R&D Report **2002:2**
Research – Methods – Development

Statistiska centralbyrån Statistics Sweden

Model-based calibration for survey estimation, with an example from expenditure analysis[1]

by

Claes Cassel, Peter Lundquist
and Jan Selén

Statistics Sweden

---

[1] This paper is based on the results of a development project at Statistics Sweden, supported by Eurostat under the heading SUP.COM 1996, Lot 23. Detailed results of this project are given in Cassel et al (1997).

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 2002:2. Model-based calibration for survey estimation, with an example from expenditure analysis / Claes Cassel m.fl.
Digitalt skapad fil, anpassad efter de digitaliserade delarna i serien. Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-2002-X101OP0202

R&D Report **2002:2**
Research – Methods – Development

Model-based calibration for survey estimation, with an example from expenditure analysis[1]

by

Claes Cassel, Peter Lundquist and Jan Selén

Statistics Sweden

---

[1] This paper is based on the results of a development project at Statistics Sweden, supported by Eurostat under the heading SUP.COM 1996, Lot 23. Detailed results of this project are given in Cassel et al (1997).

R&D Report 2002:2

Research - Methods - Development

Model-based calibration for survey estimation, with an example from expenditure analysis[1]

Från trycket	Oktober 2002
Producent	Statistiska centralbyrån, *Statistics Sweden,* metodenheten Box 24300, *SE*-104 51 STOCKHOLM

Förfrågningar

Claes Cassel,
claes.cassel@scb.se
Telefon 08- 506 943 71

Peter Lundquist
peter.lundquist@scb.se
Telefon 08 – 506 949 18

Jan Selén
jan.selen@scb.se
Telefon 08 – 506 949 39

ISSN 0283-8680

Printed in Sweden
SCB-Tryck, Örebro 2002

---

[1] This paper is based on the results of a development project at Statistics Sweden, supported by Eurostat under the heading SUP.COM 1996, Lot 23. Detailed results of this project are given in Cassel et al (1997).

1. Introduction

Calibration is, basically, a method to improve estimation in survey sampling when auxiliary information is available. Calibration works by modifications of survey weights such that known population characteristics, in practice totals, are reproduced from the sample. In a sample survey of a human population natural auxiliary variables are age and sex. The population structure by age and sex is often known from population statistics and by proper modifications of the survey weights, the population structure may be exactly reproduced by the sample. For variables in the survey correlated with the auxiliary information, higher precision in estimated totals is obtained by the new weights usually. The favourable situation for calibration is access to registers, such that the units in the sample are identified in the registers. Then auxiliary data is easily obtained as well as population aggregates, such as totals. Without registers, the possibility to improve estimation by calibration has been regarded as limited. In this paper, we will try to widen the applicability of calibration by considering a wider concept of auxiliary information. We will consider estimated auxiliary aggregates. Specifically we will see how a major variable can be used to strengthen the estimation of characteristics of other variables by the use of statistical models, such as econometric models. By a major variable is meant a variable measured with better precision than other variables, an example is total expenditure compared to expenditure in specific categories in a household budget survey. The approach is called model-based calibration.

Calibration has its origin in the raking procedures introduced by Deming and Stephan (1940). In later years, the calibration technique has received considerable attention from official statistical authorities. The interest for the approach has grown since Deville and Särndal (1992) showed the asymptotic equivalence of calibration to the generalised regression estimator, Cassel, Särndal and Wretman (1976), thereby providing a way to establish the statistical properties of calibration estimators.

In the following, a short presentation of the generalised regression estimator and of calibration is given before model-based calibration is introduced. Estimated auxiliary totals are introduced in the generalised regression estimator. Some alternative definitions of these estimated totals are given. The use of a major variable and a statistical model to estimate auxiliary totals is a main alternative. Estimator properties for the model-based calibration estimator are discussed and approximated. An application estimating average consumer expenditures is presented, utilising Swedish household budget surveys and a model of consumer demand, the AIDS-model. A discussion concludes.

2. Generalised regression estimation

In this and the following section we will recapitulate generalised regression estimation and calibration, respectively. These will serve as prerequisites for the model-based calibration.

Let the N elements in the population U be identified by subscript k;

$$U=\{1,\ldots,k,\ldots,N\}$$

Denote the variable under study, say consumption, by z and suppose that we wish to estimate the total summing over the entire population U,

$$T_z = \sum_U z_k \text{ ,}$$

from a sample s selected from U with probability $p(s)$ and whose elements have the inclusion probabilities π_k .

The standard Horvitz-Thompson estimator t_z is defined as

$$t_z = \sum_{k \in s} \left(\frac{1}{\pi_k} \right) z_k = \sum_{k \in s} d_k z_k$$

where the inverse probability d_k is the (expansion) weight for element k.

Suppose now that there exists some auxiliary information x_k with a known total

$$T_x = \sum_U x_k$$

This is utilised in the generalised regression estimator (GREG)

$$t_{GR}(z) = t_z + (T_x - t_x)' \hat{B}_{zx}$$

where t_x is the Horvitz-Thompson estimator of T_x from the sample, and $\hat{B}_{zx}$ is the regression coefficient

$$\hat{B}_{zx} = \left(\frac{\sum_{k \in s} d_k x_k z_k}{\sum_{k \in s} d_k x_k^2} \right)$$

If x and z are correlated then this is exploited in GREG such that the precision of $t_{GR}(z)$ is better than that for the simpler t_z .

3. Calibration

The original idea in calibration is to modify the weights so that known totals are reproduced for the sample. More precisely for one known total T_x we calibrate by constructing weights w_k such that

$$t_x^{'} = \sum_{k \in s} w_k x_k = T_x \,,$$

where the new weights w_k are as close as possible to the old weights d_k.

Minimising a quadratic distance measure

$$\sum_{k \in s} \left(w_k - d_k \right)^2 / d_k \,,$$

the solution giving the new weights is

$$w_k = d_k \left(1 + \frac{(T_x - t_x)}{T_s} x_k \right),$$

where

$$T_s = \sum_{k \in s} d_k x_k^2$$

Deville and Särndal (1992) showed the asymptotic equivalence of calibration to the generalised regression estimator, Cassel, Särndal and Wretman (1976), thereby providing a way to establish the properties of calibration estimators.

There are some distinct advantages with calibration. Auxiliary information is incorporated in the weights. The obtained estimates are "consistent" with known information; usually this applies to totals. There is a correction for non-response, if non-response is related to the auxiliary information, and there is a reduction of variances for totals if variables are correlated.

Among the problems of the approach is that negative weights may appear. This is against our intuition. Modifications are possible at the expense of a more complicated procedure. Variances in ratios for the estimation of e.g. averages may increase for a calibrated estimator.

4. Model based calibration

The modification of the GREG estimator is obtained if an estimator or predictor t_x^* of T_x is used in the GREG estimator, thus

$$t_c(z) = t_z + (t_x^* - t_x)'\hat{B}_{zx}$$

There are many possible definitions for this extension, depending on the definition t_x^*, here we will consider the main alternatives

- $t_x^* = R_m t_y$, where a ratio R_m is given by a supporting statistical model, for example an econometric model as in the application below

- $t_x^* = \hat{R} t_y$, an estimate of the average ratio obtained from the current sample and constructed using a so called separate ratio estimator.

A basic idea behind these choices is that we try to borrow support from the estimate of a major variable t_y in the sample, hopefully measured with better precision than t_z. The specifications also rest on the assumption that units in subgroups or domains are similar as regards their ratio R_m or $\hat{R}$.

To more concrete, let us consider household expenditure estimation, according to the application more fully described in the following sections. Suppose x_k denotes expenditure for a certain category, for example for clothing, and y_k total expenditure for household k. The ratio (x_k / y_k) thus is the proportion of total expenditure spent on clothing by household k. If we consider households of the same type, for example cohabitants with two children, we may expect the proportion of expenditure spent on clothing to be roughly equal within this group, at least over a longer period. Then we may also expect the that the average $\hat{R}$, where

$$\hat{R} = \frac{\left(\sum_s \left(\frac{x_k}{y_k}\right) / \pi_k\right)}{\left(\sum_s 1/\pi_k\right)}$$

is of the separate ratio type, carries auxiliary information. We thus form t_x^* as

$$t_x^* = \sum_s \left(\frac{y_k \hat{R}}{\pi_k}\right) = \hat{R} \sum_s \left(\frac{y_k}{\pi_k}\right) = \hat{R} t_y .$$

Turning to the properties of the estimator; the expected value of $t_c(z)$ generally is

$$E\left(t_c(z)\right) = E[t_z + (t_x^* - t_x)' \hat{B}_{zx}] \approx T_z + [E(t_x^*) - T_x]' B_{zx} ,$$

where the latter product indicate a possible bias. If the model is true such that the expression within the parentheses is zero then there is no bias; therefore the estimator is model dependent. $B_{zx} = 0$ also results in a zero bias.

Specifically for $t_x^* = \hat{R} t_y$ we have

$$E(t_x^*) = T_y \left(\frac{1}{N} \sum_1^N \left(\frac{x_k}{y_k}\right)\right) = T_y \overline{R}$$

where $\overline{R}$ is the average proportion. The bias then becomes

$$\left(T_y \overline{R} - T_x\right)' B_{zx}$$

which is small if $x_k \approx \overline{R} y_k$.

The variance of $t_c(z)$, $V(t_c(z))$, is written

$$V\left(t_z - (t_x - t_x^*)' \hat{B}_{zx}\right) =$$
$$= V\left(\sum_s \left\{\frac{z_k - (x_k - x_k^*)' \hat{B}_{zx}}{\pi_k}\right\}\right) = V\left(\sum_s \left(\frac{e*}{\pi_k}\right)\right)$$

where

$$e_k^* = (z_k - (x_k - x_k^*) \hat{B}_{zx})$$

and

$$x_k^* = \begin{cases} \hat{R} y_k \\ R_m y_k \end{cases}$$ for the two main alternatives above respectively.

Under the approximations $\hat{B}_{zx} \approx B_{zx}$ and $\hat{R} \approx \overline{R}$ we get

$$e_k^* = (z_k - (x_k - \overline{R} y_k) B_{zx})$$

An alternative way of writing the estimator is

$$t_c(z) = \sum_s \left(\frac{e_k^*}{\pi_k} \right) = \sum_U I_k \left(\frac{e_k^*}{\pi_k} \right),$$

summing over the entire population, and where I_k is an indicator which indicates whether k is included in the sample or not.

To estimate the variance it would be possible to use the expression

$$\hat{V}(t_c(z)) = \sum \sum \frac{1}{\pi_{kl}} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) e_k^* e_l^*,$$

provided that the sample is large enough for the approximations $\hat{B}_{zx} \approx B_{zx}$ and $\hat{R} \approx \overline{R}$

Note finally that when the variable under study z is identical to the auxiliary variable x we get

$$t_c(z) = t_x + (t_x^* - t_x)1 = t_x^*$$

and under the model $t_x^* = R_m t_y$ we find that

$$E\left(t_c(z)\right) = R_m T_y$$
$$V\left(t_c(z)\right) = R_m^2 V(t_y)$$

thus the variance decreases considerably if the ratio is small.

The information from the model assimilated in the weights. In an extended calibration we may include both register based information and model based information at the same time.

5. A model of consumer demand

The impetus for our efforts to explore this new alternative for the improvement of estimation was the difficulties with costly household budget surveys. Our application therefore concerns consumer expenditure. For a model-based calibration regarding consumption, there are different alternatives as regards the model choice. Here we will utilise the Almost Ideal Demand System; the AIDS-model for short. This is an elaborate model based on utility maximisation, which initially was specified for individual households observed over time, Deaton and Muellbauer (1980). In Deaton (1985) the model is specified for cross sections of domain means and thus interpretable as a modelling of the behaviour of an average household in a domain. The model can be used to predict the proportion of expenditure spent on certain goods. The model is specified as a system of equations with typical member

$$P_{gdt} = \alpha_g + \partial_d + \sum_{g_0} \gamma_{g_0} I_{g_0 t} + \beta_g x^*_{dt} + \varepsilon^*_{gdt} \qquad (1)$$

where P_{gdt} is the proportion of expenditure spent on category *g* for domain or type of household *d* at time *t*. Subscripts g and g_0 denote categories of expenditure and

$$x^*_{dt} = \frac{1}{n_{dt}} \sum_{h \in d} \log\left(\frac{x_{ht}}{y_{ht} I^*_t} \right)$$

where x_{ht} is the total amount of expenditure and y_{ht} is the number of consumer units for household h at time t. The overall price index I^*_t is in our application approximated with Stones index at time *t;* $\log I^*_t = \sum_g \overline{P}_{gt} \log I_{gt}$. Further notations

I_{gt} or $I_{g_0 t}$ is the value of the price index for category *g* (or g_0) at time t

α_g, β_g and γ_{g_0} are parameters, ∂_d is a domain effect and $\varepsilon^*_{gdt} = \frac{1}{n_{dt}} \sum_{h \in d} \varepsilon_h$ is the residual.

Even though the model is defined on a domain level it can be deduced from the micro level assuming that the parameters α, β and γ are the same for households within the domain *d*. Using dummy variables for the domains it can be estimated on the entire sample of households. For the estimation restrictions on α, β and γ are required, see Deaton and Muellbauer (1980).

The AIDS-model (1) is estimated for Swedish HBS data using the three surveys[2] 1985, 1988 and 1992. Seven categories of expenditure were identified; food, clothing and footwear, furniture and household articles, transportation, recreation and cultural services, spirits and tobacco and other expenditures. In the study we used month as time-unit, hence T=36. The domains ∂_d were type of household, specified in 10 categories according to table 1.

---

[2] Samples drawn from the register of the total population. The initial sample sizes sums to about 6000 households. The response rates were about 63 percent.

Table 1: **Household (hh) types and number of households in the analyses**

		n
SFNOC64	Single woman, -64 years, no children	666
SMNOC64	Single men, -64 years, no children	804
SNOC65_	Single persons, 65- years, no children	314
CONOC64	Cohabitant hh:s , -64 years, no children	1880
CONOC65_	Cohabitant hh: , 65- years, no children	920
CO1CH	Cohabitant hh:s, 1 child	1296
CO2CH	Cohabitant hh:s, 2 children	2343
CO3_CH	Cohabitant hh:s, 3- children	1250
ONOCH	Other hh:s, no children	939
OCH	Other hh:s, 1- children	1512

Thus 359 aggregated observations were available for the analysis. Some of these consisted of rather few sampled households; at one time-point an empty type of household occurred and some observations were based on less than ten initial households.

The residuals in (1) are aggregated means for the domains at time *t*. We assume that

$$Var\left(\varepsilon^*_{gdt}\right) = \frac{\sigma^2}{n_{dt}} \qquad d=1,2,\ldots,10 \text{ and } t=1,2,\ldots,36$$

$$Cov\left(\varepsilon^*_{g_1 d_1 t}, \varepsilon^*_{g_2 d_2 t}\right) = 0 \quad \text{when } g_1 \neq g_2 \text{ or } d_1 \neq d_2$$

where n_{dt} is the number of households in domain d at time t. Heteroscedasticity caused by the aggregation was taken into account.

For the estimation we also need the total expenditure for the households and price indexes for the expenditure categories as well as for the total expenditure.

In Table 2 below estimation results are shown. For some categories of expenditures such as for food, -spirits and tobacco- and other expenditures the coefficient of determination is high. As a reference group we selected the group consisting of cohabitant households without children -64 years. The other household types are compared to the reference group. A negative value of a parameter indicates a lower proportion of the specific group of expenditure and a positive value indicates a higher proportion. The standard errors can be used for testing the hypothesis that the parameter estimates are zero.

Table 2: **Parameter estimates for the AIDS-model estimated on the Swedish HBS**

	Food		**Clothing/ footwear**		**Furniture/ Househ. articles**		**Transport**		**Recreation/ cultural services**		**Spirits/ tobacco**		**Other expenditures**	
R^2	0.75		0.54		0.42		0.54		0.51		0.64		0.75	
Root MSE	0.081		0.074		0.056		0.111		0.100		0.039		0.124	
Dep mean	0.239		0.066		0.043		0.122		0.142		0.029		0.358	
Variables	Param. est.	Std. error	Param. est.	Std. error	Param. est.	Std. error	Param. est.	Std. error	Param. est.	Std. error	Param. est.	Std. error	Param. est.	Std. error
α_g	1.042	0.055	-0.197	0.047	-0.248	0.043	-0.232	0.072	-0.183	0.065	0.079	0.028	0.738	0.073
γ_1	0.014	0.071	-0.019	0.016	0.208	0.085	0.043	0.063	-0.060	0.046	-0.055	0.024	-0.130	0.045
γ_2	-0.055	0.024	-0.012	0.009	0.0004	0.027	-0.012	0.038	0.064	0.026	0.042	0.019	-0.027	0.025
γ_3	-0.019	0.016	0.089	0.012	0.061	0.015	-0.006	0.024	-0.093	0.023	-0.012	0.009	-0.020	0.017
γ_4	0.208	0.085	0.061	0.015	-0.415	0.121	0.300	0.074	-0.014	0.045	0.0004	0.027	-0.139	0.047
γ_5	0.043	0.063	-0.007	0.024	0.299	0.074	-0.411	0.129	-0.114	0.066	-0.012	0.038	0.202	0.082
γ_6	-0.060	0.046	-0.093	0.023	-0.014	0.045	-0.114	0.066	0.224	0.073	0.064	0.026	-0.006	0.046
γ_7	-0.130	0.045	-0.020	0.017	-0.139	0.047	0.202	0.082	-0.006	0.046	-0.027	0.025	0.121	0.059
β_g	-0.132	0.009	0.045	0.008	0.042	0.007	0.063	0.012	0.057	0.011	-0.006	0.005	-0.069	0.012
SFNOC64	0.030	0.004	0.028	0.003	-0.0001	0.003	-0.052	0.005	0.020	0.005	-0.001	0.002	0.035	0.005
SMNOC64	-0.016	0.003	-0.018	0.003	-0.019	0.002	0.008	0.005	0.026	0.004	0.021	0.002	-0.002	0.005
SNOC65_	-0.026	0.006	0.004	0.005	-0.001	0.004	-0.052	0.008	0.009	0.007	-0.014	0.003	0.081	0.008
CONOC65_	-0.001	0.004	0.005	0.004	0.016	0.003	-0.008	0.006	0.009	0.005	-0.012	0.002	-0.009	0.006
CO1CH	-0.014	0.003	0.008	0.003	0.003	0.002	0.001	0.004	-0.018	0.004	-0.008	0.002	0.028	0.004
CO2CH	-0.029	0.004	0.019	0.003	0.003	0.003	0.001	0.005	-0.011	0.004	-0.014	0.002	0.031	0.005
CO3_CH	-0.033	0.005	0.032	0.004	0.008	0.004	0.014	0.007	-0.011	0.006	-0.016	0.003	0.006	0.007
ONOCH	-0.018	0.004	0.017	0.003	-0.001	0.003	0.022	0.005	0.011	0.005	-0.006	0.002	-0.026	0.005
OCH	-0.024	0.004	0.032	0.004	0.003	0.003	0.003	0.006	0.003	0.005	-0.011	0.002	-0.005	0.006

The system was estimated with the SAS procedure SYSLIN (ver 6.10). To take the restrictions on the parameters within and between the equations in consideration we used the SRESTRICT function. We also used the WEIGHT command to avoid heteroscedasticity due to the aggregation.

This specific model may not the best choice when the purpose is to analyse the household consumption behaviour. Our more modest purpose with the estimation is to obtain estimated proportions

$$\hat{P}_{gdt} = \hat{\alpha}_g + \hat{\partial}_d + \sum_{g_0} \hat{\gamma}_{g_0} I_{g_0 t} + \hat{\beta}_g x^*_{dt}$$

which we can use as auxiliary information in the calibration. In section 5 the auxiliary information is specified as totals, T_x. We then have to transform the estimated proportions. One possibility is

$$t^*_x = R_m t_y,$$

where the ratio $R_m = \hat{P}_{gdt}$ is given by the supporting model and t_y is the estimated total expenditure. For the 1992 HBS this would mean 12 months × 7 groups of expenditures × 10 types of households = 840 auxiliary variables. To reduce this set we decided to aggregate over the year, thus $R_m = \hat{\bar{P}}_{gd}$. Consequently 7 groups of expenditures × 10 types of households = 70 auxiliary variables were used in the calibration. Compared to an ordinary least squares regression, where a smaller number of explanatory variables often is desirable, this could be still regarded as too many variables. But the purpose is to examine survey estimators when an econometric model provides auxiliary information. For a consumption model that satisfies econometric requirements, we are able to produce consistent consumption proportions for different kinds of households and goods.

6. Calibration results

Finally we are ready to evaluate the different calibration alternatives. Results for the Swedish 1992 HBS are compared using four methods for the estimation of average expenditures, namely

- unadjusted ratio estimates
- CRD, traditional calibration using register data (population size by age and sex, number of households in four geographical regions, number of owner-occupied dwellings and other dwellings)
- CMD, calibration using model data or model-based calibration (given the results in section 5)
- CMRD, calibration using both model and register data

The unadjusted ratio estimates are based on simple random sampling assumptions; these have been found to well approximate the actual HBS design, the survey consists on random samples of individuals. No account for the about 37 percent non-response is taken in the unadjusted ratio estimates which are a base for a comparison of the three calibration estimators

Results are presented as ratios in figures 1, 2 and 3. The ratios are the results of either calibration method divided by the corresponding result for the unadjusted ratio estimates. Thus, the methods are only compared to each other, and not to any "true values". In figure 1 the ratios between the averages for all households for different expenditure categories are compared, as well as coefficient of variance ratios. The ratios of averages show very small differences for the estimation alternatives; the impression is that there is no non-response bias since the ratios are close to one, or rather that the auxiliary information is unable to correct for non-response effects. There is no sign of any additional bias following from the model-based calibration.

The CV-ratios are, however, different. While the differences for total expenditure and for some categories such as food, household services and non-durables are small; we seem to obtain gains for other expenditures. Interestingly, only one of those expenditure categories with small differences was used explicitly in the AIDS-model, that is, the auxiliary model-based information does usually not concern these categories specifically. For five out of the ten consumption categories in figure 1, the CV-ratios are at least halved for the model-based calibrations, CMD or CMRD. The differences between these two model-based alternatives are small. There seems to be little or no gain of including register information in the calibration both as regards the unadjusted estimator, the CRD-ratios are close to one, and the CMRD-ratios are close to the CMD-ratios, usually.

In figures 2 and 3 CV-ratios are given for some different household types. The pattern is very similar to what we saw for all households, gains for the same expenditure categories of about the same size for the model-based calibrations. There are some notable exceptions in figure 2. For non-durable goods the model-based estimates show a CV-gain for single persons age 65 and above, while for household services the model-based estimates are inferior with CV-ratios larger than one for cohabitant households age 65 and above. This is different from the results for all and for other household types. Thus, there is no guarantee of an improvement even if this was observed for a higher level of aggregation, and improvements may occur for an expenditure category not included explicitly in the econometric model. Figure 3 shows CV-ratios for households in geographical regions not accounted for in the calibrations. Notable here is that calibration improvements are smaller than in figures 1 and 2. Estimation for domains not considered in the calibrations thus benefits less than domains explicitly included seems to be the conclusion.

7. Discussion

In this paper model-based calibration to improve precision is explored. In an application from expenditure analysis, the auxiliary information obtained from a rather elaborate model of consumer demand seems to have the potential of improving precision in estimates of average expenditure. The same model-based calibration procedure has also been applied to the Finnish HBS. Lacking some information on the survey design and on price indices, the explanatory power for the equations in the AIDS-model was lower and the gains from the calibration smaller than for the Swedish case. Still a gain was observed for most expenditure categories. A drawback of the model-based calibration here is the reliance on the assumptions of a "true" model. We also made assumptions on similarity between units. More research is needed on the effects of models of different quality and on the importance of these assumptions for more definite recommendations on the use of this kind of calibration.

References

Andersson, C. and L. Nordberg (1998). *A User's Guide to CLAN 97 - a SAS-program for computation of point- and standard error estimates in sample surveys*, Statistics Sweden.

Cassel, C.M., Granström, F., Lundquist, P and J. Selén (1997). Cumulating Data from Household Budget Survey. Some Results for Model Based Calibration Techniques Applied to Swedish Data. *Report financed the by European Communities, LOT 23.*

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

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30, 109-126.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *American Economic Review*, 70, 312-126.

Deming,W.E. and F. F. Stephan(1940). On a least squares adjustment of a sampled frequency when the expected marginals are known. *The Annals of Mathematical Statistics, 11, 427-444*

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Statistics Sweden: *The family expenditure survey*, different years

0(17)

Figure 1: Ratios of averages and CV's, different calibrations and expenditures, all.

//insert from triergrafer.doc//

Figure 2: Ratios of CV's, different calibrations, expenditure categories and household groups.

//insert from triergrafer.doc//

Figure 3: Ratios of CV's, different calibrations, expenditure categories and regions.

//insert from triergrafer.doc//

Figure 1: Ratios of averages and CV's, different calibrations and expenditures, all.

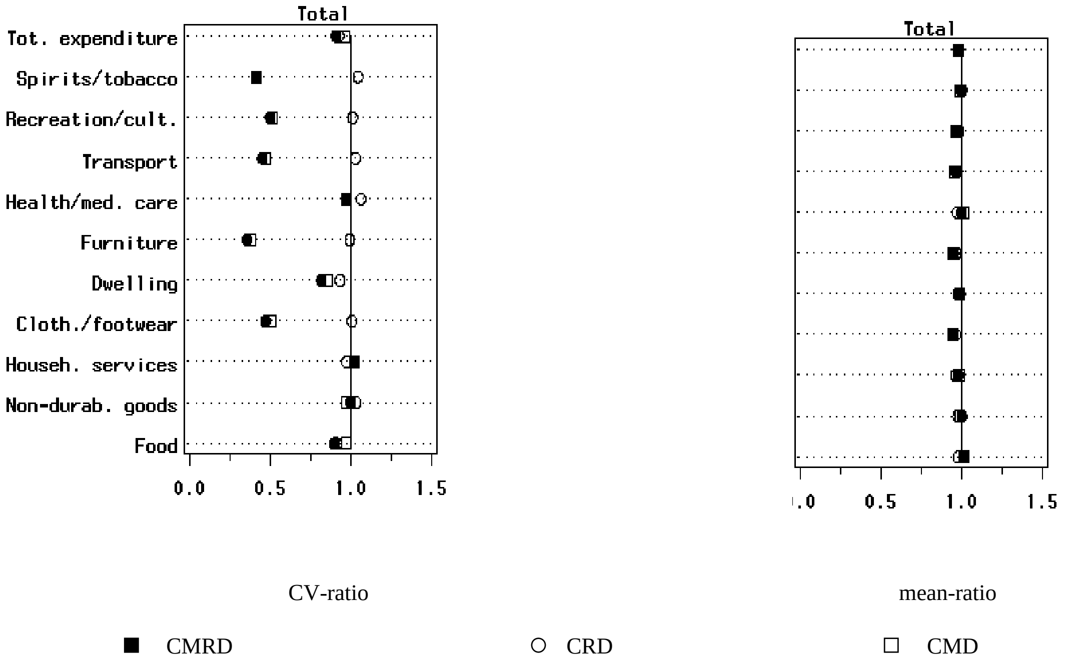


Figure 2: Ratios of CV's, different calibrations, expenditure categories and households.

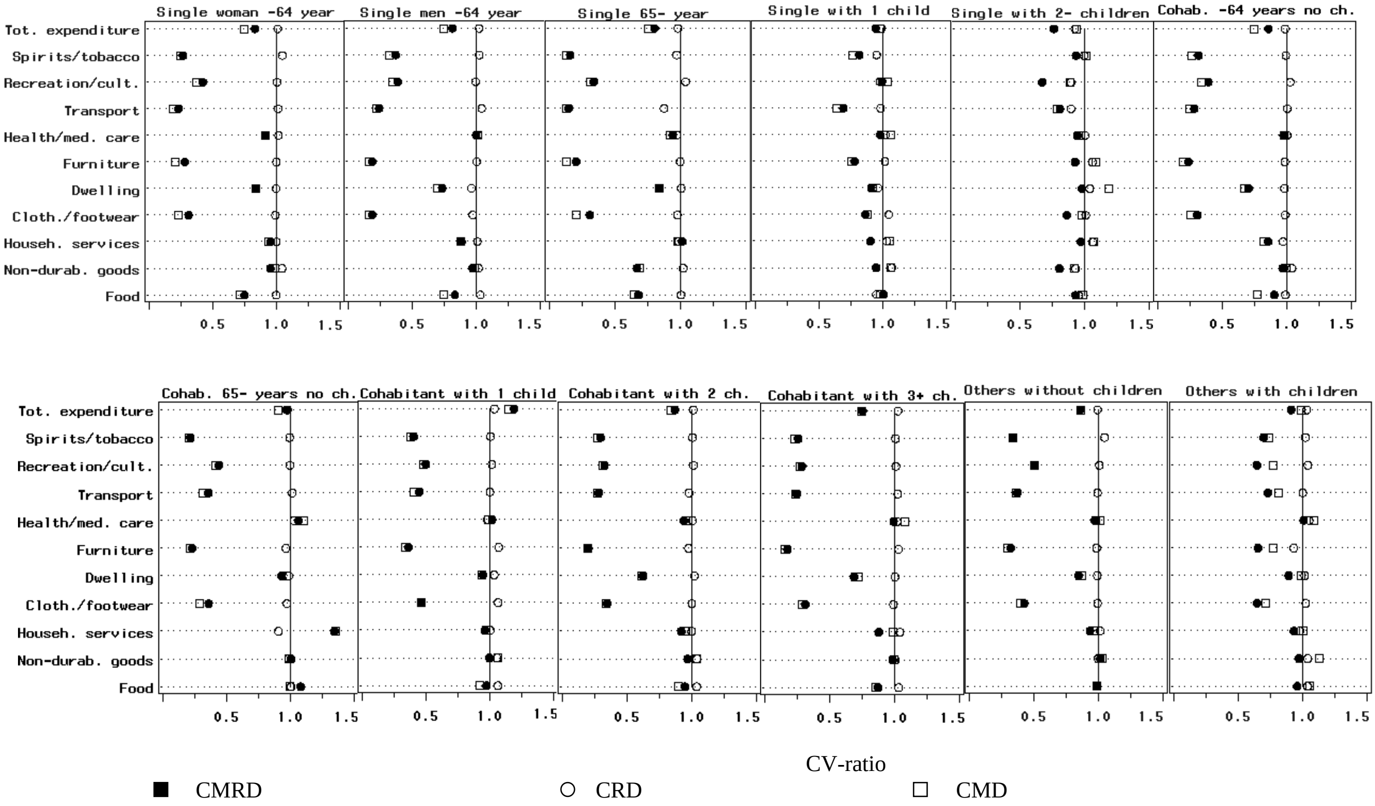
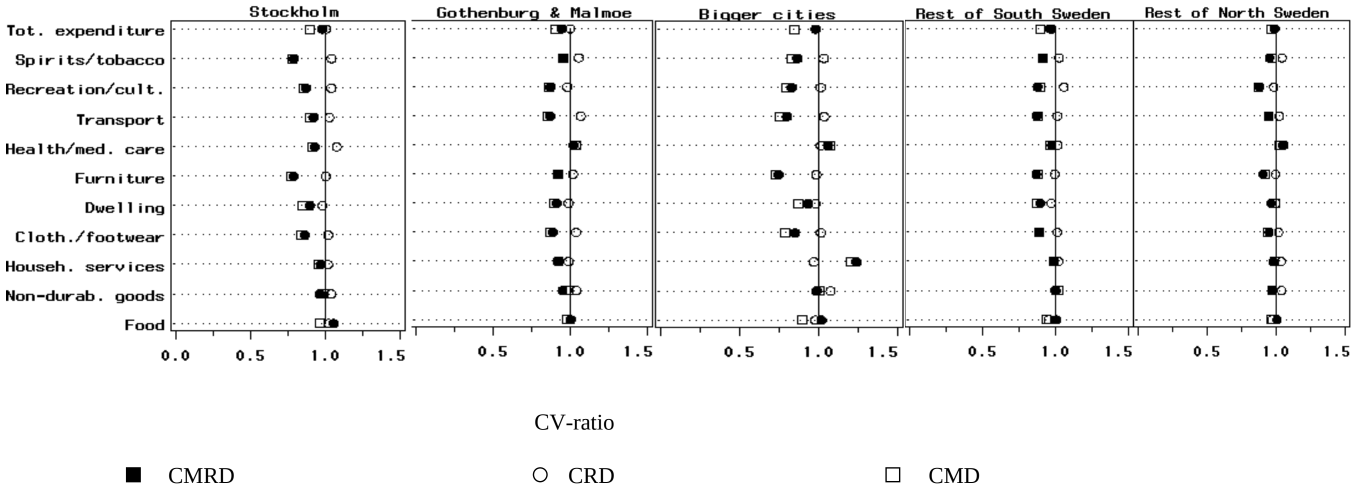


Figure 3: Ratios of CV's, different calibrations, expenditure categories and regions.



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 1999 and onwards:

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)*
2001:1	Det statistiska registersystemet. Utvecklingsmöjligheter och förslag *(SCB, Registerprojektet)*
2001:2	Order πps Inclusion Probabilities Are Asymptotically correct (*Bengt Rosén)*
2002:1	On the Choice of Sampling Design under GREG Estimation in Multiparameter Surveys *(Anders Holmberg)*
2002:2	Model-based calibration for survey estimation, with an example from expenditure analysis Surveys *(Claes Cassel, Peter Lundquist and Jan Selén)*

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 1988-1998 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).