Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price Index

Esbjörn Ohlsson

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1990:6

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.
Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

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

Efterföljare:

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

R & D Report 1990:6. Sequential poisson sampling from a business register and its application to the Swedish Consumer Price Index / Esbjörn Ohlsson.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1990-X101OP9006

Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price Index

Esbjörn Ohlsson

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1990:6

Från trycket	Maj 1990
Producent	Statistiska centralbyrån, Utvecklingsavdelningen
Ansvarig utgivare	Åke Lönnqvist
Förfrågningar	Esbjörn Ohlsson, tel. 08-783 45 13

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1990

April 1990

SEQUENTIAL POISSON SAMPLING FROM A BUSINESS REGISTER AND ITS APPLICATION TO THE SWEDISH CONSUMER PRICE INDEX

by Esbjörn Ohlsson
Department of Enterprise Statistics
Statistics Sweden

ABSTRACT. Brewer, Early & Joyce(1972) suggested a simple system for the co-ordination of several pps (probability proportional to size) samples from the same register. With this system it is also possible to have a large overlap between subsequent samples for a recurrent survey, in combination with rotation and updating for deaths, births and classification changes.

In the Brewer et al. system samples are drawn by the Poisson sampling procedure. In the present paper an alternative procedure is introduced, called *sequential Poisson sampling*. With this procedure it is possible to obtain samples with fixed size from strata which are not recognizable in the register.

Sequential Poisson sampling was developed for the Swedish Consumer Price Index, where it heavily reduced the great differences between the desired and the obtained effective size of the outlet sample. In particular, the obtained samples were almost free from out-of-scope units.

The so called SAMU system at Statistics Sweden is used to co-ordinate simple random samples from the central business register. Sequential Poisson sampling may be regarded as a pps extension of the sampling technique used in the SAMU system.

Keywords: Co-ordinated sampling, sequential sampling, Poisson sampling, pps sampling, Consumer Price Index.

Statistics Sweden
Dept. of Enterprise Statistics
Esbjörn Ohlsson

SEQUENTIAL POISSON SAMPLING FROM A BUSINESS REGISTER AND ITS APPLICATION TO THE SWEDISH CONSUMER PRICE INDEX

0. INTRODUCTION AND OUTLINE OF THE PAPER

Variables of interest in business surveys usually have a highly skewed distribution over the population. Often a few very large units balance thousands of small units for most variables. Hence it is necessary to make use of some measure of size as auxiliary information in the sampling and estimation design of a business survey. When sampling from the central business register at Statistics Sweden the size measure is commonly used to stratify by size; then a simple random sample is drawn within each stratum. In some cases the size measure is used to sample the units with inclusion probabilities proportional to size (pps).

For recurrent business surveys we usually want a large overlap between subsequent samples (positive co-ordination in time). On the other hand a population of enterprises is subject to rapid changes - deaths, births, changes in size and classification. Hence, our samples should be updated at least every year. Furthermore, there are often several surveys having the same target population. To have a more evenly distributed response burden, it is required to reduce the overlap between samples for different surveys (negative co-ordination in space) and to have a certain amount of controlled rotation. A simple system for pps sampling from a register which meets all these requirements was suggested by Brewer, Early & Joyce (1972), and is also advocated in Sunter (1977a) and Sunter (1986b). In this system, henceforth called *the BEJ system,* the samples are drawn by so called Poisson sampling.

Since the mid 1970's, pps sampling has been employed to sample outlets for the Swedish Consumer Price Index (CPI). The sampling technique which has been used up to now offers no real solution to the problem of updating samples which are positively co-ordinated in time. Therefore, in the sampling and estimation redesign for the 1990 Swedish CPI, it was decided to exchange the present technique for the BEJ system. Some problems remained, though, one being that the Poisson technique yields a random sample size. A more serious problem is that the retail trade section of the business register contains lots of units which are out-of-scope for the CPI. The conclusion was that the effective (net) sample size would be very hard to control with a straight-forward application of the BEJ system. Indeed, there would be a substantial risk for not getting any price measurements at all for some of the commodities of interest in the CPI.

To cope with these problems we have developed an alteration of Poisson sampling, which we call *sequential Poisson sampling*. It can be used in the BEJ system without loosing the advantages of the system. The sequential technique yields a fixed sample size. It also allows stratification according to classifications not available in the register. In the CPI case this feature was used, inter alia, to obtain a fixed number of units which are in scope for the survey.

A general description of the BEJ system and sequential Poisson sampling is given in sections 1.1-1.5; in section 1.4 we also discuss connections to the so called SAMU system for co-ordination of simple random samples from the Swedish central business register. In fact, the BEJ system with sequential Poisson sampling can be regarded as a pps extension of the sequential sampling technique by Fan, Muller & Rezucha (1962), which is used in the SAMU system.

The application to the CPI case is discussed in sections 2.1-2.5; in particular, we give the reasons for using pps sampling instead of stratification by size in this case (section 2.3).

An alternative to the rotation technique suggested by Brewer et al. (1972) is given in appendix 1. In appendix 2 we present numerical studies on the bias and variance of ordinary and sequential Poisson sampling. Appendix 3, finally, contains an analytic result on the bias of sequential Poisson sampling.

1. SEQUENTIAL POISSON SAMPLING FROM A BUSINESS REGISTER

1.1 POISSON SAMPLING AND THE BEJ SYSTEM

Let us assume that we have available some measure of size, denoted z_i for the i:th unit, for all the N units in our register. Let π_i denote the probability that the i:th unit is included in our sample of desired size n, and let

$$Z = \sum_{i=1}^{N} z_i \ .$$

A sampling procedure is said to be (strictly) pps, with desired sample size n, if it yields

$$\pi_i = n \frac{z_i}{Z} \quad i=1,2,\ldots,N. \tag{1.1}$$

In the sequel we will exclude the specification 'i=1,2,...,N' when there is no risk for confusion. (1.1) is impossible to achieve unless we have

$$n \frac{z_i}{Z} \le 1 \ , \tag{1.2}$$

for all the N units. In the rest of the paper we will assume that, when necessary, the largest units in the population are designated to a 'take-all' stratum so that (1.2) is fulfilled. The 'take-all' stratum will not be further discussed here.

Various problems are encountered when trying to construct 'without replacement' sampling procedures which (approximately) fulfill (1.1), and are accompanied by efficient estimators. In the literature there is an abundance of different pps procedures, see e.g. the monograph by Brewer & Hanif (1983).

Next we give a description of the well-known pps procedure called Poisson sampling. To each unit in the frame, a random number ξ_i is assigned. The ξ_i's are drawn from the uniform distribution on the interval (0,1) and are mutually independent. Unit i is in our sample if and only if, for some preassigned constant α_i, $0 \le \alpha_i \le 1$,

$$\xi_i \le \alpha_i \ . \tag{1.3}$$

Then trivially $\pi_i = \alpha_i$. Hence (1.1) will be fulfilled if we choose our α_i's as

$$\alpha_i = n \frac{z_i}{Z} \ . \tag{1.4}$$

Note that the actual sample size, m, is random, with expectation n.

Ogus & Clark (1971) report that Poisson sampling has been used at the Bureau of the Census since (at least) 1959. To the best of our knowledge, the name 'Poisson sampling' appears first in papers by Hájek (1960 and 1961).

In business surveys we usually have a 'basic' stratification by industry and/or region; the above description of pps/Poisson sampling then applies within each such stratum.

In the BEJ system for Poisson sampling from a register, the random numbers ξ_i are permanently associated with the units. This means that the same set of random numbers is used in the drawing of the (Poisson) samples for all surveys in the system at every point in time. Indeed, we may let ξ_i be a variable in the register. New units are assigned random numbers as they enter the register. Each time we take a sample we use the most up-to-date population, as regards births and deaths, classifications and size measures z_i.

The use of permanent random numbers will render positive co-ordination in time, in spite of the fact that we use an up-to-date frame. (According to Sunter, 1986b, we will have a maximal overlap between samples for the same survey taken at different points in time.) We may conduct redesigns of a survey, such as alteration of sample size or changes in industrial stratification, and still have a considerable overlap between subsequent samples. Negative co-ordination between the samples for two different surveys is obtained by simply shifting the random numbers before taking the second sample. I.e., for the second sample we use $\xi_i^* = \xi_i+b$ (modulo 1), where b is some constant. The ξ_i^*'s are of course uniform random numbers on (0,1), just like the ξ_i's. The two surveys will then have negatively co-ordinated samples even if they have different industrial/regional stratification or if they use different (but correlated) size measures. If desired, we may also obtain positive co-ordination ('in space') of samples for two surveys with different sampling designs by using identical random numbers.

Finally, it is possible to introduce a controlled rotation into this system by manipulating the random numbers between the sampling occasions. For a discussion of rotation techniques see appendix 1.

In summary, the BEJ system offers a simple solution to the serious problems with co-ordination and updating in business surveys. Note that this solution is intimately connected with the use of Poisson sampling; in particular, pps sampling with replacement is not a serious alternative, even if our sampling fractions should be small.

1.2 ESTIMATION FROM POISSON SAMPLES

Henceforth, we operate within a single basic (industrial/regional) stratum, unless otherwise stated. Let y be the target value in the survey, taking on the value y_i for unit i, and let Y denote the total of this variable over the population. Then an unbiased ('Horvitz-Thompson') estimator of Y is

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i} = \frac{Z}{n} \sum_{i \in s} \frac{y_i}{z_i} \quad . \tag{1.5}$$

where $i \in s$ means that the summation is over units in the sample s. The variance of $\hat{Y}_{HT}$ can be found in, e.g., Brewer & Hanif (1983, formula 4.2.23). As an alternative, Brewer et al. (1972) suggested the use of a ratio estimator, which for observed sample size m>0 is

$$\hat{Y}_R = \frac{n}{m} \hat{Y}_{HT} = \frac{Z}{m} \sum_{i \in s} \frac{y_i}{z_i} \; . \tag{1.6}$$

and for empty samples (m=0) is assigned the value 0. Let P_0 denote the probability of an empty sample. Then the variance of this estimator is approximately

$$V(\hat{Y}_R) \approx \frac{1}{n} \sum_{i=1}^{N} \frac{z_i}{Z}\left(1-n\,\frac{z_i}{Z}\right)\left(Z\,\frac{y_i}{z_i} - Y\right)^2 + P_0 Y^2 \; . \tag{1.7}$$

For a proof, see Brewer & Hanif (1983), where it is also claimed that the ratio estimator is more efficient than the unbiased estimator $\hat{Y}_{HT}$. This claim is strongly supported by a simulation study in Sunter (1977a), and by our numerical examples in appendix 2; in fact $\hat{Y}_{HT}$ performs much worse than pps with replacement (ppswr) in these examples. If the P_0-term is negligible, then (1.7) can be regarded as the variance formula for ppswr, modified by the 'finite population corrections' $(1-nz_i/Z)$.

The fact that the expected sample size n appears in (1.7), rather than the actual sample size m is unsatisfactory. If n is small, two samples may have quite different sizes but still be assigned the same precision by (1.7), viz. if they were drawn with the same expected sample size. Indeed, if we happen to get an empty sample, our imputed zero value for $\hat{Y}_R$ is claimed to have the precision given by (1.7) with a non-zero n!

It definitely seems more proper in this case to make our inferences conditionally on the outcome of the random sample size m. Then, $\hat{Y}_R$ still seems a reasonable estimator of Y. In Hájek (1964 and 1982) it is pointed out that Poisson sampling conditionally on m is equivalent to so called *rejective sampling* of size m. Hájek also makes the observation that in the special case when all the sizes z_i are equal, conditioning on m yields simple random sampling without replacement (srswor) of size m, and $\hat{Y}_R$ reduces to the ordinary unbiased estimator in srswor. In the general case, the probability distribution of rejective sampling is not by far as easy to handle; see Hájek (1964 and 1982) for some asymptotic results. We will not pursue this matter further, but merely conclude that proper estimation and variance calculation is not as simple in the Poisson case as it may seem at first sight.

1.3 SEQUENTIAL POISSON SAMPLING

The randomness of the actual sample size m in Poisson sampling could be a problem if we want an estimate at a fixed cost and at a fixed precision. If we are to make our inferences conditional on the outcome of m, however, we could take repeated, independent Poisson samples until we get a reasonable sample size. However, this is not possible in the BEJ system since we are to use a permanent set of random numbers in the drawing of our Poisson samples.

Let us say we want our sample to have size n_0, and hence take a Poisson sample with expected size $n=n_0$. Suppose we get a much smaller sample size m than n_0. If we are bothered by this, a natural way to obtain a larger sample without altering the random numbers would be to take a new Poisson sample with *the same set of random numbers* but with a larger expected sample size, say $n=n'>n_0$. We could continue in this way with different expected sample sizes until, for $n=n''$ say, we get a sample s'' with exactly n_0 units in it. This sample is the one which we would have obtained, with the current set of permanent random numbers, if we had been lucky enough to 'guess' the value n'' as the proper value of n in (1.4) from the beginning.

The procedure of successively adjusting n is of course impractical when sampling from a large register; next we present a quicker way to obtain the same sample. By the definition of Poisson sampling with expected sample size n, unit i is in the sample s if the random number ξ_i fulfills

$$\xi_i \le n \frac{z_i}{Z} . \tag{1.8}$$

Let us introduce the 'normed random numbers'

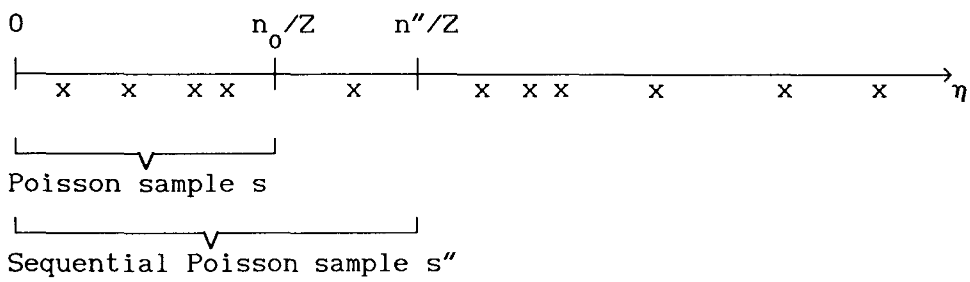
$$\eta_i = \frac{\xi_i}{z_i} . \tag{1.9}$$

Then $i \in s$ if

$$\eta_i \le n/Z . \tag{1.10}$$

Next note that the right-hand side in (1.10) is a constant and that adjusting n to n' just means adjusting this constant. It should be clear that adjusting n until we get a sample s'' with size n_0, is equivalent to taking the n_0 units with the smallest normed random numbers η_i as our sample. We call the latter sampling procedure *sequential Poisson sampling.* The relation between ordinary and sequential Poisson sampling is illustrated in figure 1 below.

Figure 1. Illustration of ordinary Poisson and sequential Poisson sampling, when the desired sample size is $n_0=5$. The x's represent the normed random numbers η_i on the real line.



> *<u>Definition of Sequential Poisson Sampling</u>. A sample is said to be drawn with sequential Poisson sampling of size n, if it consists of the n (unique) units with the smallest 'normed random numbers' η_i, defined in (1.9).*

Sequential Poisson sampling from a business register is very simple to perform in practice, whether we are interested in using permanent random numbers ξ_i or not. In one pass of the file we can generate the ξ_i's with some pseudo random number generator and divide them by the size measure z_i, to obtain η_i. Then the file is sorted by η_i and the first n (n_0) units in the sorted file are selected as our sample. Note that knowledge of the sum of the sizes, Z, is not required in this sampling procedure, as long as we feel secure that (1.2) is fulfilled. Sorting may be (CPU-) time-consuming in a huge register; a possibility is to sort only a 'gross' sample, see section 2.4.

What are the statistical properties of sequential Poisson sampling? We have already observed that the obtained sample s″ is the same as the one we would have got if we had taken an ordinary Poisson samle with the 'right' expected sample size n″ from the beginning; hence we conjecture that the statistical properties of a sequential sample of size n are approximately those of a Poisson sample conditionally on m=n. In particular we conjecture that the inclusion probabilities are approximately as in (1.1). If so, the following analogue of $\hat{Y}_R$ would be approximately unbiased,

$$\hat{Y}_S = \frac{Z}{n} \sum_{i \in s''} \frac{y_i}{z_i} \quad . \tag{1.11}$$

From the connection to Poisson sampling we also conjecture that the variance of this estimator can be approximated by (1.7), less the P_0-term, i.e.

$$V(\hat{Y}_S) \approx \frac{1}{n} \sum_{i=1}^{N} \frac{z_i}{Z}\left(1-n\,\frac{z_i}{Z}\right)\left(Z\,\frac{y_i}{z_i} - Y\right)^2 \quad . \tag{1.12}$$

In appendix 3 we give a crude upper bound for the bias of $\hat{Y}_S$, viz. of order $1/\sqrt{n}$. Of course we would like a bound of order 1/n just like we have for the ratio estimator in Poisson sampling, see Sunter (1986b). However, we have not been able to derive such a bound nor to prove (1.12). In appendix 3 we also give an example showing that sequential Poisson sampling is not in general strictly pps, i.e. (1.1) is not strictly fulfilled.

In lack of analytic results, our confidence in sequential Poisson sampling rests on a simulation study which is presented in appendix 2. The results of this study give strong support for the above conjectures. The observed bias of $\hat{Y}_S$ is very small and the variance is close to that given by (1.12). The differences in bias and variance between $\hat{Y}_S$ and $\hat{Y}_R$ (the latter based on ordinary Poisson sampling) are in favour of the sequential technique, but in most cases small. As regards inclusion probabilities, we have observed significant, but not large, bias in some very large units having inclusion probabilities close to 1; for the other units there is no evidence of bias in the inclusion probabilities. For details, see appendix 2.

By the close relations between the two procedures, cf. figure 1, exchanging ordinary Poisson for sequential Poisson should not induce any substantial changes in the co-ordination features of the BEJ system. Note that all shifts of random numbers for co-ordination of sequential samples should be made on the original ξ_i's, not on the 'normed' η_i's.

In large-scale surveys we usually should not bother that much about the random sample size in Poisson sampling. The substantial gain in using the sequential technique is when, like in the Swedish CPI, we have lots of rather easily detected out-of-scope units in our frame. This case is discussed in section 1.5 below.

1.4 CO-ORDINATION OF PPS SAMPLES AND SIMPLE RANDOM SAMPLES

Suppose that all the size measures z_i are equal. Then sequential Poisson sampling, unlike ordinary Poisson sampling, reduces to simple random sampling without replacement (srswor). The idea of drawing an srswor sample in this sequential way originates from Fan, Muller & Rezucha (1962, 'Method 4') and was adopted in the so called JALES technique developed at Statistics Sweden in the early 1970's, see Atmer et al. (1975, unfortunately in Swedish). In this special case, the estimator $\hat{Y}_S$ reduces to the ordinary unbiased srswor estimator and the variance formula (1.12) is exact, up to a factor N/(N-1).

The basic idea in the JALES technique is the same as in the BEJ system, i.e., to let the random numbers involved in the sampling be permanently associated with the units. In the BEJ system this idea is used in combination with Poisson sampling; in the JALES technique it is used in combination with the Fan et al. manner of drawing srswor or with equal probability Poisson sampling. Both systems, apparently developed independently, have more or less the same features as to updating, negative co-ordination etc., described in section 1.1.

The JALES technique is implemented in the so called SAMU system at Statistics Sweden which serves a majority of our business surveys with stratified srswor samples. In fact, the BEJ and JALES systems can be integrated, simply by using the same set of random numbers for the pps (Poisson or sequential Poisson) as well as srswor samples. The result is a system in which we may have positive and negative co-ordination between both pps and srswor samples. Such integration is now taking place at Statistics Sweden where the samples for the 1990 CPI, drawn with sequential Poisson sampling, use the same random numbers as the SAMU system. Sequential Poisson sampling can be regarded either as a fixed sample size development of the BEJ system or as a pps development of the JALES technique, cf. figure 2.

Figure 2. A scheme of permanent random number techniques.

	Equal prob.	Unequal prob.
Poisson technique, random n	BEJ & JALES	BEJ
Sequential technique, fixed n	JALES	*

Sequential Poisson sampling is an attempt to fill the empty box marked * in figure 2. As far as we know, no other pps procedure in the literature fits into that box. A very interesting sequential pps procedure is presented in Sunter (1977b) and further developed in Sunter (1986a). However, it has no connection to the use of permanent random numbers and is not meant to solve the updating and co-ordinating problems in business surveys. In Sunter (1977b) there is a proof of the fact that there can be no *strict* pps extension of the JALES-technique.

1.5 SEQUENTIALLY DELETING OUT-OF-SCOPE UNITS

Assume that we are sampling from a register with a substantial number of out-of-scope units. Suppose we can cheaply detect and delete most of those units from our sample, e.g. by making a telephone call. Still, it is likely to be very costly to detect all the out-of-scopes in the entire register. If we just take a sample with a fixed number of units from the register, we will have a large variability in the *effective* sample size, i.e. the number of 'in-scope' units. A well-known remedy is to use two-phase sampling, where the out-of-scopes are deleted from a large first-phase sample and a fixed number of in-scopes are drawn in the second phase. Unfortunately, in the BEJ (or JALES) system two-phase sampling would destroy the co-ordination between subsequent samples and the other features obtained by using permanent random numbers.

On the other hand, with sequential sampling the out-of-scope problem is very simple to handle. After having listed our units in the order given by the normed random numbers η_i, we simply go through the list, deleting the out-of-scopes as they come, until we have found enough (n) units for our net sample. Since the random numbers are independent, our sample will be (probabilistically) the same in the presence of out-of-scopes as it would have been in a 'clean' frame.

Hence, we can work with our sequentially drawn samples as if the out-of-scopes were not there. We must keep in mind, however, that all quantities in our estimators should then relate to in-scopes only. In particular Z in (1.11) should be the sum of z_i over all in-scope units. It is very likely that this quantity is unknown, although it could be estimated quite simply. Of course, this estimation would deteriorate our over-all precision; in particular formula (1.12) would not be valid. In some cases we might get a cheap and efficient estimate of Z from some other source. In other cases, such as the CPI, we are actually not interested in Y but rather in Y/Z. Then we do not have to estimate Z. Note that this is the pps analogue to the fact that in order to estimate a mean from an srswor sample we do not have to know the population size N.

In the equal probability (srswor) case Fan et al. (1962) notes that we do not even have to stratify our population before we start to sample, we can just take out the proper number of units *for the sample from each stratum in the random order in which they* appear. The same is true in the unequal probability case with sequential Poisson sampling. Of course, sequentially deleting out-of-scopes is just a special case of this kind of stratification.

In cases with no out-of-scopes, our simulations in appendix 2 indicate that ordinary and sequential sampling yield equally efficient estimation. Hence, with a large amount of out-of-scopes in the frame sequential Poisson sampling could be expected to be more efficient than ordinary Poisson sampling. If we need to stratify according to classifications not found in the register and want to use permanent random numbers, sequential Poisson sampling may be the only choice.

2. SEQUENTIAL POISSON SAMPLING OF OUTLETS IN THE SWEDISH CPI

2.1 THE SWEDISH CPI

The Consumer Price Index for Sweden is a measure of the average change in prices paid by Swedish consumers. It is based on price measurements for a judgmental sample of commodities. This sample will be regarded as fixed in what follows. Prices for the commodi-tites are measured in several different ways. The part of the price collection which is considered in the present paper is cal-led the Local Price System (LPS). It contains about 180 commodi-ties, which represent about 25% of the Swedish private consump-tion. For the LPS a sample of outlets is taken from the central business register. In these outlets, the prices of the LPS commo-dities are measured by interviewers, usually on the spot but oc-casionally by telephone.

2.2 THE PARAMETER

For the LPS, the outlet population is stratified according to industrial classification in the business register (department stores, grocers stores, shoe stores, restaurants, etc.). This 'basic' stratification aims at producing a decent number of price measurements for all selected commodities and will not be ques-tioned here. We shall discuss how to choose the samples of outlets in these basic strata. We start by presenting the parameter which is the target of our estimation, i.e. the way we would compute (the LPS part of) the CPI if prices were measured in the entire outlet population, not only in a sample.

At the lowest level of aggregation we have the price index I for a certain commodity in an industrial stratum. Let I_i be the change of price of this commodity in outlet i over the considered period. Then I is a weighted average of the I_i's,

$$I = \frac{\sum_{i=1}^{N} a_i \, I_i}{\sum_{i=1}^{N} a_i} . \qquad (2.1)$$

Ideally the weight a_i of an outlet should be its turnover for the specific commodity; since such weights are not available, the weights are in practice some measure of the total turnover of the outlet. At present the weight for outlet i, a_i, is the number of employees according to the business register, added by 1 (the shop-owner).

In the further calculations (the LPS part of) the CPI is obtained as a weighted average of such indexes I over all industrial strata and commodities, the weights being taken from sources independent of the LPS samples, e.g. the National Accounts.

In reality, the parameter which is used in the Swedish CPI is a bit more complicated than I in (2.1). However, the simplification (2.1) should serve well for our discussion of sampling design. For a detailed description and discussion of the CPI parameter see Dalén (1989).

2.3 PPS SAMPLING OR STRATIFICATION BY SIZE?

The weights a_i described above are known for (almost) all outlets in the business register. In the redesign of the outlet sampling procedure for the 1990 CPI the question was raised whether this 'auxiliary information' should be utilized for pps sampling or for stratification by size. It was decided that pps sampling was preferable for reasons which will now be given.

Among the units (locations) with classification 'retail trade' in the business register there are lots which are out-of-scope for the CPI, such as outlets which only trade commodities that are not measured in the CPI, or warehouses. Even if we would get rid of every out-of-scope unit, we are still left with lots of units which do not trade all the commodities we look for in a stratum; such units will be said to be 'partially out-of-scope'. Note that they should *not* be classified as (partial) non-respondents. In some of our industrial strata the sample sizes are as small as 10. In such a stratum we will, for some commodities, get just a few price measurements. In this case there is no room for further stratification. Even in industrial strata with larger sample size, such as n=40, extensive size stratification will render so small sample sizes that we are likely to get several strata with effective sample size 0 for some commodities. We conclude that when there is any room at all for stratification by size, we can have just a few strata.

Note that in practice, the summation in (2.1) should only be over outlets which trade the commodity in question.

Let us assume that the pps procedure under consideration is sequential Poisson sampling. Note though, that the arguments given below for pps sampling would apply for any reasonable pps sampling procedure. With $y_i=a_iI_i$ and size measure $z_i=a_i$ our parameter in (2.1) can be written I=Y/Z, capital Y and Z denoting totals as before. The estimator of this quantity is by (1.11)

$$\hat{I}_S = \frac{1}{n} \sum_{i \in s} I_i \ . \qquad (2.2)$$

Here, and in the sequel, we disregard the fact that in a few industrial strata we have a small take-all stratum. By (1.12) we have

$$V(\hat{I}_S) \approx \frac{1}{n} \sum_{i=1}^{N} \frac{a_i}{A}\left(1-n\cdot\frac{a_i}{A}\right) (I_j - I)^2 \ . \qquad (2.3)$$

The price indexes I_i are rarely outside the interval (0.5,2.0). On the other hand the number of employees, a_i, may vary by a factor 100, or more, in an industrial stratum. With stratified srswor we will get a variance whith the stratum variance of the y_i's as a main component. Because of the variation among the a_i's this component will be large unless we have lots of size strata. On the other hand the pps variance in (2.3) depends only on the restricted variance of the I_i's, and not on the variation of the y_i's. Hence pps sampling could be expected to be more efficient than srswor with a few size strata, which is the only realistic alternative in the CPI case. A numerical study of these matters is presented in Ålenius (1989) for the population of department stores (population I in appendix 2). With realistic n, pps sampling is consistently more efficient than stratified srswor for any number of strata less than n in the studied cases. This remains true even if we use ratio estimators in the stratified case.

Furthermore, pps sampling is much simpler to handle than stratified srswor in this case: pps sampling yields the self-weighting estimator (2.2) from one stratum (except in the few cases with a take-all stratum). In stratified srswor we have a more complicated estimator to be computed in a number of strata and we must take much care in the choice of stratum boundaries and optimal allocation.

The conclusion is that pps sampling is both simpler and more efficient than stratified srswor in the CPI case.

2.4 SEQUENTIAL POISSON SAMPLING IN THE CPI

As already mentioned, there are lots of units in the business register which are out-of-scope for the CPI. Since most out-of-scopes are easily detected in the sample, the sequential technique for controlling the effective sample size, described in section 1.5, was very useful in this case.

In practice sequential Poisson sampling for the CPI was carried out in the following way. First, a large 'gross' ordinary Poisson sample was drawn from each industrial stratum, the expected sample size being taken so large that we would be almost sure to get at least the desired number of in-scope units in each stratum. In this step we also constructed a take-all stratum in the few cases where this was necessary because of the restriction (1.2). The gross sample was listed in the order given by the normed random numbers, defined in (1.9). In the gross sample, most out-of-scopes could be identified by their names or after a quick telephone-call and were then removed. This process started from the top of the list and was terminated when we had obtained a 'preliminary net sample' of slightly larger size than the desired net sample size. In the yearly 'assortment controll' of outlets which are new in the sample, the interviewers detected a few more out-of-scopes. Finally, the last of the remaining outlets were excluded so that the desired net sample size was obtainded.

There are also some cases in which entirely different types of outlets have the same industrial classification; an example is hamburger stands which are mixed with grocers stores. Here we wanted to split the industrial stratum in two; if this could not be done we would end up with a sample in which we had none or very few hamburger stands since these are rare in the stratum. There are 18000 units with this industrial classification; hence it is impracticable to detect the hamburger stands in the frame before sampling. With the sequential technique, however, the stratification was easily obtained by taking out the proper number of hamburger stands as well as grocers stores in the order they appeared on the list.

2.5 CONCLUSION FOR THE CPI CASE

For the sample of outlets in the Local Price System of the Swedish

CPI, sequential Poisson sampling provides a useful and apparently efficient pps sampling technique; it allows stratification by classifications not available in the business register and it almost eliminates the impact of out-of-scope units on the effective sample size. With the use of permanent random numbers, we also obtain the advantages of the BEJ system such as a large overlap between samples from subsequent years and updating according to births and classification changes. We may - and do - also co-ordinate the CPI sample with other samples from the business register, pps or srswor.

APPENDIX 1. ROTATION

Suppose that we use the BEJ system to handle samples for a number of surveys and that we renew our samples once a year. As desired, we will then have a large amount of overlap between samples drawn subsequent years, but there will also be a substantial number of new units in the sample each year due to births, deaths, classification renewal and redesign. However, a unit in the first years sample which is not subject to any vital changes is very likely to remain in sample forever. Hence, for the sake of evenly distributing the response burden over the years, we must introduce some controlled rotation into the system.

In this appendix we shall only discuss rotation techniques for the BEJ system with Poisson sampling. The described techniques are expected to have approximately the same effect with sequential Poisson sampling or in the JALES system.

Brewer et al.(1972) suggested the following rotation technique which is also put forward in Sunter(1977a). Between the years the random numbers are shifted by a fix quantity b. If ξ_i denotes the random number the first year and ξ_i' the random number the second year, then $\xi_i'=\xi_i-b$ (modulo 1). This procedure is repeated every year. With this technique, small units having inclusion probabilities $\pi_i \le b$ will only be in sample for one year. The larger a unit, the more time it will stay in sample once it is drawn. (In the description of this technique given in the mentioned papers it is not the random numbers that are shifted, but rather their 'in sample region' $(0,\pi_i)$ which is shifted to $(0+b,\pi_i+b)$. It is readily seen that the result will be the same in both cases.)

In practice we would like to be able to specify a maximum number of years that small and medium sized units will stay in the sample, let us say 5 years. Furthermore, in order to keep a large amount of overlap between the years we do not want too much rotation, say at most 20% which is consistent with the rule '5 years in sample'. With the technique suggested by Brewer et al. it will be very hard to choose the constant b so as to obtain these goals. Indeed, most of the units are small and thus b must be very small in order to ensure a maximum 20% overall rotation. On the other hand this will mean that medium sized units will remain in sample for a very long time. Even if there exists a sufficient choice of b, it will not be a simple task to find it. The most severe problem with this rotation technique is that a choice of b which is sufficient for one survey will not in general be good for another survey, due to the fact that we have different inclusion probabilities π_i for the surveys. On the other hand, we must make one simultaneous choice of b for our entire system, or else our coordination pattern will be destroyed.

As an alternative, we suggest the following rotation technique. Each unit in the entire register is designated to one out of five rotation groups by a random trial giving 20% probability to each of the five groups. After year 1 the random numbers ξ_i *for units in rotation group 1* are shifted by a constant b as described above; next year those in group 2 are shifted etc. Among the units with $\pi_i \le b$ we will then have an expected amount of rotation of 20%; such units will stay in sample for at most 5 years. The constant b should be chosen quite large in order to give the required rotation to most of the units. On the other hand, if the shift is to large there is a risk that units will rotate out of one sample and into the sample of another (negatively correlated) survey. In the case of the SAMU system and the CPI at Statistics Sweden b was chosen as 0.1; this gave an (expected) rotation pattern '5 years in sample 1 - 5 years out - 5 years in sample 2 - etc.' for units with $\pi_i \le 0.1$.

Of course, the parameter choices '5 years in sample' and b=0.1 is just a special case; the above discussion applies equally well for any feasible choice of parameters.

It should be noted that the inclusion probability of a unit may increase over the years due to e.g. growth or change of industry. Hence, unfortunately, no strict guarantee can be given that a unit will remain in sample for at most 5 years.

Appendix 2. Numerical Illustrations of the Properties of Ordinary and Sequential Poisson Sampling

In this appendix we present numerical computations of the mean and variance of the suggested estimator in the case of sequential Poisson sampling as well as the Horvitz-Thompson estimator and the ratio estimator in ordinary Poisson sampling. Since our main interest is in the CPI we will use the notation introduced in sections 2.2-2.3. As a complement to this notation, we introduce the companions of (1.5) and (1.6) in the CPI case, viz.

$$\hat{I}_{HT} = \hat{Y}_{HT}/Z , \qquad \hat{I}_R = \hat{Y}_R/Z . \tag{A2.1}$$

For Poisson sampling we have analytic (approximative) variance formulas, see section 1.2. For the estimator $\hat{I}_S$ in the sequential case some conjectures were made in section 1.3, which in the CPI notation can be restated in the following way:

(i) $\hat{I}_S$ is an approximately unbiased estimator of I.

(ii) The variance of $\hat{I}_S$ can be approximated by the formula (2.3).

In order to investigate the validity of these conjectures, especially in the CPI case, we have performed a simulation ('Monte-Carlo') study. In this study we also investigate the bias of $\hat{I}_R$ and the validity of the approximation for $V(\hat{I}_R)$ obtained from (1.7). Throughout this appendix, we assume that there are no out-of-scope units in the frame.

We study three different populations, denoted I-III.

I. The first one is the complete population of Swedish department stores, obtained from the frame for the 1989 CPI sample (viz. the central business register of May 1989). Each unit in the frame has a notation of its number of employees which (added by 1) is the size measure $a_i=z_i$ in the case of the CPI. The 4 units in the 1989 'take-all' stratum were excluded from the study. Furthermore, 11 units with less than 15 employees were deleted since they are likely to be out-of-scope. The resulting population consists of 260 units. For I_i we chose the increase of the price of the item 'mens socks' from December 1987 to December 1988, as measured for the department stores in the 1987 CPI sample. A matching of the 1989 population and the 1987 sample gave I_i-values for 76 of the 260 units. In order to get I_i for the remaining units a probability distribution was fitted to the 76 I_i's. This distribution takes on the value 1 with probability 21/76 and is normally distributed with probability 55/76. It was checked that a_i and I_i were uncorrelated. Then the missing I_i-values were generated from the fitted distribution, independent of the a_i's. Some characteristics of the resulting population are given in table A2.1. A selection of 20 units from the population is listed in table A2.4 at the end of this appendix.

II. To obtain a very small population, every 10:th unit of population I was selected. The largest one of these units was dropped in order to ensure $\pi_i \leq 1$.

III. At Statistics Sweden there is an annual survey of financial accounts of enterprises. As the basis for our third population we used the take-all part of the industry 'manufacturing of machinery' in this survey. The 13 largest enterprises were removed to ensure $\pi_i \leq 1$, which left us with a population of 477 units. As size measure a_i we used the number of employees (+1) from the survey. The target variable y_i was 'investments', which was chosen in order to get a set-up which is not as well suited for pps-sampling as the CPI is. In order to adjust to the CPI notaion, we also introduce the variable $I_i=y_i/a_i$. Note that this set-up is completely 'natural', i.e., no variables have been artificially generated.

For each of the three populations we have computed the ratio between the variance for ppswr (pps with replacement) and the variance for srswr (simple random sampling with replacement) when the conventional estimators are used and we have the same sample size in both cases. Note that this ratio is independent of the chosen sample size. It serves as a measure of how well suited pps-sampling is for the particular population, and is denoted 'ppswr/srswr' in table A2.1 below. In the table some other characteristics of the three populations are also given; $\rho(a,y)$ denotes the population correlation betweeen a_i and y_i and analogously for $\rho(a,I)$.

Table A2.1. Population characteristics.

Set-up	I	II	III
N	260	25	477
Range a_i	22-522	33-283	1- 990
Range I_i	0.4-2.1	0.4-1.7	0-1487*
ppswr/srswr	0.08	0.13	0.40
$\rho(a,y)$	0.95	0.94	0.61
$\rho(a,I)$	0.03	0.29	-0.03

*Note: * Next largest value is 360.*

The populations derived from the CPI, I and II, are very well suited for pps-sampling. Population III is less well suited, but there is still a large gain in using ppswr instead of srswr.

Population I, in which we are mainly interested here, was examined for two different sample sizes, n=41 which is the actual (net) size of the 1989 department store sample and n=5 which was chosen since it is about the smallest size we could allow in any CPI stratum. We will refer to these cases as Ia and Ib respectively.

All the sample sizes n are given in table A2.2; note that for Poisson sampling n is the expected sample size. For reference, table A2.2 contains the analytic standard deviations (σ) for srswr and ppswr. We have computed the standard deviation $\sigma(\hat{I}_{HT})$ according to Brewer & Hanif(1983, formula 4.2.23), the approximation for $\sigma(\hat{I}_R)$ obtained from (1.7), and the *conjectured* value of $\sigma(\hat{I}_S)$ given by (2.3). We also give the probability P_0 that Poisson sampling should result in an empty sample. P_0 is negligible except for the case Ib; hence, in all other cases $\sigma(\hat{I}_R)$ and $\sigma(\hat{I}_S)$ coincide. For I-II the standard deviations have been multipied by 100, here as well as in table A2.3.

Table A2.2. Analytic standard deviations.

Set-up	Ia	Ib	II	III
N	260	260	25	477
n	41	5	9	50
σ(srswr)	12.00	34.36	25.96	6.10
σ(ppswr)	3.34	9.55	9.22	3.85
$\sigma(\hat{I}_{HT})$	14.94	48.08	25.46	4.16
$\sigma(\hat{I}_R)$	2.98	12.68	6.87	3.52
$\sigma(\hat{I}_S)$	2.98	9.43	6.87	3.52
P_0	$2.6\cdot10^{-21}$	$6.3\cdot10^{-3}$	$7.9\cdot10^{-7}$	$6.4\cdot10^{-30}$

As expected with Poisson sampling, the 'Horvitz-Thompson' estimator $\hat{I}_{HT}$ is of extremely poor precision in the CPI populations I-II: in population I it is even worse than srswr! This is consistent with the simulation results in Sunter(1977a). In population III, which is not as well suited for pps sampling as I and II, the precision of $\hat{I}_{HT}$ is not as poor but still worse than ppswr.

Conclusions about $\hat{I}_R$ and $\hat{I}_S$ should preferrably be drawn from table A2.3 below, in which we present the results from our simulation study. For each population we drew samples both by ordinary Poisson sampling (PO) and sequential Poisson sampling (SEQ), using the same set of random numbers for both samples. This procedure was iterated thousands of times, each time with a new, independent set of random numbers.

In table A2.3 we specify the number of iterations used for each set-up. For bias considerations the parameter I is given, followed by the average of $\hat{I}_R$ and $\hat{I}_S$ over the iterations. For easy reference, we repeat the analytic approximations for $\sigma(\hat{I}_R)$ and $\sigma(\hat{I}_S)$ from table A2.2 before presenting the empirical standard deviations over the iterations for both cases.

Table A2.3. Simulation results.

Set-up	Ia	Ib	II	III
# iterations	6000	1500	7020	6000
I=parameter	106.93	106.93	106.46	18.822
PO: Simu mean $\hat{I}_R$ (m>0)	106.93	106.00 (106.85)	106.71	18.887
SEQ: Simu mean $\hat{I}_S$	106.92	107.02	106.38	18.838
PO: Anal $\sigma(\hat{I}_R)$	2.98	12.68	6.87	3.52
PO: Simu $\sigma(\hat{I}_R)$ (m>0)	3.03	14.26 (10.66)	7.36	3.57
SEQ: Anal $\sigma(\hat{I}_S)$	2.98	9.43	6.87	3.52
SEQ: Simu $\sigma(\hat{I}_S)$	3.01	9.27	7.01	3.54

Note: For set-up Ib we got an empty Poisson sample 12 times out of 1500. In this case the estimator was given the value 0 as suggested by Brewer&Hanif(1983), though the value 1 seems more proper in this case. We also repeated the calculations amongst the non-empty samples (m>0); this figure is given in paranthesis.

In set-up Ia and Ib the observed bias is far from significant, except in Ib for $\hat{I}_R$ when empty samples are allowed. In case II and III there is a significant but small positive bias for $\hat{I}_R$. For $\hat{I}_S$ we have a significant but very small bias in II, and a non-significant bias in III (5% significance level). In all cases, the observed $(\text{bias})^2$, significant or not, is negligible beside the variance.

As comes to the standard deviations, formula (2.3) yields a very good approximation of $\sigma(\hat{I}_S)$ in the studied cases. The approximation of $\sigma(\hat{I}_R)$ is also quite accurate except in case Ib; in this case we have quite a large standard deviation due to the empty samples. Even if we condition on $m>0$, $\hat{I}_R$ is in fact inferior to ppswr in this case, as seen by comparing tables A2.2 and A2.3.

In fact, a comparison of tables A2.2 and A2.3 gives at hand that sequential Poisson sampling with the estimator $\hat{I}_S$ has consistently the least variance of all studied estimators in all the set-ups. With exception for set-up Ib the difference between $\hat{I}_S$ and $\hat{I}_R$ is negligible, though.

From the simulations, the relative frequences of inclusion in the samples can be computed for each unit and compared with the desired inclusion probabilities of (1.1). This has been done for all set-ups. In table A2.4, at the end of this appendix, we present the inclusion frequences for 20 arbitrarily selected units in set-up Ia.

As a complement to the investigatons of bias in table A2.3 the inclusion frequences can be used to investigate the bias of the sequential technique. The relative frequences for PO are unbiased estimates of the π_i's in (1.1); in general the relative frequences for SEQ are quite close to the desired π_i's and in most cases PO and SEQ show the same 'deviation pattern' from the π_i's. However in a few cases we have clearly significant bias for SEQ, namely for the two largest units in set-up II and the largest unit in set-up III. The worst bias here is from 96% to 91%. There are also 'analytical reasons', which will not be given here, for SEQ not to behave as desired for units with very large inclusion probabilities. We conclude that it might be a good idea to move such units to the take-all stratum, especially in a small population like II. With the exception of the largest units, however, sequential Poisson sampling seems to yield the desired inclusion probabilities with good approximation. In the CPI case, the measurement errors in the registred number of employees is likely to yield a much larger bias to the π_i's than the 'technical bias' discussed here.

Conclusions. The approximations in (i) and (ii) at the beginning of this appendix are quite accurate for sequential Poisson sampling in the studied cases. In most cases, there is little to choose between ordinary and sequential Poisson sampling as regards bias and variance. However, sequential Poisson sampling is without doubt the best choice in set-up Ib where the sample size is small.

From a practical point of view, the fixed sample size is an advantage of the sequential technique. In the CPI case we can expect large gains in precision due to the fact that sequential sampling yields a fixed number of units which are not out of scope; this has not been taken into account in the above simulations. Hence, for the Swedish CPI sequential Poisson sampling seems to be the most adequate technique.

Table A2.4. A selection of 20 out of 260 department stores in set-up Ia, with the number of employed a_i*, price index* I_i*, desired inclusion probability* π_i *(1.1) and the relative frequences of inclusions in the 6000 iterations for ordinary and sequential Poisson sampling,* PO *and* SEQ*, respectively.*

NR	a_i	I_i	π_i	PO	SEQ
1	22	0.89	0.031	0.032	0.031
2	23	0.98	0.033	0.034	0.032
3	26	1.00	0.037	0.038	0.037
4	26	1.00	0.037	0.041	0.042
5	29	1.08	0.041	0.043	0.043
.					.
.					.
101	78	0.77	0.111	0.115	0.112
102	78	0.75	0.111	0.112	0.110
103	79	0.79	0.112	0.117	0.115
104	79	1.00	0.112	0.104	0.106
105	80	1.19	0.114	0.111	0.111
.					.
.					.
201	144	0.94	0.204	0.195	0.198
202	146	1.19	0.207	0.208	0.206
203	146	1.22	0.207	0.209	0.209
204	148	1.00	0.210	0.216	0.218
205	149	0.78	0.211	0.205	0.204
.					.
.					.
256	401	1.30	0.569	0.559	0.564
257	401	1.05	0.569	0.565	0.573
258	406	1.08	0.576	0.580	0.583
259	451	1.00	0.640	0.639	0.650
260	522	1.00	0.741	0.746	0.754

Appendix 3. Analytic Properties of Sequential Poisson Sampling

In this appendix we derive a rather crude upper bound for the bias of the estimator (1.11) in sequential Poisson sampling. We also give an example showing that this procedure is not in general strictly pps.

Theorem A. Assume that we have drawn a sample s″ by sequential Poisson sampling. Then the relative bias of the corresponding estimator estimator $\hat{Y}_S$ of Y, defined in (1.11), is bounded as follows.

$$\frac{|E(\hat{Y}_S) - Y|}{Y} \le \frac{1}{\sqrt{n}} \cdot \frac{\max_i \left\{ \frac{y_i}{z_i} - \frac{Y}{Z} \right\}}{Y/Z} . \qquad \text{(A3.1)}$$

Typically, we use pps sampling when we believe that y_i is roughly proportional to z_i. If so, the 'max' value in the right-hand side of (A3.1) should be small. In any case, the bound in Theorem A is of order $n^{-1/2}$, so that the bias decreases with n. However, we would like to have the order n^{-1}, which would allow us to neglect the bias beside the variance. The simulations in appendix 2 indicate that the bias of $\hat{Y}_S$ is indeed negligible and that the bound in Theorem A is very pessimistic.

Proof of Theorem A: Let s be an ordinary Poisson sample drawn with *the same set of random numbers* as s″ is, and $\hat{Y}_{HT}$ the corresponding unbiased estimator defined in (1.5). Let m be the random size of s and

$$\gamma = \max_i \left\{ \frac{y_i}{z_i} - \frac{Y}{Z} \right\} . \qquad \text{(A3.2)}$$

With $1\{\cdot\}$ denoting the indicator of the event $\{\cdot\}$, we have

$$|E(\hat{Y}_S) - Y| = |E(\hat{Y}_S - Y) - E(\hat{Y}_{HT} - \frac{m}{n} Y)| =$$

$$= \frac{Z}{n} \left| E\left(\sum_{i \in s''} \left(\frac{y_i}{z_i} - \frac{Y}{Z} \right) \right) - E\left(\sum_{i \in s} \left(\frac{y_i}{z_i} - \frac{Y}{Z} \right) \right) \right| =$$

$$= \frac{Z}{n} \left| E\left(\sum_{i=1}^{N} \left(\frac{y_i}{z_i} - \frac{Y}{Z} \right) (1\{i \in s''\} - 1\{i \in s\}) \right) \right| \le$$

$$\le \frac{Z}{n} E \left(\sum_{i=1}^{N} \left| \left(\frac{y_i}{z_i} - \frac{Y}{Z} \right) (1\{i \in s''\} - 1\{i \in s\}) \right| \right) \le$$

$$\le \gamma \frac{Z}{n} E \left(\sum_{i=1}^{N} \left| (1\{i \in s''\} - 1\{i \in s\}) \right| \right) . \tag{A3.3}$$

We shall prove that

$$\sum_{i=1}^{N} |(1\{i \in s''\} - 1\{i \in s\})| = |m-n| . \tag{A3.4}$$

The following argument may be easier to follow after a glance at figure 1 in section 1.3. If $m=n$ then s and s'' contain the same units and (A3.4) is trivially fulfilled. If $m>n$ then all the units in s'' are also in s; hence the absolute values in the left-hand side of (A3.4) are all zero, except for the $(m-n)$ units i which are in s but not in s''; in the latter case the absolute value is 1. Hence the left-hand side of (A3.4) equals $(m-n)$ and, since $m>n$, (A3.4) is fulfilled. Finally, if $m<n$ then all units in s are also in s''; by arguing as in the case $m>n$ we find that the left-hand side of (A3.4) equals $(n-m)=|m-n|$ in the case $m<n$, too. We conclude that (A3.4) is valid in any case.

From (A3.3), (A3.4) and the Cauchy-Swartz inequality we obtain

$$|E(\hat{Y}_S)-Y| \le \gamma \frac{Z}{n} E|m-n| \le \gamma \frac{Z}{n} \sqrt{E|m-n|^2} . \tag{A3.5}$$

Finally,

$$E|m-n|^2 = V(m) = \sum_{i=1}^{N} \pi_i(1-\pi_i) \le \sum_{i=1}^{N} \pi_i = n \; . \qquad (A3.6)$$

From (A3.5), (A3.6) and (A3.2) we get (A3.1) and the theorem is proved. □

We shall end this technical appendix by giving a simple example showing that sequential Poisson sampling is not strictly pps, i.e. that (1.1) is not always valid. Note, though, that the simulations in appendix 2 indicate that (1.1) is a good approximation in most cases.

Example A. Let N=2 and n=1 and assume that $z_1 \le z_2$. Note that η_i, defined in (1.9) follows an $R(0,1/z_i)$ probability distribution and that η_1 and η_2 are independent. By conditioning on the outcome of η_1 we find that

$$\pi_1 = P(1 \in s'') = \int_{x=0}^{z_1^{-1}} P(\eta_2 > \eta_1) z_1 dx = \int_{x=0}^{z_2^{-1}} (1-z_2 x) z_1 dx = \frac{z_1}{2z_2}. \qquad (A3.7)$$

The right-hand side of (1.1) is $z_1/(z_1+z_2)$ in this case. Hence (1.1) is fulfilled only if $z_1=z_2$ and we conclude that sequential Poisson sampling is not strict pps in this case. □

REFERENCES

Atmer, J., Thulin, G. & Bäcklund, S. (1975). *Samordning av urval med JALES-metoden.* Statistisk Tidskrift 13, 443-450.

Brewer, K.R.W., Early, L.J. & Joyce, S.F. (1972). *Selecting several samples from a single population.* Austral. J. Statist. 14, 231-239.

Brewer, K.R.W. & Hanif, M. (1983). *Sampling with unequal probabilities.* Springer, New York.

Dalén, J. (1989). *Definitions- och parameterproblem i KPI:s lokalprissystem.* Memo, Statistics Sweden.

Fan, C.T., Muller, M.E. & Rezucha, I. (1962). *Development of sampling plans by using sequential (item by item) techniques and digital computers.* J. Amer. Statist. Ass. 57, 387-402.

Hájek,J. (1960). *Limiting distributions in simple random sampling from a finite population.* Publ. Inst. Hung. Acad. Sci. 5, 361-371.

Hájek,J. (1961). *Some extensions of the Wald-Wolfowitz-Noether theorem.* Ann. Math. Statist. 14, 333-362.

Hájek,J. (1964). *Asymptotic theory of rejective sampling with varying probabilities from a finite population.* Ann. Math. Statist. 35, 1491-1523.

Hájek,J. (1982). *Sampling from a finite population.* Dekker, New York.

Ogus, J.L. & Clark, D.F. (1971). *The annual survey of manufacturers: a report on methodology.* Tech. Paper No. 24, US Bureau of the Census, Washington DC.

Sunter, A.B. (1977a). *Respons burden, sample rotation, and classification renewal in economic surveys.* Int. Stat. Rev. 45, 209-222.

Sunter, A.B. (1977b). *List sequential sampling with equal or unequal probabilities without replacement.* Appl. Statist. 26, 261-268.

Sunter, A.B. (1986a). *Solutions to the problem of unequal probability sampling without replacement.* Int. Stat. Rev. 54, 33-50.

Sunter, A.B. (1986b). *Implicit longitudinal sampling from administrative files: A useful technique.* J. Official Statist. 2, 161-168.

Ålenius, M. (1989). *Storleksstratifiering eller pps-urval? En jämförande studie för KPI-data.* Memo, F-metod Nr 25, Statistics Sweden.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna. I serien ingår även **Abstracts** (sammanfattning av metodrapporter från SCB).

R & D Reports Statistics Sweden, are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers. In addition, abstracts are published three times a year (light brown /beige/ covers).

Report published earlier during 1990:

1990:1 (grön)	Calibration Estimators and Generalized Raking Techniques in Survey Sampling **(Jean-Claude Deville, Carl-Erik Särndal)**
1990:2 (grön)	Sampling, Nonresponse and Measurement Issues in the 1984/85 Swedish Time Budget Survey **(Ingrid Lyberg)**
1990:3 (grön)	Om justering för undertäckning vid undersökningar med urval i "rum och tid" **(Bengt Rosén)**
1990:4 (gul)	Data Processing at the Central Statistical Office - Lessons from recent history **(M. Jambwa)**
1990:5 (beige)	Abstracts I - sammanfattning av metodrapporter från SCB

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Elisabet Klingberg, U/STM, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 78.

Kvarvarande **gula** exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 47.