An Application of Generalized Precision Functions in the 1985 Swedish Family Expenditure Survey

Håkan L Lindström and Peter Lundquist

R&D Report
Statistics Sweden
Research-Methods-Development
1989:10

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 1989:10. An application of generalized precision functions in the 1985 Swedish family expenditure survey / Håkan L Lindström; Peter Lundquist.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1989-X101OP8910

An Application of Generalized Precision Functions in the 1985 Swedish Family Expenditure Survey

Håkan L Lindström and Peter Lundquist

SCB

R&D Report
Statistics Sweden
Research-Methods-Development
1989:10

Från trycket	Juni 1989
Producent	Statistiska centralbyrån, Utvecklingsavdelningen
Ansvarig utgivare	Staffan Wahlström
Förfrågningar	Peter Lundquist, tel. 08-783 49 18

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1989

1989 - 06 - 01

AN APPLICATION OF GENERALIZED PRECISION FUNCTIONS IN THE 1985 SWEDISH FAMILY EXPENDITURE SURVEY

Håkan L Lindström and Peter Lundquist

Content Page

1 AIM OF THE STUDY

In surveys, measures of precision are not calculated for all survey estimates, although precision measures are important. Users do not always understand the importance of precision measures, such measures can be expensive, and sometimes it would be hard to present and handle all the information if the precision of each estimate was calculated.

Presenting all the estimates of coefficients of variation (cv), standard errors (se) or confidence intervals in special supplementary tables or diagrams would provide approximations of the precision even of those estimates for which the precision is not explicitly calculated. This study aims to find out to what degree such approximations can be used in the Swedish 1985 Family Expenditure Survey (FEX).

2 BACKGROUND

For the past several years, Statistics Sweden has been calculating the precision of totals, averages, and percentages routinely and at a fairly low cost. Software developed by Statistics Sweden (SMED83) allows the calculation of variances for data collected with sampling designs such as stratified sampling, cluster sampling, p.p.s. sampling and combinations of these. SMED83 can also handle nonresponse, both when subsampling is done in a nonresponse stratum and when estimates are weighted by strata or subgroups. SMED83 allows for the calculation of variances, standard errors, confidence intervals and coefficients of variation.

The estimates of precision have their statistical uncertainty, which is seldom estimated and might be of great magnitude. The use of variance functions might have a stabilizing effect on the estimates. Model based precision estimates might be more accurate than traditional precision estimates.

Variance functions can be presented concisely and used more easily than the abundance of data sheets obtained from the regular calculations. Variance functions can give better understanding of the variability of the estimates than lots of unorganized calculated values. When results are reported, analysis is planned, and results are evaluated, concise information will be very useful. It is especially important when there is a great number of variables and domains of study and the regular precision estimates are not made available for all estimates.

The 1985 Family Expenditure Survey has a simple design; we hoped that the simplicity would allow for the calculation of stable precision functions. These functions are based on simple parameters like the sample size and the average of each domain of study. A more complicated design with disproportional allocation on strata, cluster sampling or multistage sampling would make the calculation of precision functions much more involved.

3 The FAMILY EXPENDITURE SURVEY (FEX) - sampling and estimation.

Our work is based on the 1985 Family Expenditure Survey. Another FEX was conducted in 1988 and FEXs are now planned triennially. Before 1985 the intervals between two FEXs were longer - surveys were done in 1958, in 1969 and 1978.

In the 1985 FEX a simple random sample of individuals aged 0 - 74 was drawn. A household was included in the survey when one of its members was selected. The net sample consisted of 6004 households. The households were contacted for a preliminary interview in which certain characteristics were established. About 27% of the sample were nonrespondents - most of whom were refusers. In all, 4354 households participated in the survey.

Each participating household was asked to keep diaries of its expenditures for one month. The period of bookkeeping was decided through the sampling procedure to secure a random and equal distribution of households over the year. In an additional mail survey (telephone interviewing was used for a small part of the sample) expenditures for specified expensive goods and services were reported for the entire year.

The estimation was based on three types of weights. In the first, each household's expenditures were summed into prespecified aggregates and weighted to the level of yearly expenditures, denoted x_j. In the second, the varying sampling probabilities for households were adjusted for by dividing the estimate of the entire year's expenditures for each household with the number of household members (p_j).

In the third, the sample was poststratified into 45 strata by size, region, and age of the head of the household. The size of each stratum had been calculated from the sampling frame. The poststratification is meant to reduce the nonresponse bias by compensating for nonresponse rates varying between strata. Poststratification also leads to some reduction in the variance.

Since we intend to conduct all subsequent FEXs in mainly the same way as the 1985 FEX, the estimated precision is useful not for only this survey, but also in the planning of the FEXs that will follow. Coefficients of variation (cv) are better measures of precision than standard errors since they are less sensitive to inflation or other price fluctuations. The design and the results of the survey are presented in [6] which is referred to as the FEX Report throughout this paper.

4 ESTIMATED COEFFICIENTS OF VARIATION FOR AVERAGES

4.1 VARIABLES AND DOMAINS OF STUDY

A coefficient of variation was calculated for each of the ten expenditure aggregates and their sum for 130 domains of study. The ten aggregates are shown in the tables that follow. The aggregates are mutually exclusive and sum to "all expenditure."

The domains of study chosen for this study are both important and representative of the domains used in the survey. Since they are not a random sample from a well defined universe, generalizations must be made with great caution, if at all. The domains of study are shown in Tables 1-8 of the FEX Report. The domains of study are obtained when the sample is divided up by:

* type of household, Table 1
* households stage in life cycle, Table 2
* households with children by age group of children, Table 3
* socioeconomic group, Table 4
* type of household and socioeconomic group, Table 5
* type of household and degree of employment, Table 6
* region, Table 7
* population density area and type of household, Table 8.

The domains of study of each classification are in Appendix 1

4.2 ESTIMATION AND REPRESENTATION OF CVs

In this study, we devote our attention to coefficients of variation (cv) and not to standard errors (se), as already mentioned. Coefficients of variation may vary less by variable than standard errors; when this is true, coefficients of variation are also more easily modelled and more concisely presented. The cvs can be approximately the same from year to year. This is true even when the standard errors (ses) are influenced by inflation and other price fluctuations. For this reason, the cvs makes it possible to plan the subsequent surveys with great accuracy.

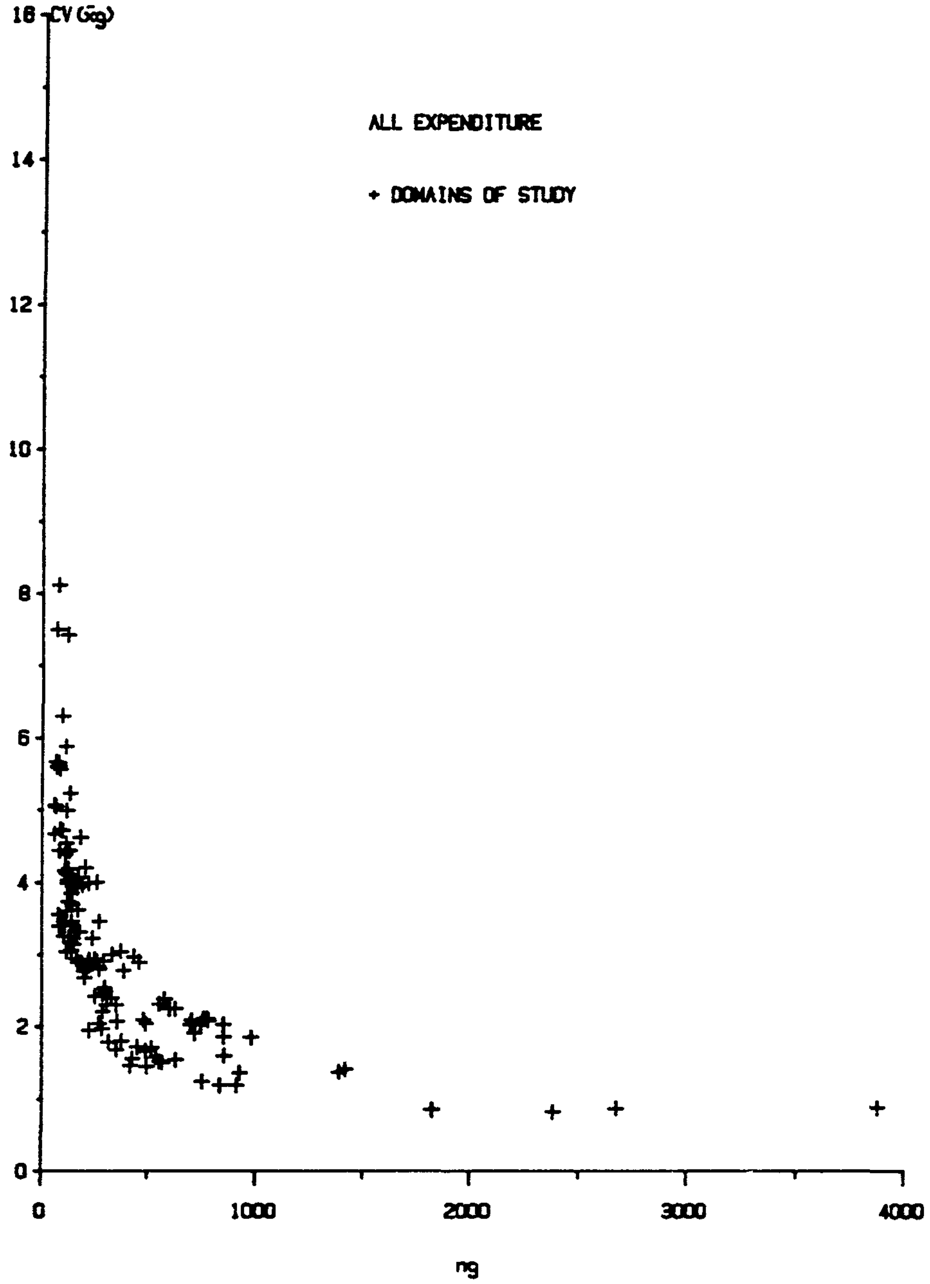
Coefficients of variation were calculated for the averages for all eleven expenditure aggregates and for the domains of study. The entire sample was included in the calculations. The full stratification is described in Section 3.

The cvs are slightly overestimated. One reason is that each household's yearly expenditures are estimated from one sampled period. Since the sample is not designed to measure this effect, an ultimate cluster technique is in fact used.

For each aggregate, the cvs of the 130 domains of study were plotted against the number of households in the given domain of study. In Figures 1-4 we present the results for the following four aggregates: "All expenditure," is an obvious choice; "Recreation" has the smallest variation and "Clothing and footwear" the largest. "Dwelling" has a variation of intermediate magnitude. We illustrate our work with only four aggregates because the results were what we had expected, and were similar for all aggregates.

The plots below exhibit the following pattern. The cvs of the domains of study decrease as the sample sizes increase. Although there is some variation for each level of n_g, the observations are rather concentrated. There are no obvious outliers, and it appears that an averaging function can be used for variance approximations. The five largest domains of study (n_g>1500) have a cv of almost the same magnitude as the entire sample. Group homogeneity may compensate for smaller sample size. Figures 1-4 depict cvs in 130 domains of study plotted against sample size.

Figure 1



CV(Ḡg)
ALL EXPENDITURE
+ DOMAINS OF STUDY
16
14
12
10
8
6
4
2
0
0
1000
2000
3000
4000
ng

Figure 2

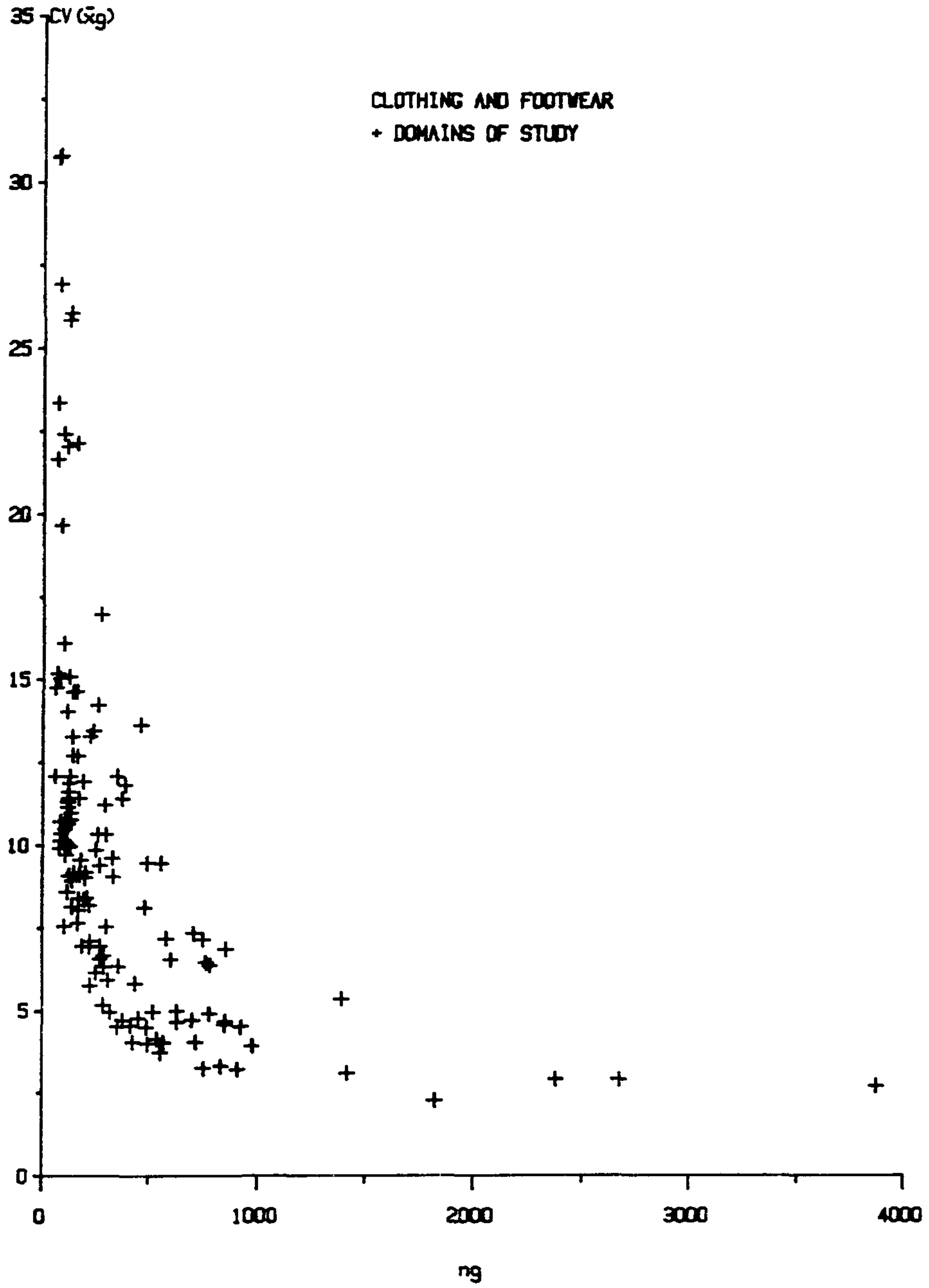

CV(x̄g)
CLOTHING AND FOOTWEAR
+ DOMAINS OF STUDY
35
30
25
20
15
10
5
0
0
1000
2000
3000
4000
ng

Figure 3

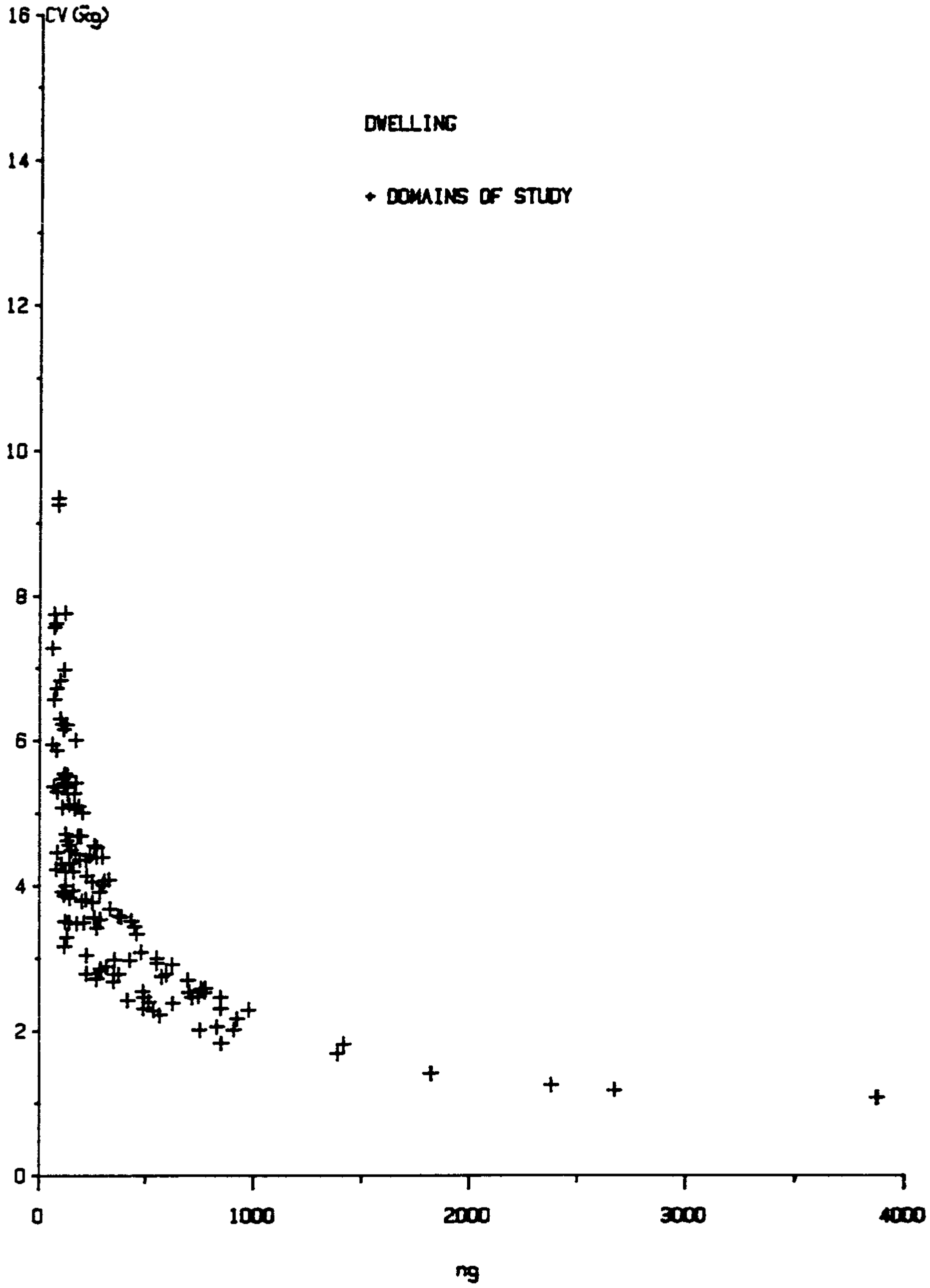
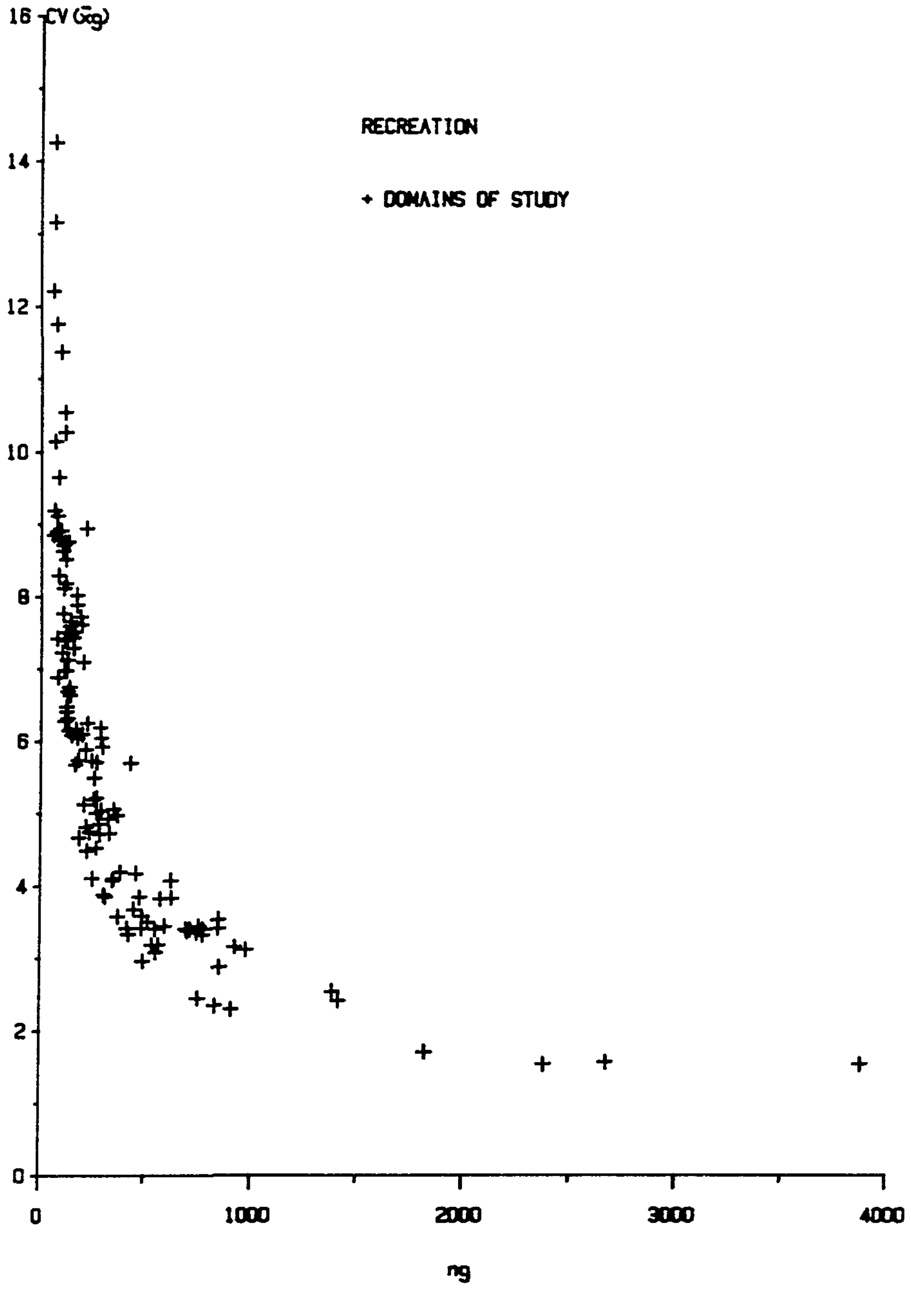

16
14
12
10
8
6
4
2
0
CV (Gg)
DWELLING
+ DOMAINS OF STUDY
0
1000
2000
3000
4000
ng

Figure 4


CV (x̄g)
16
14
12
10
8
6
4
2
0
RECREATION
+ DOMAINS OF STUDY
0
1000
2000
3000
4000
ng

5 CONTENT OF THE STUDY

This study covers two aspects of the measurement of an estimate's precision. The first is the random variation of the all sample cv studied by use of the random groups technique. The second is a search for good approximations of cvs for averages in domains of study for expenditure variables. General variance functions for the cvs were tried; these were calculated as unweighted and weighted linear regressions.

Another idea was to group variables with similar cv functions. The criteria used for grouping could be, for instance, the variable's distribution as measured by its skewness, the number of zero values, the level of the expenditures, etc. Nevertheless, this suggestion did not lead to any obvious or useful results and was abandoned.

As both the means of the aggregates of expenditure and their dispersions are at very different levels for the different domains of study, the results are not easy to compare and summarize. To make comparisons easier, two measures were calculated. One is the square root of the ratio between the total sample size (n) and the sample size in the domain of study (n_g). The other is the ratio between cv(x_g), the coefficient of variation in the domain of study, g, and cv(x), the coefficient of variation of the entire sample. Formally the two measures are

$$R_c = \frac{cv(\bar{x}_g)}{cv(\bar{x})} \quad (5.1) \qquad \text{and} \qquad R_n = \sqrt{\frac{n}{n_g}} \quad (5.2)$$

R_c , (5.1), will be referred to as the cv ratio. For all aggregates of expenditure and each domain of study the cv ratios were plotted against the R_n value. The central aim of this study is finding a relationship between R_c and R_n that will yield reliable cv estimates in the domains of study.

Cvs in domains of study were estimated by the mainframe computer, but all other calculations were done by PC SAS.

6 THE RANDOM GROUP STUDY

6.1 THE MODEL

Often no reason is given for the choice of a particular generalized variance function (GVF). What is given, however, is empirical evidence on how the function fits the data. This may be justified as long as the function is useful in practice. Nevertheless, all practitioners would appreciate having a stronger theoretical basis that explains why and when variances follow a certain law. This would help preclude improper applications of GVFs.

For a simple one-stage sample design, as the 1985 FEX, it is reasonable to examine whether the cv in a domain of study and for a given variable is a function of both the sample's cv for that variable and of the domain sample size. If each domain of study was a simple random sample from the whole sample, the points (Rn, Rc) would concentrate around the line $R_c=R_n$, when plotted in a figure.

However, there are theoretical reasons for not expecting this hypothesis to be completely true. Considering first the variance, one would expect the unit variance in a domain of study to be smaller than the corresponding population variance with some homogeneity factor.

The domains of study that we have examined are not independently and randomly chosen, thus the estimates are correlated since the domains of study are overlapping. That kind of correlation tends to reduce the dispersion compared to independent observations.

Even when a model is successfully fitted to the main body of empirical data, there are cases when the model should not be expected to be a good approximation. This might happen when the pattern of expenditure in a domain of study deviates from the average or when the distribution of household size of the domain of study deviates from the entire sample distribution. The sampling probabilities depend on the number of household members. For example, the level of precision of domains of study consisting of one person households might deviate from the level of precision in households with more than one person.

6.2 ESTIMATION OF THE PRECISION OF A CV

To decide whether the dispersion of the domain of study cvs in the 1985 FEX is random or not, we need to know the standard deviation of the estimated cvs. Reference [2] demonstrates that for a variable Y_j with an average $\bar{Y}$, a simple random sample the variance of a variance estimator, s^2 is:

$$\mathrm{Var}(s^2) = 2\sigma^4/(n-1) + [E(Y_j - \bar{Y})^4 - 3\sigma^4]/n \quad . \qquad (6.1)$$

This variance is applicable on the 1985 FEX if the stratification is adapted to 6.1 and (x_j / p_j) is used as the calculating unit.

It would have been laborious to estimate $Var(s^2)$ or still more complicated to estimate $sd(cv(\bar{x}_g))$ in a direct way and little would be known of the reliability of these estimates. Instead the well-known random groups technique described in detail in [7] has been used.

The 4354 responding households in the 1985 FEX were randomly and without replacement divided into 20 "random groups" each consisting of 217 or 218 households. The same weights that were used in sampling were then used in the estimation for all random groups.

For the random group study of precision, the full stratification scheme was abandoned. Instead, four strata were used. Household size was the stratification variable (1, 2, 3 and 4 or more members according to the population register). This was found to be the most powerful single stratification variable.

A more finely divided stratification mainly serves to improve estimates of totals in regions and for types of households. Highly detailed stratification has little effect on the studied estimates of averages for both bias and precision. As a consequence, the estimators used in the random group study are in practice consistent with the regular estimators used for domains of study even if they are not theoretically so.

For the eleven aggregates of expenditure and each random group, the $cv(\bar{x}_{rg})$ was calculated as was the standard deviation between the estimated cvs of the twenty random groups. Estimates and sample sizes of random groups are denoted by the subscript rg to distinguish them from those of real domains of study which are denoted by the subscript g.

The standard formula of variance estimation for the random group is the well-known

$$var(z_1) = \frac{1}{t-1} \sum_{k}^{t} (z_k - \bar{z})^2 \quad . \qquad (6.2)$$

This formula is applied when

$$z_k = cv(\bar{x}_{rg}), \text{ and } t = 20.$$

Note that this is the variance for one observation of $cv(\bar{x}_{rg})$ and not for the average of the random groups estimate.

This estimator is unbiased if sampling is done with replacement. Without replacement, the estimator is likely to have a slight positive bias. For equal sized random groups, for small or medium size samples, and for small sampling fractions in a large population the bias can generally be overlooked. As very few studies of nonlinear estimators exist, there is little that has been proved empirically about their biases.

Both the variance estimation method and the number of groups were chosen with practical considerations in mind. Since cvs are ratios, their biases may not be negligible when sample sizes are small. Reference [2] and others warn that the cv of the term in the denominator must be less than 0.10 (10%). For the sample size 4354/20 most aggregates fulfil or almost fulfil that condition, as is demonstrated by Table 2.

The sizes of the random groups were chosen so that no stratum would have zero observations in any calculation (except for very small domains of study). Furthermore the random group sample size yields an $R_n=\surd 20$ which comes close to the midpoint of the interval 1 $<R_n<$ 9 which covers all domains of study included in this report.

Additional studies might indicate possible improvements in the accuracy of the estimation. One such study would be the repeated random groups method (RRG), possibly in combination with another random group size.

6.3 OUTCOME OF THE RANDOM GROUP STUDY

The standard deviations of the random groups cvs are given in Table 1 below together with the averages, the minimum and maximum values, and the ranges for all eleven aggregates. The calculated cvs are expressed as percentages and not as fractions.

Table 1 **Distribution of estimated cvs (%) in 20 random groups**

Aggregate	Minimum value	Maximum value	Average	Standard error	Range
All expenditure	3.16	4.69	3.77	0.39	1.53
Food	3.14	5.14	3.56	0.45	2.00
Non-durable goods	5.53	9.51	6.80	0.82	3.98
Household services	7.58	14.27	10.26	1.71	6.69
Clothing and footwear	7.40	11.45	8.91	1.07	4.05
Dwelling	3.81	5.94	4.77	0.63	2,13
Furniture and household articles	9.01	19.54	12.20	2.52	10.53
Health- and medical care	11.82	20.88	16.12	2.55	9.06
Transportation	7.26	11.43	9.15	1.06	4.17
Recreation etc.	5.24	9.24	6.39	0.92	4.00
Spirits and tobacco	7.93	12.04	9.17	1.09	4.11

The estimated standard errors of the cvs vary between 0.39 (all expenditure) and 2.55 (health care) and their ranges vary from 1.53 to 10.53. This does not necessarily mean that the random group method is less reliable for those expenditure aggregates have large standard errors (ses). The mean of the eleven relative sds of the estimated cvs (standard error/averages) varies little around 0.14. The size of the sds and the average of the estimated cvs are also highly correlated (r^2=0.86).

The distributions of the estimated cvs seem to be positively skew as the averages fall closer to the minimum then to the maximum value. There were not any extremely deviant values.

6.4 PREDICTION INTERVALS

When prediction intervals around the line $R_c=R_n$ are approximated, n_{rg} is treated as a constant. In that case the variance of R_c is

$$\mathrm{Var}(R_c) = E_1[\mathrm{Var}_2(R_c|s_n)] + \mathrm{Var}_1[E_2(R_c|s_n)] = A + B . \quad (6.3)$$

In Var_2 and E_2 of (6.3), R_c is conditional on s_n, the full sample.

In part B of the variance formula, $cv(\bar{x})$, the denominator of R_c, is constant when conditioned on s_n. $cv(\bar{x}_{rg})$ in the nominator of Rc is a ratio estimator. Such an estimator is approximately unbiased according to reference [2] when the cv of its denominator [i.e. of $\bar{x}_{rg}$] is less then 0.10 (10 %). Then

$$E_2[cv(\bar{x}_{rg})]/[cv(\bar{x})] \approx R_n \quad \text{and} \quad B = \mathrm{Var}_1[E_2(R_c|s_n)] \approx 0 \quad (6.4)$$

The values of $cv(\bar{x}_{rg})$ were calculated - for all 11 aggregates of expenditure - as if the sample size had been 218 and are given in Column 3 of Table 2. For most cases (6.4) is valid since the "10% condition" on the denominator is fulfilled or almost fulfilled. Only for "Furniture" and "Medical and Health Care" the critical limit is violated at the sample size used for the random groups.

In part A of $\mathrm{Var}(R_c)$

$$\mathrm{Var}_2(R_c|s_n) = \mathrm{Var}_2[cv(\bar{x}_{rg})|s_n]/[cv(\bar{x})]^2$$

$$\text{and then} \quad \mathrm{Var}(R_c) \approx E_1\left(\frac{\mathrm{Var}_2[cv(\bar{x}_{rg})]}{[cv(\bar{x})]^2}\right)$$

To estimate a confidence region around the hypothetical average line $R_c = R_n$, the standard error of a cv ratio conditional on the ratio n/n_g, has to be calculated for values of n_{rg} within the range of the n_g of the included domains of study.

For the ten aggregates and their sum

var[cv($\bar{x}_{rg}$)] = w^2 is estimated for n_g = 217 or 218 according to formula (6.2).

The values of w (the standard errors for R_cs of the random groups) are given in the first column of table 2. In the second column the all sample cvs for averages in the 1985 FEX are given. They are the cv($\bar{X}$) in the calculations.

The standard errors of cv($\bar{x}_{rg}$) at other levels of n_{rg} are finally approximated with

se[cv($\bar{x}_{rg}$) | n_{rg}] = $w * \sqrt{218/n_{rg}}$ and the standard errors

of R_c, conditional on n_{rg}, with $\dfrac{w * \sqrt{218/n_{rg}}}{cv(\bar{x})}$.

The degree of approximation is not studied. It might vary with the sample size.

In Columns 4 and 5 of Table 2 we have approximations for two levels of n_{rg}. They were used to construct the prediction intervals around the line R_c=R_n presented in Figures 5-8. The values in Column 2 were recalculated based on a hypothetical sample size of 218. These hypothetical values are found in Column 3. These calculations were made to examine whether the condition on the denominator of R_c is fulfilled or else to determine to what degree it is violated.

Table 2 Calculation of ses for random group R_cs (cvs as %)

Aggregate	se of cv for x_{rg}	cv_of x n=4354	cv of x if n=218	standard error for R_c if n_{rg} = 100	800
All expenditure	0.39	0.80	3.57	0.72	0.25
Food	0.45	0.75	3.35	0.89	0.31
Non-durable goods	0.82	1.49	6.66	0.81	0.29
Household services	1.71	2.61	11.66	0.97	0.34
Clothing and footwear	1.07	2.57	11.48	0.61	0.22
Dwelling	0.63	1.01	4.51	0.92	0.33
Furniture and household articles	2.52	2.99	13.36	1.24	0.44
Health- and medical care	2.55	3.89	17.38	0.97	0.34
Transportation	1.06	2.04	9.12	0.77	0.27
Recreation etc.	0.92	1.43	6.39	0.95	0.34
Spirits and tobacco	1.09	2.01	8.98	0.80	0.28

To judge whether the R_cs in the domains of study behave like the R_cs calculated for random groups, the tabulated values in Columns 4-5 are used to construct 95% prediction intervals around the line $R_c = R_n$ for normally distributed R_cs in the random groups. The prediction intervals are depicted in Figures 5-8 where the observed R_cs in domains of study also are plotted. Since the potential bias increases as the sampling fraction increases, extrapolation is a bit risky and prediction intervals are drawn only for $2 < R_n < 9$ which approximately corresponds to $50 < n_g < 1500$. All most all the selected domains of study will be found in this interval.

A look at the plots shows, that, as anticipated, the "simple model" does not fit the observations. The points of the scatter plots fall outside the prediction intervals so often that the idea of a random distribution around R_c=R_n must be abandoned.

The concentration of the observations is below the line R_c= R_n. Some values fall high above the upper prediction limit and the distribution is obviously non-normal. Instead of the "simple model" approach, one must try exploratory methods. Figures 5-8 present 95 % prediction intervals for R_c and observed R_c in 130 domains of study.

Figure 5

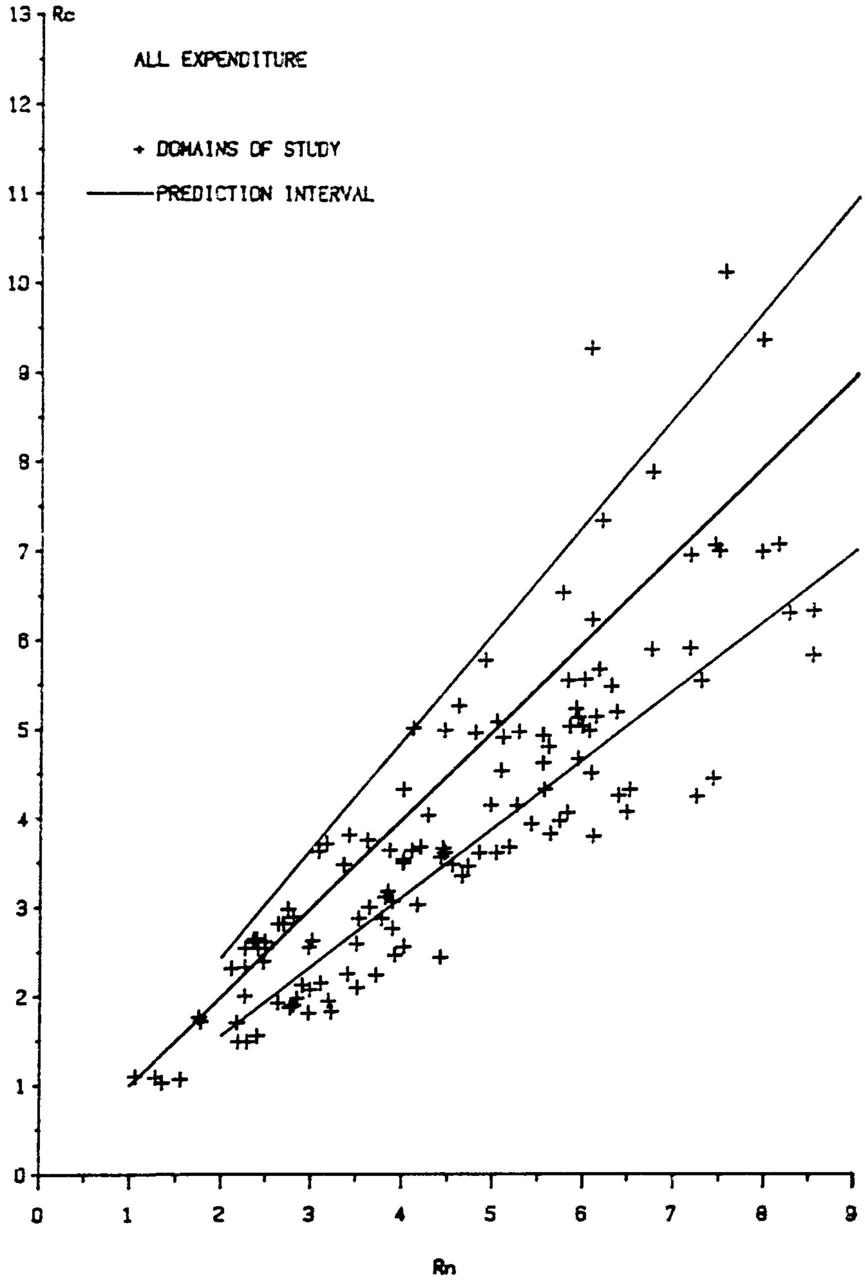

Rc
ALL EXPENDITURE
+ DOMAINS OF STUDY
PREDICTION INTERVAL
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

Figure 6

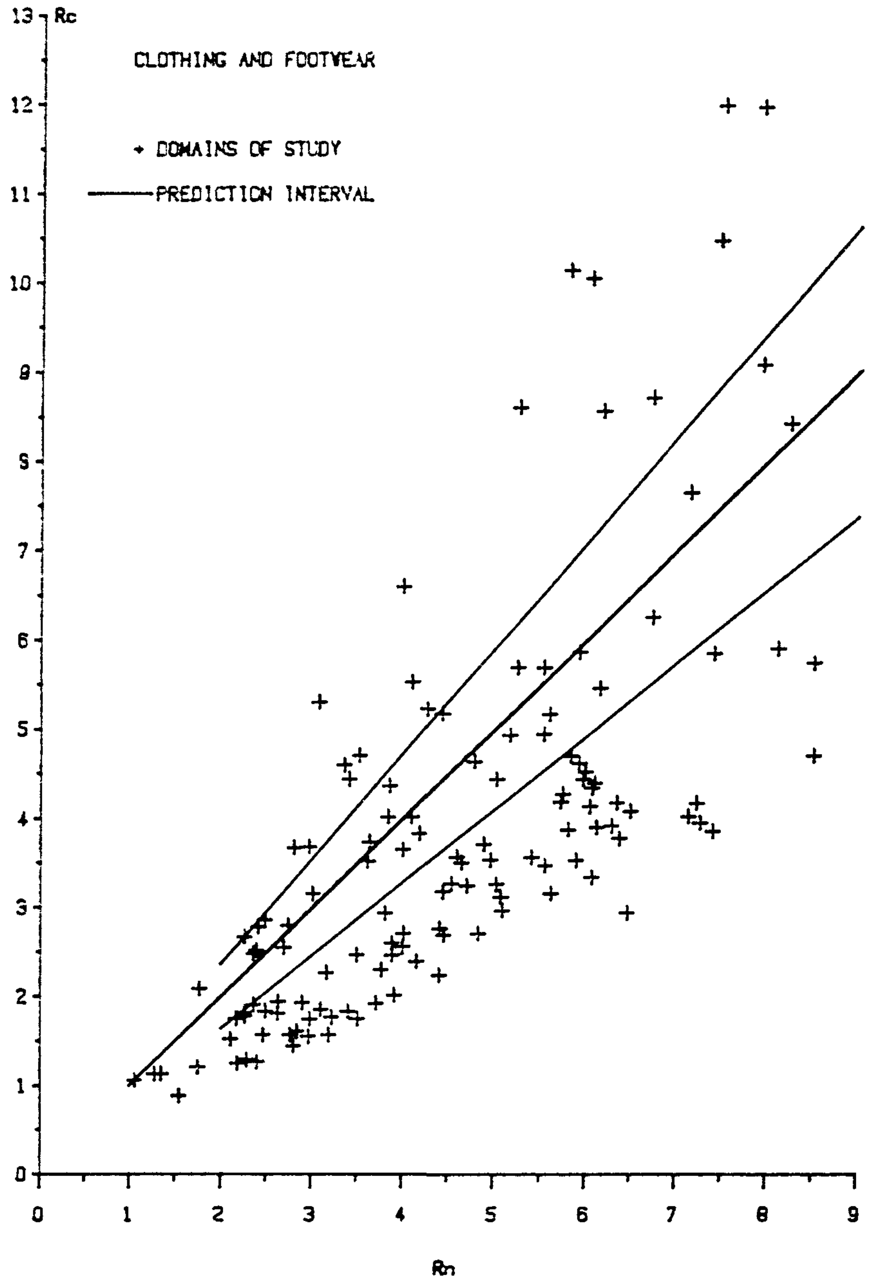

Rc
CLOTHING AND FOOTYEAR
+ DOMAINS OF STUDY
PREDICTION INTERVAL
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

Figure 7

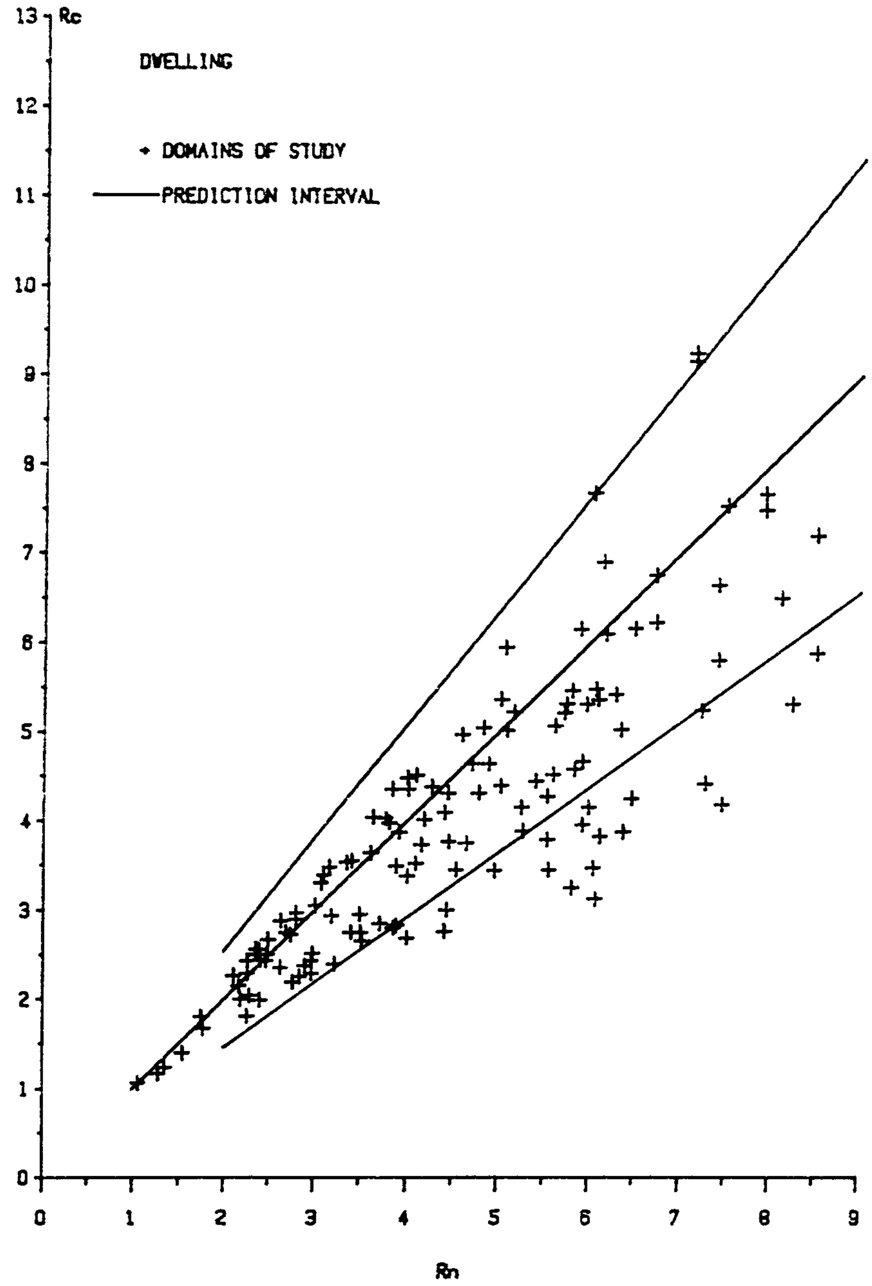
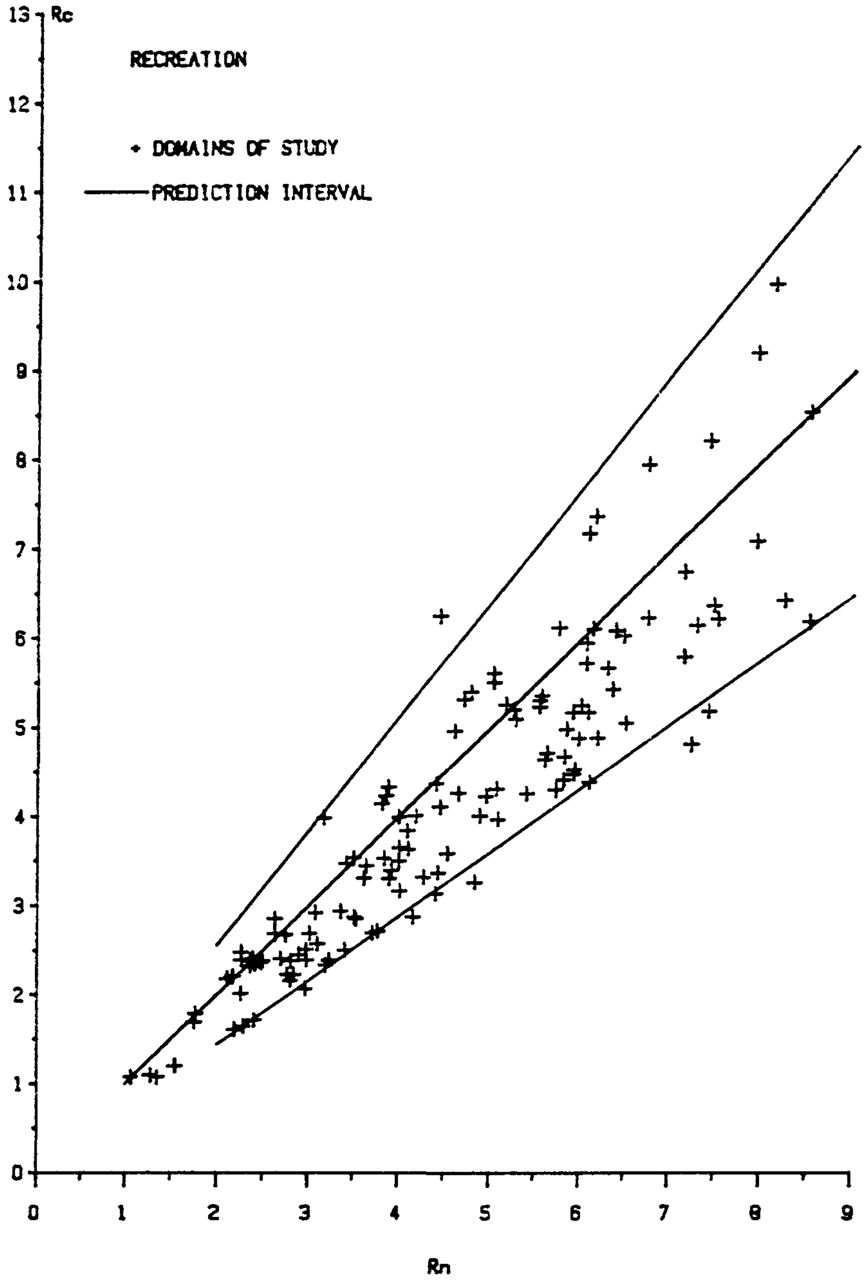

Rc
DWELLING
+ DOMAINS OF STUDY
PREDICTION INTERVAL
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

Figure 8


13
Rc
RECREATION
12
+ DOMAINS OF STUDY
11
PREDICTION INTERVAL
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

7 WEIGHTED AND UNWEIGHTED REGRESSIONS

7.1 THE PRESENTATION

For the eleven expenditure aggregates, linear mean square regressions of R_n on R_c are calculated and reported. The relationship between R_c and R_n is essentially linear for all eleven variables. Very little, if any, gain would have come from applying nonlinear relationships.

Results are reported for all eleven aggregates in Appendix 2. In all cases, the dispersion of R_cs around the regression line increases when R_n increases. This is even more obvious when the residuals are plotted. Some of the observations seem to be outliers.

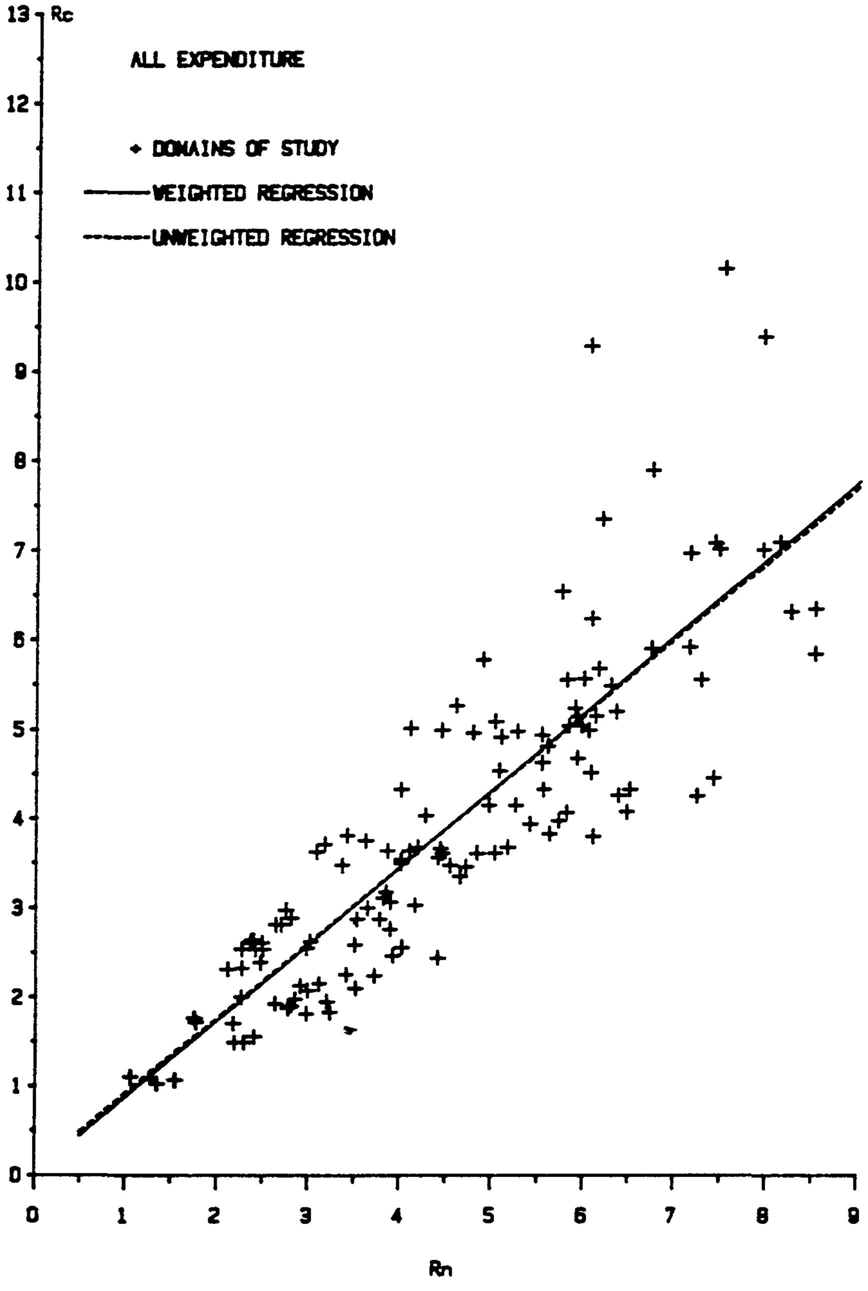
It is well known that the precision of the estimated regression suffers when the variance is not constant and that the best precision is obtained when the observations are weighted inversely to their variances and covariances. Attempts have been made to improve the fit of the regressions by identifying outliers and by weighting the observations. These weights are derived from the observations' variances, which are assumed to be of the type $x^{\alpha} * \sigma^2_{y|x}$. Covariances were not available.

The results of the calculations are summarized and presented in the figures and in Tables 3-5 in this section. For each aggregate of expenditures the scatter plot of (R_n, R_c) is presented together with unweighted and weighted linear regressions. All eleven figures are presented in Appendix 2. The figures of the four selected aggregates are given also in this section.

7.2 ALL DOMAINS OF STUDY INCLUDED

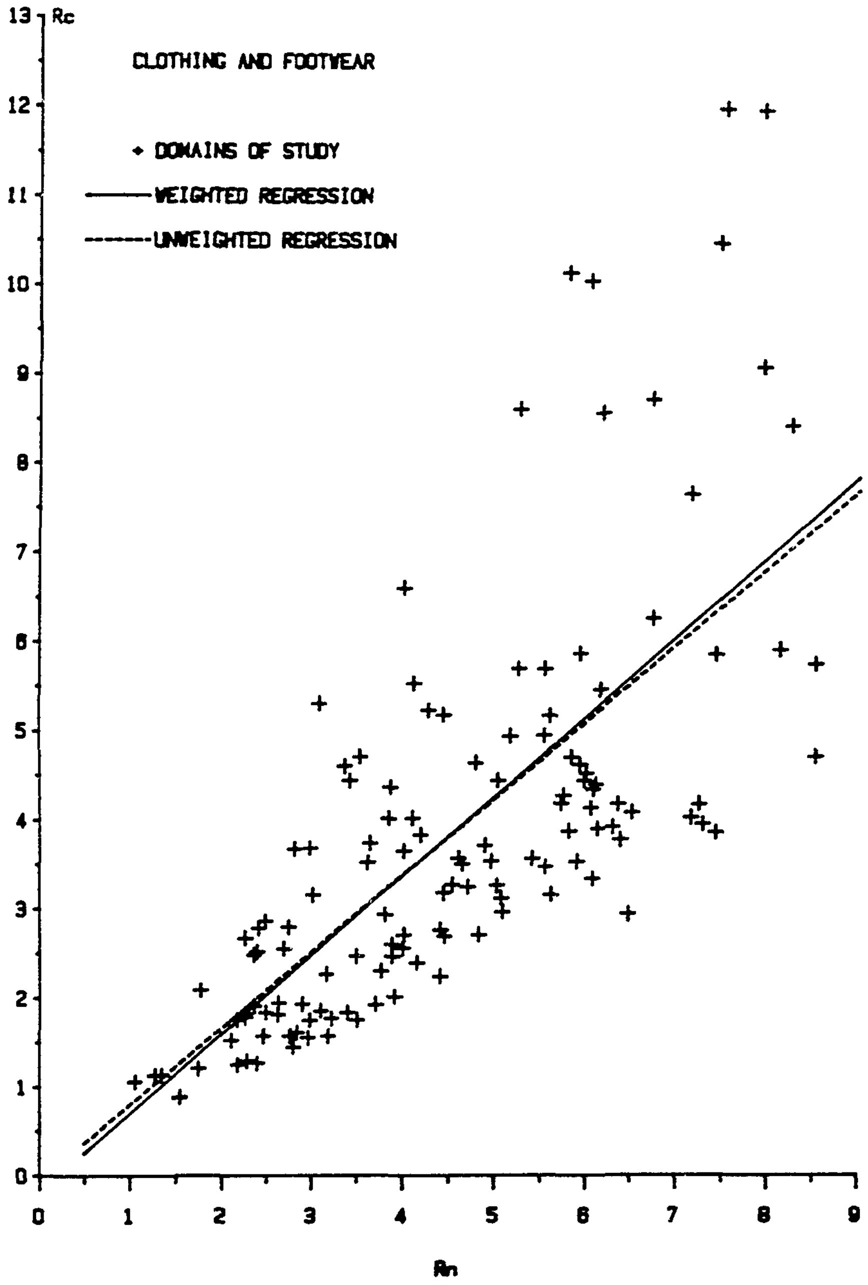
First, unweighted linear regressions were calculated for the eleven aggregates of expenditure. The results are reported in Table 3. When the regression is depicted in the same figure as the observations, we see that the dispersion of the R_cs around the fitted line increases as R_n increases. It was the same for all aggregates of expenditure and was still more obvious when residual plots (not reproduced here) were regarded. Figures 9-12 presents plots of (R_n,R_c) in 130 domains of study and their unweighted (------) and weighted (————) regressions.

Figure 9



Rc
ALL EXPENDITURE
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
1
2
3
4
5
6
7
8
9
Rn

Figure 10



CLOTHING AND FOOTWEAR
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
Rc
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

Figure 11

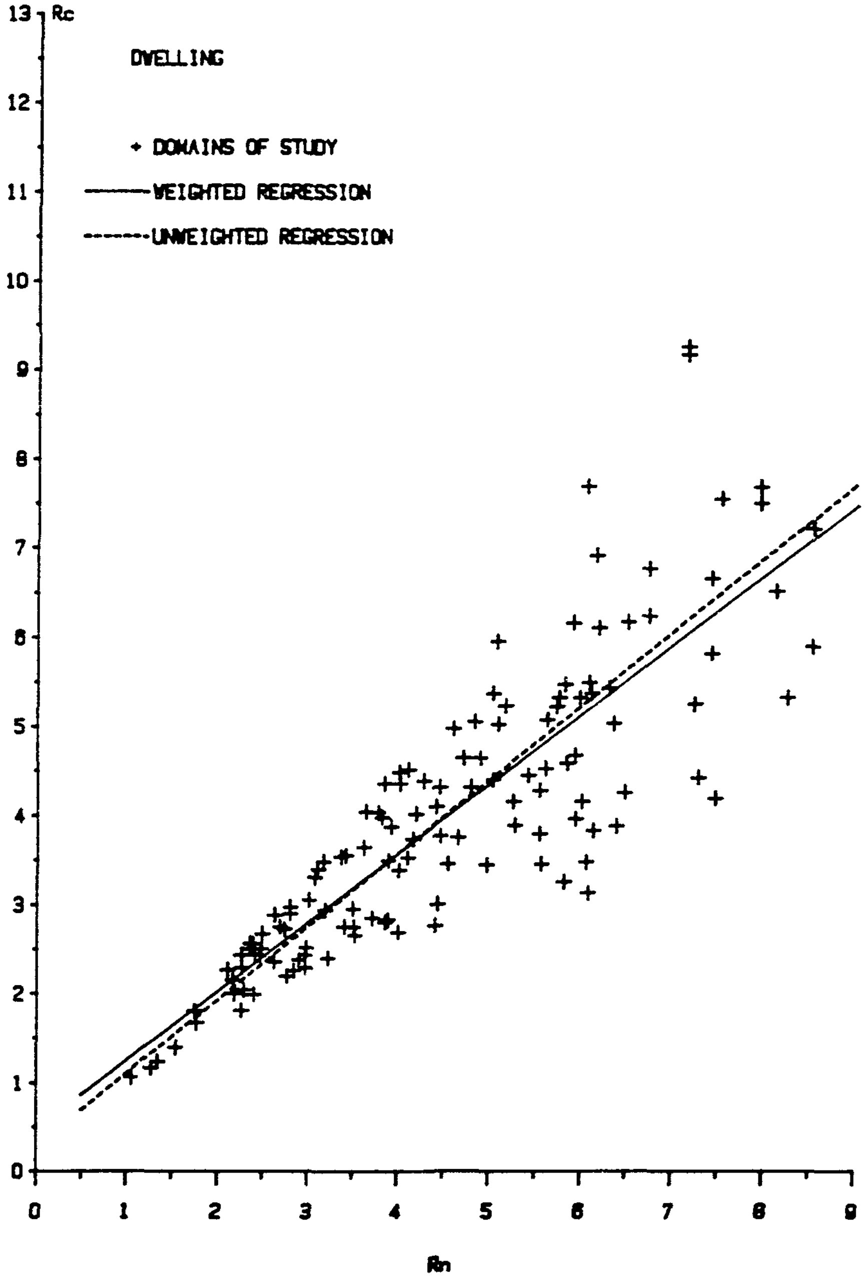
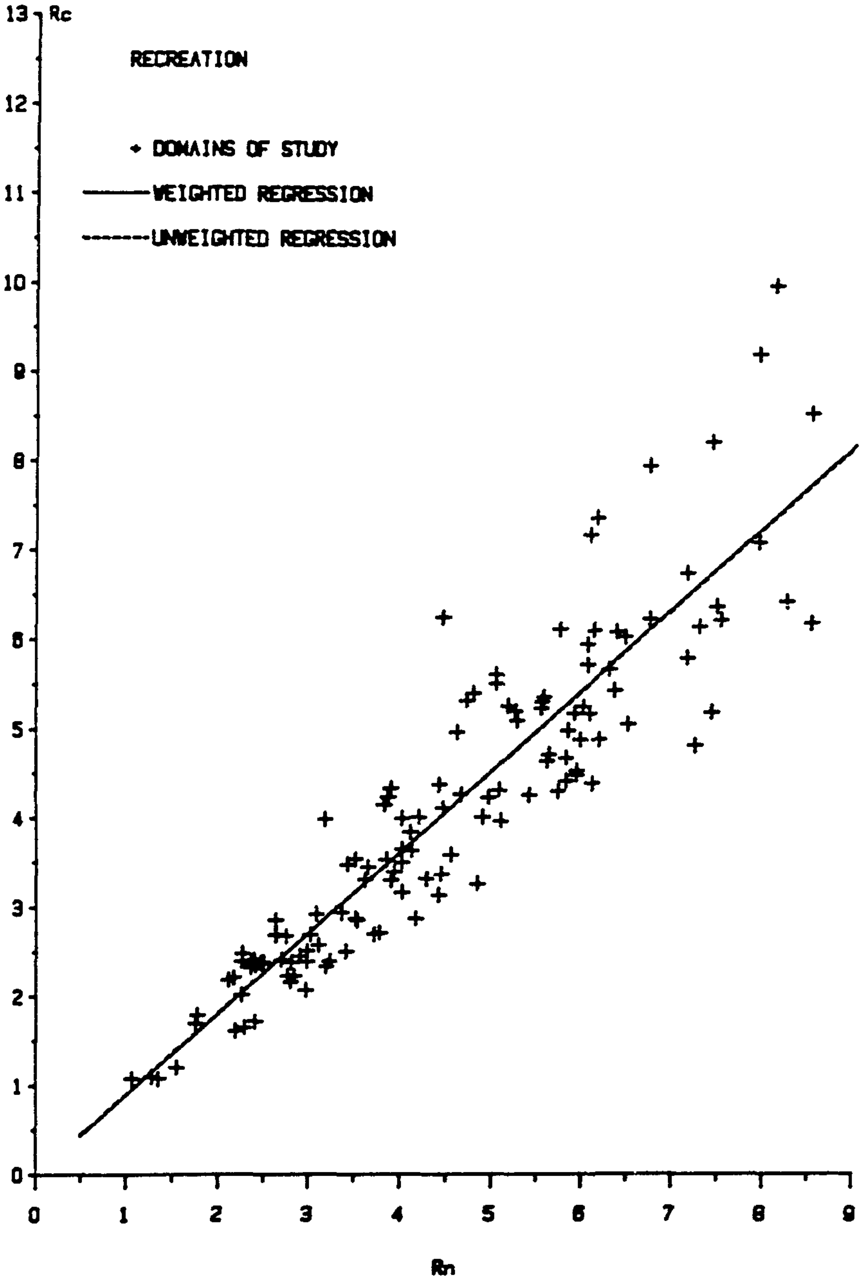

Rc
DWELLING
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

Figure 12


Rc
RECREATION
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

The table headings employ standard notation, i.e., intercept (b_o), slope (b_1) and correlation coefficient (r). The standard errors se(b_o) and se(b_1) are only descriptive measures.

Table 3 **Unweighted regression**

Aggregate	b_o	se(b_o)	b_1	se(b_1)	r^2
All expenditure	0.01	0.21	0.86	0.04	0.74
Food	0.20	0.18	0.78	0,04	0.76
Non-durable goods	-0.01	0.28	0.95	0.06	0.67
Household services	0.05	0.34	0.89	0.07	0.56
Clothing and footwear	-0.19	0.37	0.89	0.08	0.51
Dwelling	0.47	0.20	0.76	0.04	0.73
Furniture and household articles	0.71	0.26	0.71	0.05	0.59
Health- and medical care	0.07	0.25	0.85	0.05	0.67
Transportation	-0.23	0.29	0.99	0.06	0.68
Recreation etc.	-0.02	0.17	0.91	0.04	0.84
Spirits and tobacco	0.12	0.19	0.87	0.04	0.79

Since R_c depends on the sample size, the variance around the regression line is expected to vary with the value of R_n. This is further confirmed by an inspection of the residuals. To get regression estimates with the smallest variances possible, one has to weight the observations with respect to their variances and covariances.

We did several experiments to find an α-value to use in a weighting function of type

$$w = 1/(x^{\alpha} * \sigma^2_{y|x}) \quad . \qquad (7.1)$$

The objective of (7.1) was to minimize the residual variance and make the residuals more evenly distributed. Among the α values tested was the conventional choice of $\alpha = 2$, i.e., the variance proportional to $x^2\sigma^2$. We finally chose $\alpha = 2$ since no single choice of α was found to be uniformly best for all of the eleven aggregates. The results of these calculations are given in Table 4.

Other weights did not improve the fit of the regression lines considerably or consistently. Similar results are reported in [1].

Table 4 **Weighted linear regression**

Aggregate	b_o	$se(b_o)$	b_1	$se(b_1)$	r^2
All expenditure	0.06	0.11	0.85	0.04	0.82
Food	0.13	0.11	0.80	0.04	0.80
Non-durable goods	-0.09	0.13	0.97	0.04	0.81
Household services	-0.05	0.17	0.92	0.05	0.71
Clothing and footwear	-0.07	0.19	0.86	0.05	0.63
Dwelling	0.28	0.10	0.83	0.03	0.86
Furniture and household articles	0.44	0.14	0.78	0.04	0.72
Health- and medical care	0.06	0.13	0.85	0.03	0.78
Transportation	-0.25	0.16	1.00	0.05	0.77
Recreation etc.	-0.01	0.08	0.91	0.03	0.90
Spirits and tobacco	0.03	0.09	0.89	0.03	0.89

The result of the weighting was that: the squared correlations increased by about 10 percentage points; the residuals assumed more of a uniform distribution (but not perfectly uniform); and the standard errors of the intercepts were reduced. This study demonstrated that the gains in precision are robust against deviations from optimal weighting. The estimated regressions were not influenced much by the weighting as can be seen from Figures 5-8 and in Appendix 2.

In most cases, the intercepts are not two standard errors away from zero. The only exceptions are "Dwelling" and "Furniture and Household Articles." The slopes range from 0.78 to 1.00 and are estimated with better precision than the intercepts. Their standard errors are at most 0.05. The slope is at least two standard errors below 1.00 except in three cases: "Non-durable goods," Household services," and "Transportation." It was not possible to identify classes of expenditures that had separate regressions. But it can be noted that "the simple model" $R_c=R_n$ almost always gave conservative estimates in the interval studied.

There is also one important reason not to carry the weighting to far. The cv functions are going to be used for an unknown universe of domains of study appearing in tabulations to come. As both the aggregates and the domains of study are purposely chosen, we were happy to find a procedure that was robust, even though we did not find any variance minimizing weights.

7.3 EXCLUSION OF LARGE DOMAINS OF STUDY.

Since Figures 1-4 indicated that the largest domains of study and the entire sample had cvs of almost equal sizes, the usefulness of a composite approximation rule was also studied. The cvs of "large domains of study" should be approximated by the entire sample cv. In other domains of study, a linear regression of the cvs was tried.

We decided from the figures that domains of study with a sample of more than 1500 households should be regarded as large. Five domains of study were large according to this criteria and excluded from the calculations. The linear mean square regressions were calculated on the remaining domains of study and without any restrictions on the intercept. The principle of weighting was the same as reported in Section 7.2. The results are given in Table 5.

Table 5 **Weighted regression when $n_g < 1500$**

Aggregate	b_o	$se(b_o)$	b_1	$se(b_1)$	r^2
All expenditure	0.18	0.16	0.82	0.04	0.74
Food	0.36	0.16	0.75	0.04	0.71
Non-durable goods	-0.02	0.19	0.95	0.05	0.74
Household services	0.00	0.24	0.90	0.07	0.61
Clothing and footwear	-0.06	0.27	0.86	0.07	0.53
Dwelling	0.48	0.13	0.78	0.04	0.78
Furniture and household articles	0.49	0.20	0.77	0.05	0.62
Health- and medical care	0.14	0.18	0.83	0.05	0.70
Transportation	-0.23	-0.22	0.99	0.06	0.69
Recreation etc.	0.07	0.12	0.89	0.03	0.85
Spirits and tobacco	0.12	0.13	0.87	0.03	0.84

Compared with the regression using all observations in Table 4, neither the fit of the line nor the fit of the standard errors were improved by the exclusion of the largest domains of study. The absolute value of the intercept tended to be somewhat large. This is reasonable since the values closest to the origin were excluded.

As a consequence, we did not use a composite rule for common use. Nor would it have been clear which part of the rule to use when the regression line did not reach the value of the constant for a value of n_g close to 1500. Furthermore, there were very few large domains of study to base any conclusions on.

7.4 OUTLIERS

The scatter plots and residual plots show that some of the observations might be outliers. If they are, they should be excluded and the regressions should be recalculated using the remaining observations. To aid the identification of outliers we must decide when the approximations of precision should be used and when they should not.

Outliers are something that should be expected in this study. The distribution of the responding households on the domains of study cannot be assumed to be random. Nor can the pattern of expenditure be assumed to be similar in all domains of study. Some types of expenditure might be frequent in some domains of study and totally absent in others.

We used the standard criteria for determining whether an observation was an outlier. Observations deviating more than three standard deviations from the unweighted regression were considered outliers. Of all 11 * 130 scrutinized R_Cs only 23 (1.6 per cent) were outliers according to this criteria. The outliers are listed in Appendix 3. All of the observed outliers were found above the mean square regressions.

The outliers were further scrutinized to see if there was some external criteria by which potential outliers could be identified. However, no such criteria could be found. No single aggregate of expenditure had more than four outliers. Nor was there any obvious concentration in certain domains of study. Domains of study consisting of "single persons" or described as "others" were slightly over-represented.

Most outliers appeared in domains of study with small sample sizes. Only for four sets of outliers were the sample sizes greater than 200. In fact, the outliers scrutinized may not have been outliers at all. The distribution of a cv might be more positively skew than normal and even more so for small sample sizes.

On the whole, there was no obvious concentration of outliers, neither to certain domains of study nor to certain aggregates of expenditure. Since no obvious criteria for excluding outliers was found, all observations were included when the regressions were estimated.

Even if no rules for outlier identification could be given, one should be specially cautious to use the cv functions in some cases - specially when a domain of study is small and heterogeneous.

8 SUMMARY AND CONCLUSIONS

Under a random group model, the estimated values of (R_n, R_c) should group around the line Rc = Rc. Yet, the observations in this study were not expected to do so. The domains of study were non-random groups and the estimates were mutually dependent due to partial overlaps. The "simple model" also proved inaccurate.

It was obvious from the plots and residual plots that the variance was not constant around the unweighted regression line. After experimentation, weighting according to the conventional model $Var(y|x) = x^2*\sigma^2$ was performed. It made the squared correlation coefficient, r^2, increase by 0.10, on the average, compared to the unweighted regression, but was not uniquely the best. There might be other weighting rules which perform equally well.

Except for three aggregates, the observations clustered around lines with slopes less than 1.00. In two cases, the intercepts were different from zero.

Thus, for the 1985 FEX, we suggested that the cv of the average in a domain of study can be approximated according to the following formula

$$cv(\bar{x}_g) = (b_0 + \sqrt{\frac{4353}{n_g}} * b_1) * cv(\bar{x}) \quad , \qquad (8.1)$$

where b_0 and b_1 are given in Table 4. This formula should be used for all levels of n_g. Calculations of the prediction error were not fully reliable since the domains of study were not randomly selected.

We saw indications that (8.1) tended to over-estimate the cvs of large domains of study. There were some large deviations in small domains of study that fell above the regression line. Ratios deviating from the average line could occur in:

* very small domains of study,
* domains of study whose composition of households was different from the all sample composition,
* domains of study with a pattern of expenditure different from the average.

Although it had been decided that the random group model

$$cv(\bar{x}_g) = (\sqrt{4353/n_g}) * cv(\bar{x})$$

was inaccurate, it can be used when more detailed studies cannot be made. This suggestion is more conservative than the regressions recommended above. (See Figures 5-12.)

Our study was restricted to cvs of averages. Nevertheless, regular estimation of precision has shown that for certain expenditure variables the cv for the total is mainly the same as the cv of the average in the Swedish FEXs.

APPENDIX 1

LIST OF DOMAINS OF STUDY.

Table 1 Type of Household

Domain	Number of Respondents
Single persons -64 years without children	
- women	221
- men	271
- all	492
Single persons 65- years without children	139
Single persons with children	
- 1 child	119
- more than 1 child	104
- all	223
Cohabitant households without children	
- up to 64 years	630
- 65 years and older	295
Cohabitant households with children	
- 1 child	490
- 2 children	910
- more than 2 children	427
- all	1827
Other without children	373
Other with children	375

Table 2 Stage in Life Cycle

Domain	Number of Respondents
Single persons -24 years without children	
- women	60
- men	64
- all	124
Cohabitant households -24 years without children	79
Single persons 25-44 years without children	
- women	83
- men	128
- all	211
Cohabitant households 25-44 years without children	196
Single persons 45-64 years without children	
- women	78
- men	79
- all	157
Cohabitant households 45-64 years without children	355

Table 3 Households with Children by Age of Children

Domain	Number of Respondents
Single persons with children	
- children -6 and maybe 7-17 years	82
- children only 7-17 years	141
- children 18-24 and maybe -17 years	69
- all	292
Cohabitant households with children	
- children only -6 years	518
- children only 7-17 years	754
- children -6 and 7-17 years	555
- children -17 and 18-24 years	315
- children only 18-24 years	238
- all	2380
All households with children	2672

Table 4 Socioeconomic Group

Domain	Number of Respondents
Worker	
- unskilled	699
- skilled	717
- all	1416
Assistant non-manual employees	
- lower level	168
- higher level	386
- all	554
Intermediate non-manual employees	851
Higher non-manual employees	258
Upper-level executives	85
Self-employed professionals	182
Farmer	118
Entrepreneurs	220
Pensioner	480
Other	190

Table 5 Type of Household and Socioeconomic Group

Domain	Number of Respondents
Single persons	
- unskilled worker	201
- skilled worker	108
- assistant non-manual employees	129
- intermediate and higher non-manual employees and upper-level executives	122
- pensioner and other	259
- all	854
Cohabitant households without children	
- unskilled worker	103
- skilled worker	138
- assistant non-manual employees	117
- intermediate and higher non-manual employees and upper-level executives	172
- self-employed professionals, farmer and entrepreneurs	66
- pensioner and other	329
- all	925
Cohabitant households with children	
- unskilled worker	270
- skilled worker	353
- assistant non-manual employees	223
- intermediate non-manual employees	495
- higher non-manual employees and upper-level executives	186
- self-employed professionals, farmer and entrepreneurs	287
Other	
- unskilled worker	125
- skilled worker	118
- assistant non-manual employees	85
- intermediate non-manual employees	142
- higher non-manual employees and upper-level executives	77
- self-employed professionals, farmer and entrepreneurs	132
- pensioner and other	69
- all	748

Table 6 Type of Household and Degree of Employment

Domain	Number of Respondents
Single persons without children	
- up to 74%	142
- 75-100%	350
Single persons with children	
- up to 74%	107
- 75-100%	116
Cohabitant households without children	
- up to 49%	60
- 50-74%	133
- 75-79%	149
- 80-100%	288
Cohabitant households with children	
- up to 74%	451
- 75-99%	833
- 100%	539
- all	1823
Other	
- up to 49%	119
- 50-74%	333
- 75-99%	158
- 100%	96
- all	706
All	3874

Table 7 Country Areas

Domain	Number of Respondents
Stockholm	780
East central	778
South central	434
Southern Sweden	578
Western Sweden	849
North central	459
Lower north	205
Upper north	271

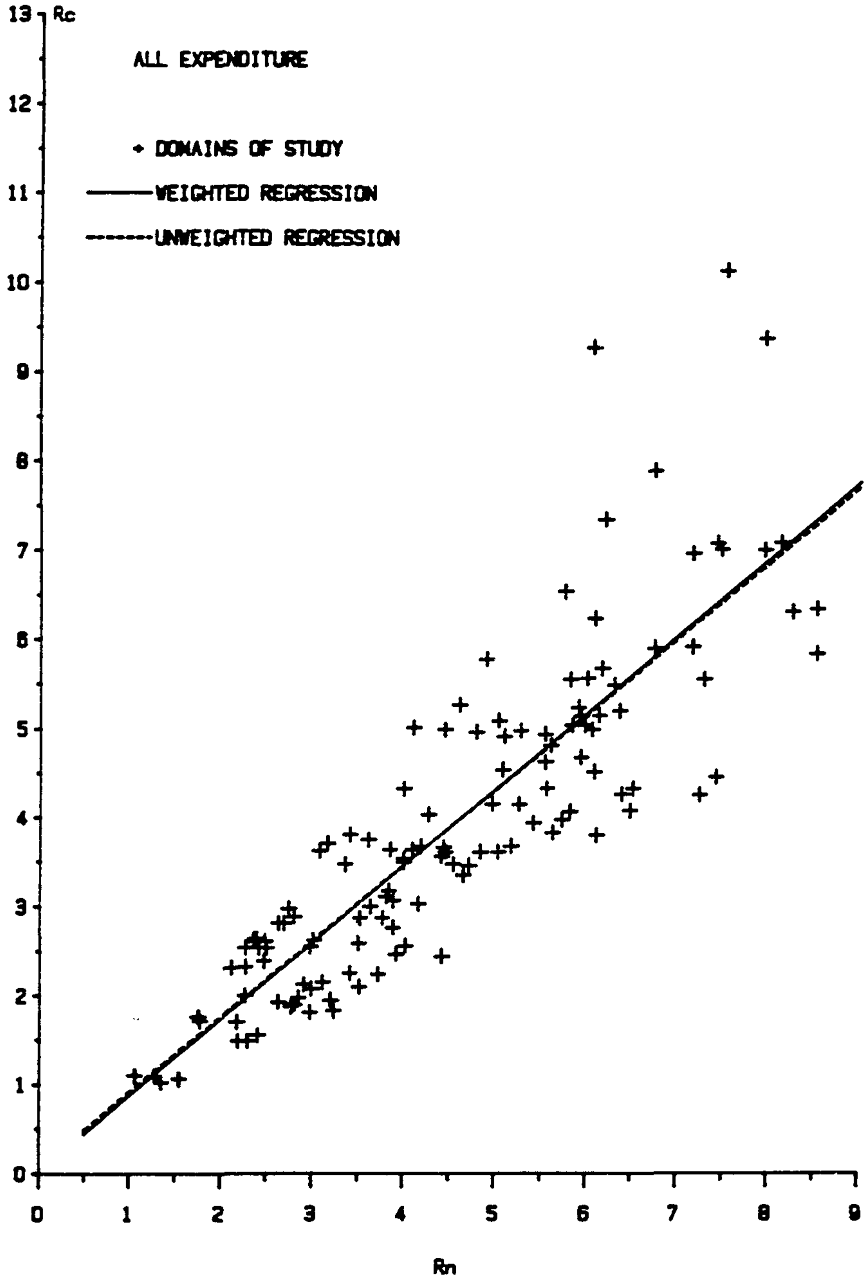
Table 8 H-region and Type of Household

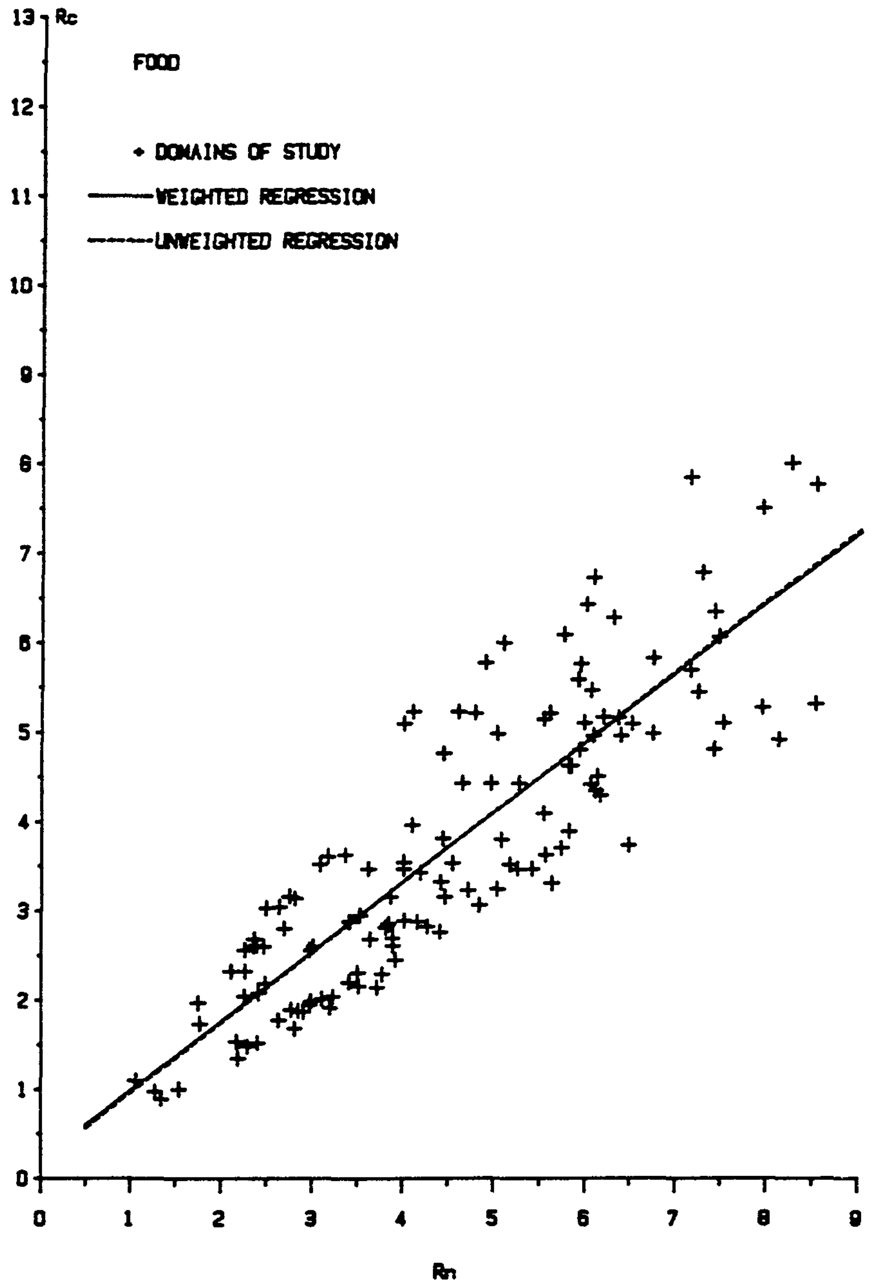
Domain	Number of Respondents
Stockholm	
- single persons	177
- cohabitant households without children	163
- cohabitant households with children	306
- other	114
- all	760
Göteborg and Malmö	
- single persons	124
- cohabitant households without children	129
- cohabitant households with children	251
- other	96
- all	600
Major towns	
- single persons	271
- cohabitant households without children	299
- cohabitant households with children	569
- other	248
- all	1387
Southern areas	
- single persons	172
- cohabitant households without children	219
- cohabitant households with children	418
- other	169
- all	978
Northern areas	
- single persons	110
- cohabitant households without children	115
- cohabitant households with children	283
- other	121
- all	629

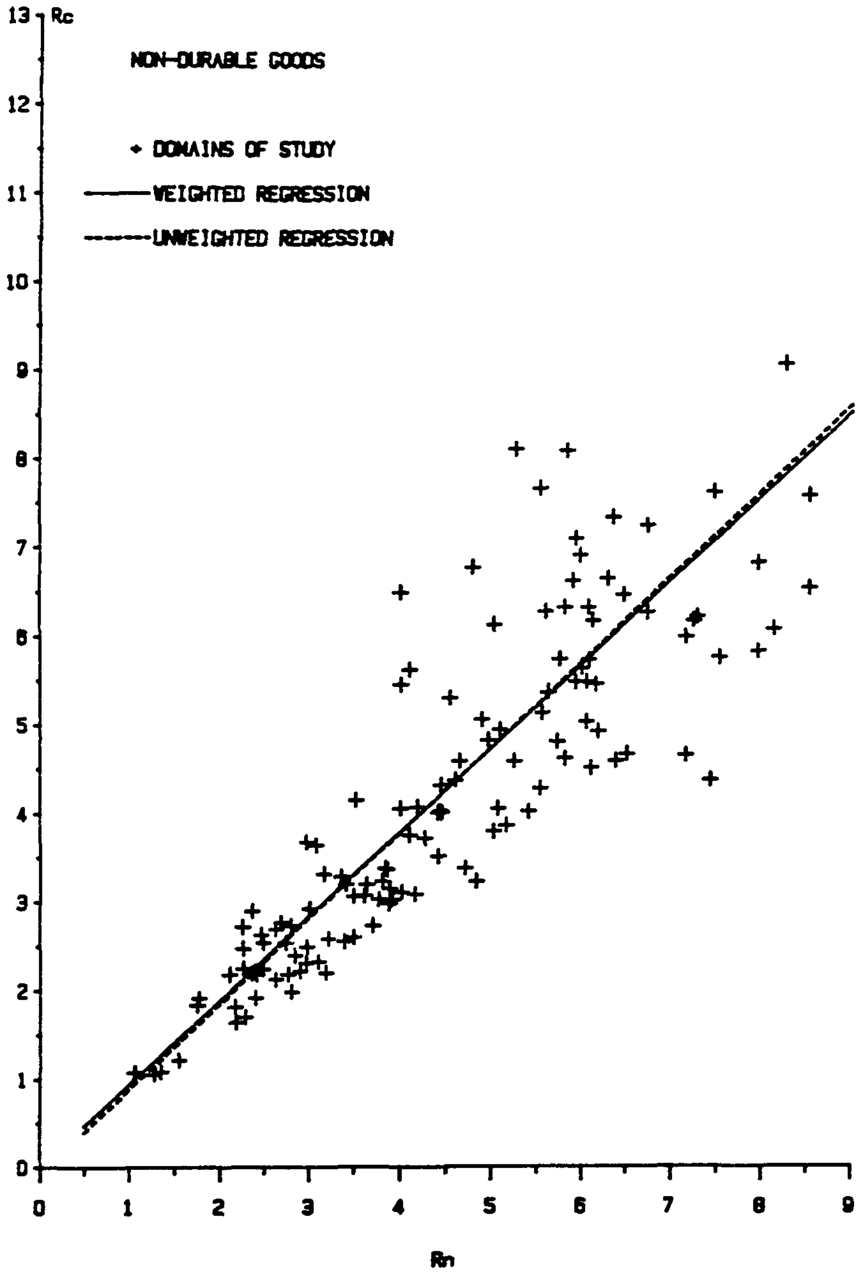
APPENDIX 2

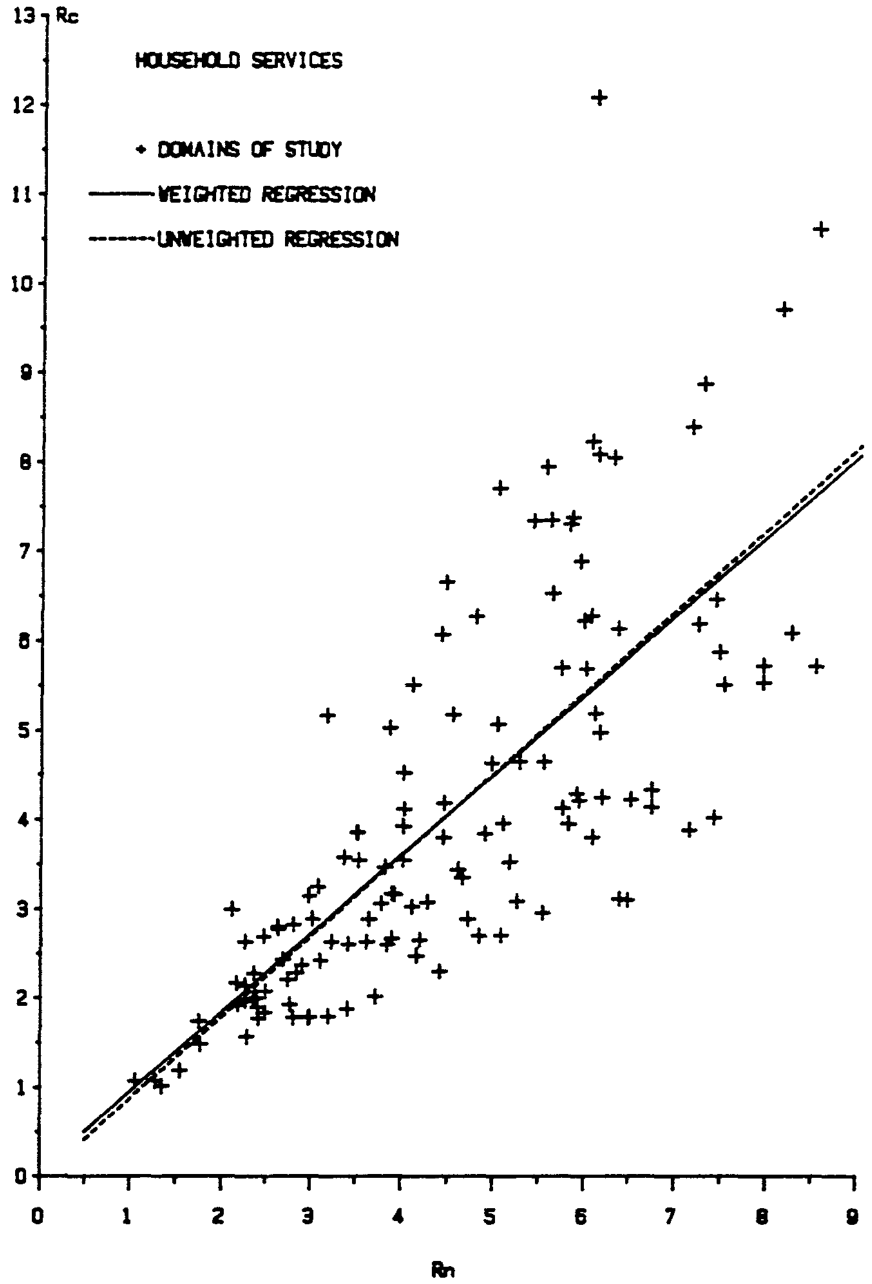
PLOTS OF CV RATIOS WITH WEIGHTED AND UNWEIGHTED REGRESSIONS

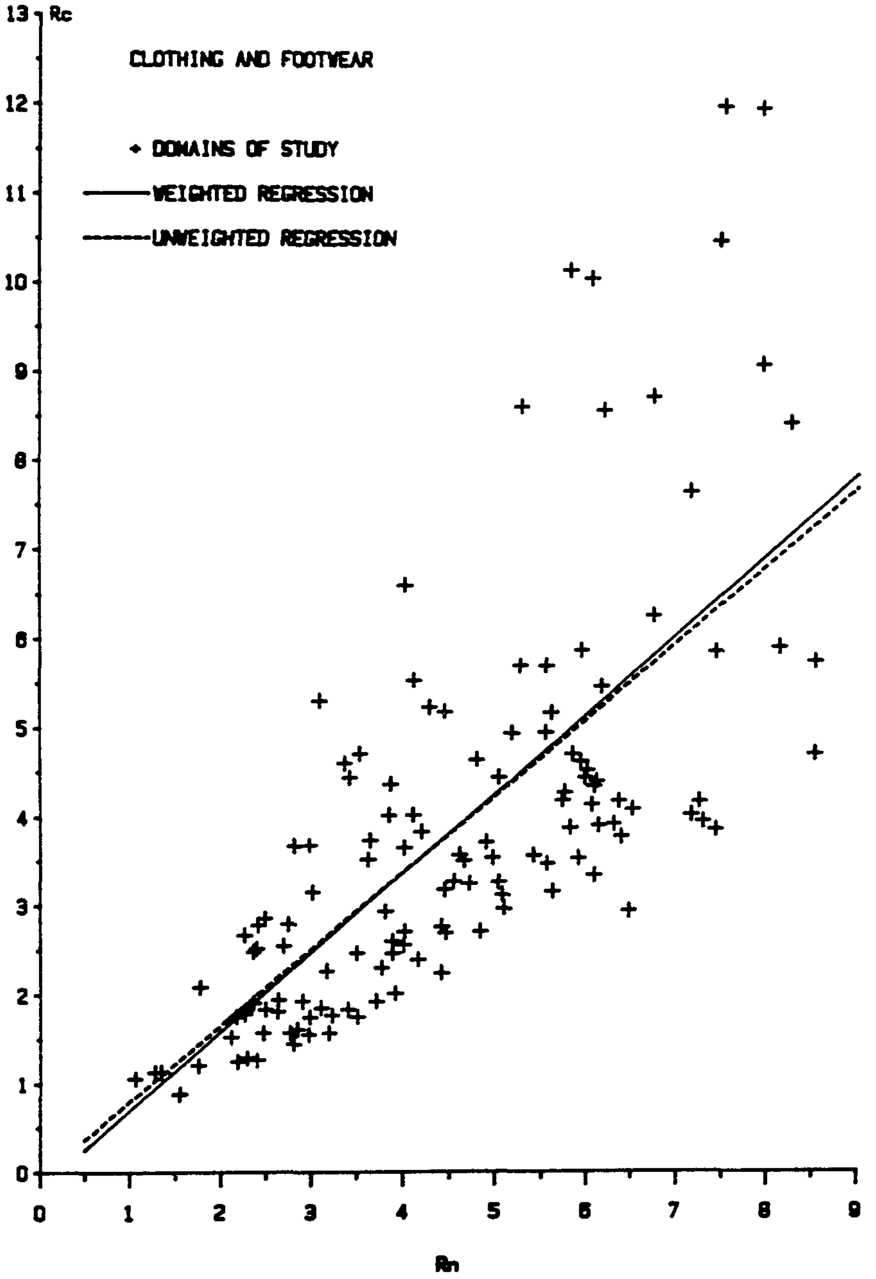
Plots of (R_n, R_c) aggregates from 130 domains of study and their unweighted (-----) and weighted (———) regressions.

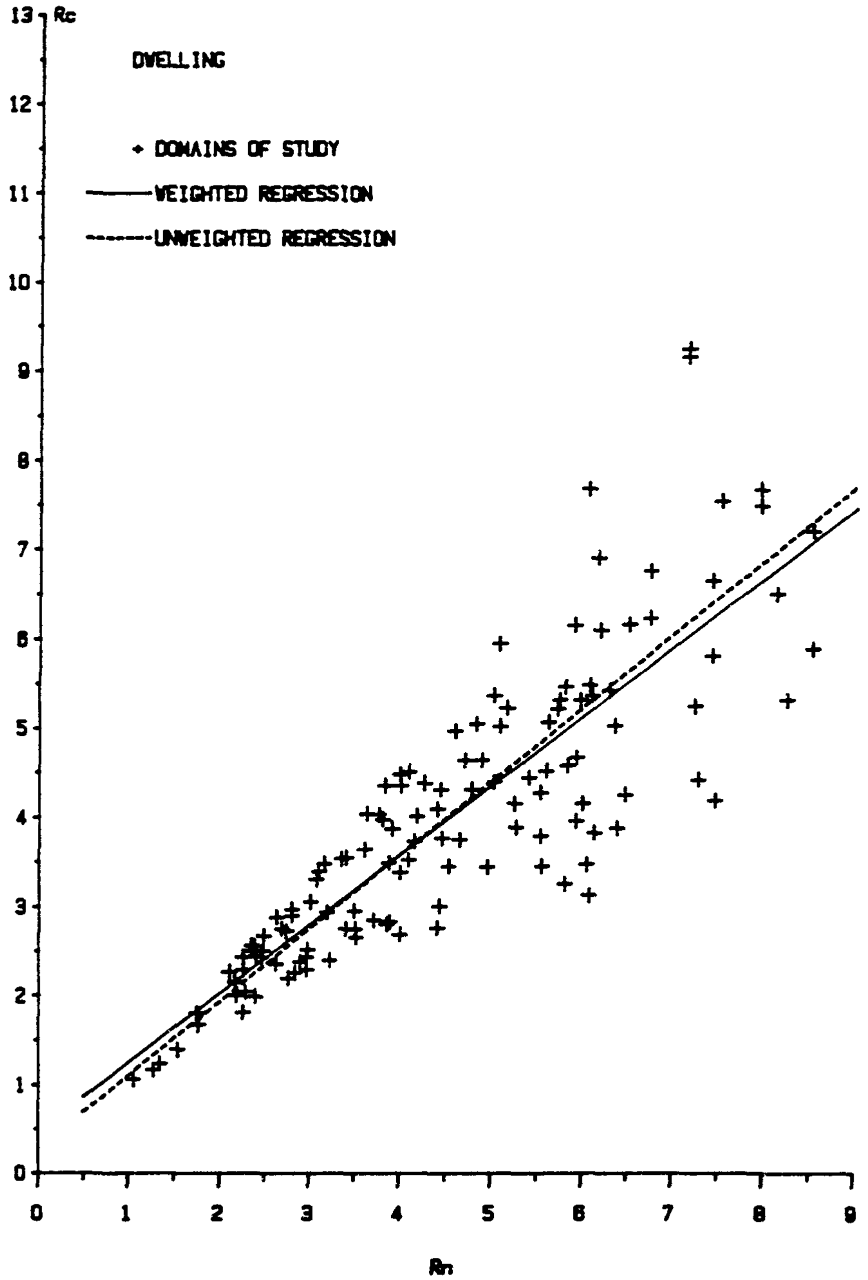



Rc
FOOD
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Rn

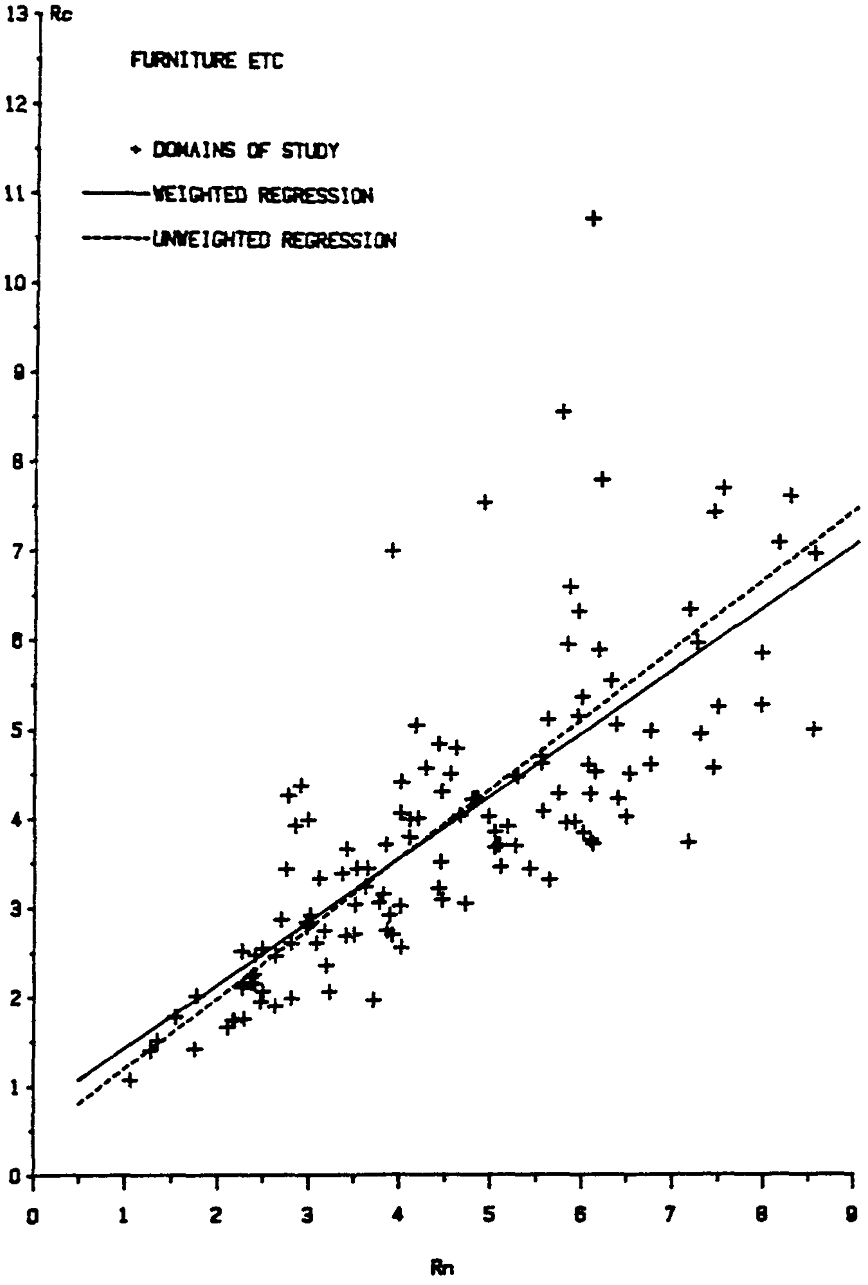

13
12
11
10
9
8
7
6
5
4
3
2
1
0
Rc
NON-DURABLE GOODS
+ DOMAINS OF STUDY
——— WEIGHTED REGRESSION
------ UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
Rn

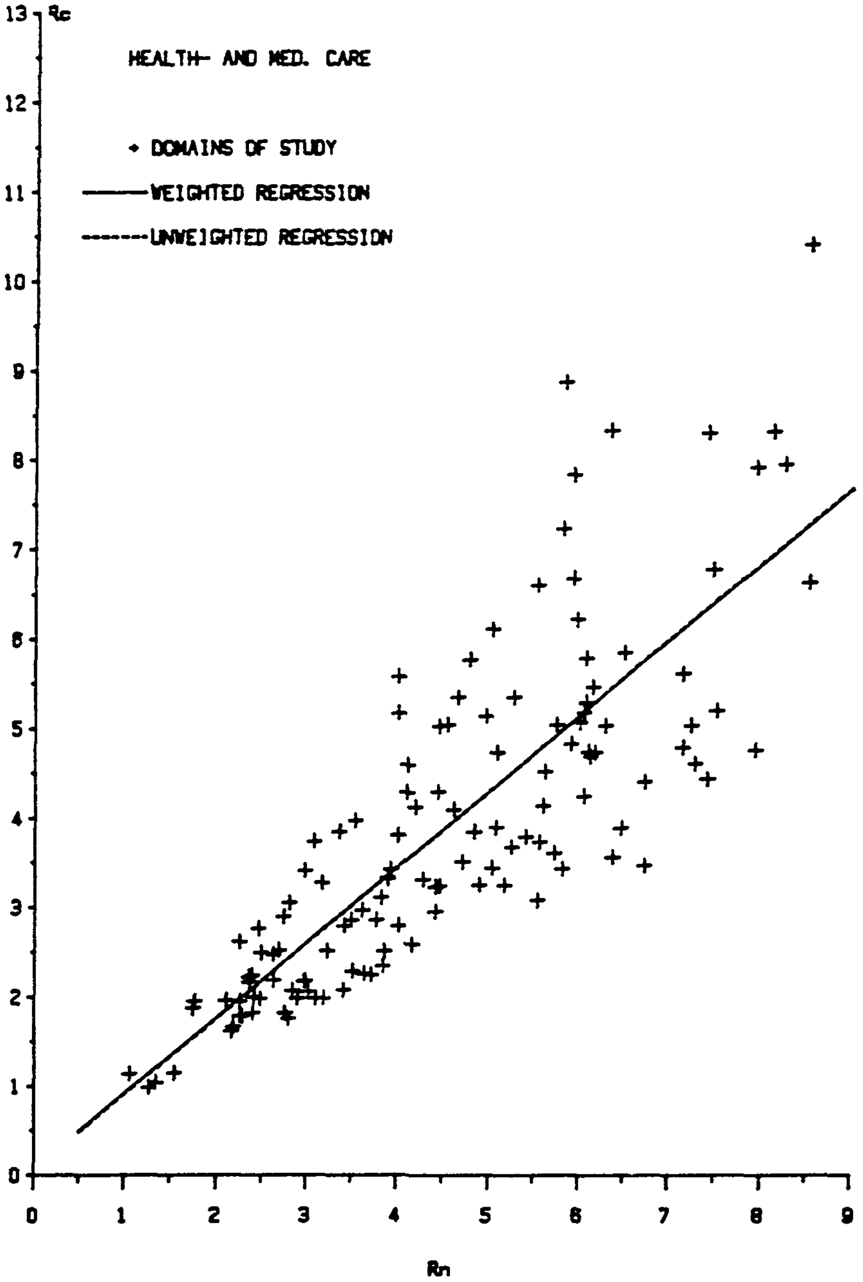

13
12
11
10
9
8
7
6
5
4
3
2
1
0
Rc
HOUSEHOLD SERVICES
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
Rn


Rc
CLOTHING AND FOOTWEAR
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn

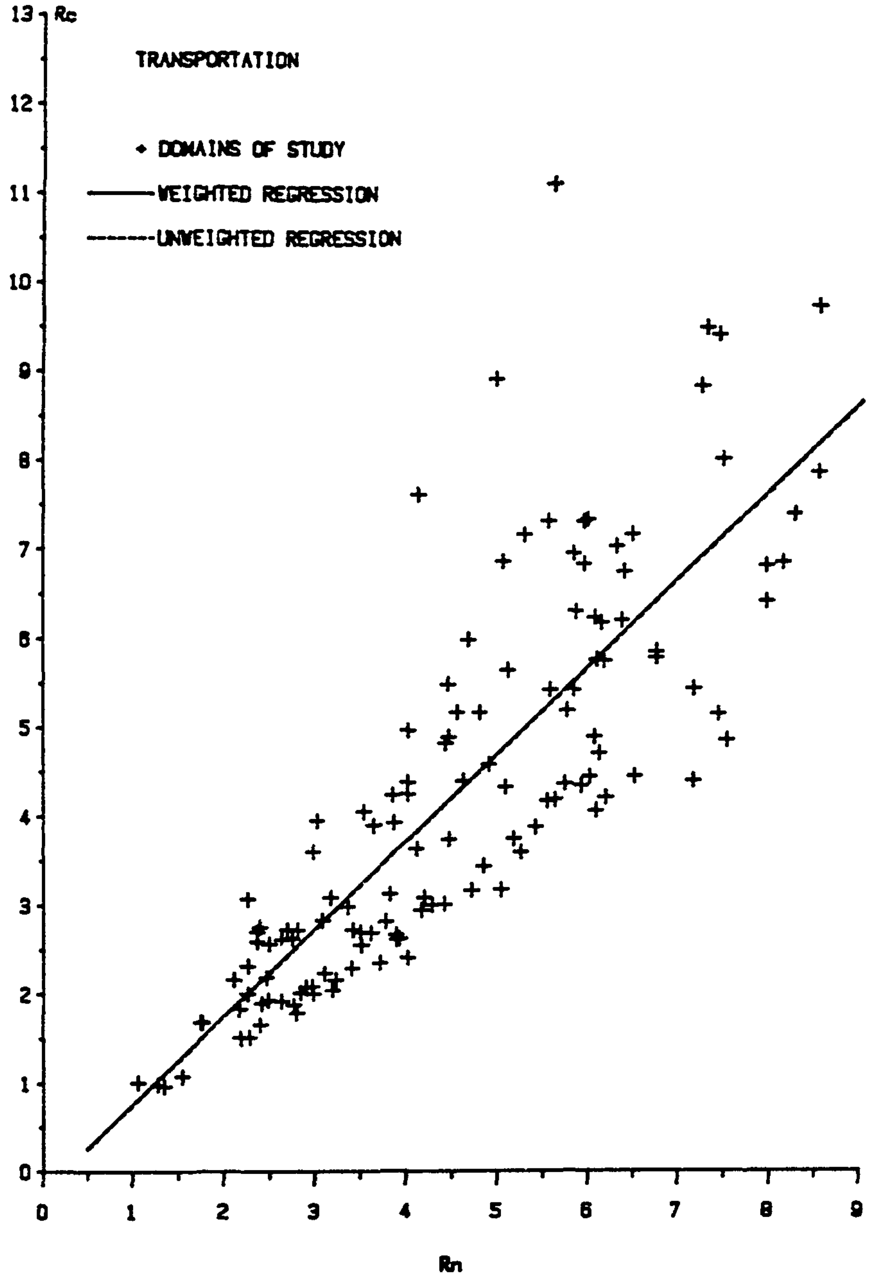


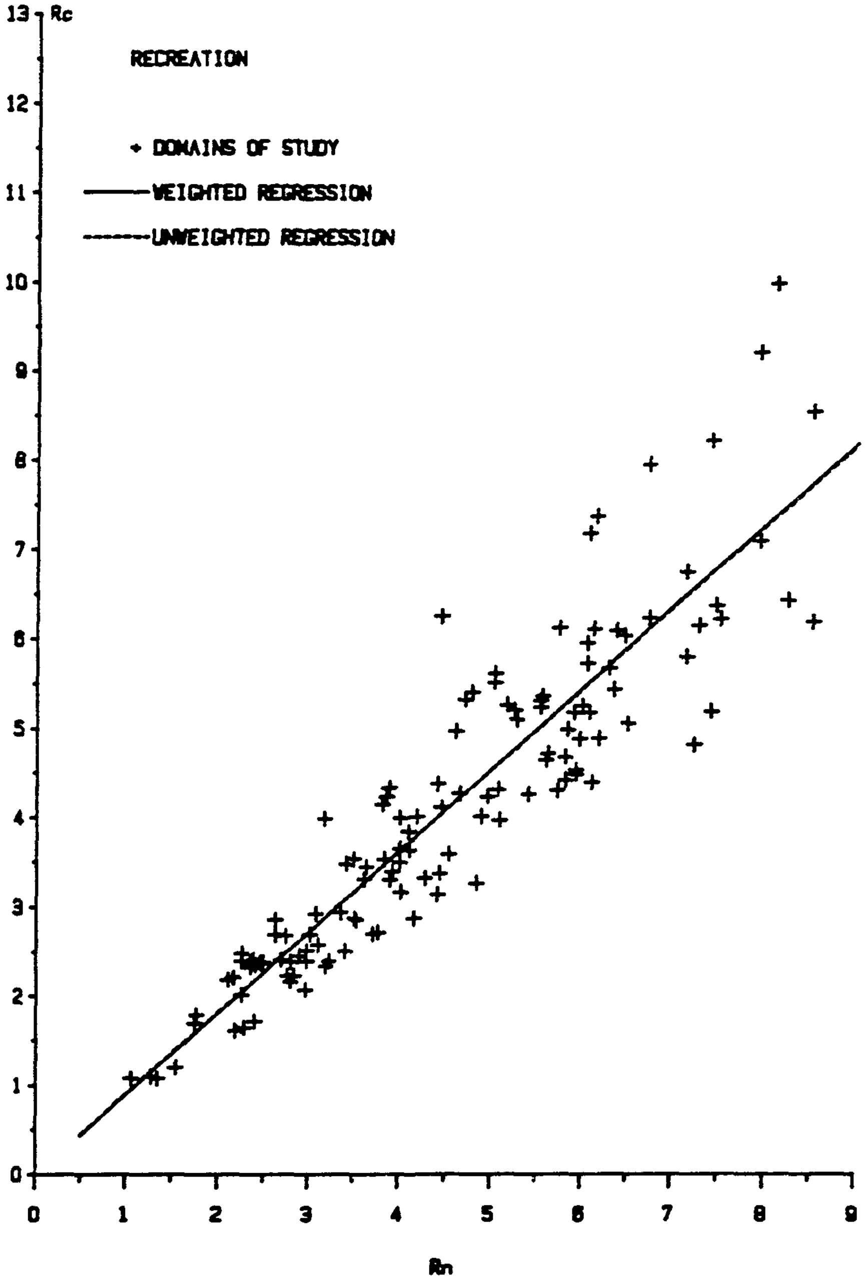
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Rc
DWELLING
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
Rn

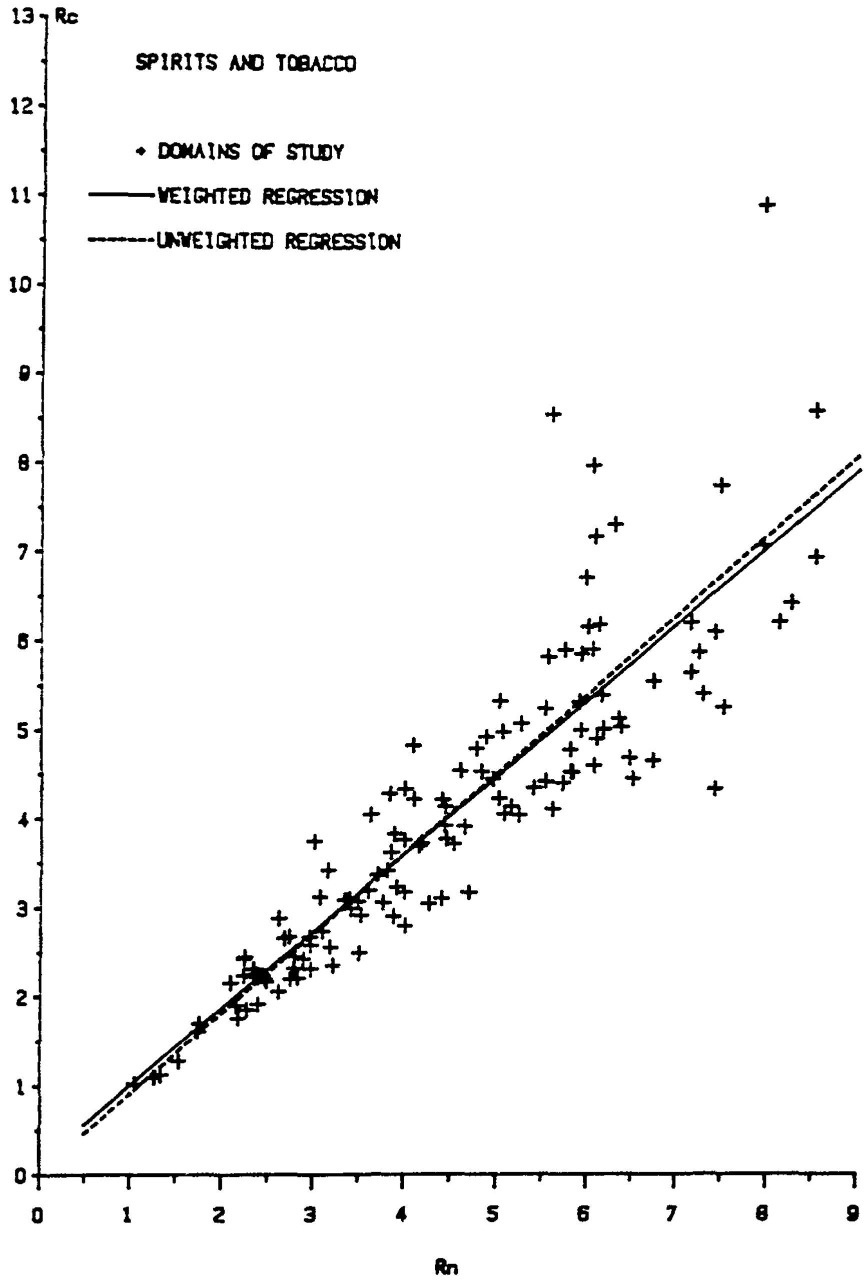

Rc
FURNITURE ETC
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
1
2
3
4
5
6
7
8
9
Rn



13
Rc
HEALTH- AND MED. CARE
12
+ DOMAINS OF STUDY
11
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
Rn


13
12
11
10
9
8
7
6
5
4
3
2
1
0
Rc
TRANSPORTATION
+ DOMAINS OF STUDY
——WEIGHTED REGRESSION
------UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
Rn


Rc
RECREATION
+ DOMAINS OF STUDY
——WEIGHTED REGRESSION
------UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
1
2
3
4
5
6
7
8
9
Rn


Rc
SPIRITS AND TOBACCO
+ DOMAINS OF STUDY
WEIGHTED REGRESSION
UNWEIGHTED REGRESSION
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Rn

APPENDIX 3 LIST OF OUTLIERS

		All	Food	Non-	House	Cloth	Dwell	Furn	Heal	Trp	Rec	Spi
	ng	exp		dura	serv	footw		art	med			
Table 1												
Single persons 65- years without children	139									Trp		Spi
Table 2												
Single men 25-44 years without children	128								Heal			
Single men 45-64 years without children	79			Non-								
Table 3												
Sing pers with children 18-24 and maybe -17	69					Cloth						
Table 4												
Upper-level executives	85						Dwell					
Self-empl professionals	182							Furn				
Farmer	118				House							
Entrepreneurs	220										Rec	
Table 5												
Single persons												
Assist non-manual empl	129					Cloth						
Pensioners and other	259									Trp		
Cohab househ without ch												
Self-empl profession , farmer and entrepren	66										Rec	
Cohab househ with ch												
Self-empl profession , farmer and entrepren	287							Furn				
Other												
Assist non-manual empl	85						Dwell					
Higher non-manual empl and upper-level exec	77	All				Cloth						
Self-empl profession , farmer and entrepren	132							Furn				
Pensioners and other	69											Spi
Table 6												
Other (degree of empl) up to 49%	119	All				Cloth		Furn				Spi
Table 8												
Stockholm single persons	177									Trp		

APPENDIX 4

OTHER EXAMPLES OF GENERALIZED PRECISION FUNCTIONS

At Statistics Sweden's Research Institute of Living Conditions (where the FEXs are conducted) variance functions have been found useful in several cases.

Twenty years ago a set of tables for binary variables was produced. Standard errors for percentages, differences between percentages, and for totals were calculated and presented in these tables. The entries were sample size and the percentage. The design was simple random sampling, s.r.s, except in one table were clustering effects were calculated. The set of tables was much in use specially before the introduction of personal computers made it easy for everyone to design such precision tables to their own needs. It is still in use occasionally but will not be revised any more.

More complex variances were evaluated for the Survey of Living Conditions. These variables were also binary but the sampling was not s.r.s. of individuals but of clusters consisting of one or two persons. The clustered persons were man and wife (married or not, but cohabiting). Standard errors were studied both theoretically and empirically. It was possible to establish theoretical limits for the standard errors both among individuals and households. The limits were expressed as functions of the sample size and the percentage. Whether the variance of a domain of study approached its lower or the upper limit depended on the variance's distribution on one-person and two-person clusters.

Empirical studies of a great number of domains of study and variables demonstrated that the theoretical interval could be narrowed for practical purposes with small risk for underestimating the confidence intervals. As a consequence, the Survey of Living Conditions used modeled standard errors in most standard applications. Since 1980, the survey has turned to s.r.s of individuals, making simple variance functions an obvious choice.

At least two studies have been performed on quantitative variables. In the Household Income Survey up to 1987, there was a highly stratified sample with much disproportionate allocation. Variances in domains of study were studied as a function of the entire sample's variance and the distribution of the domain of study on the strata. This attempt at modeling was not successful, but the survey has since turned to a simpler and more efficient design. At present, promising attempts are going on.

The first FEX to attempt to use generalized precision functions was the 1978 survey. The goal was to identify simple ways to approximate cvs in domains of study as a function of the number of observations and the entire sample cv of the same variable. Average R_cs in domains of study were calculated for both separate expenditures and for aggregates of expenditures and plotted against R_n.

Average R_cs for aggregates were slightly above the line R_c=R_n in all domains of study. The range of values for the separate aggregates was fairly wide. We were not able to analyze the degree to which the variation was random and could be explained by differences in the distribution of the variables studied. The average cv ratios for separate expenditures came closer to the line R_c=Rn, but the range was still larger than for the averages.

We concluded that we should use

$s_{x_g} = s_x * \sqrt{n/n_g}$ only as a first approximation of the standard error of a domain of study (when regular calculations are not possible). It would not be possible to use the approximation with any confidence without additional studies to further develop the method.

Since the 1978 FEX had a different and more complicated design than the 1985 survey, there was no reason to expect that the present study would produce the same results and the 1978 study. The two studies also differed in their choices of aggregates.

9 REFERENCES

[1] Bloch, D.A. and Moses, L.E.: (1988) Nonoptimally Weighted Least Squares. The American Statistician, Vol 43,no 1.

[2] Cochran,W.G. (1963). Sampling Techniques. John Wiley and Sons, New York, second edition.

[3] Draper, N.R and Smith,H. (1981). Applied regression analysis, Second edition, Wiley and Sons, New York.

[4] Draper,N.R. and John J.A. (1981) Influential observations and Outliers in Regression, Technometrics, Vol 23, No 1.

[5] Hansen,M.H., Hurwitz,N.H. and Madow, W.G. (1966): Sample Survey Methods and Theory. John Wiley and Sons, New York, 7th printing.

[6] Hushållens utgifter 1985 (1987), Sveriges officiella statistik, SCB (Family Expenditure in 1985, The official statistics of Sweden, Statistics Sweden)

[7] Wolter,K.M. (1985): Introduction to Variance Estimation. Springer-Verlag, New York.

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).

Reports published earlier during 1989 are:

1989:1 (grön)	Går det att mäta produktivitetsutvecklingen för SCB? **(Rune Sandström)**
1989:2 (grön)	Slutrapporter från U-avdelningens översyn av HINK och KPI **(flera författare)**
1989:3 (grön)	A Cohort Model for Analyzing and Projecting Fertility by Birth Order **(Sten Martinelle)**
1989:4 (beige)	**Abstracts** I - Sammanfattningar av metodrapporter från SCB
1989:5 (gul)	On the use of Semantic Models for specifying Information Needs **(Erik Malmborg)**
1989:6 (grön)	On Testing for Symmetry in Business Cycles **(Anders Westlund och Sven Öhlén)**
1989:7 (grön)	Design and quality of the Swedish Family Expenditure Survey **(Håkan L Lindström, Hans Lindkvist och Hans Näsholm)**
1989:8 (grön)	Om utnyttjande av urvalsdesignen vid regressionsanalys av surveydata **(Lennart Nordberg)**
1989:9 (grön)	Variations in the Age-Pattern of Fertility in Sweden Around 1986 **(Michael Hartmann)**

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.

Dito GULA exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 Stockholm, eller per telefon 08-783 41 47.