*,\mathbf{b}^*)$ is a local minimum, then $\nabla f(\mathbf{n}^*,\mathbf{b}^*) + \sum_{h=1}^{H+1} l_h \nabla g_h(\mathbf{n}^*,\mathbf{b}^*) = \mathbf{0}$ for a vector $l \in \mathbf{R}^{H+1}$ with $l \geq \mathbf{0}$ and $l_h g_h(\mathbf{n}^*,\mathbf{b}^*) = 0$, $h = 1, 2 \ldots H+1$. The H first components in the Kuhn-Tucker equations, which are associated with the stratum sample sizes n_h, give the following set of equations:

$$
l_h + l_{H+1} = \left(\frac{N_h s_h}{n_h}\right)^2, \; h = 1, 2 \ldots H. \qquad \text{(A.4)}
$$

By hypothesis there are at least two strata from which less than all units are sampled. Denote the indices of two such strata by *s* and *t*, and we have $l_s = l_t = 0$,

$l_{H+1} = (N_s s_s)^2 n_s^{-2}$, $\forall s$ with $n_s < N_s$. Hence

$$\left(\frac{N_s s_s}{n_s}\right)^2 = \left(\frac{N_t s_t}{n_t}\right)^2 \text{, } \forall\ s \text{ and } t \text{ where } n_s < N_s \text{ and } n_t < N_t . \qquad \text{(A.5)}$$

Thus (12) is proven. Now, for one particular stratum boundary, b_h, where

$h = 1, 2, ..., H-1$, we obtain

$$Nf(b_h)$$

$$\times\left[(b_h - m_h)^2\left(\frac{N_h}{n_h} - 1\right) + \frac{N_h s_h^2}{n_h} - (b_h - m_{h+1})^2\left(\frac{N_{h+1}}{n_{h+1}} - 1\right) + \frac{N_{h+1} s_{h+1}^2}{n_{h+1}}\right] \qquad \text{(A.6)}$$

$$+ Nf(b_h)(l_{h+1} - l_h) = 0, \quad h = 1, 2, \ldots H - 1.$$

By hypothesis $f(b_h) \neq 0$ and (13) is proven. Note that if all strata are predetermined genuine sampling strata, then $l_h = 0$, $h = 1, 2 \ldots H$, but if this constraint is not imposed then $l_h \geq 0$.

APPENDIX B. PROOF OF THEOREM 1

Equation (14) follows from Lemma 1. To prove (15), first note as the constraints $g_1, g_2, \ldots, g_{H-1}$ in (7) are predetermined to be satisfied with strict inequality, l_h and l_{h+1}, $h = 1, 2, \ldots H-2$, in (13) both vanish. After a little algebra, (15) is obtained from (13) and (14). To prove (16), set $h = H-1$ in (13) and note that $l_{H-1} = 0$, whereas l_H is derived as follows. Proceeding as in the proof of Lemma 1, use (A.4) twice with $h = H$ and $h = H-1$ to obtain

$l_H + l_{H+1} = (N_H s_H)^2 n_H^{-2}$ and $l_{H+1} = (N_{H-1} s_{H-1})^2 n_{H-1}^{-2}$. Since $n_H = N_H$ we have

$$l_H = s_H^2 - \left(\frac{N_{H-1} s_{H-1}}{n_{H-1}} \right)^2 . \qquad \text{(B.1)}$$

Insert (B.1) into (13) with $h = H–1$ and $n_H = N_H$, to obtain

$$\left(b_{H-1} - m_{H-1}\right)^2 \left(\frac{N_{H-1}}{n_{H-1}} - 1 \right) = \left(\frac{N_{H-1}}{n_{H-1}} s_{H-1} \right)^2 - \frac{N_{H-1}}{n_{H-1}} s_{H-1}^2$$

Divide both sides by $N_{H-1}/n_{H-1} - 1$, which by (7) is greater than zero, and (16) is obtained.

There is some ambiguity in the representation of l_H in (B.1) as we could have focused on another genuine sampling stratum than *H*–1. Any of the other possible choices lead to conditions equivalent to (16), although less appealing.

ACKNOWLEDGMENTS

A major part of this work was done when the author was at Statistics Sweden. The author thanks colleagues at Statistics Sweden, in particular Bengt Rosén, for constructive criticism of earlier versions of this paper.

REFERENCES

BAXI, H.R.S. (1995). Approximately optimum stratified design for a finite population – II. *Sankhya* B, 57, 391-404.

COCHRAN, W.G. (1961). Comparison of methods for determining stratum boundaries. *Bulletin de l'Institut International de Statistique*, 38, 345-357.

COCHRAN, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: Wiley.

DALENIUS, T. (1950). The problem of optimum stratification. *Skandinavisk Aktuarietidskrift*, 33, 203-213.

DALENIUS, T. (1952). The problem of optimum stratification in a special type of design. *Skandinavisk Aktuarietidskrift*, 35, 61-70.

DALENIUS, T. and GURNEY, M. (1951). The problem of optimum stratification. II. *Skandinavisk Aktuarietidskrift*, 34, 133-148.

DALENIUS, T. and HODGES, J.L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54, 88-101.

DETLEFSEN, R.E. and VEUM, C.S. (1991). Design issues for the retail trade sample surveys of the U.S. Bureau of the Census. *American Statistical Association Proceedings of the Survey Research Methods Section*, 214-219.

EKMAN, G. (1959). An approximation useful in univariate stratification. *The Annals of Mathematical Statistics*, 30, 219-229.

GLASSER, G.J. (1962). On the complete coverage of large units in a statistical study. *Review of the International Statistical Institute*, 30, 28-32.

HEDLIN, D. (2000). A procedure for stratification by an extended Ekman rule. *Journal of Official Statistics*, 16, 15-29.

HESS, I., SETHI, V.K., and BALAKRISHNAN, T.R. (1966). Stratification: A practical investigation. *Journal of the American Statistical Association*, 61, 74-90.

HIDIROGLOU, M.A. (1986). The construction of a self-representing stratum of large units in survey design. *The American Statistician*, 40, 27-31.

HIDIROGLOU, M.A. and SRINATH K.P. (1993). Problems associated with designing subannual business surveys. *Journal of Business & Economic Statistics*, 11, 397-405.

LAVALLÉE P. and HIDIROGLOU, M.A. (1988). On the stratification of skewed populations. *Survey Methodology*, 14, 33-43.

LUENBERGER, D.G. (1973). *Introduction to linear and nonlinear programming.* Reading, Massachusetts: Addison-Wesley.

MEHTA, S.K., SINGH, R., and KISHORE, L. (1996). On optimum stratification for allocation proportional to strata totals. *Journal of Indian Statistical Association*, 34, 9-19.

MURTHY, M.N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

PANDHER, G.S. (1996). Optimal sample redesign under GREG in skewed populations with application. *Survey Methodology*, 22, 199-204.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SCHNEEBERGER, H. (1985). Maxima, Minima und Sattelpunkte bei optimaler Schichtung und optimaler Aufteilung. *Allgemeines Statistisches Archiv*, 69, 286-297 (in German).

SERFLING, R.J. (1968). Approximately optimal stratification. *Journal of the American Statistical Association*, 63, 1298-1309.

SETHI, V.K. (1963). A note on optimum stratification of populations for estimating the population means. *Australian Journal of Statistics*, 5, 20-33.

SIGMAN, R. and MONSOUR, N. (1995). Selecting samples from list frames of businesses. In *Business Survey Methods*. (Eds. Cox, B., Binder, D., Chinappa, N., Christianson, A., Colledge, M. and Kott, P.) New York: Wiley.

SINGH, R. (1971). Approximately optimum stratification on the auxiliary variable. *Journal of the American Statistical Association*, 66, 829-833.

SLANTA, J.G. and KRENZKE, T.R. (1996). Applying the Lavallée and Hidiroglou method to obtain stratification boundaries for the Census Bureau's Annual Capital Expenditures Survey. *Survey Methodology*, 22, 65-75.

SWEET, E.M. and SIGMAN R.S. (1995). Evaluation of model-assisted procedures for stratifying skewed populations using auxiliary data. *Proceedings of the Survey Research Methods, Vol I. American Statistical Association*, 491-496.

UNNITHAN V.K.G. and NAIR N.U. (1995). Minimum-variance stratification. *Communications in Statistics – Simulation and Computation*, 24, 275-284.

WRIGHT, R.L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, 78, 879-884.