R&D Report
Research – Methods – Development
2000:6

Statistiska centralbyrån Statistics Sweden

A User's Guide to Pareto πps Sampling

Bengt Rosén

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 2000:6. A user's guide to Pareto πps sampling / Bengt Rosén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-2000-X101OP0006

R&D Report **2000:6**
Research – Methods – Development

A User's Guide to Pareto πps Sampling

Bengt Rosén

R&D Report 2000:6
Research - Methods - Development

A User's Guide to Pareto πps Sampling

Från trycket December 2000
Producent Statistiska centralbyrån, *Statistics Sweden,* metodenheten
Box 24300, *SE*-104 51 STOCKHOLM

Förfrågningar Bengt Rosén
bengt.rosen@scb.se
telefon 08- 506 944 90
telefax 08- 667 77 88

ISSN 0283-8680

Printed in Sweden
SCB-Tryck, Örebro 2000

December 2000

A User's Guide to Pareto πps Sampling

Bengt Rosén

ABSTRACT

A vehicle for utilization of auxiliary information is to employ a πps scheme, i.e. to sample with inclusion probabilities proportional to given size values. Pareto πps is a scheme for selection of πps list samples, having a number of attractive properties, in particular the following. It has predetermined sample size. It works with negligible point estimator bias. As regards point estimate accuracy it is, in our best understanding, optimal among schemes which admit objective assessment of sampling errors. Simple procedures for variance estimation and non-response adjustment are available. The scheme admits sample coordination by permanent random numbers. A sampling-estimation strategy with particularly good properties is obtained by combining Pareto πps and generalized regression estimation.

Aim with the report

This report contains a paper presented at the International Conference on Establishment Surveys – II, Buffalo June 17-21, 2000, which will appear in the forthcoming conference proceedings. The chief reason for printing it also as a Stat Sweden R&D Report is as follows. As the title indicates, the paper advices on practical application of the Pareto πps scheme. This scheme is already in use at various Stat Sweden's surveys, and is perhaps considered for use in others. It is believed to be of value for existing and potential users at Stat Sweden to have a User's Guide easily available as a Stat Sweden R&D Report. It has the same contents and layout as the conference proceeding paper, but for some minor changes to the effect that detected printing errors have been corrected and some references have been updated.

CONTENTS

A User's Guide to Pareto πps Sampling

Bengt Rosén

Statistics Sweden, Box 24 300, S-104 51 Stockholm, Sweden. bengt.rosen@scb.se

ABSTRACT

A vehicle for utilization of auxiliary information is to employ a πps scheme, i.e. to sample with inclusion probabilities proportional to given size values. Pareto πps is a scheme for selection of a list sample with predetermined size. It has a number of attractive properties, in particular the following. As regards point estimate accuracy it is, in our best understanding, optimal among schemes which admit objective assessment of sampling errors. Simple procedures for variance estimation and non - response adjustment are available. The scheme admits efficient sample coordination by permanent random numbers. A sampling - estimation strategy with particularly good properties is obtained by combining Pareto πps and generalized regression estimation.

Key words: Pareto πps, order πps, point and variance estimation, non-response adjustment.

1 Introduction

The following situation will be considered. Information about a population characteristic is to be gained from observations on a probability sample from the population U = (1, 2 ,... , N). List sampling is used, from a frame that one - to - one corresponds with the units in U, and also contains unit - wise auxiliary information. Until Section 7 the auxiliary data are assumed to be size values $\mathbf{s} = (s_1, s_2, \dots, s_N)$, $s_k > 0$, which typically are positively correlated with the study variable $\mathbf{y} = (y_1, y_2, \dots, y_N)$. (Example: unit = enterprise, y = sales during Mach 2000, s = number of employees.) As is well known, estimation precision for a population total (or mean) often benefits from use of a ***πps scheme***, i.e. a scheme with sample inclusion probabilities $\pi_1, \pi_2, \dots, \pi_N$ such that;

π_k is proportional to s_k, $k = 1, 2, \dots, N.$ (1.1)

A well - known πps scheme is Poisson sampling, which has simple sample selection and estimation procedures. However, it and various other πps schemes have the drawback of random sample size. It is generally desirable that a sampling scheme has ***fixed sample size***, and we confine to schemes with that property. Then (1.1) leads to the following ***desired inclusion probabilities*** $\lambda_1, \lambda_2, \dots, \lambda_N$, where n is sample size and $\tau(\mathbf{s}) = s_1 + s_2 + \dots + s_N$;

$$\lambda_k = n \cdot s_k / \sum_{j=1}^{N} s_j = n \cdot s_k / \tau(\mathbf{s}), \quad k = 1, 2, \dots, N. \tag{1.2}$$

Formula (1.2) may yield λ:s exceeding 1, which is incompatible with being probabilities. If so, some adjustment has to be made, usually by introducing a "take all" stratum. In the sequel is presumed that $\lambda_k < 1$, $k = 1, 2, \dots, N$.

A scheme with inclusion probabilities according to (1.2) has the following Horvitz - Thompson estimator for the population total $\tau(\mathbf{y}) = y_1 + y_2 + \dots + y_N$;

$$\hat{\tau}(\mathbf{y}) = \sum_{k \in \text{Sample}} y_k / \lambda_k = (\tau(\mathbf{s})/n) \cdot \sum_{k \in \text{Sample}} y_k / s_k \,. \tag{1.3}$$

A "perfect" πps scheme shall satisfy $\pi_k = \lambda_k$, $k = 1, 2, \dots, N$. We will be a bit more generous, though, and accept a sampling scheme as a ***πps schemes*** if (1.4) below is met;

$\pi_k \approx \lambda_k$ holds with good approximation for $k = 1, 2, \dots, N.$ (1.4)

The literature offers several πps schemes, and the statistician must choose. The main ***desired properties of a πps scheme***, besides having fixed sample size, are listed below.

- The scheme has ***simple sample selection***. (1.5)
- The scheme yields ***good estimation accuracy***. (1.6)
- The scheme ***admits objective assessment of sampling errors*** (consistent variance estimation). (1.7)
  - Variance estimators are available, the simpler the better.
  - Variance estimates never become negative.
- The scheme ***admits sample coordination*** over time and between surveys. (1.8)

The πps scheme which is most frequently used in practice is ***systematic πps***. This is in fact is a whole family of schemes generated by different frame ordering rules, whereby ***random frame order*** (rfo) and ***frame ordered by size*** (sfo) are chief possibilities. ***Sunter πps***, Sunter (1977), is highlighted in Särndahl et al. (1992).

This paper focuses on ***Pareto πps***, a member in the family of ***order πps schemes*** specified in Section 2.2. The aim is to (hopefully) demonstrate that Pareto πps meets the desires formulated above particularly well and, accordingly, to recommend it for practical use. Rigorous justifications of subsequent claims require quite sophisticated theory, though, which is not presented in this paper. We confine to just giving references.

2. Order πps schemes, notably Pareto πps

2.1. Pareto πps

DEFINITION 2.1 : ***Pareto πps*** with ***size values*** $\mathbf{s}=(s_1, s_2, \ldots, s_N)$ and ***sample size*** n generates a sample as follows.

1. ***Desired inclusion probabilities*** $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ are computed by (1.2).
2. Independent random variables $R_1, R_2, \ldots, R_N$ with uniform distributions on the interval [0, 1] are realized, and ***ranking variables*** Q are computed as follows;

$$Q_k = \frac{R_k \cdot (1-\lambda_k)}{\lambda_k \cdot (1-R_k)}, \quad k=1,2,\ldots,N. \tag{2.1}$$

3. ***The sample*** consists of the units with the ***n smallest Q - values.***

The somewhat fancy name Pareto is explained in next section. When all s_k (or equivalently all λ_k) agree, Pareto πps is nothing but simple random sampling (SRS).

All order πps schemes and, hence, also Pareto πps are based on asymptotic considerations. They are approximate πps schemes in the (1.4) sense, having perfect πps properties only for "infinite" samples. In particular, desired and factual inclusion probabilities do not agree exactly for finite samples. However, as discussed in Section 5, these imperfections have negligible effects in most practical survey situations, even for quite small sample sizes.

Pareto πps certainly meets desire (1.5), as is illustrated by the SAS - program below. It selects a Pareto πps sample with sample size SAMPZ from the records in the SAS data set FRAME, equipped with desired inclusion probabilities in the variable LAMB.

```
Data RANKING; Set FRAME; R= ranuni(SEED);
Q=R*(1 - LAMB)/(LAMB*(1 - R)); Proc SORT; By Q; Run;
Data SAMPLE; Set RANKING; If _N_ <= SAMPZ then output; Run;
```

2.2 Order πps

As stated, Pareto πps is a particular member in the family of order πps schemes introduced in Rosén (1997b), which is defined below. H(·) denotes the distribution function of a probability distribution with density.

DEFINITION 2.2 : ***Order πps*** with ***size values*** $\mathbf{s} = (s_1, s_2, \ldots, s_N)$, ***sample size*** n and ***shape distribution*** H(·) generates a sample by the same type of steps as in Definition 2.1, with the following modification of step 2. The ***ranking variables*** Q are computed by (2.2) below, where H^{-1} denotes inverse function;

$$Q_k = H^{-1}(R_k) / H^{-1}(\lambda_k), \quad k = 1, 2, \ldots, N. \tag{2.2}$$

It is by no means obvious that Definition 2.2 leads to πps schemes even in the (1.4) sense. However, Rosén (2000) proves that this is the case under general conditions on the shape distribution H.

Pareto πps is the particular order πps scheme given by the shape distribution;

$$H(t) = t/(1+t), \quad \text{which has density } h(t) = 1/(1+t)^2, \ 0 \le t < \infty. \tag{2.3}$$

Definition 2.2 introduces a whole family of sampling schemes, different H yield different schemes (although not in an entirely one - to - one fashion). To distinguish order πps schemes they are named by their shape distribution. The distribution in (2.3) seems to have no well - established name, though. After literature consultation we follow Feller (1966) and call it a Pareto distribution. Hence, the name Pareto πps.

Hitherto studies of order πps schemes have paid special attention to, besides Pareto πps, ***uniform order πps*** given by the uniform shape distribution $H(t) = \min(t, 1)$, $0 \le t < \infty$, and ***exponential order πps*** given by the exponential shape distribution $H(t) = 1 - e^{-t}$, $0 \le t < \infty$. The former scheme was first studied by Ohlsson (1990, 1998). He calls it ***sequential Poisson sampling***. Ohlsson's work provided background for the generalized notion "order πps". As regards Pareto πps, the author and Saavedra (1995) independently came across that scheme. Saavedra calls it ***odds ratio sequential Poisson sampling***. Exponential order sampling is considered in the literature under the name ***successive sampling***, see e.g. Rosén (1997a).

3 Estimation from observations on Pareto πps samples

Procedures for point and variance estimation are, of course, crucial in practical application of a sampling scheme. The following discussion is confined to the key estimation problem, estimation of the characteristic "population total", denoted $\tau(y) = y_1 + y_2 + ... + y_N$. As is well known, if this problem can be handled the estimation problem is solved for the vast majority of practically interesting characteristics, including ratios and domain characteristics. We first consider the ideal situation when all sampled units respond.

3.1 Estimation under full response

3.1.1 Point estimation

Since a πps scheme is presumed to satisfy (1.4) it is natural to use the estimator (1.3), which is re-stated in (3.1) below. However, under (1.4) it is not a "perfect", unbiased HT - estimator, rather a "quasi" HT - estimator, afflicted with some bias. The bias issue for Pareto πps is discussed in Section 5, with the conclusion that the bias is negligible almost always in practice.

$$\hat{\tau}(y) = \sum_{k \in \text{Sample}} y_k / \lambda_k \tag{3.1}$$

3.1.2 Estimator variance

At least in survey planning it is of interest to have an expression for the theoretical estimator variance. The following approximate variance formula, derived in Rosén (1997a, b), is asymptotically correct;

$$V[\hat{\tau}(y)] \approx \frac{N}{N-1} \cdot \sum_{k=1}^{N} \left(\frac{y_k}{\lambda_k} - \sum_{j=1}^{N} y_j \cdot (1-\lambda_j) \Big/ \sum_{j=1}^{N} \lambda_j \cdot (1-\lambda_j) \right)^2 \cdot \lambda_k \cdot (1-\lambda_k). \tag{3.2}$$

3.1.3 Assessment of sampling error

Formula (3.2) is theoretical and involves **y**-values for all population units. In practice an estimator $\hat{V}[\hat{\tau}(y)]$ of $V[\hat{\tau}(y)]$ must be exhibited, which together with approximate normal distribution for the estimator justifies the following type of (approximate) ***level 100 · (1 - α) % confidence interval*** for $\tau(y)$, $\delta_{\alpha/2}$ denoting the standard normal $1-\alpha/2$ fractile;

$$\hat{\tau}(y) \pm \delta_{\alpha/2} \cdot \sqrt{\hat{V}[\hat{\tau}(y)]}. \tag{3.3}$$

Consistent estimation of $V[\hat{\tau}(y)]$ is provided by, see Rosén (1997b);

$$\hat{V}[\hat{\tau}(y)] = \frac{n}{n-1} \cdot \sum_{k \in \text{Sample}} \left(\frac{y_k}{\lambda_k} - \sum_{j \in \text{Sample}} \frac{y_j \cdot (1-\lambda_j)}{\lambda_j} \Big/ \sum_{j \in \text{Sample}} (1-\lambda_j) \right)^2 \cdot (1-\lambda_k). \tag{3.4}$$

Formula (3.4) may look cumbersome at first glance. However, it is quite innocent from a computation point of view, as is seen from the formulas below. Note that only "single-summations" are involved.

$$\hat{V}[\hat{\tau}(y)] = \frac{n}{n-1} \cdot (A - B^2/C), \quad \text{where} \tag{3.5}$$

$$A = \sum_{k \in \text{Sample}} (y_k/\lambda_k)^2 \cdot (1-\lambda_k), \quad B = \sum_{k \in \text{Sample}} (y_k/\lambda_k) \cdot (1-\lambda_k), \quad C = \sum_{k \in \text{Sample}} (1-\lambda_k). \tag{3.6}$$

Formulas (3.4) - (3.6) show that Pareto πps meets both desires in (1.7), simple and non-negative variance estimation. Moreover, in Rosén (1997b) is proved that $\hat{\tau}(y)$ is asymptotically normally distributed under Pareto πps.

3.2 Estimation when non-response occurs

Practical surveys are seldom ideal, in particular non-response almost always occurs. Then formulas (3.1) and (3.4) cannot be used straight away, some adjustment for non-response has to be made. The following results are taken from Rosén & Lundqvist (1998), where adjustment procedures for uniform, exponential and Pareto πps are presented, with theoretical as well as numerical justifications.

Non-response adjustment must be based on some model (= assumption) about response behavior. The simplest, and most commonly used, is the simple MAR (Missing At Random) model stated below.

Simple MAR model: Sampled units respond independently, all with the same response propensity. (3.7)

Under (3.7) adjustment is achieved by a recipe which somewhat sweepingly can be formulated as follows.

Use (3.1) and (3.4) with "number of sampled units" exchanged for "number of responding units". (3.8)

More precisely, under (3.7) point estimation is carried out as follows. With

n' = number of responding units, (3.9)

modified inclusion probabilities: $\lambda'_k = (n'/n)\cdot\lambda_k = n'\cdot s_k/\sum_{j=1}^{N} s_j$, $k = 1, 2, \ldots, N$, (3.10)

$\mathcal{R}$sample = the collection of responding units, (3.11)

the following counterpart to the estimator in (3.1) works with negligible bias;

$$\hat{t}(y) = \sum_{k \in \mathcal{R}\text{sample}} y_k/\lambda'_k . \tag{3.12}$$

Moreover, the confidence interval (3.3) works with the variance estimator;

$$\hat{V}[\hat{t}(y)] = \frac{n'}{n'-1}\cdot\sum_{k\in\mathcal{R}\text{sample}}\left(\frac{y_k}{\lambda'_k} - \sum_{j\in\mathcal{R}\text{sample}}\frac{y_j\cdot(1-\lambda'_j)}{\lambda'_j}\Big/\sum_{j\in\mathcal{R}\text{sample}}(1-\lambda'_j)\right)^2\cdot(1-\lambda'_k) . \tag{3.13}$$

Formulas (3.5) and (3.6) have the following counterparts;

$$\hat{V}[\hat{t}(y)] = \frac{n'}{n'-1}\cdot(A'-B'^2/C') , \quad \text{with A', B' and C' as in (3.5)-(3.6) with } \lambda \text{ changed to } \lambda'. \tag{3.14}$$

A more elaborate response model runs as follows.

MAR model with several response homogeneity groups: With the population partitioned into (known) disjoint groups $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_G$, the following holds. Sampled units respond independently, and units in the same group have the same response propensity, which may vary between groups. (3.15)

Under response model (3.15), adjustment for non-response is achieved by the following post-stratification procedure. For $g = 1, 2, \ldots, G$, n_g denotes the number of sampled units from $\mathcal{G}_g$, n'_g the number of responding units and $\mathcal{R}_g$sample the set of responding units. Set $\mathcal{R}$sample $= \mathcal{R}_1$sample $\cup\, \mathcal{R}_2$sample $\cup \ldots \cup\, \mathcal{R}_G$sample. Modify the λ:s according to (3.16) below, where $g_{(k)}$ is the index for the group to which unit k belongs. The somewhat intriguing operation min(· , 1) is introduced for the following reason. Without it, it can happen, although only in exceptional cases, that one or more λ'_k becomes greater than 1.

$$\lambda'_k = \min\Big(n'_{g(k)}\cdot\lambda_k\Big/\sum_{j\in\mathcal{G}_{g(k)}}\lambda_j\,,1\Big) = \min\Big(n'_{g(k)}\cdot s_k\Big/\sum_{j\in\mathcal{G}_{g(k)}}s_j\,,1\Big) , \quad k = 1,2,\ldots,N. \tag{3.16}$$

After these λ-modifications, (3.12) works and also the confidence interval (3.3) with the variance estimator;

$$\hat{V}[\hat{t}(y)] = \sum_{g=1}^{G}\frac{n'_g}{n'_g-1}\cdot\sum_{k\in\mathcal{R}_g\text{sample}}\left(\frac{y_k}{\lambda'_k} - \sum_{j\in\mathcal{R}_g\text{sample}}\frac{y_j\cdot(1-\lambda'_j)}{\lambda'_j}\Big/\sum_{j\in\mathcal{R}_g\text{sample}}(1-\lambda'_j)\right)^2\cdot(1-\lambda'_i) . \tag{3.17}$$

Numerical computation of variance estimates by (3.17) is simplified by expansion of the squares, which leads to formulas analogous to (3.5) and (3.6). Moreover, formulas (3.13) and (3.17) show that both desires in (1.7) are met also under non-response adjustment.

4 Estimation accuracy

The accuracy of an estimator depends on its variance and possible bias. As already mentioned, $\hat{t}(y)$ in (3.1) is afflicted with some bias which, however, is negligible in almost all practical contexts. Therefore, we confine the discussion of estimation accuracy to estimator variance. In the sequel notions as "optimal", "better", etc. relate to $V[\hat{t}(y)]$. The following result from Rosén (1997b) explains why Pareto πps is of special interest.

Pareto πps is asymptotically (as $n \to \infty$) optimal among order πps schemes with the same size values and sample size, uniformly over study variables **y**. Moreover, optimality holds not only for estimation of totals, it holds for all "usual" types of characteristics, including ratios and domain characteristics. (4.1)

The result (4.1) tells that Pareto πps at least asymptotically performs better than other order πps schemes. However, there are other "competing" schemes, notably those mentioned at the end of Section 1, as well as small sample situations. To compare schemes we use the measure ***relative*** (to Pareto πps) ***variance increase*** (RVI);

$$\textit{RVI}(\text{for scheme } \mathcal{S}) = \frac{V[\hat{t}(y)] \text{ under } \pi\text{ps scheme } \mathcal{S}}{V[\hat{t}(y)] \text{ under Pareto } \pi\text{ps}} - 1 . \tag{4.2}$$

Factors that affect RVI significantly are: (i) The sampling fraction n/N. (ii) The relation between the size and study variables **s** and **y**. As regards this relation, ideal for πps is when **s** and **y** are exactly proportional, then $\tau(y)$ is in fact estimated without error. In that case the (**s**, **y**)-values lie along a straight line through the origin. In practice this never occurs, though, the (**s**, **y**)-values ***scatter*** more or less around a ***trend***, which is assumed to be increasing. Chief possibilities for trend type are ***proportional*** trend (= linear through the origin), ***convex*** and ***con-***

cave trend (when **y** grows relatively faster respectively slower than **s**). If the trend is flat or decreasing, πps sampling is non-favorable compared with SRS.

Table 4.1 presents examples of RVI-values for different schemes in certain sampling situations (N, **s** and **y**). As seen, the compared schemes are uniform and exponential order πps, Sunter πps, systematic πps(rfo) and πps(sfo) and SRS. Pareto πps is covered implicitly with all RVI = 0. SRS is included as a benchmark. The figures in the table, which come from Rosén (1997b), are based on Monte Carlo simulations. The considered sampling situations were generated with (**s**, **y**)-relation according to model (4.3) below, where α determines the trend shape (proportional for $\alpha = 1$, convex for $\alpha > 1$, concave for $\alpha < 1$) and σ the magnitude of scatter.

$$s_k = k, \quad y_k = c \cdot (s_k^{\alpha} + \sigma \cdot Z_k \cdot \sqrt{s_k^{\alpha}}), \quad c > 0, \quad \sigma \geq 0, \quad \text{the } Z_k \text{ being iid } N(0,1), \quad k = 1, 2, ..., N. \tag{4.3}$$

Table 4.1	**RVI - values (in %) for situations (4.3) with N = 100 and α, σ as stated below**											
	Uniform πps			**Exponential πps**			**Sunter πps**			**Systematic πps(rfo)**		
Sampling fraction	**$\alpha = 1.5$ $\sigma = 2$**	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**	**$\alpha = 1.5$ $\sigma = 2$**	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**	**$\alpha = 1.5$ $\sigma = 2$**	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**	**$\alpha = 1.5$ $\sigma = 2$**	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**
0.1	0.2	$4 \cdot 10^{-4}$	0.1	0.1	$1 \cdot 10^{-4}$	0.03	5.8	2.6	25	2.2	2.4	3.4
0.2	1.3	$2 \cdot 10^{-3}$	0.7	0.3	$6 \cdot 10^{-4}$	0.2	62	58	43	1.0	-3.5	-0.4
0.3	4.2	$6 \cdot 10^{-3}$	1.9	0.9	$2 \cdot 10^{-3}$	0.5	262	181	103	0.7	0.6	-0.6
0.4	12	0.02	4.7	2.3	$7 \cdot 10^{-3}$	1.7	653	280	141	21	-2.2	7.3
0.5	39	0.04	12	4.7	0.03	2.0	1457	401	180	29	-3.3	13

Systematic πps(sfo)			**SRS**		
$\alpha = 1.5$ $\sigma = 2$	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**	**$\alpha = 1.5$ $\sigma = 2$**	**$\alpha = 1$ $\sigma = 2$**	**$\alpha = 0.7$ $\sigma = 0.5$**
-74	+3.6	-34	1083	450	246
-75	-16	-52	1095	446	229
-73	+88	-9.2	1130	440	213
-86	-21	-52	1222	432	197
-73	+12	-47	1538	421	186

The main conclusions from the full collection of numerical findings in Rosén (1997b) are stated in (4.4) - (4.6).

> Under proportional (**s**, **y**)-trend the order πps schemes perform very similarly, with a slight edge for Pareto πps. Under non-proportional trend Pareto πps performs better than the other, the edge is small, though, for small sampling fractions but may be considerable for high ones. (4.4)

This result tells: (i) The optimality result (4.1) holds not only asymptotically, in essence it holds for quite small samples. (ii) Among order πps schemes, Pareto πps provides an "insurance without premium", it never performs worse than other order πps schemes and in some situations considerably better.

Next comparison is made with schemes outside order πps, which admit objective assessment of sampling error (as all order πps schemes do), with Sunter πps and systematic πps(rfo). The chief finding is as follows.

> (4.4) holds with "order πps" exchanged for "πps scheme which admits consistent variance estimation". (4.5)

The findings from comparison with systematic πps(sfo) (i.e. systematic πps with frame ordered by the **s**-values) are more complex. The strong side of this scheme is that it sets two variance reducing forces in action, πps and "implicit stratification". Its well-known weakness is that it does not admit assessment of sampling errors. The following rather confusing comparison picture arose.

> Systematic πps(sfo) often yields dramatically better estimation accuracy than Pareto πps, notably in situations with non-proportional (**s**, **y**)-trend,. In situations with fairly proportional (**s**, **y**)-trend the ranking between systematic πps(sfo) and Pareto πps seems to be erratic, they take turn to be best. (4.6)

To the best of our understanding it is hard to tell in advance which of Pareto πps and systematic πps(sfo) that is advantageous in a particular sampling situation.

5 Approximation accuracy, notably estimator bias

5.1 Some basic notions

As mentioned several times, since Pareto πps is based on asymptotic considerations desired and factual inclusion probabilities do not agree exactly, which in turn afflicts $\hat{\tau}(y)$ in (3.1) with some bias. When discussing these issues we use the performance measures in (5.1) and (5.2) below. First a comment on notation. Inclusion probabilities depend on population size N, size values $\mathbf{s} = (s_1, s_2, \ldots, s_N)$ and sample size n. In the sequel these parameters often are exhibited in notations like $\pi_k(n;N;\mathbf{s})$ and $\lambda_k(n;N;\mathbf{s})$.

Maximal absolute relative error for inclusion probabilities:

$$\Psi(n;N;\mathbf{s}) = \max_k |\pi_k(n;N;\mathbf{s})/\lambda_k(n;N;\mathbf{s}) - 1| \,. \tag{5.1}$$

Absolute relative estimator bias: $\mathrm{AREB}[\hat{\tau}(y)] = |E[\hat{\tau}(y)]/\tau(y) - 1| \,.$ (5.2)

The above concepts are related as follows, which is demonstrated e. g. in Rosén (2000 a);

$$\mathrm{AREB}[\hat{\tau}(y)] \le \Psi(n;N;\mathbf{s}) \cdot \left(\sum_{k=1}^{N} |y_k| \Big/ \tau(y) \right). \tag{5.3}$$

If the ***study variable is non-negative***, i.e. if $y_k \ge 0$, $k = 1, 2, \ldots, N$, which is the case in most practical surveys, the last factor in (5.3) equals 1, and (5.3) takes the simple form;

$$\mathrm{AREB}[\hat{\tau}(y)] \le \Psi(n;N;\mathbf{s}) \,. \tag{5.4}$$

The AREB bounds in (5.3) and (5.4) are often fairly conservative, as discussed in Rosén (2000 a).

$\Psi(n;N;\mathbf{s})$ is defined in terms of approximation accuracy for inclusion probabilities, a rather theoretical topic. However, by (5.3) and (5.4) Ψ also provides information about estimator bias, which makes it interesting also from survey practical point of view. Before entering bias questions we present some results about the asymptotic behavior of inclusion probabilities.

5.2 Asymptotic behavior of inclusion probabilities

Rosén (2000 a) proves that (5.5) below holds under general conditions for Pareto, uniform and exponential πps;

$$\max_k |\pi_k(n;N;\mathbf{s})/\lambda_k(n;N;\mathbf{s}) - 1| \text{ is (at most) of order } O(\log n/\sqrt{n}) \,. \tag{5.5}$$

Results of type (5.5) are used to study the asymptotics of inclusion probabilities, in the usual frame - work for finite population asymptotics: A sequence of populations with sizes tending to infinity. In particular, the result (5.6) below is proved for the schemes mentioned above. It tells that desired and factual inclusion probabilities agree asymptotically. It is conjectured that (5.6) holds generally for any order πps scheme.

$$\pi_k(n;N;\mathbf{s})/\lambda_k(n;N;\mathbf{s}) \to 1 \text{ as } n \to \infty \text{ (and hence also } N \to \infty) \,. \tag{5.6}$$

5.3 On estimator bias

5.3.1 Factors that affect the bias magnitude

When considering use of Pareto πps in practice, a crucial question for the statistician is;

Will the Pareto πps point estimator bias be negligible in my particular survey situation? (5.7)

In search for answers to (5.7), available theoretical bounds of type (5.5), regrettably only add to the general experience that theoretical error bounds seldom are sharp enough to yield practically valuable information on the small sample performance of a large sample procedure. The bound $O(\log n/\sqrt{n})$ in (5.5) is too crude for that purpose. In our understanding, practically useful information can only be gained by numerical investigations, exact computations or/and Monte Carlo simulations. The results presented in the sequel come from Aires & Rosén (2000), which reports on an extensive numerical study of exactly computed Pareto πps inclusion probabilities.

Answers to (5.7) are with necessity a bit involved, since the bias depends on several factors. The study variable is of course one of them. On this point we confine to the case with non - negative study variables, which is the most common in practice. By (5.4), Ψ can then be interpreted as an AREB bound. Other factors that affect whether the bias is negligible or not are: (i) The tolerance limit for "negligible". (ii) The variation of the size values. (iii) The population size. (iv) The sample size. These factors are discussed below.

Tolerance limit for negligibility

There is of course no unanimous answer to how large a "negligible" bias may be. This depends on the intended use of the statistic and on the magnitude of sampling errors and other survey errors. We believe that most survey statisticians regard 1%, and even 2%, as a negligible relative bias.

Dependence on size values

When all size values are equal, Pareto πps is SRS with $\pi_k = \lambda_k = n/N$. Hence, for bias to be at hand the size values must show variation. In the following, the size values **s** are presumed to be ***normed*** so that average size is 1, i.e. so that (5.8) below is met. A normed **s** is referred to as a ***size pattern***.

$$(1/N) \cdot \sum_{k=1}^{N} s_k = 1 . \tag{5.8}$$

The maximal and minimal normed s-values are denoted s_{max} and s_{min}. The ***size pattern range*** is specified by the interval $[s_{min}, s_{max}]$. Another aspect on s-value variation is the ***size pattern shape***, which concerns how size values spread over $[s_{min}, s_{max}]$. Following Aires & Rosén (2000), where precise definitions are given, three "extremal" shape types are considered. (i) The s-values are fairly ***evenly spread*** over $[s_{min}, s_{max}]$. (ii) The majority of s-values lie in the ***middle*** of $[s_{min}, s_{max}]$. (ii) The majority of s-values lie at the ***boundaries*** of $[s_{min}, s_{max}]$.

Figures 5.1 illustrates how $\Psi(\cdot; N; \mathbf{s})$-sequences may differ for different pattern shapes with the same N, s_{min} and s_{max}. In particular it illustrates the following general circumstance. When the sample size is not "very small", the boundary shape is most adverse to good agreement between desired and factual inclusion probabilities.

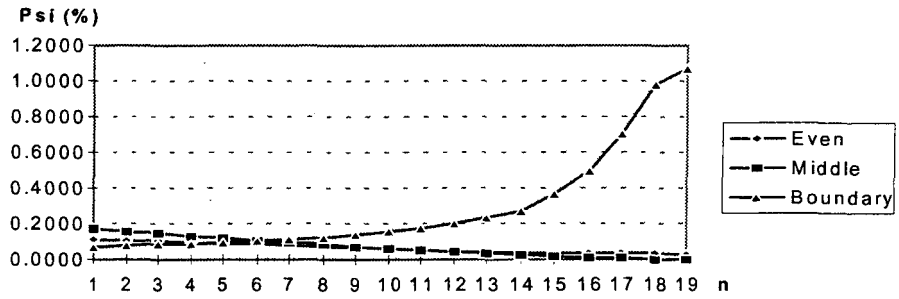

Figure 5.1 Psi-sequences for N=100, s_{min} =0.1, s_{max}=5

The statements in (5.9) and (5.10) below are based on our experience of size value patterns met in practice.

We believe that s_{max} seldom is larger than 5 and that s_{min} seldom is smaller than 0.1. (5.9)

The boundary shape is very unusual in practice. Most practical size pattern shapes resemble the even spread shape, in the sequel referred to as "lying ***in the vicinity of even spread***". (5.10)

Some background for (5.9) and (5.10) is as follows. (i) The surveyor disposes of the size values, preliminary values may be modified. If the frame comprises units with very small preliminary size values, such units are often either definition-wise excluded from the survey population or given larger **s**-values in the sample selection.

(ii) If size values vary very much over the entire population, there are often subject matter grounds for stratification by size before sampling, followed by drawing independent samples from the strata. (An example is an enterprise survey with "number of employees" as size. It is usually natural to divide into strata of type "very big", "big" and "small" enterprises. Mostly the "very big" stratum is totally inspected.) The strata then take population roles, and s_{max} and s_{min} in the strata are usually considerably smaller / larger than in the entire population.

Table 5.1 below introduces, for later use, a broad categorization of size value patterns.

Table 5.1 Some categories of size value patterns

	Category A	Category B	Category C	Category D
Size pattern shape	In the vicinity of *even spread*.	In the vicinity of *even spread*.	No restriction.	No restriction.
s_{max}	≤ 5	≤ 10	≤ 5	≤ 10
s_{min}	≥ 0.1	≥ 0.05	≥ 0.1	≥ 0.05
Comments on occurrence in practice	Most practical situations are believed to fall in Category A.	"Normal" pattern shape, while s_{max} and/or s_{min} may be extreme	"Normal" s_{max} and s_{min}, while shape may be extreme (e.g. of boundary type).	Pattern shape as well as s_{max} and/or s_{min} may be extreme.

Dependence on population size

The numerical findings show that for given size value pattern **s** and sample size n, $\Psi(n; N; \mathbf{s})$ decreases as the population size N increases, i.e. desired and factual inclusion probabilities come closer to each other.

Dependence on sample size

The assumption $\lambda_k < 1$ constrains sample sizes as stated in (5.11), where [·] denotes integral part, and - "less than". The quantity n_m is called the ***maximal sample size***, and an n which satisfies (5.11) is said to be ***admissible.***

$$n \le n_m = n_m(N; \mathbf{s}) := [N/s_{max} -]. \tag{5.11}$$

Since Pareto πps is based on asymptotic considerations, one expects in the first round that conditions for small Ψ (hence, for small bias) would be of the type "provided that n is *at least* ...". However, as illustrated in Figure 5.1, conditions for small Ψ, which encompass all types of size pattern shapes, inclusive the unpleasant boundary shape type, rather are of the form "provided that n is *at most* ..." (For pattern shapes in the vicinity of even spread, though , Ψ typically decreases as n increases .) This aspect is handled technically as follows. An α is specified, $0 < \alpha < 1$, and is used to determine an ***α-maximal sample size*** $n_{m,\alpha}$ as follows;

$$n_{m,\alpha}(N;\mathbf{s}) := [\alpha \cdot N/s_{max}]. \tag{5.12}$$

Conditions to the effect "sample size is *at most* ..." are formulated by stating that n must not exceed $n_{m,\alpha}$.

5.3.2 Conditions which imply negligible estimator bias

The numerical findings in Aires & Rosén (2000) are condensed in Table 5.2 below. More detailed information is provided in the full report. Population sizes smaller than 25 were not considered in that study.

Table 5.2. Sample sizes that imply negligible bias

N = population size, α states that sample size must not exceed $n_{m,\alpha}$ in (5.12). n_0 specifies a sufficiently large sample size, under the α-restriction, for negligible bias with specified tolerance. Study variables are presumed to be non-negative.

Size Pattern category See Table 5.1	Tolerance limit for negligibility								
	2%			1%			0.5%		
	N	α	n_0	N	α	n_0	N	α	n_0
A	≥ 25	1	1	≥ 40	1	1	≥ 80	1	1
				[25, 40)	1	3	[50, 80)	1	3
							[40, 50)	1	4
B	≥ 80	1	1	≥ 80	1	1	≥ 125	1	1
	[25, 80)	1	2	[40, 80)	1	3	[100, 125)	1	3
							[80, 100)	1	4
C	≥ 100	1	1	≥ 125	1	1	≥ 175	1	1
	≥ 80	0.9	1	≥ 100	0.9	1	≥ 150	0.9	1
	≥ 40	0.8	1	≥ 80	0.8	1	≥ 100	0.8	1
	≥ 25	0.5	1	≥ 40	0.5	1	≥ 80	0.5	1
D	≥ 150	0.9	1	≥ 175	0.8	1	≥ 125	0.5	1
	≥ 125	0.8	1	≥ 80	0.5	1	[100, 125)	0.5	3
	≥ 80	0.5	1						
	[50, 80)	0.5	2						

Earlier remarks imply that the sufficient sample sizes n_0 in Table 5.2 in most practical situations are conservative and, hence, "overly safe". In particular, one should not conclude that the bias necessarily is larger than "guaranteed" for sample sizes that are smaller then stated n_0 - values. However, even with the above conservative bounds the conclusion is that the bias in almost all practical situations is negligible for all admissible sample sizes.

6 On sample coordination and adjustment for overcoverage

Ohlsson (1990, 1995, 1998) emphasizes that uniform order πps has the attractive properties which are discussed below. These properties are shared by all order πps schemes, hence also by Pareto πps.

For an order πps scheme positive ***coordination of samples*** (to achieve great sample overlap) drawn at different occasions in time from the "same" (but updated) frame is obtained by associating ***permanent random numbers*** to the frame units, to be used at successive draw occasions, i.e. by letting the R_k in Definitions 2.1 and 2.2 be permanent. Similar technique can be used for positive or negative coordination of simultaneously drawn samples to different surveys. Negative coordination is achieved for example if R_k in one sample selection is exchanged for $1 - R_k$ in another selection.

When the frame contains ***overcoverage*** (out - of - scope units), a sample of predetermined size from the (unobservable) list of in - scope units can be selected as follows. Order the frame units by the Q : s in Definition 2.2, and start observing them in that order. Exclude successively encountered out - of - scope units until a sample of in-scope units of prescribed size is obtained. This procedure yields an order πps sample from the in - scope units. One looses full control of the inclusion probabilities, though, since the size sum over the in - scope list is unknown. However, if the task is to estimate a ratio $\tau(\mathbf{y})/\tau(\mathbf{x})$, as is the case in e.g. price index surveys, this does not matter since the unknown size sums in the estimates of nominator and denominator cancel. In the general case the unknown size sum is readily estimated.

7 Pareto πps as component in optimal sampling - estimation strategies

So far available auxiliary information has consisted of size values $\mathbf{s} = (s_1, s_2, \dots, s_N)$. Here we turn to a more elaborate situation, with auxiliary information in conjunction with a superpopulation model. The general version of the simple model below provides background for generalized regression estimation, as described in Chapter 6 in Särndal et al. (1992). The study variable **y** relates to auxiliary data $x_1, x_2, \dots, x_N$ (here one - dimensional) according to the following superpopulation model;

$$y_k = \beta \cdot x_k + \varepsilon_k, \quad k = 1, 2, \dots, N, \tag{7.1}$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ satisfy the following conditions, with $\mathcal{E}, \mathcal{V}$ and $\mathcal{C}$ for superpopulation expectation, variance and covariance: $\mathcal{E}[\varepsilon_k] = 0$, $\mathcal{V}[\varepsilon_k] = \sigma_k^2$ and $\mathcal{C}[\varepsilon_k, \varepsilon_l] = 0$, $k \neq l$, $k, l = 1, 2, \dots, N$. The parameters $\sigma_1, \sigma_2, \dots, \sigma_N$ are part of the auxiliary information, and are regarded as known modulo a proportionality factor.

Some notation. Subscript HT signifies Horvitz - Thompson estimators. Algebraic operations on variables shall be interpreted as component - wise. For $\mathbf{y} = (y_1, y_2, \dots, y_N)$ and $\mathbf{x} = (x_1, x_2, \dots, x_N)$, $\mathbf{y} \cdot \mathbf{x} = (y_1 \cdot x_1, y_2 \cdot x_2, \dots, y_N \cdot x_N)$, $\mathbf{y}/\mathbf{x} = (y_1/x_1, y_2/x_2, \dots, y_N/x_N)$, $\mathbf{x}^2 = (x_1^2, x_2^2, \dots, x_N^2)$.

Next we state a result due to Cassel et al (1976). An admissible (sampling - estimation) ***strategy*** is a pair $[P, \hat{\tau}(\mathbf{y})]$ of a sample design P and a linear, design unbiased estimator $\hat{\tau}(\mathbf{y})$. They showed that ***optimal strategies*** relative to minimization of the anticipated estimator variance $\mathcal{E}(V[\hat{\tau}(\mathbf{y})])$ are characterized by the following properties;

P is a πps scheme with size values proportional to $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$. (7.2)

The estimator $\hat{\tau}(\mathbf{y})$ is of the form $\tau(\mathbf{y})_{HT} + \beta \cdot [\tau(\mathbf{x}) - \tau(\mathbf{x})_{HT}]$. (7.3)

Since the Cassel et al. paper much effort has been devoted to the estimator part of the optimal strategy, leading to the generalized regression estimator (GREG);

$$\hat{\tau}(\mathbf{y})_{GREG} = \hat{\tau}(\mathbf{y})_{HT} + \hat{B} \cdot [\tau(\mathbf{x}) - \hat{\tau}(\mathbf{x})_{HT}] \quad \text{where} \quad \hat{B} = \tau(\mathbf{y} \cdot \mathbf{x}/\boldsymbol{\sigma}^2)_{HT} / \tau(\mathbf{x}^2/\boldsymbol{\sigma}^2)_{HT}. \tag{7.4}$$

However, only little attention has been paid to the design part, the πps scheme. A possible reason may be shortage of πps schemes with good properties. Since Pareto πps provides a nice πps scheme, at least in the author's opinion, it is of interest to revisit the optimal strategy problem by studying the performance of the strategy [Pareto πps($\boldsymbol{\sigma}$), $\tau(\mathbf{y})_{GREG}$]. The Cassel et al. result gives background for the conjecture that this strategy is close to "universally optimal" under the above superpopulation model.

Since Pareto πps does not admit exact HT - estimation, the GREG estimator in (7.4) has to be modified a bit. In the sequel πps(**s**) indicates estimation in accordance with (3.1). The modified GREG estimator is;

$$\hat{\tau}(\mathbf{y})_{GREG}^{\pi ps(\boldsymbol{\sigma})} = \hat{\tau}(\mathbf{y})_{\pi ps(\boldsymbol{\sigma})} + \hat{B} \cdot [\tau(\mathbf{x}) - \hat{\tau}(\mathbf{x})_{\pi ps(\boldsymbol{\sigma})}] \quad \text{where} \quad \hat{B} = \hat{\tau}(\mathbf{y} \cdot \mathbf{x}/\boldsymbol{\sigma}^2)_{\pi ps(\boldsymbol{\sigma})} / \hat{\tau}(\mathbf{x}^2/\boldsymbol{\sigma}^2)_{\pi ps(\boldsymbol{\sigma})}. \tag{7.5}$$

Theoretical and numerical results on comparison between [Pareto πps($\boldsymbol{\sigma}$), $\tau(\mathbf{y})_{GREG}^{\pi ps(\boldsymbol{\sigma})}$] and various "competing" strategies are presented in Rosén (2000 b). To make a long story short, the findings support the conjecture that the strategy in fact is close to being universally optimal when the superpopulation model is correct.

8 Summarizing conclusions

In Section 4 is argued that Pareto πps should be preferred among πps schemes which admit objective assessment of sampling error. The choice of πps design then stands between Pareto πps and systematic πps(sfo) (= with frame ordered by size). The latter scheme has the following pros and cons. Often it yields more accurate point estimates than Pareto πps, but the opposite may also occur, and it is hard to tell in advance which will be the case in a specific survey situation. On the (very) negative side stands that systematic πps(sfo) deprives assessment of sampling error as well as sample coordination by (permanent) random numbers. We believe that under these premises Pareto πps is seen as the best alternative in most practical survey contexts.

References

Aires, N. (1999). Algorithms to Find Exact Inclusion Probabilities for Conditional Poisson Sampling and Pareto πps Sampling Designs. *Methodology and Computing in Applied Probability* **4** *459-473.*

Aires, N. & Rosén, B. (2000). On Inclusion Probabilities and Estimator Bias for Pareto πps Sampling. Statistics Sweden R&D Report 2000:2.

Cassel, C. M., Särndal C.-E. & Wretman, J. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63** *615-620.*

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Volume II.* Wiley, New York.

Ohlsson, E. (1990). Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price index. Statistics Sweden R&D Report 1990:6.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In *Business Survey Methods.* John Wiley & Sons, New York.

Ohlsson, E. (1998). Sequential Poisson Sampling. *Journal of Official Statistics*, **14** *149-162.*

Rosén, B. (1997a). Asymptotic Theory for Order Sampling. *J Stat Plan Inf*, **62** *135-158.*

Rosén, B. (1977b). On Sampling with Probability Proportional to Size. *J Stat Plan. Inf.*, **62** *159-191.*

Rosén, B. (2000a). On Inclusion Probabilities for Order Sampling. *J Stat Plan. Inf., 62 117–143.*

Rosén, B. (2000b). Generalized Regression Estimation and Pareto πps. Statistics Sweden R&D Report 2000:5.

Rosén, B. & Lundqvist P. (1998). Estimation from Order πps Samples with Non - Response. Statistics Sweden R&D Report 1998:5.

Saavedra (1995). Fixed Sample Size PPS Approximations with a Permanent Random Number. *1995 Joint Statistical Meetings American Statistical Association.* Orlando, Florida.

Sunter A. B. (1977). List Sequential Sampling with Equal or Unequal Probabilities without Replacement. *Applied Statistics* **26** *261-268.*

Särndal, C-E, Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer, New York.

Förteckning över utkomna R&D Reports

R&D Reports är en för IT-enheten och Metodenheten gemensam publikationsserie, som 1988-01-01 ersatte de tidigare "gula" och "gröna" serierna. I serien ingick fram till årsskiftet 1992-93 även **Abstracts** (sammanfattning av metodrapporter från SCB).

Reports published during 1998 and onwards:

1998:1 Preliminär statistik: Nybyggnadskostnader - en simuleringstudie *(Catarina Elffors)*

1998:2 On Inclusion Probabilities for Order Sampling *(Bengt Rosén)*

1998:3 On the Stratification of Highly Skewed Populations *(Dan Hedlin)*

1998:4 Bortfallsbarometer nr 13 *(Per Nilsson, Antti Ahtiainen, Stefan Berg, Mats Bergdahl, Monica Rennermalm och Marcus Vingren)*

1998:5 Estimation from Order πps Samples with Non - Response *(Bengt Rosén and Pär Lundqvist)*

1998:6 On variance estimation for measures of change when samples are coordinated by a permanent random numbers technique *(Lennart Nordberg)*

1999:1 Täckningsproblem i Registret över totalbefolkning RTB. Skattning av övertäckning med en indirekt metod *(Jan Qvist)*

1999:2 Bortfallsbarometer nr 14 *(Per Nilsson, Antti Ahtiainen, Mats Bergdahl, Tomas Garås, Jan Qvist och Charlotte Strömstedt)*

1999:3 Att mäta statistikens kvalitet *(Claes Andersson, Håkan L. Lindström och Thomas Polfeldt)*

2000:1 Kalibrering av vikter – beskrivning av tekniken och de SCB-fall den prövats i *(Sixten Lundström et al)*

2000:2 On Inclusion Probabilities and Estimator Bias for Pareto πps Sampling *(Nibia Aires and Bengt Rosén)*

2000:3 Bortfallsbarometer nr 15 *(Per Nilsson, Ann-Louise Engstrand, Sara Tångdahl, Stefan Berg, Tomas Garås och Arne Holmqvist)*

2000:4 Bortfallsanalys av SCB-undersökningarna HINK och ULF *(Jan Qvist)*

2000:5 Generalized Regression Estimation and Pareto πps *(Bengt Rosén)*

2000:6 A User's Guide to Pareto πps Sampling *(Bengt Rosén)*

ISSN 0283-8680

Tidigare utgivna **R&D Reports** kan beställas genom Katarina Klingberg, SCB, MET, Box 24 300, 104 51 STOCKHOLM (telefon 08-506 942 82, fax 08-506 945 99, e-post katarina.klingberg@scb.se).
R&D Reports from 1988-1997 can - in case they are still in stock - be ordered from Statistics Sweden, attn. Katarina Klingberg, MET, Box 24 300, SE-104 51 STOCKHOLM (telephone +46 8 506 942 82, fax +46 8 506 945 99, e-mail katarina.klingberg@scb.se).

SCB-Tryck, Örebro 2000.12