Some Optimality Problems When Estimating Household Data on the Basis of a "Primary" Stratified Sample of Individuals

Bengt Rosén

SCB

R & D Report
Statistics Sweden
Research - Methods - Development
U/STM - 36

Från trycket	August 10, 1987
Producent	Statistiska centralbyrån, Enheten för statistiska metoder
Ansvarig utgivare	Staffan Wahlström
Förfrågningar	Bengt Rosén, tel. 08 7834490

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1987

INLEDNING

TILL

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Föregångare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

Efterföljare:

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report, U/STM 1987:36. Some optimality Problems When Estimating Household Data on the Basis of a “Primary” Stratified Sample of Individuals / Bengt Rosén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-RnD-USTM-1987-36

August 10, 1987

SOME OPTIMALITY PROBLEMS WHEN ESTIMATING HOUSEHOLD DATA ON THE BASIS OF A "PRIMARY" STRATIFIED SAMPLE OF INDIVIDUALS

Bengt Rosén

Statistics Sweden

Paper presented at the 1987 ASA Annual Meeting, San Francisco, California

SOME OPTIMALITY PROBLEMS WHEN ESTIMATING HOUSEHOLD DATA ON THE BASIS OF A "PRIMARY", STRATIFIED SAMPLE OF INDIVIDUALS

Bengt Rosén

Statistics Sweden

Abstract. Statistics Sweden conducts a yearly sample survey (called HINK) with the objective to describe income conditions for different domains of households, household being determined by factual cohabitation and not by marital status.

The national population register is used as a frame for drawing a primary, stratified sample of adults. By interviews, the households of the sampled individuals are identified. Then various income (and expenditure) data are collected for the "entire" households, and used to achieve desired estimates for domains of households.

Within the framework of this sampling design, the statistician has various options; how to form strata in the population of adults, how to allocate the sample among strata, which estimator weights to use, to mention the most important ones. The multitude of objectives for the survey will also be an essential feature of the problem. We present some theoretical results for this type of optimization situation as well as numerical illustrations related to the HINK survey.

1 Background

Since 1975 Statistics Sweden has conducted a yearly survey, called HINK, with the main purpose of providing data on income for different classes of (cohabitation) households, the socioeconomic classification being of chief interest. The survey can also be, and is, used to yield data for classes of individuals. However, in this paper we shall confine ourself to the most important aspect, i.e. the household aspect.

In a recent revision of the HINK survey, we examined the efficiency of its design-estimation strategy. This paper reports on some general findings from that study. The design-estimation procedure in HINK is somewhat complicated, a main reason for this is that no sampling frame (in the form of a register) exists for cohabitation households (i.e. households determined by factual cohabitation and not by marital status). Therefore the sampling procedure which is used has sampling of adults as its "kernel". Further discussion of sampling frame, the sampling design etc., is given in Section 3.

Hence, efficiency problems concerning HINK fall under the following general heading; "Optimization of household surveys, where households are sampled via a stratified sample of adults", and this general topic will be our main theme. The presentation will be linked to the HINK survey, though, for the following reasons. The general problems and results will hopefully become more comprehensible if they are given a concrete background and moreover, a fairly concrete application will enable illustration of the orders of magnitude of the effects under consideration. We shall confine ourselves, though, to an "idealized" version of the HINK survey and work under the following simplifying assumptions; (i) No population changes occur during the survey period. (ii) The population is sampled without under- as well as overcoverage. (iii) All sampled units respond. It can be shown, though, that the analysis of a factual survey (as e.g. HINK) can be conceptualized as the simplified case.

2 Some terminology and notation

In the HINK survey a household is defined as follows. Its "core" is its "adult part" (adult = individual $\geq$ 18 years), which is either a cohabitation couple of opposite sexes (be they married or not) or a single adult. The complete household also includes the children (< 18 years) under "everyday care" of the adult(s). Let

U^H denote the population of households, and let d denote a generic element in U, (2.1)

$\underline{x} = \{x_d; d \in U\}$ denote a household variable, (2.2)

G denote a domain of study (i.e. a subset of U^H). (2.3)

The x-total over G, the size of G and the x-mean over G are denoted as follows, where $1_G(\cdot)$ stands for the indicator of the set G and $\underline{1}$ for the household variable $\underline{1} = \{\equiv 1; d \in U^H\}$,

$$\tau(\underline{x};G) = \sum_{d \in U^H} x_d \cdot 1_G(d), \tag{2.4}$$

$$g(G) = \sum_{d \in U^H} 1_G(d) = \tau(\underline{1};G), \tag{2.5}$$

$$\mu(\underline{x};G) = \tau(\underline{x};G)/g(G) = \tau(\underline{x};G)/\tau(\underline{1};G). \tag{2.6}$$

3 Chief aims and main features of the sampling procedure

3.1 Chief aims

A rough formulation of the problem we shall consider is as follows.

Let $G_1, G_2, \ldots, G_R$ be a specified set of disjoint household domains. Achieve, under prevailing constraints, the best possible estimates of $\{\mu(\underline{x};G_r), g(G_r), \tau(\underline{x};G_r);\ r=1,2,\ldots,R\}$, for a specified collection of $\underline{x}$-variables. (3.1)

Here the order μ, g, τ should be regarded as an ordering according to importance, domain means being of greatest interest while domain totals are of comparatively less interest.

In the HINK survey the $\underline{x}$-variable of greatest interest is disposable income, which roughly is defined as income from work and capital plus social benefits minus tax. The study domains of chief interest are the socioeconomic classes of households listed in Table 1 below. (There are of course rules for classifying a household when partners belong to different socioeconomic classes.)

Notation	Domain	Approximate group size, g(G)
G_1	Unskilled worker households	805 000
G_2	Skilled worker households	525 000
G_3	Junior salaried employee households	395 000
G_4	Intermediate salaried employee households	425 000
G_5	Senior salaried employee households	275 000
G_6	Enterpreneur households	145 000
G_7	Farmer households	75 000
G_8	Pensioner households	1 065 000

Table 1. The major socioeconomic classes in the HINK survey.

3.2 Main features of the sampling procedure

Since the chief aim of the HINK survey is to describe household conditions, the sampling procedure would ideally involve a properly designed sample from a frame containing all households. However, although we have many registers in Sweden, there is no register of (cohabitation) households. Lacking such an ideal frame, the HINK survey uses the register of the total population (RTB), which includes adults as well as children. RTB does contain information on marriages, but the frequency of nonmarital cohabitation is quite high in Sweden and, therefore, individuals are chosen as the "primary" sampling units.

A sample of households is generated in the following way. In the first round a "primary", stratified sample of adults is drawn. Then, by interviewing the primary individuals, the composition of their households is determined and thereby the sample of households is obtained. Once the individuals in the sampled households are identified, data for each household member are collected, mainly from various public agencies (tax authorities, different social welfare agencies etc.). Let

V^I denote the population of adults i.e. the adults in the RTB-register. (RTB contains information on age.) (3.2)

Next we discuss methods for drawing an efficient sample from V^I. Suppose one drew a simple random sample. Then, to the first order of approximation, the estimates of the domain means, $\mu(\underline{x};G_1),\mu(\underline{x};G_2),\ldots,\mu(\underline{x};G_R)$ will have variances which are roughly inversely proportional to the sizes of the domains, i.e. to $g(G_1),g(G_2),\ldots,g(G_R)$. If there is great variation among domain sizes, this type of picture would be nonconcordant with essentially any design principle for comparison of domain means. Even if design principles often disagree, there seem to be rough concensus on the rule of thumb that, when the aim is to compare means, one should strive for fairly equal precisions in the estimates of the means of interest, and this rule of thumb will be a guide for future considerations.

As is seen in Table 1, in HINK the domain sizes differ considerably. The largest domain (pensioners) contains roughly 15 times as many households as the smallest one (farmers). One way to adjust for this unbalance, at least as a "first step", is to introduce strata A_1, A_2, A_3 and A_4 in the sampling population V^I, which have the following properties.

A_1 is "directed" towards the smallest domain G_7 of farmer households, in the sense that there is (at least one hopes) a great chance that an individual from Stratum A_1 leads to a farmer household. Similarly, assume that A_2 is directed towards the (next smallest) domain G_6 and A_3 towards the largest domain G_8. Finally let A_4 be the remaining part of the population V^I.

This type of stratification should then be followed by a sample allocation structure of the following type. Sample "high" (i.e. with a sample fraction above average) in the strata A_1 and A_2 which are directed towards small domains and sample "low" in the stratum A_3 which is directed towards the large domain.

If the directing of the strata is good (to be discussed in more detail later on) and if the sample allocation is as just described, the following will occur. Extra observations (compared with simple random sampling) are "pumped" into the domains G_7 and G_6, thereby improving estimation precision in these domains as compared with "inversely proportional to domain size", while A_3 stears away observations from the large domain G_8, thereby avoiding resource waste by an "overly" good estimation precision for this domain.

Hence, we have presented a main motivation (but others exist) for stratification of the sampling population V^I of individuals. We pursue the matter in a more general setting.

Let $A_1, A_2, \ldots, A_k$ denote a stratification (i.e. a partitioning) of the sampling population V^I, and let the corresponding stratum sizes be denoted by $N_1, N_2, \ldots, N_k$. We assume that the primary sample of adults consists of independent, simple random samples from the different strata, with sample sizes $n_1, n_2, \ldots, n_k$. The corresponding sampling fractions are denoted by

$$f_h = n_h/N_h, \qquad h = 1,2,\ldots,k. \qquad (3.3)$$

4 Estimators and their variances

To estimate the quantities τ, g and μ in (2.4)-(2.6) we follow the "ordinary route" by letting estimates $\tau(\underline{x};G)$ of domain totals be the fundamental building blocks. Domain sizes and domain means are then estimated as the special case $\hat{g}(G) = \hat{\tau}(\underline{1};G)$ and by the ratio estimator $\hat{\mu}(\underline{x};G) = \hat{\tau}(\underline{x};G)/\hat{\tau}(\underline{1};G)$.

As estimators of domain totals we consider the following type of statistics (explanation of new notation is given afterwards)

$$\hat{\tau}(\underline{x};G;\underline{\alpha}) = \sum_{h=1}^{k} \frac{N_h}{n_h} \sum_{i \in A_h} x_{d(i)} \cdot \alpha_i \cdot 1_G(d(i)) \cdot I_i, \qquad (4.1)$$

where

$d(i)$ = the <u>household</u> to which individual i belongs, (4.2)

$m(i)$ = the <u>partner</u> of individual i, when i is cohabiting, (4.3)

$\underline{\alpha} = \{\alpha_i; i \epsilon V^I\}$ is a set of numbers, called <u>estimation weights</u>. (4.4)

I_i is the <u>sample inclusion indicator</u> for individual i. (4.5)

The following result is fairly straightforward.

<u>LEMMA 4.1</u>:. The statistic $\hat{\tau}(\underline{x};G;\underline{\alpha})$ in (4.1) yields unbiased estimation of $\tau(\underline{x};G)$ if, and only if, the estimation weights satisfy the following condition (4.6), which we call <u>household balancedness</u>,

$$\alpha_i + \alpha_{m(i)} = 1, \text{ if individual i cohabits, and } \alpha_i = 1, \text{ if individual i is single,} \qquad (4.6)$$

<u>Remark 4.1</u>: As a special case of the lemma we have that $\hat{g}(G;\underline{\alpha})=\hat{\tau}(\underline{1};G;\underline{\alpha})$ yields unbiased estimation of $g(G)$ as soon as $\underline{\alpha}$ is household balanced.

Furthermore, if we neglect the bias of the ratio estimator (as usually can be done), $\hat{\mu}(\underline{x};G;\underline{\alpha};\underline{\beta})=\hat{\tau}(\underline{x};G;\underline{\alpha})/\hat{\tau}(\underline{1};G;\underline{\beta})$ yields unbiased estimation of $\mu(\underline{x};G)$ as soon as $\underline{\alpha}$ and $\underline{\beta}$ both are household balanced.

In the sequel, estimation weights are presumed to be household balanced. •

<u>Remark 4.2</u>: The present estimation situation can be regarded as a special case whithin the general framework known as "network sampling", in particular "stratified network sampling", and the following papers treat problems which are related to ours; Birnbaum & Sirken(1965), Sirken(1972) and Levy(1977). Their considerations do not cover our situation, though, for the following main reason. We allow a wider class of estimator weights in (4.1) than is done in the mentioned papers, where the interest is confined to so called multiplicity estimators. A crucial step in our analysis will be to derive optimal weights within our wider class, and the weights which turn out to be optimal, see (6.6), yield in fact an estimator outside the class of multiplicity estimators. Moreover, one of the aims in this paper is to show that optimal weights can lead to considerable efficiency gains compared with the multiplicity estimator.

In our context the <u>multiplicity estimator</u> corresponds to the following α-weights, which are readily seen to be household balanced,

$$\alpha_i = \alpha_{m(i)} = 1/2, \text{ for i cohabiting.} \qquad (4.7)$$

We shall refer to this weighting system as <u>half-weighting</u>. •

Remark 4.3: A household should contribute twice in (4.1) if both adults in a cohabitation household happen to be sampled. However, in the HINK survey such double counting is omitted for practical reasons (and the omittance is adjusted for). In the sequel we neglect this complication, which in fact is practically negligible when sampling fractions are as small as in HINK (of the order 0.1 per cent).

Another matter which relates to the question of "simple or double counting of households" is the following. Let $\hat{\tau}^*$ denote the Horvitz-Thompson estimator of $\tau(\underline{x};G)$ based on the household sample which is generated by the sample of adults i.e., I_d^* and π_d^* denoting the inclusion indicator respectively the inclusion probability for household d,

$$\hat{\tau}^*(\underline{x};G) = \sum_{d \epsilon U^H} x_d \cdot 1_G(d) \cdot \frac{I_d^*}{\pi_d^*} . \tag{4.8}$$

The following claim is fairly straightforward to check, and we omit details. Under the assumption that sampling fractions are such that the frequency of "two adults from the same household" is low, $\hat{\tau}^*(\underline{x};G)$ is, with very good approximation, an estimator within the class (4.1), namely the one given by the weighting system which is introduced in Section 6, notably in (6.6). ●

Next we turn to the variances of the estimators. The general structure of the estimator $\hat{\tau}$ in (4.1) is quite simple. It is a domain total estimator based on a stratified sample. By employing this fact, variance formulas for τ, g and μ can be reached in a fairly straightforward way, the details of which we omit. We shall adapt our formulas to a further assumption on the estimation weights which we introduce next, and which we assume to be in force in the rest of the paper. Set

$$h(i) = \text{the stratum to which individual i belongs, } i \epsilon V^I. \tag{4.9}$$

The estimation weights $\underline{\alpha}$ are said to be <u>stratum combination constant</u> if the following relation holds true,

$$\alpha_i=\alpha_j \quad \text{as soon as } (h(i),h(m(i)))=(h(j),h(m(j))),\ i,j \epsilon V^I \text{ and have partners.} \tag{4.10}$$

When (4.10) is in force we change the α-parameters to a-parameters as follows,

$$a_{h\ell} = \text{is the common value for the } \alpha\text{-weights of individuals in stratum } A_h \text{ which have partner in stratum } A_\ell. \tag{4.11}$$

The previous household balancing condition, (4.6), then takes the form,

$$a_{h\ell} + a_{\ell h} =1, \qquad h,\ell =1,2,\ldots,k. \tag{4.12}$$

For a (fixed) domain G in U^H, set

B_h = the set of single-adult households in G for which the adult belongs to stratum A_h in V^I, h=1,2,...,k, (4.13)

$B_{h\ell}$ = the set of two-adult households in G for which one of the adults belongs to stratum A_h and the other one to stratum A_ℓ , h,ℓ=1,2,...,k. (4.14)

Note the relation

$$B_{h\ell} = B_{\ell h}\ , \qquad h,\ell=1,2,\dots,k. \tag{4.15}$$

Set, with # denoting the number of elements in a set,

$$g_h = \#B_h\ ,\quad g_{h\ell}=\#B_{h\ell}\ ,\qquad h,\ell=1,2,\dots,k. \tag{4.16}$$

Furthermore, for a household variable $\underline{x}$ let

$$\mu_\varrho = \text{the } \underline{x}\text{-mean over } B_\varrho\ ,\ \varrho = h \text{ or } (h,\ell),\ h,\ell=1,2,\dots,k, \tag{4.17}$$

$$\sigma_\varrho^2 = \text{the } \underline{x}\text{-variance over } B_\varrho\ ,\ \varrho = h \text{ or } (h,\ell),\ h,\ell=1,2,\dots,k, \tag{4.18}$$

$$\chi_\varrho^2 = \sigma_\varrho^2+(\mu_\varrho-\mu(\underline{x};G))^2,\ \varrho = h \text{ or } (h,\ell),\ h,\ell=1,2,\dots,k. \tag{4.19}$$

<u>Remark 4.4</u>: Note that the quantities in (4.13), (4.14) and (4.16) depend on the domain G, while the quantities in (4.17)-(4.19) depend on the domain G as well as on the variable $\underline{x}$, although we have surpressed this dependence in the notation. ●

<u>Remark 4.5</u>: The following relations are straightforward consequences of (4.15);

$$g_{h\ell} = g_{\ell h}\ ,\quad \mu_{h\ell} = \mu_{\ell h}\ ,\quad \sigma_{h\ell}^2 = \sigma_{\ell h}^2\ ,\quad \chi_{h\ell}^2 = \chi_{\ell h}^2\ ,\ h,\ell=1,2,\dots,k \tag{4.20}$$

●

We are now prepared to write down the desired variance formulas. Let us state that, for the sake of simplicity, we have made some approximations of the following types; finite population corrections are neglected, N-1 and N are regarded as equal, etc.. In view of (4.11) we change the α-parameter in the previous notation to an a-parameter. Below and henceforth V denotes variance.

$$V(\hat{\tau}(\underline{x};G;\underline{a})) = \sum_{h=1}^{k}\frac{N_h}{n_h}\cdot\{g_h(\sigma_h^2+\mu_h^2)+\frac{1}{2}\cdot g_{hh}\cdot(\sigma_{hh}^2+\mu_{hh}^2) + \sum_{\ell\neq h} a_{h\ell}^2\cdot g_{h\ell}\cdot(\sigma_{h\ell}^2+\mu_{h\ell}^2)\}$$

$$-\sum_{h=1}^{k}\frac{1}{n_h}\cdot(g_h\cdot\mu_h+g_{hh}\cdot\mu_{hh} + \sum_{\ell\neq h} a_{h\ell}\cdot g_{h\ell}\cdot\mu_{h\ell})^2. \tag{4.21}$$

As special case of (4.21), obtained by setting $\underline{x}=\underline{1}$ (which yields $\mu_\varrho=1$ and $\sigma_\varrho^2=0$) we get,

$$V(\hat{g}(G;\underline{a})) = \sum_{h=1}^{k} \frac{N_h}{n_h}\cdot(g_h+\frac{g_{hh}}{2}+\sum_{\ell\neq h} a_{h\ell}^2\cdot g_{h\ell}) - \sum_{h=1}^{k}\frac{1}{n_h}\cdot(g_h+g_{hh}+\sum_{\ell\neq h} a_{h\ell}\cdot g_{h\ell})^2. \tag{4.22}$$

Next, by applying the usual approximation formula for the variance of a ratio estimator and adopting the following approximation assumption,

the "squared mean" part of $V(\hat{\mu})$ is negligible compared with the "mean of squares" part, (4.23)

we arrive at the following formula,

$$V(\hat{\mu}(\underline{x};G;\underline{a};\underline{a})) = \frac{1}{g(G)^2}\sum_{h=1}^{k}\frac{N_h}{n_h}\cdot(g_h\cdot x_h^2+\frac{1}{2}\cdot g_{hh}\cdot x_{hh}^2+\sum_{\ell\neq h} a_{h\ell}^2\cdot g_{h\ell}\cdot x_{h\ell}^2). \tag{4.24}$$

<u>Remark 4.6</u>: There is no general guarantee that (4.23) should hold, but it can be expected to hold in "many" (maybe even in "most") situations. We have checked (4.23) empirically for HINK, and found it to hold with very good approximation there. ●

5 On the directing of strata

As discussed in Section 3, the main idea behind the stratification of the population of individuals is that the strata should serve as "directors" towards certain domains of study. In the following discussion, we regard G as a "target" domain and assume that stratum A_q is directed towards G. For simplicity we assume that A_q is the only stratum which is directed towards G. A stratification will, however, usually not be perfectly directed. "Misses" will occur, and we shall distinguish between two types of misses; a <u>miss of type I</u> if a household in G has no adult in A_q, and a <u>miss of type II</u> if an individual in A_q leads to a household outside G. (Misses of types I and II can be viewed as respectively under- and overcoverage when sampling G via A_q.)

Quantification of the number of misses of the two types can be given as follows,

$\sum_{h\neq q} g_h + \sum_{\substack{1\leq h\leq \ell\leq k \\ h,\ell\neq q}} g_{h\ell}$ tells the number of households in G which are misses of type I, (5.1)

while

$N_q - (g_q + 2g_{qq} + \sum_{\ell\neq q} g_{q\ell})$ tells the number of individuals in stratum A_q which yield misses of type II. (5.2)

We shall later on give a more quantitative account of how the efficiency of a stratification depends on its "missing" (or positively formulated "hitting") properties. At this stage we confine ourselves to the following qualitative claim, which we believe to sound intuitively very plausible.

If, ceteris paribus, misses of type I and/or type II are reduced then estimation precision for target domain characteristics are improved. (5.3)

6 Optimization of the survey

When planning a survey with the general structure outlined in Section 3, the statistician has (at least) the following options;

- choice of stratification $\mathcal{S}$ (definitions of strata as well as the number of strata),
- choice of sample alloction $\mathcal{A}$ (among the strata decided upon),
- choice of estimation procedure $\mathcal{E}$. (In our setting this means choosing estimation weights.)

Note that the above choices are "chained"; in the practical situation they must be carried out in the order $\mathcal{S}$, $\mathcal{A}$, $\mathcal{E}$.

Our previous "roughly" formulated problem (cf. (3,1)) can now be given the label "How to optimize the chain ($\mathcal{S}$, $\mathcal{A}$, $\mathcal{E}$)?" When seeking to give this problem a precise formulation we encounter the well known obstacle of "multipurposeness". We refer to Section 7.3 in Kish(1987) for a general discussion of multipurpose design problems, where also further references can be found. We adopt the approach of minimizing the "total imprecision" under given survey resourses. Hence, to make the optimization problem mathematically well posed we notably have to specify an overall criterion for estimation precision, but also to give precise specifications of constraints. Since the last point is the simplest, we start with that.

We lay the following constraints on $\mathcal{S}$, $\mathcal{A}$ and $\mathcal{E}$.

- For $\mathcal{S}$ we make no other assumption than "realizability", i.e. the information which is needed for a stratification should actually be available in the sampling frame. (6.1)
- For the sample allocation $\mathcal{A}$, we assume, for simplicity, fixed sample size, i.e.,

  $$n_1+n_2+\ldots+n_k=n \text{ is given.} \quad (6.2)$$

  In subsequent considerations, the assumption (6.2) could easily be changed to a fixed cost constraint with a cost function which is linear in stratum sample sizes.
- For $\mathcal{E}$ we stick to the assumptions which have been introduced previously; the estimation weights $\underline{a}$ should be stratum combination constant (see (4.11)) and household balanced (see 4.12)) (6.3)

Next we turn to the specification of an overall criterion for estimation precision. Regarding the precision for a single estimator, we employ the usual criterion; the smaller the variance, the better the precision. In our situation we meet the "multipurpose complication" in that we are interested in several domains of study $G_1, G_2, \ldots, G_R$ and (at least possibly) in many different study variables $\underline{x}$. Moreover, we are concerned with different types of population characteristics; μ, g and τ. We shall consider measures of overall estimation imprecision of the following type (recall notation introduced in Remark 4.1),

$$\Psi(V(\hat{\mu}(\underline{x};G_r;\underline{a},\underline{b}))),V(\hat{g}(G_r;\underline{\alpha}^*)),V(\hat{\tau}(\underline{x};G_r;\underline{\alpha}^{**})); \; r=1,2,\ldots,R). \tag{6.4}$$

The choice of a specific overall function Ψ is intricate and probably also controversal. However, for the time being we regard Ψ as decided upon. Thereby our problem is well posed at least from a mathematical point of view, and it runs as follows.

Find the tripple $(\mathcal{S},\mathcal{A},\mathcal{E})$ which minimizes the quantity in (6.4) under the contraints (6.1)-(6.3). (6.5)

In general such an optimization problem is quite messy. In particular we have;

For the optimal strategy $(\mathcal{S}_0,\mathcal{A}_0,\mathcal{E}_0)$ all the quantities $\mathcal{S}_0$, $\mathcal{A}_0$ and $\mathcal{E}_0$ will in general depend on

- the $\underline{x}$-variable,
- the domains of study $G_1, G_2, \ldots, G_R$,
- the overall precision criterion Ψ. (6.5)

In our situation, though, by a strike of good luck the optimization problem simplifies considerably, and the salient result to that effect is presented below. Although this result does not give our optimization problem a one stroke solution, it brings it down to "manageable".

<u>(APPROXIMATE) OPTIMALITY THEOREM</u>: Assume that the sampling fractions, f_h, are small. Then, under general conditions on $\underline{x}$ and G the following estimation weights

$$\tilde{\underline{a}}_{h\ell}(=\tilde{\underline{b}}_{h\ell})=f_h/(f_h+f_\ell), \quad h,\ell=1,2,\ldots,k, \tag{6.6}$$

simultaneously minimize all three variances

$$V(\hat{\mu}(\underline{x};G;\underline{a},\underline{b})), \; V(\hat{g}(G;\underline{a})), \; V(\hat{\tau}(\underline{x};G;\underline{a})). \tag{6.7}$$

<u>Remark 6.1</u>: The estimation weights $\tilde{\underline{a}}$ according to (6.6) will be referred to as the (sampling fraction) <u>proportional weights</u>. ●

Remark 6.2: As indicated in the naming of the above theorem, it is not true in a strict mathematical sense. However, the relative differences between the V:s for $\underline{a} = \tilde{\underline{a}}$ and $\underline{a}$ = the truly optimal weights, are so small that the result can be regarded as true from a practical point of view, at least over a wide range of $\underline{x}$:s and G:s. We have checked this claim in the HINK situation, and found the approximation to be good there.

One exception should be pointed to, though. The weights in (6.6) can be distinctly non-optimal for estimation of domain totals, $\tau(\underline{x};G)$ and $g(G)$, for domains G of the following type. G contains a great number of two-adult households with one adult in a low-sample stratum and the other adult in an average/high-sample stratum. ●

Remark 6.3: Proofs of the above approximation results can be given along the following lines. Minimize the expressions for $V(\tau(\hat{\underline{x}};G;\underline{\alpha}))$, $V(\hat{g}(G))$ and $V(\hat{\mu}(\underline{x};G;\underline{\alpha},\underline{b}))$ (cf. (4.21)-(4.23)), which are quadratic functions of $\underline{a}$ and $\underline{b}$, under the constraint (4.12), which is linear in $\underline{a}$ and $\underline{b}$. Lagrange's multiplier method leads to a system of linear equations. Then it can be shown that $\underline{a} = \tilde{\underline{a}}$ not only is an approximate solution to the linear system, but also a good approximation to the original optimization problem. We do not give details. ●

The most pertinent conclusion we shall draw from the approximation theorem, thereby using it as an "exact" theorem, is stated in (6.10) below. We start with the following observation.

The optimal weights $\tilde{\underline{a}}$ do not depend on the "nuisance" parameters $\underline{x}$ and G. (6.8)

Next, even if statisticians may disagree on what should be the "proper" choice of the overall criterion Ψ, we presume they do agree that any reasonable Ψ has the following property,

Ψ is (strictly) increasing in each of its arguments. (6.9)

Under the assumption (6.9), (6.8) leads to the following conclusion.

When seeking the optimal tripple $(\mathcal{S}, \mathcal{A}, \mathcal{E})$, the estimation part $\mathcal{E}$ can be "factored out" since it has a "universal" solution (which is independent of $\underline{x}$, $G_1, G_2, \ldots, G_R$ and Ψ), namely the solution given by (6.6). (6.10)

In the rest of this paper we assume that (4.23) is satisfied, and hence that (4.24) applies. By inserting $\underline{a} = \tilde{\underline{a}}$ into (4.24) and paying regard to (6.10) the following result is obtained after some straightforward algebra.

For a given stratification and a given sample allocation, the variance of the (universally) optimal domain mean estimator $\hat{\mu}(\underline{x};G;P) := \hat{\mu}(\underline{x};G;\underline{a};\underline{a})$ (P for proportional) is

$$V(\hat{\mu}(\underline{x};G;P)) =$$

$$= \frac{1}{g(G)^2} \left\{ \sum_{h=1}^{k} \frac{g_h \cdot x_h^2 + \frac{1}{2} g_{hh} \cdot x_{hh}^2}{\frac{n_h}{N_h}} + \sum_{1 \leq h < \ell \leq k} \frac{g_{h\ell} \cdot x_{h\ell}^2}{\frac{n_h}{N_h} + \frac{n_\ell}{N_\ell}} \right\}. \qquad (6.11)$$

For HINK, there has been uncertainty and debate how the P-versions of the estimators compare with the half weighted versions, in the sequal denoted H-versions (see (4.8)). The above approximation theorem tells that the P-versions never perform worse than the H-versions, but so far we have not given any quantitative measure of how much optimality pays. It is therefore of interest to have an expression also for the variance of $\hat{\mu}(\underline{x};G;H)$. Such an expression is obtained by setting $a_{h\ell}=1/2$ in (4.24). We give the resulting formula in a somewhat implicite fashion, which has the merit that it clearly shows that $\hat{\mu}(\underline{x};G;P)$ is superior to $\hat{\mu}(\underline{x};G;H)$ (as it should be according to the approximation theorem). It also gives a quantitative expression of the amount of variance reduction the P-version gives compared with the H-version. The following formula is readily obtained from (4.24) and some algebra, which we omit.

$$V(\hat{\mu}(\underline{x};G;H)) = V(\hat{\mu}(\underline{x};G;P)) +$$

$$+ \frac{1}{4g(G)^2} \sum_{1\leq h<\ell\leq k} \frac{\left(\frac{N_h}{n_h} - \frac{N_\ell}{n_\ell}\right)^2}{\frac{N_h}{n_h} + \frac{N_\ell}{n_\ell}} g_{h\ell} \cdot x_{h\ell}^2 . \qquad (6.12)$$

Analogous formulas can be obtained for the P- and H-versions of $V(\hat{\tau})$ and $V(\hat{g})$ by insertion into (4.21) and (4.22).

The variance formulas which are written out, respectively indicated, above provide tools for theoretical analyses of optimal allocation and optimal stratification under the present design. However, as the paper is already long, we obstain from persuing a theoretical line any further. Instead we shall try to indicate the usefulness of the formulas by considering some numerical illustrations.

7 Some numerical illustrations

In this concluding section we shall present some numerical findings related to the HINK survey. Our main aims are as follows; (i) To illustrate the use of the formulas in Section 4-6. (ii) To shed some light on the following general questions;

- How is a good sample allocation found?
- How is a good stratification found?
- How do proportional and half weighted estimators compare with each other?

We start with a short discussion of stratification possibilities in the RTB register of relevance for HINK, and a description of the stratification which is presently used in HINK.

As the most interesting domains of study are the socioeconomic classes, the most helpful information for forming directed strata concerns occupation and age, the latter to identify pensioners. As already mentioned, age data are available in RTB. When it comes to occupation, RTB as well as other registers are meagre. What is available in RTB is taxation information (although two years old relative to the investigation year). One aspect of this taxation information is the type(s) of taxation form(s) the person used, and there are three possibilities; farming income forms, entrepreneur income forms and for "other" incomes (the latter is the one used by the vast majority). Although meagre as information on occupation in general, it gives relevant information for the domains which are of particular interest because of their relative smallness, i.e. farmers and entrepreneurs.

Below we indicate the four strata which presently are used in HINK. (See also Remark 7.3 for some further information on HINK's stratification.)

A_1: A special register over farmers (called LBR), which is matched onto RTB. (7.1)

A_2: Individuals outside A_1, who are "predicted" to be entrepreneurs or farmers on the basis that they used the special tax forms for entrepreneurs respectively farmers. This stratum is chiefly directed towards entrepreneur households but also towards farmer households. (7.2)

A_3: Individuals outside A_1 and A_2 who are $\geq$ 65 years old. (7.3)

A_4: The remaining adults in RTB. (7.4)

The sizes of the strata are roughly as follows;

$$\#A_1=88\ 000,\quad \#A_2=293\ 000,\quad \#A_3=1\ 334\ 000,\quad \#A_4=4\ 785\ 000. \qquad (7.5)$$

Until further notice the domains G_1-G_8 are as in Table 1, and $\underline{x}$ stands for disposable income. In order to apply formulas as (6.11) one needs estimates of the quantities in (4.16)-(4.19) for the domains under consideration. Such estimates were derived by pooling estimates based on the HINK surveys for 1982, 1983 and 1984. We omit the details of this estimation, instead we regard the necessary quantities as known. (As usual in design contexts, errors in estimates of population characteristics do not lead to bias, at worst to nonoptimal choices.) By insertion into (6.11) we obtain expressions for $V(\hat{\mu}(\underline{x};G_1)),V(\hat{\mu}(\underline{x};G_2)),\ldots,V(\hat{\mu}(\underline{x};G_8))$. To exemplify we present the following formula, which is a slightly simplified version of (6.11) which, however, works well over the range of sample allocations which are of practical interest,

$$V(\hat{\mu}(\underline{x},G_7;P)) = \frac{420}{n_1} + \frac{430}{n_1+70} + \frac{730}{n_2} + \frac{250}{n_2+230} + \frac{470}{n_1+0.3n_2} + \frac{1700}{n_4}. \qquad (7.6)$$

Let us point to the following, somewhat unusual feature of the above formula, which is related to the fact that we employ the proportionally weighted estimator. In variance formulas for mean estimators, usually only n_h:s turn up in the denominators. Here linear functions of the n_h:s appear.

7.1 On sample allocation

We continue with the situation described above. We regard the stratification as decided upon, and our objective is to find a good sample allocation when domain means are regarded as the "all important" characteristics. Note that, in order to accomplish a "fair" comparison of possible sample allocations, each allocation should have the "right" to be followed by the optimal estimation procedure for the given stratification and sample allocation. Hence, in view of the optimality theorem, the estimator variances to be compared are those which correspond to proportional weighting, i.e. the formula (6.11) is the relevant variance formula.

In our general formulation of the problem of finding the optimal tripple $(\mathcal{J},\mathcal{A},\mathcal{E})$ we introduced an overall criterion function Ψ. In the previous section we showed, though, that the estimation part, $\mathcal{E}$, of the problem could be "factored out" without specifying a particular choice of Ψ. If we want a "strictly algorithmic" solution to the optimal allocation problem we must fix Ψ, though, because then the solution will effectively depend on the choice of Ψ. However, we shall here report on a more naive approach. Once explicit formulas for estimator variances, like (7.6), have been derived (and the laborious part of the derivation is the estimation of the quantities in (4.16)-(4.19)),it is quite easy to compute numerical variance values for a multitude of different allocations. Having the numerical values one can let the eye act as a Ψ-function. A further argument for this type of approach is that the variance functions are quite flat in a wide vicinity around "reasonably optimal" allocations. To exemplify, we present in Table 2 below the variances of the mean estimators for the domains G_1-G_8 under some different allocations which all satisfy the following sample size condition,

$$n_1+n_2+n_3+n_4=5000. \quad (7.7)$$

	Allocation						
Domain	a	b	c	d	e	f	g
G_1	1.63	1.35	1.85	3.24	1.71	1.62	1.62
G_2	2.77	2.30	3.15	5.51	2.91	2.76	2.76
G_3	4.37	3.71	5.16	8.86	4.81	4.57	4.57
G_4	4.74	4.07	5.27	9.24	4.90	4.66	4.64
G_5	10.01	[9.02]	12.46	20.56	11.63	11.11	11.09
G_6	19.29	31.01	[8.86]	13.24	10.39	12.48	10.67
G_7	16.75	25.01	14.08	[2.37]	5.07	5.08	4.51
G_8	0.63	2.27	4.14	4.32	4.14	4.14	4.13

Table 2. Values of $V(\hat{\mu}(\underline{x};G;P))$ (in monetary units) for some allocations which satisfy (7.7). The allocations are specified, and commented upon below. The rectangles exhibit global minima for domains G_5-G_7.

a: n_1=70, n_2=225, n_3=1025, n_4=3680. Allocation proportional to stratum size, i.e. essentially equivalent with simple random sampling.

b: n_1=55, n_2=100, n_3=410, n_4=4435. Optimal allocation for domain G_5.

c: n_1=50, n_2=100, n_3=200, n_4=3250. Optimal allocation for domain G_6.

d: n_1=1450, n_2=1500, n_3=200, n_4=1850. Optimal allocation for domain G_7, under the provision n_3=200.

e: n_1=300, n_2=800, n_3=200, n_4=3700.

f: n_1=400, n_2=500, n_3=200, n_4=3900.

g: n_1=400, n_2=700, n_3=200, n_4=3700. This allocation was judged to be "overall optimal", and is now in use in HINK.

7.2 On stratification and its relation to estimation

Next we shall consider examples which are intended to shed some light on the following questions.

How does the gain from a directed stratum depend on

- the hitting properties of the stratum?
- the estimation procedure? (7.8)

7.2.1 An example

To make the first example as "clean" as possible we shall confine ourselves to the very simplest case with only two domains of study, G_1 and G_2, and two strata, A_1 and A_2, and we assume that A_1 is directed towards G_1. A pertinent feature of the example should be that G_1 is small compared with G_2, and as a consequence of this (unless the hitting properties would be very poor) that A_1 is small compared with A_2. The example is meant to model the following part of the HINK context; G_1 = farmer households and G_2 = "other" households. We assume throughout the example that

$$N_1+N_2=6\ 500\ 000. \qquad (7.9)$$

As discussed in Section 5, the hitting properties of stratum A_1 are described by $N_1, g_1, g_2, g_{11}, g_{12}$ and g_{22} (where the g:s are computed with respect to G_1). In Table 3 below we list six different g-cases under two different situations, called Q and Z, regarding the size N_1 of the directed stratum. In both situations, Q and Z, the number of misses of type I decreases as the case number increases (g_2 and g_{22} both decrease). In situation Q the reduction of misses of type I is accompanied by a reduction in the number of misses of type II. (This is a consequence of the fact that N_1 is constant throughout the cases). In situation Z, though, the number of misses of type II remains (essentially) constant throughout the cases, and hence only type I hitting properties are improved as the case number increases. Note that the starting case, i.e. Case 1, is common to the two situations.

	Case					
	1	2	3	4	5	6
g_1	8000	12000	12000	18000	23000	24000
g_2	16000	12000	12000	6000	1000	0
g_{11}	8000	15000	21000	21000	24000	25000
g_{12}	8000	10000	17000	23000	24000	25000
g_{22}	34000	25000	12000	6000	2000	0
N_1, situation Q	105000	105000	105000	105000	105000	105000
N_1, situation Z	105000	125000	144000	156000	168000	172000

Table 3. Hitting characteristics for a spectrum of stratifications.

In accordance with the conjecture made in (5.3) (and other intuitive feelings), we expect the following qualitative behaviour of the estimation precision as cases and situations vary. In both situations, Q and Z, estimation precision should improve as the case number increases, lesser in Situation Q than in Situation Z, though. This picture should be present irrespective of the type of estimation weighting procedure, proportional or half. The proportional weighting should come out better than the half weighting, though. In Table 4 below we present findings which give quantitative information on how estimation precision varies with case, situation and type of estimator. Then we comment on the findings.

	Ratios between estimator variances in per cent											
	Situation Q						Situation Z					
Case	1	2	3	4	5	6	1	2	3	4	5	6
$V(\hat{\mu}(P))/V(\hat{\mu}(S))$	80	64	50	35	23	20	80	66	55	43	34	31
$V(\hat{\mu}(H))/V(\hat{\mu}(S))$	84	68	58	45	34	30	84	70	61	50	41	39
$V(\hat{\mu}(P))/V(\hat{\mu}(H))$	96	94	88	78	70	65	96	95	90	85	82	81
$V(\hat{g}(P))/V(\hat{g}(S))$	79	60	46	26	9	4	79	62	50	33	20	16
$V(\hat{g}(H))/V(\hat{g}(S))$	82	65	54	37	21	16	82	66	57	42	29	25
$V(\hat{g}(P))/V(\hat{g}(H))$	96	93	86	70	41	22	96	94	89	79	68	62
$V(\hat{\tau}(P))/V(\hat{\tau}(S))$	80	62	40	25	12	7	80	63	45	33	24	20
$V(\hat{\tau}(H))/V(\hat{\tau}(S))$	85	68	52	40	29	25	85	69	55	45	36	33
$V(\hat{\tau}(P))/V(\hat{\tau}(S))$	94	91	78	61	41	29	94	92	83	72	65	60

Table 4: Ratios between estimator variances relative to domain G_1 for g-values according to Table 3 and values of μ_Q, σ_Q^2 and χ_Q^2 as estimated for farmers when $\underline{x}$=disposable income. $n_1=500$ and $n_2=4500$. P and H indicate stratified samples with proportional respectively half weighted estimators, while S indicates estimates based on a simple random sample of size $n=n_1+n_2=5000$ from the entire population.

Below we give some remarks on the contents of Table 4.

Remark 7.1: The choice of allocation in Table 4 ($n_1=500, n_2=4500$) is of course rather arbitrary. However, the general picture from a qualitative point of view, as summarized in (7.10) and (7.11) below, is essentially the same for any sample allocation with $f_1>f_2$. ●

Remark 7.2: As shown in Table 4, the directed stratum A_1 can lead to considerable improvement of the estimation precision for domain G_1. This gain of course has a price, the estimation precision in domain G_2 is not as good as it would be under simple random sampling. In the above situation, estimator variances for domain G_2 will be roughly 9 per cent higher than they would be under simple random sampling from the entire population. ●

Remark 7.3: In the HINK survey, a major change of the stratification was made a couple of years ago. A main ingredient in the change was the introduction of the LBR register as the chief stratum directed towards farmer households, but also some other changes were made. Before that, the farmer households were directed at by a stratum which was constructed only with the aid of the taxation information mentioned above. The following holds roughly. The "before" situation is modelled by Situation Q, Case 3, while the "after" situation is modelled by Situation Q, Case 5. As indicated by the findings in Table 4, the change was quite beneficial for estimation of farmer characteristics. ●

Below we stress some pertinent features of the findings in Table 4, which in fact can be shown to hold quite generally.

Reduction of misses of type I as well as of type II are beneficial to estimation precision in the corresponding target domain. The former type of reductions have a more pronounced gainful effect than the latter type. This claim holds true for proportional as well as half weighting. (7.10)

Proportional weighting is better than half weighting. Its degree of superiority increases as the hitting properties of the directed stratum improve. Hence, if one does introduce a directed stratum, this step should be followed up by using the proportionally weighted estimator. (7.11)

7.3 Another example

The previous example, on which we based the conclusions (7.10) and (7.11), is a bit artificial (at least relative to the HINK survey) to the effect that only two domains of study are considered.

To give a more practical exemplification of the claim in (7.11), we present in Table 5 some findings from the actual HINK survey, with its eight domains, concerning the relative efficiencies of the proportional and the half weighted estimators for domain means and domain sizes.

Domain	$1-V(\hat{\mu}(P))/V(\hat{\mu}(H))$	$1-V(\hat{\vartheta}(P))/V(\hat{\vartheta}(H))$
Unskilled worker households	2.9 %	2.4 %
Skilled worker households	0.2 %	0.1 %
Junior salaried employee households	1.4 %	1.1 %
Intermediate salaried employee households	1.0 %	1.3 %
Senior salaried employee households	1.2 %	1.8 %
Entrepreneur households	15.0 %	22.1 %
Farmer households	40.2 %	53.5 %
Pensioner households	4.4 %	-17.2 %

Table 5: Variance reduction by the proportionally weighted estimator relative to the half weighted. Allocation according to g) in Table 2.

Remark 7.4: The figures in Table 5 show that the proportional weights, are considerably better than the half weights for the small domains with a directed stratum (i.e. farmers and entrepreneurs). For the large domain with a directed low-sampling stratum (i.e. pensioners) the P-estimator performs a bit better than the H-estimator for domain mean, as it should according to (6.12), but worse for domain size. The last finding illustrates the exception which was pointed to in Remark 6.2. For domains without a directed stratum, the gain is next to negligible. ●

References:

Birnbaum, Z. W. & Sirken, M. G. (1965). Design of sample surveys to estimate the prevalance of rare diseases: Three unbiased estimators. U.S. Department of Health, Education and Welfare 12, no 11.

Kish, L. (1987). Statistical design for researchers. John Wiley & Sons, New York.

Levy, P. S. (1977). Optimum allocation in stratified random network sampling for estimating the prevalence of attributes in rare populations. JASA 72, 758-763.

Sirken, M. G. (1972). Stratified sample surveys with multiplicity. JASA 67, 224-227.

Tidigare nummer av Promemorior från U/STM:
NR

1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)

2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)

3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)

4 Handledning i AID-analys. Anders Norberg (1980-10-22)

5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)

6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)

7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)

8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)

9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)

10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)

11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)

12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)

13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)

14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)

15 Granskning och evalvering av surveymodeller, tiden före 1960. Gösta Forsman (1985-05-30)

16 Variance estimators of the Gini coefficient - simple random sampling. Arne Sandström, Jan Wretman och Bertil Waldén (Memo, Februari 1985)

17 Variance estimators of the Gini coefficient - probability sampling. Arne Sandström, Jan Wretman och Bertil Waldén (1985-07-05)

18 Reconciling tables and margins using least-squares. Harry Lütjohann (1985-08-01)

19 Ersättningens och uppgiftslämnarbördans betydelse för kvaliteten i undersökningarna om hushållens utgifter. Håkan L. Lindström (1985-11-29)

20 A general view of estimation for two phases of selection. Carl-Erik Särndal och Bengt Swensson (1985-12-05)

21 On the use of automated coding at Statistics Sweden. Lars Lyberg (1986-01-16)

22 Quality Control of Coding Operations at Statistics Sweden. Lars Lyberg (1986-03-20)

23 A General View of Nonresponse Bias in Some Sample Surveys of the Swedish Population. Håkan L Lindström (1986-05-16)

24 Nonresponse rates in 1970 - 1985 in surveys of Individuals and Households. Håkan L. Lindström och Pat Dean (1986-06-06)

25 Two Evaluation Studies of Small Area Estimation Methods: The Case of Estimating Population Characteristics in Swedish Municipalities for the Intercensal Period. Sixten Lundström (1986-10-14)

26 A Survey Practitioner's Notion of Nonresponse. Richard Platek (1986-10-20)

27 Factors to be Considered in Developing a Reinterview Program and Interviewer Debriefings at SCB. Dawn D. Nelson (1986-10-24)

28 Utredning kring bortfallet i samband med den s k Metropolitdebatten. Sixten Lundström (1986-11-10)

29 Om HINK-undersökningens design- och allokeringsproblematik. Bengt Rosén (1986-10-06)

30 Om HINK-undersökningens estimationsförfarande. Bengt Rosén (1987-01-28)

31 Program för minskat bortfall i SCBs individ- och hushållsundersökningar - förslag från aktionsgruppen för uppgiftslämnarfrågor. Lars Lundgren (1987-02-20)

32 Practical estimators of a population total which reduce the impact of large observations. Jörgen Dalén (87-03-03)

33 A Multivariate Approach to Social Inequality. M Ribe (1987-05-05)

34 BORTFALLSBAROMETERN NR 2. Bortfall vid I-avdelningen, AKU och F-avdelningen t o m 1986. Håkan L. Lindström, Peter Lundquist och Tomas Garås (1987-08-01)

35 Use of Laboratory Methods and Cognitive Science for the Design and Testing of Questionnaires. Handbook of Methods. Judith T. Lessler (June 1987)

Kvarvarande exemplar av ovanstående promemorior kan rekvireras från Elseliv Lindfors, U/STM, SCB, 115 81 Stockholm, eller per telefon 08 7834178