Statistiska centralbyrån Statistics Sweden

Regression Estimators in Theory and in Practice

Tomas Garås

R&D Report
Research - Methods - Development
1996:3

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 1996:3. Regression estimators in theory and in practice / Tomas Garås
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1996-X101OP9603

Statistiska centralbyrån Statistics Sweden

Regression Estimators in Theory and in Practice

Tomas Garås

R&D Report
Research - Methods - Development
1996:3

Från trycket November 1996
Producent Statistiska centralbyrån, utvecklingsavdelningen
Ansvarig utgivare Lars Lyberg

Förfrågningar Tomas Garås
tel 08-783 49 57
telefax 08-783 48 12

ISSN 0283-8680

Printed November, 1996
Producer Statistics Sweden, Department of Research and Development
Publisher Lars Lyberg

Inquiries Tomas Garås
telephone +46 8 783 49 57
telefax +46 8 783 48 12

ISSN 0283-8680

1996-11-01

REGRESSION ESTIMATORS IN THEORY AND IN PRACTICE

Statistics Sweden, AM/UTV, Tomas Garås

ABSTRACT :

Regression estimators are often an effective way to use auxiliary information. There are several registers and censuses at Statistics Sweden that could be used as sources of auxiliary information.

The regression estimator is in theory very efficient for estimates of levels. For the SRS-design you get

$\hat{Y}_{reg} = \hat{Y} + \hat{\beta}\left(Z - \hat{Z}\right)$ with the variance $V\left(\hat{Y}_{reg}\right) \cong V\left(\hat{Y}\right)\left[1 - \rho_{yz}^2\right]$,

where z is the auxiliary variable in the frame. The variance-efficiency is then $\left[1 - \rho_{yz}^2\right]$ compared with the usual SRS-estimator.

The main problem discussed in this paper, is that estimators of change (such as ratios or differences between regression estimators of levels) do not uniformly have the good properties as described above in the case of levels.

Other problems discussed in this paper are, what happens with the regression estimates in practice when you have non-sampling errors as non-response, overcount and so on?
Or what happens when the estimate of the coefficients of regression are unstable?
Will the theoretical properties remain in practice?

This paper will deal with these problems along with practical examples from a survey at Statistics Sweden.

1996-11-01

REGRESSION ESTIMATORS IN THEORY AND IN PRACTICE

Statistics Sweden, AM/UTV, Tomas Garås

1996-11-01

REGRESSION ESTIMATORS IN THEORY AND IN PRACTICE

Statistics Sweden, AM/UTV, Tomas Garås

1. INTRODUCTION

Regression estimators are in theory and often in practice very efficient when there is good auxiliary information available.

More and more auxiliary information are becoming available at Statistics Sweden, for example the traditional registers for statistical purposes, censuses made at Statistics Sweden and external registers not for statistical purposes but never the less useful.

The ordinary sample designs at Statistics Sweden are stratified (one-stage) SRS with the usual expansion estimators. The efficiency of these estimates can often be greatly improved with help of auxiliary information through regression estimators.

This study will be in two blocks, regression estimators in theory and in practice. Both estimators of levels and of changes will be considered.
The theoretical situation is briefly when there are no disturbances as non-response, frame problems and so on from nonsampling errors and the variance properties are known.
The practical situation is when the survey is actually done with usual disturbances from nonsampling errors:

---- *A design not fit for regression estimates.*
---- *The stability of the estmates of the coefficients of regression. This can be considered as the classical problem of regression estimates.*
---- *Non-response, frame problems and so on.*

So, the question is, what happens with the regression estimates (practical situation) versus the theoretical case according to properties as expectation value right (EVR), variance and variance efficiency compared to the ordinary expansion estimators ?

In the case with regression estimators of changes with the efficiency compared to the usual ratio estimator the picture is more unclear than the case with estimators of levels. These are uniformly very efficient for regression estimators. This is not the case for estimators of change and also the formulas with regression estimators are very messy here.

The efficiency properties of the estimators of changes versus the ordinary ratio estimators will be dealt with in this paper.

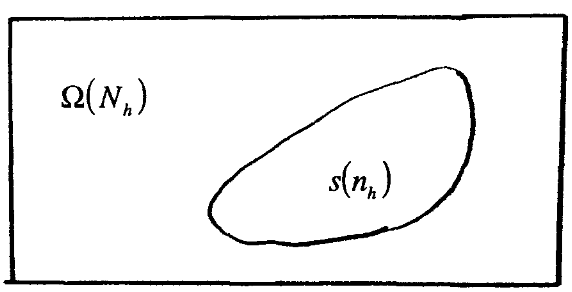
Practical examples will be given from a survey of retail business at Statistics Sweden that will illustrate both the power and the problems with regression estimates. Both estimates of levels and changes will be illustrated as well as the usual estimates.

The following persons have contributed to this paper:
Kajsa Lindell (AM/UTV) has looked over the English. Esbjörn Ohlsson and Patrik Öhagen (ES/SES) have given advise concerning section 4.1.1, Correlation matrix.

2. PRELIMINARIES

The design will be stratified (one-stage) SRS (STSRS) with complete panels over time and the population under study will be non-dynamic. A non-dynamic population has no in- or outflow of objects over time and this will also be valid for the strata populations. It will often be enough to study various properties for the SRS design.

Population- and sample notations will be as below.



Design for all time periods under study.

x^t variable under study, time period t.
z^t auxiliary variable, time period t.

($x^1 = x, x^2 = y, z^1 = m, z^2 = z$)

t time period under study
h stratum

$\Omega(N_h)$ Population with N_h objects, stratum h.
$s(n_h)$ Sample with n_h objects, stratum h.
The formula notations are as follows for the design STSRS.

A. ORDINARY EXPANSION ESTIMATOR FOR LEVELS/TOTALS.

$$\hat{Y} = \sum_h \frac{N_h}{n_h} \sum_{i \in s}^{n_h} y_{hi} \quad \text{with the variance}$$

$$V(\hat{Y}) = \sum_h N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2 \quad \text{where} \quad S_h^2 = \frac{\sum_{i \in \Omega}^{N_h} \left(y_{hi} - \overline{Y}_h \right)^2}{N_h - 1} \quad .$$

B. SEPARATE REGRESSION ESTIMATOR FOR LEVELS/TOTALS.

$$\hat{Y}_{reg} = \sum_h \left[\hat{Y}_h + b_{hyz} \left(Z_h - \hat{Z}_h \right) \right], Z_h = \sum_{i \in \Omega}^{N_h} z_{hi} \quad \text{with the variance}$$

$$V\left(\hat{Y}_{reg}\right) \cong \sum_h V\left(\hat{Y}_h\right)\left[1 - \rho_{hyz}^2\right] \quad \text{where the estimator of the}$$

coefficient of regression is $b_{hyz} = \dfrac{\sum_{i \in s}^{n_h} (y_{hi} - \bar{y}_h)(z_{hi} - \bar{z}_h)}{\sum_{i \in s}^{n_h} (z_{hi} - \bar{z}_h)^2}$

and the coefficient of correlation is $\rho_{hyz} = \dfrac{\sum_{i \in \Omega}^{N_h} \left(y_{hi} - \overline{Y}_h \right)\left(z_{hi} - \overline{Z}_h \right)}{\sqrt{\sum_{i \in \Omega}^{N_h} \left(y_{hi} - \overline{Y}_h \right)^2 \sum_{i \in \Omega}^{N_h} \left(z_{hi} - \overline{Z}_h \right)^2}}$.

C. ORDINARY COMBINED RATIO ESTIMATOR FOR CHANGES.

$\hat{R} = \dfrac{\hat{Y}}{\hat{X}}$ with the variance $V(\hat{R}) \cong \dfrac{1}{X^2}\left[V(\hat{Y}) + R^2 V(\hat{X}) - 2R\,Cov(\hat{Y}, \hat{X})\right] =$

$= R^2 \left[\dfrac{V(\hat{Y})}{Y^2} + \dfrac{V(\hat{X})}{X^2} - \dfrac{2Cov(\hat{Y}, \hat{X})}{YX} \right]$ where R = Y/X and the

covariance is $Cov(\hat{Y}, \hat{X}) = \sum_h N_h^2 \left(\dfrac{1}{n_h} - \dfrac{1}{N_h} \right) S_{hy} S_{hx} \rho_{hyx}$.

D. REGRESSION COMBINED RATIO ESTIMATOR FOR CHANGES.

$$\hat{R}_{reg} = \frac{\hat{Y}_{reg}}{\hat{X}_{reg}} = \frac{\sum_h \left[\hat{Y}_h + b_{hyz}\left(Z_h - \hat{Z}_h\right)\right]}{\sum_h \left[\hat{X}_h + b_{hxm}\left(M_h - \hat{M}_h\right)\right]}$$ with the variance

$$V\left(\hat{R}_{reg}\right) \cong \frac{1}{X^2}\left[V\left(\hat{Y}_{reg}\right) + R^2 V\left(\hat{X}_{reg}\right) - 2RCov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)\right] =$$

$$= R^2\left[\frac{V\left(\hat{Y}_{reg}\right)}{Y^2} + \frac{V\left(\hat{X}_{reg}\right)}{X^2} - \frac{2Cov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)}{YX}\right]$$ where the covariance is

$$Cov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right) = \sum_h N_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{hy} S_{hx} \rho_{hyx}\left[1 - \frac{\left(\rho_{hxm}\rho_{hym} + \rho_{hyz}\rho_{hxz} - \rho_{hyz}\rho_{hxm}\rho_{hzm}\right)}{\rho_{hyx}}\right] .$$

E. DOUBLE RATIO ESTIMATORS.

The case with the double ratio estimators will only be dealt with briefly, but the formulas will be given in section 4.1.3 .

Ordinary estimator: $$\hat{Q} = \frac{\hat{Y}/\hat{Z}}{\hat{X}/\hat{M}} .$$

Regression estimator: $$\hat{Q}_{reg} = \frac{\hat{Y}_{reg}/\hat{Z}_{reg}}{\hat{X}_{reg}/\hat{M}_{reg}} .$$

Here four arbitrary auxiliary variables are possible.

The formulas for the design SRS follows from the design STSRS given here.

Proofs here and onwards will be given in APPENDIX.

3. REGRESSION ESTIMATORS OF LEVELS

3.1 THEORY

The separate regression estimator for levels with the design STSRS is

$\hat{Y}_{reg} = \sum_h \left[\hat{Y}_h + b_{hyz}\left(Z_h - \hat{Z}_h\right)\right]$ with the variance $V\left(\hat{Y}_{reg}\right) \cong \sum_h V\left(\hat{Y}_h\right)\left[1 - \rho^2_{hyz}\right]$

as mentioned earlier.

Without further loss of generality the SRS design can be studied,

$\hat{Y}_{reg} = \hat{Y} + b_{yz}\left(Z - \hat{Z}\right)$ with the variance $V\left(\hat{Y}_{reg}\right) \cong V\left(\hat{Y}\right)\left[1 - \rho^2_{yz}\right]$.

The expected value of $\hat{Y}_{reg}$ conditioning on b_{yz} , $E\left(\hat{Y}_{reg} \mathrm{I} b_{yz}\right) = Y$, is expected value right. The unconditioned expected value of $\hat{Y}_{reg}$ have a bias of order 1/n. The bias of the sampling error $\sqrt{V\left(\hat{Y}_{reg}\right)}$ is of order $\frac{1}{\sqrt{n}}$. If $\hat{Y}_{reg} = \hat{Y} + \beta_{yz}\left(Z - \hat{Z}\right)$ the variance $V\left(\hat{Y}_{reg}\right)$ is exact (β_{yz} is the true coefficient of regression) .

Then for large samples $E\left(\hat{Y}_{reg}\right) \cong Y$ and with the variance $V\left(\hat{Y}_{reg}\right) \cong V\left(\hat{Y}\right)\left[1 - \rho^2_{yz}\right]$.

The efficiency of the regression estimator will be compared versus the ordinary expansion estimator. With the SRS design you get $Eff\left(\hat{Y}_{reg}\right) = \frac{V\left(\hat{Y}_{reg}\right)}{V\left(\hat{Y}\right)} = 1 - \rho^2_{yz}$ (y is the variable under study and z the auxiliary variable), which makes the regression estimator very efficient if the coefficient of correlation, ρ_{yz} , is large and it is always true that $V\left(\hat{Y}_{reg}\right) \le V\left(\hat{Y}\right)$. The efficiency is illustrated below.

$$Eff\left(\hat{Y}_{reg}\right) = \frac{V\left(\hat{Y}_{reg}\right)}{V\left(\hat{Y}\right)} = 1 - \rho^2_{yz}$$

ρ	0	0.20	0.50	0.70	0.80	0.90	0.99
Eff	1	0.96	0.75	0.51	0.36	0.19	0.02
$\sqrt{Eff}$	1	0.98	0.87	0.71	0.60	0.44	0.14

As can be seen a correlation of at least 0.70 makes the regression estimator very efficient compared with ordinary estimator.

So, what then happens in practice ?

3.2 PRACTICE

In practice, as always when a survey is actually done, there are a lot of disturbances on the theoretical properties from non-sampling errors as for example non-response, overcount and so on.
With a regression estimator there are to begin with in the theoretical properties slight biases in the estimator, its variance and the variance estimator; which biases are small with a proper large sample design. One practical problem is then a non-fit design.

The main problem is when the estimates of the coefficients of regression are unstable. As the formulas are conditioned on these estimates, they must be stable enough to make the formulas approximately right. This is the hard core if the regression estimates are sufficient or not.

That is, in practice disturbances occur on the theoretical properties that leads to biases and reduction of efficiency in the estimates. Non-sampling errors can also lead to stochastic errors with likewise efficiency reduction.

Below the following points are described when a regression estimator is used in practice :

3.2.1 Non-fit design.
3.2.2 Unstable estimates of coefficients of regression.
3.2.3 Overcount and undercount objects with respect to the auxiliary information.
3.2.4 Non-response.

Hold the following formulas in mind, the regression estimator of levels for the SRS design :

$\hat{Y}_{reg} = \hat{Y} + b_{yz}\left(Z - \hat{Z}\right)$ with the variance estimator $v\left(\hat{Y}_{reg}\right) \cong v\left(\hat{Y}\right)\left[1 - \hat{\rho}_{yz}^2\right]$.

y is the variable under study and z is the auxiliary variable.

3.2.1 NON-FIT DESIGN

The regression estimator is a large sample formula, that is for sufficiently large samples the estimates and their variance estimates are approximately expectation value right. Since the bias of the sampling errors is of order $1/\sqrt{n}$, a sample of at least 100 objects is needed both for the SRS design and the STSRS design . In practice many domains contains less than 100 objects in the sample, especially in enterprise statistics. It is mostly in enterprise statistics that there are good auxiliary information available.

With too few objects in the sample the regression estimator does not work but mostly with sample sizes over 50 objects it does as will be illustrated in the practical examples in section 5.

With few objects in the sample and a STSRS design some of the strata will contain very few objects according to optimal sample allocation. With non-response and perhaps some overcount objects the regression estimates in these strata will be merely nonsense. What causes this will be discussed below.

3.2.2 ESTIMATES OF COEFFICIENTS OF REGRESSION

The classical problem of regression estimates are the stableness of the estimates of coefficients of regression.

An extreme case is for example with a sample of two objects. This always leads to an estimate of the coefficient of correlation equal to one (a straight line can always go through two points), with an estimated variance equal to zero. As mentioned above this can occur in small sample strata in the STSRS design.

What is used here, is the separate regression estimator with an estimate of the coefficient of regression, b_{hyz} , in each stratum. This is the most efficient regression estimator for a proper design.
Also large sample estimators, but more robust for lesser sample sizes, are the combined regression estimator (a combined estimate of the coefficient of regression weighted from individual strata coefficients) and the separate ratio estimator for levels (the combined ratio estimator for levels is of course even more robust).
Still another alternative in time-series surveys is to estimate the coefficient of regression in the separate regression estimator by a model. The model can for example be that the coefficients of regression are stable over time, but that they in each time period are unstable estimates. Time series information can

then by a model-dependent estimate produce stable coefficients. If the model is eligible this is a good solution for "small" large sample designs and if not the other estimators mentioned earlier,which use auxiliary information, are suitable.

3.2.3 OVERCOUNT AND UNDERCOUNT OBJECTS WITH RESPECT TO THE AUXILIARY INFORMATION

Objects within the frame that do not belong to the population are called overcount objects. When estimating a total these objects will have a zero value. For the ordinary expansion estimator this will lead to a larger variance. The auxiliary information exists for all objects in the frame. In practice this means that that an overcunt object can have a non-zero auxiliary variable value. Also, the objects in the frame which belong to the population can miss auxiliary information which means a zero value of the auxiliary variable.

All this leads to reduction in efficiency of the regression estimator by the resulting smaller coefficient of correlation and hence the variance.

3.2.4 NON-RESPONSE

Non-response is more critical for the regression estimates than the ordinary estimates since the non-response can "ruin" the estimates of the coefficients of regression and correlation. As previously mentioned small strata samples with non response can lead to mere nonsense.

That is, the ordinary expansion estimates are more robust for non response errors than the regression estimates. Also, imputation methods become more complex for the regression estimates because of the more complex estimator.

3.3 CONCLUSION ON PRACTICAL PROBLEMS

a. Conclusion on practical survey problems, which always occur more or less, is that more complex estimators are more sensitive to disturbances from non-sampling errors.

b. To use a more complex estimator the practical problems shall be relatively small and the more complex estimator shall produce significantly more efficient estimates.

c. Often, but not always, it is possible to use regression estimates for levels in the surveys of Statistics Sweden.

d. The use of regression estimators and other estimators that uses auxiliary information is surprisingly low just now at Statistics Sweden.

e. If the practical problems with the non-sampling errors are too large to use complex estimators who uses auxiliary information, it is possible to use the more robust (to non-sampling errors) ordinary expansion estimators. The auxiliary information can then be used for alternative stratification, non-response treatment, modelling and so forth.

But;
with few non-sampling errors, a proper large sample-design and a good auxiliary variable the regression estimator of level is very efficient compared to the ordinary expansion estimator!

4. REGRESSION ESTIMATORS OF CHANGES

Regression estimators of changes, such as ratios or differences between regression estimators of levels, don´t uniformly have the good theoretical properties as for regression estimators of levels. Also, the formulas here are messy and not so easy to penetrate. This will be dealt with here.

In the surveys of Statistics Sweden there are always both estimates of levels and changes. For short time surveys (monthly, quarterly for example) the estimates of changes are the most important. In yearly surveys it is the opposite. Often the estimates of changes are calculated as a by-product of a design suited for the estimators of levels.

4.1 THEORY

The regression estimator of change with the STSRS design is

$$\hat{R}_{reg} = \frac{\hat{Y}_{reg}}{\hat{X}_{reg}} = \frac{\sum_h \left[\hat{Y}_h + b_{hyz}\left(Z_h - \hat{Z}_h\right)\right]}{\sum_h \left[\hat{X}_h + b_{hxm}\left(M_h - \hat{M}_h\right)\right]} \quad \text{with the variance}$$

$$V\left(\hat{R}_{reg}\right) \cong \frac{1}{X^2}\left[V\left(\hat{Y}_{reg}\right) + R^2 V\left(\hat{X}_{reg}\right) - 2RCov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)\right] =$$

$$= R^2\left[\frac{V\left(\hat{Y}_{reg}\right)}{Y^2} + \frac{V\left(\hat{X}_{reg}\right)}{X^2} - \frac{2Cov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)}{YX}\right] =$$

$$= R^2 \sum_h N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\left\{\frac{S_{hy}^2\left(1-\rho_{hyz}^2\right)}{Y^2} + \frac{S_{hx}^2\left(1-\rho_{hxm}^2\right)}{X^2} - \right.$$

$$\left. - \frac{2S_{hy}S_{hx}\rho_{hyx}}{YX}\left(1 - \frac{\left(\rho_{hyz}\rho_{hxz} + \rho_{hxm}\rho_{hym} - \rho_{hyz}\rho_{hxm}\rho_{hzm}\right)}{\rho_{hyx}}\right)\right\}.$$

The properties of this variance is shown better, and without loss of generality, for the SRS design:

$$V\left(\hat{R}_{reg}\right) \cong R^2\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\frac{S_y^2\left(1-\rho_{yz}^2\right)}{\overline{Y}^2}+\frac{S_x^2\left(1-\rho_{xm}^2\right)}{\overline{X}^2}-\right.$$

$$\left.-\frac{2S_yS_x\rho_{yx}}{\overline{Y}\overline{X}}\left(1-\frac{\left(\rho_{yz}\rho_{xz}+\rho_{xm}\rho_{ym}-\rho_{yz}\rho_{xm}\rho_{zm}\right)}{\rho_{yx}}\right)\right\}=\left[if\,\frac{S_y}{\overline{Y}}=\frac{S_x}{\overline{X}}\right]=$$

$$= R^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_y^2}{\overline{Y}^2}\left[2\left(1-\rho_{yx}\right)-\rho_{yz}^2-\rho_{xm}^2+2\rho_{yz}\rho_{xz}+2\rho_{xm}\rho_{ym}-2\rho_{yz}\rho_{xm}\rho_{zm}\right]=$$

$$= R^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_y^2}{\overline{Y}^2}\left[2\left(1-\rho_{yx}\right)-\rho_{yz}^2-\rho_{xm}^2+2\rho_{yz}\rho_{xm}\left\{\frac{\rho_{ym}}{\rho_{yz}}+\frac{\rho_{xz}}{\rho_{xm}}-\rho_{zm}\right\}\right]\ .$$

The ordinary estimator of change, $\hat{R}=\hat{Y}/\hat{X}$, has in the above fashion the variance

$$V\left(\hat{R}\right)\cong R^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_y^2}{\overline{Y}^2}2\left[1-\rho_{yx}\right]\ .$$

Then, what happens with the efficiency of the regression estimator of change compared with the ordinary estimator of change ?

$$Eff\left(\hat{R}_{reg}\right)=\frac{V\left(\hat{R}_{reg}\right)}{V\left(\hat{R}\right)}=1-\frac{\left[\rho_{yz}^2+\rho_{xm}^2-2\rho_{yz}\rho_{xm}\left\{\frac{\rho_{ym}}{\rho_{yz}}+\frac{\rho_{xz}}{\rho_{xm}}-\rho_{zm}\right\}\right]}{2\left(1-\rho_{yx}\right)}\ .$$

y and x are variables of study for time periods 1 and 0, respectively.
z and m are auxiliary variables for time periods 1 and 0, respectively.

Another thing is that the coefficients of correlation can´t vary freely since they are pairwise dependent.

4.1.1 CORRELATION MATRIX

The definition of a covariance matrix gives a clue how the correlation combinations can vary when the coefficients of correlation are pairwise dependent.

Feller II (see reference 4), pages 82-83: " *the covariance matrix of any nondegenerate probability distribution is positive definit*" .

Marsden & Tromba (see reference 10), pages 211-212: *"n x n symmetric matrix B. Consider the n square submatrices along the diagonal.*

$$B=\begin{pmatrix} x_{11} & \cdot & \cdot & x_{1n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & x_{nn} \end{pmatrix}$$

Then B is positive definite (that is, the quadratic function associated with B is positive definite) if and only if the determinants of these diagonal submatrices are all greater than zero" .

Then the covariance matrix must be greater than zero,

$\bar{x}^T C\bar{x} > 0 \Rightarrow \bar{x}^T \rho \bar{x} > 0$ and if $\bar{x} = (1,........,1)$ then $Det\rho > 0$.

ρ in our case is

$$\rho=\begin{pmatrix} 1 & \rho_{yx} & \rho_{yz} & \rho_{ym} \\ \rho_{xy} & 1 & \rho_{xz} & \rho_{xm} \\ \rho_{zx} & \rho_{zy} & 1 & \rho_{zm} \\ \rho_{mx} & \rho_{my} & \rho_{mz} & 1 \end{pmatrix}.$$

According to above ($Det\rho > 0$ and all diagonal subdeterminants > 0) you get the following criteria for how the pairwise dependent correlations can vary :

$$
\begin{cases}
\text{(i)} & 1 > 0 \,. \\[2ex]
\text{(ii)} & \rho_{yx}^2 < 1\left(\rho_{yz}^2 < 1; \rho_{ym}^2 < 1; \rho_{xm}^2 < 1; \rho_{xz}^2 < 1; \rho_{zm}^2 < 1\right) . \\[2ex]
\text{(iii)} & \begin{array}{l} \rho_{yx}^2 + \rho_{yz}^2 + \rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz} < 1 \\ \rho_{yx}^2 + \rho_{ym}^2 + \rho_{xm}^2 - 2\rho_{yx}\rho_{ym}\rho_{xm} < 1 \\ \rho_{yz}^2 + \rho_{ym}^2 + \rho_{zm}^2 - 2\rho_{yz}\rho_{ym}\rho_{zm} < 1 \\ \rho_{xz}^2 + \rho_{xm}^2 + \rho_{zm}^2 - 2\rho_{xz}\rho_{xm}\rho_{zm} < 1 \end{array} \\[2ex]
\text{(iv)} & \begin{array}{c} \left(\rho_{yx}^2 + \rho_{yz}^2 + \rho_{ym}^2 + \rho_{xz}^2 + \rho_{xm}^2 + \rho_{zm}^2\right) - \\ -2\left(\rho_{yx}\rho_{yz}\rho_{xz} + \rho_{yx}\rho_{ym}\rho_{xm} + \rho_{yz}\rho_{ym}\rho_{zm} + \rho_{xz}\rho_{xm}\rho_{zm}\right) + \\ +2\left(\rho_{yx}\rho_{ym}\rho_{xz}\rho_{zm} + \rho_{yx}\rho_{yz}\rho_{xm}\rho_{zm} + \rho_{yz}\rho_{ym}\rho_{xz}\rho_{xm}\right) - \\ -\left(\rho_{ym}^2\rho_{xz}^2 + \rho_{yx}^2\rho_{zm}^2 + \rho_{yz}^2\rho_{xm}^2\right) < 1 \ . \end{array}
\end{cases}
$$

The criteria for permitted correlation combinations are as the variance formula for the regression ratio estimator of change quite messy, but never the less useful.

4.1.2 EFFICIENCY OF REGRESSION RATIO ESTIMATOR

The efficiency formula for the regression ratio estimator (with two arbitrary auxiliary variables and SRS design) compared to the ordinary ratio estimator, is from section 4.1 :

$$
Eff\left(\hat{R}_{reg}\right) = \frac{V\left(\hat{R}_{reg}\right)}{V\left(\hat{R}\right)} = 1 - \frac{\left[\rho_{yz}^2 + \rho_{xm}^2 - 2\rho_{yz}\rho_{xm}\left\{\frac{\rho_{ym}}{\rho_{yz}} + \frac{\rho_{xz}}{\rho_{xm}} - \rho_{zm}\right\}\right]}{2\left(1 - \rho_{yx}\right)}
$$

and the correlation criterias above in section 4.1.1 tells us how the pairwise dependent correlation combinations can vary.

Let´s first study a more simple case when the auxiliary variables m and z are the same for both studied time periods, that is m = z .

As previous mentioned the regression estimator for levels is very efficient compared to the ordinary expansion estimator according to (with SRS design) from section 3.1 :

$$Eff\left(\hat{Y}_{reg}\right)=\frac{V\left(\hat{Y}_{reg}\right)}{V\left(\hat{Y}\right)}=1-\rho_{yz}^{2}\ .$$

The purpose here is to see when and if the regression estimator of change is efficient compared to the ordinary estimator of change. As seen from above the picture here is unclear and there is not the uniformly powerful result as for the estimator of levels.

4.1.2.1 ONE AUXILIARY VARIABLE FOR BOTH TIME PERIODS

The more simpler case when the auxiliary information are the same for both studied time periods, z = m, is the usual case for the short-time surveys (successive investigation periods less than a year) at Statistics Sweden.
The efficiency formula is here (from the efficiency formula with two auxiliary variables at beginning of this section) :

$$Eff\left(\hat{R}_{reg}\mathrm{I}\, z=m\right)=\frac{V\left(\hat{R}_{reg}\mathrm{I}\, z=m\right)}{\mathrm{V}\left(\hat{\mathrm{R}}\right)}=1-\frac{\left(\rho_{yz}-\rho_{xz}\right)^{2}}{2\left(1-\rho_{yx}\right)}\ .$$

The correlation matrix is here $\rho=\begin{pmatrix}1 & \rho_{yx} & \rho_{yz}\\ \rho_{xy} & 1 & \rho_{xz}\\ \rho_{zy} & \rho_{zx} & 1\end{pmatrix}$

and the criteria for allowed correlation combinations are according to section 4.1.1 :

(i) $1 > 0$.

(ii) $\rho_{yx}<1\ \ (\rho_{yz}<1;\rho_{xz}<1)$.

(iii) $\rho_{yx}^{2}+\rho_{yz}^{2}+\rho_{xz}^{2}-2\rho_{yx}\rho_{yz}\rho_{xz}<1$.

Here are some examples of the efficiency of the regression estimator of change versus the ordinary estimator of change when the auxiliary information is the same, z = m, for both studied time periods.

A. TABLES 1 - 3:

$$Eff\left(\hat{R}_{reg}\mathrm{I}\, z=m\right)=\frac{V\left(\hat{R}_{reg}\mathrm{I}\, z=\mathrm{m}\right)}{\mathrm{V}\left(\hat{\mathrm{R}}\right)}=1-\frac{\left(\rho_{yx}-\rho_{xz}\right)^{2}}{2\left(1-\rho_{yx}\right)}$$ where y, x and z are

variable of study time period 1, variable of study time period 0 and auxiliary variable (for both time periods), respectively.
Shaded area in the tables are non-allowed $\rho-$matrix according to the correlation criterias $\left(- \;= V\left(\hat{R}_{reg}\mathrm{I}\, z=m\right)<0\right)$.

TABLE A.1 : $\rho_{yx}=0.9$

ρ_{xz} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20	1	0.80	0.55	0.20	--	--	--
0.40	0.80	1	0.95	0.80	0.55	0.20	--
0.50	0.55	0.95	1	0.95	0.80	0.55	0.20
0.60	0.20	0.80	0.95	1	0.95	0.80	0.55
0.70	--	0.55	0.80	0.95	1	0.95	0.80
0.80	--	0.20	0.55	0.80	0.95	1	0.95
0.90	--	--	0.20	0.55	0.80	0.95	1

TABLE A.2 : $\rho_{yx}=0.7$

ρ_{xz} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20	1	0.93	0.85	0.73	0.58	0.40	0.18
0.40	0.93	1	0.98	0.93	0.85	0.73	0.58
0.50	0.85	0.98	1	0.98	0.93	0.85	0.73
0.60	0.73	0.93	0.98	1	0.98	0.93	0.85
0.70	0.58	0.85	0.93	0.98	1	0.98	0.93
0.80	0.40	0.73	0.85	0.93	0.98	1	0.98
0.90	0.18	0.58	0.73	0.85	0.93	0.98	1

TABLE A.3 : $\rho_{yx} = 0.5$

ρ_{xz} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20	1	0.96	0.91	0.84	0.75	0.64	0.51
0.40	0.96	1	0.99	0.96	0.91	0.84	0.75
0.50	0.91	0.99	1	0.99	0.96	0.91	0.84
0.60	0.84	0.96	0.99	1	0.99	0.96	0.91
0.70	0.75	0.91	0.96	0.99	1	0.99	0.96
0.80	0.64	0.84	0.91	0.96	0.99	1	0.99
0.90	0.51	0.75	0.84	0.91	0.96	0.99	1

As can be seen the regression estimator of change is not very efficient in this case compared with the ordinary ratio estimator, although $V\left(\hat{R}_{reg} \mathrm{I}\, z = m\right) \leq V\left(\hat{R}\right)$.

Even if some extreme correlation combinations, for example $\left(\rho_{yx}, \rho_{yz}, \rho_{xz}\right) = (0.7, 0.8, 0.2)$, lead to good efficiency for the regression estimator, these combinations are rare and not very likely to appear in practice. For most short-time surveys, the situation $\rho_{yz} \approx \rho_{xz}$ exists and leads to $Eff\left(\hat{R}_{reg} \mathrm{I}\, z = m\right) \approx 1$. If the ratio estimator for levels are used for both time periods with the same auxiliary information z you get

$$\hat{R}' = \frac{\frac{\hat{Y}}{\hat{Z}}Z}{\frac{\hat{X}}{\hat{Z}}Z} = \frac{\hat{Y}}{\hat{X}} = \hat{R} \text{ and the efficiency } Eff\left(\hat{R}'\right) = \frac{V\left(\hat{R}'\right)}{V\left(\hat{R}\right)} = 1 \text{ ,which means that}$$

the auxiliary information used two times cancels out completely, and this is what happens in lesser scale in the efficiency tables here. By the way, if

$$b_{yz} = \hat{Y}/\hat{Z} \text{ then } \hat{Y}_{reg} = \hat{Y} + \frac{\hat{Y}}{\hat{Z}}\left(Z - \hat{Z}\right) = \frac{\hat{Y}}{\hat{Z}}Z .$$

The conclusions for the ratio regression estimator with the same auxiliary information used at both time periods (z = m), are that it is slightly more efficient for most correlation combinations, very efficient for a few extreme combinations but on the whole the auxiliary information cancels out when used twice. (In practice the efficiency often are approximately one since $\rho_{yz} \approx \rho_{xz}$ are common.)

This leads us to the more general case when the auxiliary information is not the same between the two time periods, $z \neq m$. If the above case, z = m, is the situation similar to short-time surveys, the more general case, $z \neq m$, is more similar to the situation for yearly surveys.

4.1.2.2 TWO AUXILIARY VARIABLES

The efficiency formula to be studied here from the beginning of the section is

$$Eff\left(\hat{R}_{reg}\right)=\frac{V\left(\hat{R}_{reg}\right)}{V\left(\hat{R}\right)}=1-\frac{\left[\rho_{yz}^{2}+\rho_{xm}^{2}-2\rho_{yz}\rho_{xm}\left\{\frac{\rho_{ym}}{\rho_{yz}}+\frac{\rho_{xz}}{\rho_{xm}}-\rho_{zm}\right\}\right]}{2\left(1-\rho_{yx}\right)}$$ where

y is variable of study time period 1, x is variable of study time period 0, z is auxiliary variable time period 1 and m is auxiliary variable time period 0. The correlation criteria for allowed ρ-matrix are according to section 4.1.1 .

As can be seen, the principal difference between the case z = m and $z \neq m$ is the factor F,

$F=\left\{\frac{\rho_{ym}}{\rho_{yz}}+\frac{\rho_{xz}}{\rho_{xm}}-\rho_{zm}\right\}$, which is equal to one when z = m .

If $F \geq 1$ then $Eff\left(\hat{R}_{reg}\right) \geq Eff\left(\hat{R}_{reg} \mathrm{I}\ z=m\right)$, that is $\hat{R}_{reg}$ is less efficient than $\hat{R}_{reg \mathrm{I}\ z=m}$ since the denominator in both efficiency formulas are the same. Similarly if $F \leq 1$ then $Eff\left(\hat{R}_{reg}\right) \leq Eff\left(\hat{R}_{reg} \mathrm{I}\ z=m\right)$, that is $\hat{R}_{reg}$ is more efficient than $\hat{R}_{reg \mathrm{I}\ z=m}$.

Let us study correlation combinations where the factor F is less than one, since the efficiency picture of $Eff\left(\hat{R}_{reg} \mathrm{I}\ z=m\right)$ is known and less efficient in this case. Even if only positive correlations are studied in the tables, there are no restrictions on negative correlations besides the correlation criteria for the permitted correlation matrix. In practice though, positive coefficients of correlation will dominate.

Here are some examples of the $Eff\left(\hat{R}_{reg}\right)$ (with factor F less than one) for different correlation combinations when it is supposed that $\rho_{ym}=\rho_{xz}=\rho_{zm}$. This assumption of equality of the cross correlations is for simplicity of the tables and hopefully not to unrealistic in practice.
But first, let´s study values of the factor F for different correlation combinations.

$$Eff\left(\hat{R}_{reg} \mathrm{I}\, \rho_{ym} = \rho_{xz} = \rho_{zm}\right) = \frac{V\left(\hat{R}_{reg} \mathrm{I}\, \rho_{ym} = \rho_{xz} = \rho_{zm}\right)}{V\left(\hat{R}\right)} = 1 - \frac{\left[\rho_{yz}^2 + \rho_{xm}^2 - 2\rho_{yz}\rho_{xm}F\right]}{2\left(1-\rho_{yx}\right)}$$

where

$$F = \left\{\rho_{ym}\left(\frac{1}{\rho_{yz}} + \frac{1}{\rho_{xm}} - 1\right)\right\}.$$

B. TABLES 1 - 2 :

$$F = \left\{\rho_{ym}\left(\frac{1}{\rho_{yz}} + \frac{1}{\rho_{xm}} - 1\right)\right\}.$$

TABLE B.1 : $\rho_{ym} = \rho_{xz} = \rho_{zm} = 0.4$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90	
0.20	3.60	2.60	2.40	2.27	2.17	2.10	2.04	
0.40	2.60	1.60	1.40	1.27	1.17	1.10	1.04	$\forall F \geq 1$
0.50	2.40	1.40	1.20	1.07	0.97	0.90	0.84	
0.60	2.27	1.27	1.07	0.93	0.84	0.77	0.71	
0.70	2.17	1.17	0.97	0.84	0.74	0.67	0.62	$\forall F < 1$
0.80	2.10	1.10	0.90	0.77	0.67	0.60	0.54	
0.90	2.04	1.04	0.84	0.71	0.62	0.54	0.49	

TABLE B.2 : $\rho_{ym} = \rho_{xz} = \rho_{zm} = 0.2$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90	
0.20	1.80	1.30	1.20	1.14	1.09	1.05	1.02	$\forall F \geq 1$
0.40	1.30	0.80	0.70	0.64	0.59	0.55	0.52	
0.50	1.20	0.70	0.60	0.54	0.49	0.45	0.42	
0.60	1.14	0.64	0.54	0.47	0.42	0.39	0.36	
0.70	1.09	0.59	0.49	0.42	0.37	0.34	0.31	$\forall F < 1$
0.80	1.05	0.55	0.45	0.39	0.34	0.30	0.27	
0.90	1.02	0.52	0.42	0.36	0.31	0.27	0.25	

These tables show us where the factor F is less than one, for which we shall study the efficiency of the regression ratio estimator with two auxiliary variables compared with the ordinary ratio estimator. From above we have,

if $F \le 1$ then $Eff(\hat{R}_{reg}) \le Eff(\hat{R}_{reg} \mathrm{I}\, z = m)$ and if $F \ge 1$ the opposite.

C. TABLES 1 - 6 :

$$Eff(\hat{R}_{reg} \mathrm{I}\, \rho_{ym} = \rho_{xz} = \rho_{zm} \cap F \le 1) = \frac{V(\hat{R}_{reg} \mathrm{I}\, \rho_{ym} = \rho_{xz} = \rho_{zm})}{V(\hat{R}_{reg})} = 1 - \frac{[\rho_{yz}^2 + \rho_{xm}^2 - 2\rho_{yz}\rho_{xm}F]}{2(1-\rho_{yx})}$$

where $F = \left\{ \rho_{ym} \left(\frac{1}{\rho_{yz}} + \frac{1}{\rho_{xm}} - 1 \right) \right\}$. y, x, z, m are the variables of study time period 1 and 0 and the auxiliary variables time period 1 and 0, respectively. Shaded area in the tables are non-allowed ρ - matrix according to section 4.1.1 .
$(-- = V(\hat{R}_{reg} \mathrm{I}\, \rho_{ym} = \rho_{xz} = \rho_{zm}) < 0)$.

TABLE C.1 : $\rho_{yx} = 0.9$, $\rho_{ym} = \rho_{xz} = \rho_{zm} = 0.4$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40							
0.50		$F \ge 1$			0.70	0.15	--
0.60				0.75	0.28	--	--
0.70			0.70	0.28	--	--	--
0.80			0.15	--	--	--	--
0.90			--	--	--	--	--

TABLE C.2 : $\rho_{yx} = 0.9, \rho_{ym} = \rho_{xz} = \rho_{zm} = 0.2$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40		0.68	0.35	--	--	--	--
0.50		0.35	--	--	--	--	--
0.60	$F \ge 1$	--	--	--	--	--	--
0.70		--	--	--	--	--	--
0.80		--	--	--	--	--	--
0.90		--	--	--	--	--	--

TABLE C.3: $\rho_{yx} = 0.7, \rho_{ym} = \rho_{xz} = \rho_{zm} = 0.4$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40							
0.50	$F \geq 1$				0.90	0.72	0.49
0.60				0.92	0.76	0.57	0.33
0.70			0.90	0.76	0.58	0.37	0.14
0.80			0.72	0.57	0.37	0.15	--
0.90			0.49	0.33	0.14	--	--

TABLE C.4 : $\rho_{yx} = 0.7, \rho_{ym} = \rho_{xz} = \rho_{zm} = 0.2$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40		0.89	0.78	0.65	0.47	0.25	0.01
0.50		0.78	0.67	0.52	0.34	0.12	--
0.60	$F \geq 1$	0.65	0.52	0.36	0.17	--	--
0.70		0.47	0.34	0.17	--	--	--
0.80		0.25	0.12	--	--	--	--
0.90		0.01	--	--	--	--	--

TABLE C.5 : $\rho_{yx} = 0.5, \rho_{ym} = \rho_{xz} = \rho_{zm} = 0.4$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40							
0.50	$F \geq 1$				0.94	0.83	0.70
0.60				0.95	0.86	0.74	0.60
0.70			0.94	0.86	0.75	0.62	0.48
0.80			0.83	0.74	0.62	0.49	0.33
0.90			0.70	0.60	0.48	0.33	0.17

TABLE C.6 : $\rho_{yx} = 0.5, \rho_{ym} = \rho_{xz} = \rho_{zm} = 0.2$

ρ_{xm} \ ρ_{yz}	0.20	0.40	0.50	0.60	0.70	0.80	0.90
0.20							
0.40		0.94	0.87	0.79	0.68	0.55	0.40
0.50		0.87	0.80	0.71	0.60	0.47	0.32
0.60	$F \geq 1$	0.79	0.71	0.62	0.50	0.37	0.22
0.70		0.68	0.60	0.50	0.38	0.25	0.09
0.80		0.55	0.47	0.37	0.25	0.10	--
0.90		0.40	0.32	0.22	0.09	--	--

Conclusions on the efficiency of the regression estimator of change with different auxiliary information for two consecutive time periods compared with the ordinary ratio estimator will be given below, and also supplementary conclusions on the regression estimator with the same auxiliary information for the two time periods.

EFFICIENCY FORMULAS ACCORDING TO PREVIOUS:

General case: $Eff\left(\hat{R}_{reg}\right)=\dfrac{V\left(\hat{R}_{reg}\right)}{V\left(\hat{R}\right)}=1-\dfrac{\left[\rho_{yz}^{2}+\rho_{xm}^{2}-2\rho_{yz}\rho_{xm}F\right]}{2\left(1-\rho_{yx}\right)}$

where $F=\left\{\dfrac{\rho_{ym}}{\rho_{yz}}+\dfrac{\rho_{xz}}{\rho_{xm}}-\rho_{zm}\right\}$.

Studied cases:

$$Eff\left(\hat{R}_{reg}\mathrm{I}\ \rho_{ym}=\rho_{xz}=\rho_{zm}\cap F\le 1\right)=\frac{V\left(\hat{R}_{reg}\mathrm{I}\ \rho_{ym}=\rho_{xz}=\rho_{zm}\right)}{V\left(\hat{R}\right)}=1-\frac{\left[\rho_{yz}^{2}+\rho_{xm}^{2}-2\rho_{yz}\rho_{xm}F\right]}{2\left(1-\rho_{yx}\right)}\le 1$$

where $F=\left\{\rho_{ym}\left(\dfrac{1}{\rho_{yz}}+\dfrac{1}{\rho_{xm}}-\rho_{zm}\right)\right\}$ and

$$Eff\left(\hat{R}_{reg}\mathrm{I}\ z=m\right)=\frac{V\left(\hat{R}_{reg}\mathrm{I}\ z=m\right)}{V\left(\hat{R}\right)}=1-\frac{\left(\rho_{yz}-\rho_{xz}\right)^{2}}{2\left(1-\rho_{yx}\right)}\le 1\ .$$

A. If $\mathrm{F}\le 1$ then $Eff\left(\hat{R}_{reg}\right)\le Eff\left(\hat{R}_{reg}\mathrm{I}\ z=m\right)\le 1$.

If $F=1$ then $Eff\left(\hat{R}_{reg}\right)=Eff\left(\hat{R}_{reg}\mathrm{I}\ z=m\right)\le 1$.

If $F\ge 1$ then $Eff\left(\hat{R}_{reg}\right)\ge Eff\left(\hat{R}_{reg}\mathrm{I}\ z=m\right)\le 1$.

It is provided in the inequalities and equality above that in $Eff\left(\hat{R}_{reg}\mathrm{I}\ \mathrm{z=m}\right)$ $\rho_{xz}=\rho_{xm}$. The $\rho-matrix$ is more bounded in the case z = m than in $z\neq m$.

That is, for the case $F\ge 1$ $\hat{R}_{reg}$ is less or equal efficient than $\hat{R}_{reg\mathrm{I}\,z=m}$, which means that the conclusions for $\hat{R}_{reg\mathrm{I}\,z=m}$ are equal or worse for the general case $\hat{R}_{reg}$. The main conclusion for the z = m case is that $\hat{R}_{reg\mathrm{I}\,z=m}$ is slightly more efficient than $\hat{R}$, but on the whole that the same auxiliary information used twice cancels out and $Eff\left(\hat{R}_{reg}\mathrm{I}\ \mathrm{z=m}\right)\approx 1$.

The interesting area where it is possible to find great efficiency for $\hat{R}_{reg}$ is then when $F\le 1$.

B. If $F \leq 1$ then $Eff\left(\hat{R}_{reg}\right) \leq 1$. The conclusions for this case are :

-- The higher ρ_{yx} the smaller is the allowed $\rho-matrix$ area. For very high correlation between the variables of study, x and y, for example 0.9 or more the allowed $\rho-$matrix is almost nought. Here there is no use to deal with the regression estimator.

-- If ρ_{yx} are more moderate, say 0.5 - 0.7, the picture looks better for efficiency gains for the regression estimator.
For low cross correlations (between time periods; $\rho_{ym}, \rho_{xz}, \rho_{zm}$) , high auxiliary variable correlations (ρ_{yz}, ρ_{xm}) and moderate or low variables of study correlation (ρ_{yx}) the higher is the efficiency gain for the regression estimator compared to the ordinary ratio estimator.

-- The situation studied here are similar to yearly surveys, consecutive years or longer intervals. Efficiency gains for the regression estimator ought to be possible between two consecutive years, say $\rho_{yx} \approx 0.7$ or lower,
ρ_{yz} and $\rho_{xm} \approx 0.6$ or higher and the cross correlations
ρ_{ym}, ρ_{xz} and ρ_{zm} $\approx$ 0.4 or lower.
The best situation for efficiency gains for the regression esrtimator ought to be for intermittent surveys, for example a yearly survey done each third year. Then ρ_{yx} will be lower, say 0.5 with high auxiliary correlations 0.6 or more and low cross correlations, say 0.2 .

-- The main conclusion is that, if $F \leq 1$, ρ_{yx} is moderate, the auxiliary correlations are high and the cross correlations between the time periods are moderate or low; there are great efficiency gains for the regression estimator. The conclusion for $F \geq 1$ is slightly better or worse efficiency compared to the ordinary estimator of change.

If $F \leq 1$ *there might be efficiency gains, sometimes great, for the regression estimator.*

If $F \geq 1$ *there is no use for the regression estimator.*

4.1.3 DOUBLE RATIO ESTIMATOR

An even more complex case is for the double ratio estimator, which includes four regression estimators of levels and with four arbitrary auxiliary variables. The variance of this regression estimator explodes with correlation terms and regression estimator should be compared to the ordinary double ratio estimator. The efficiency formula will be given here, but this case will not be further studied here although some conclusions maybe can be drawn from the former studied cases.

1. DOUBLE REGRESSION RATIO ESTIMATOR :

$$\hat{Q}_{reg} = \frac{\hat{Y}_{reg}/\hat{Z}_{reg}}{\hat{X}_{reg}/\hat{M}_{reg}} = \frac{\left[\hat{Y} + b_{yz_1}\left(Z_1 - \hat{Z}_1\right)\right] / \left[\hat{Z} + b_{zz_2}\left(Z_2 - \hat{Z}_2\right)\right]}{\left[\hat{X} + b_{xm_1}\left(M_1 - \hat{M}_1\right)\right] / \left[\hat{M} + b_{mm_2}\left(M_2 - \hat{M}_2\right)\right]}$$

with parameter $Q = \dfrac{Y/Z}{X/M}$ and variance

$$V\left(\hat{Q}_{reg}\right) \cong Q^2\left[CV^2\left(\hat{Y}_{reg}/\hat{Z}_{reg}\right) + CV^2\left(\hat{X}_{reg}/\hat{M}_{reg}\right) - 2CV^2\left(\hat{Y}_{reg}/\hat{Z}_{reg}, \hat{X}_{reg}/\hat{M}_{reg}\right)\right] =$$

$$= Q^2\left[\left\{\frac{V\left(\hat{Y}_{reg}\right)}{Y^2} + \frac{V\left(\hat{Z}_{reg}\right)}{Z^2} - \frac{2Cov\left(\hat{Y}_{reg}, \hat{Z}_{reg}\right)}{YZ}\right\} + \right.$$

$$+\left\{\frac{V\left(\hat{X}_{reg}\right)}{X^2} + \frac{V\left(\hat{M}_{reg}\right)}{M^2} - \frac{2Cov\left(\hat{X}_{reg}, \hat{M}_{reg}\right)}{XM}\right\} -$$

$$\left. -2\left\{\frac{Cov\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)}{YX} + \frac{Cov\left(\hat{Z}_{reg}, \hat{M}_{reg}\right)}{ZM} - \frac{Cov\left(\hat{Y}_{reg}, \hat{M}_{reg}\right)}{YM} - \frac{Cov\left(\hat{Z}_{reg}, \hat{X}_{reg}\right)}{ZX}\right\}\right] =$$

$$= \left[\text{Model assumption: All population coefficients of variation equal } ; \frac{S}{\overline{\overline{\Theta}}} = \frac{S_y}{\overline{\overline{Y}}}\right] =$$

$$= Q^2\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_y^2}{\overline{\overline{Y}}^2}\left[\left\{2\left(1 - \rho_{yz}\right) - \rho_{yz_1}^2 - \rho_{zz_2}^2 + 2\rho_{yz_1}\rho_{zz_2}\left[\frac{\rho_{yz_2}}{\rho_{yz_1}} + \frac{\rho_{zz_1}}{\rho_{zz_2}} - \rho_{z_1z_2}\right]\right\} + \right.$$

$$+\left\{2\left(1 - \rho_{xm}\right) - \rho_{xm_1}^2 - \rho_{mm_2}^2 + 2\rho_{xm_1}\rho_{mm_2}\left[\frac{\rho_{xm_2}}{\rho_{xm_1}} + \frac{\rho_{mm_1}}{\rho_{xm_2}} - \rho_{m_1}\rho_{m_2}\right]\right\} -$$

$$-2\left\{\left[\rho_{yx} - \rho_{yz_1}\rho_{xm_1}\left\{\frac{\rho_{ym_1}}{\rho_{yz_1}} + \frac{\rho_{xz_1}}{\rho_{xm_1}} - \rho_{z_1}\rho_{m_1}\right\}\right] + \left[\rho_{zm} - \rho_{zz_2}\rho_{mm_2}\left\{\frac{\rho_{zm_2}}{\rho_{zz_2}} + \frac{\rho_{mz_2}}{\rho_{mm_2}} - \rho_{z_2}\rho_{m_2}\right\}\right] - \right.$$

$$\left.\left. -\left[\rho_{ym} - \rho_{yz_1}\rho_{mm_2}\left\{\frac{\rho_{ym_2}}{\rho_{yz_1}} + \frac{\rho_{mz_1}}{\rho_{mm_2}} - \rho_{z_1}\rho_{m_2}\right\}\right] - \left[\rho_{zx} - \rho_{zz_2}\rho_{xm_1}\left\{\frac{\rho_{zm_1}}{\rho_{zz_2}} + \frac{\rho_{xz_2}}{\rho_{xm_1}} - \rho_{z_2m_1}\right\}\right]\right\}\right] \bullet$$

2. DOUBLE ORDINARY RATIO ESTIMATOR :

$\hat{Q} = \dfrac{\hat{Y}/\hat{Z}}{\hat{X}/\hat{M}}$ with the parameter $Q = \dfrac{Y/Z}{X/M}$ and variance

$$V(\hat{Q}) \cong Q^2\left[CV^2(\hat{Y}/\hat{Z}) + CV^2(\hat{X}/\hat{M}) - 2CV^2(\hat{Y}/\hat{Z},\hat{X}/\hat{M})\right] =$$

$$= Q^2\left[\left\{\frac{V(\hat{Y})}{Y^2} + \frac{V(\hat{Z})}{Z^2} - \frac{2Cov(\hat{Y},\hat{Z})}{YZ}\right\} + \left\{\frac{V(\hat{X})}{X^2} + \frac{V(\hat{M})}{M^2} - \frac{2Cov(\hat{X},\hat{M})}{XM}\right\} -\right.$$

$$\left.-2\left\{\frac{Cov(\hat{Y},\hat{X})}{YX} + \frac{Cov(\hat{Z},\hat{M})}{ZM} - \frac{Cov(\hat{Y},\hat{M})}{YM} - \frac{Cov(\hat{Z},\hat{X})}{ZX}\right\}\right] =$$

$$= \left[\text{Model assumption: All population coefficients of variation equal } ; \frac{S}{\overline{\Theta}} = \frac{S_y}{\overline{Y}}\right] =$$

$$= Q^2\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_y^2}{\overline{Y}^2}\left[2(1-\rho_{yz}) + 2(1-\rho_{xm}) - 2(\rho_{yx} + \rho_{zm} - \rho_{ym} - \rho_{zx})\right] =$$

$$= Q^2\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_y^2}{\overline{Y}^2}\left[2\left\{(1-\rho_{yz}) + (1-\rho_{xm}) - (\rho_{yx} + \rho_{zm} - \rho_{ym} - \rho_{zx})\right\}\right].$$

3. THE EFFICIENCY FORMULA is then

$$Eff\left(\hat{Q}_{reg}\right)=\frac{V\left(\hat{Q}_{reg}\right)}{V\left(\hat{Q}\right)}=1-\frac{1}{2\left[\left(1-\rho_{yz}\right)+\left(1-\rho_{xm}\right)-\left(\rho_{yx}+\rho_{zm}-\rho_{ym}-\rho_{zx}\right)\right]}x$$

$$x\left\{\left[\rho^2_{yz_1}+\rho^2_{zz_2}-2\rho_{yz_1}\rho_{zz_2}\left\{\frac{\rho_{yz_2}}{\rho_{yz_1}}+\frac{\rho_{zz_1}}{\rho_{zz_2}}-\rho_{z_1z_2}\right\}\right]+\right.$$

$$+\left[\rho^2_{xm_1}+\rho^2_{mm_2}-2\rho_{xm_1}\rho_{mm_2}\left\{\frac{\rho_{xm_2}}{\rho_{xm_1}}+\frac{\rho_{mm_1}}{\rho_{mm_2}}-\rho_{m_1m_2}\right\}\right]-$$

$$-\left[2\rho_{yz_1}\rho_{xm_1}\left\{\frac{\rho_{ym_1}}{\rho_{yz_1}}+\frac{\rho_{xz_1}}{\rho_{xm_1}}-\rho_{z_1m_1}\right\}\right]-\left[2\rho_{zz_2}\rho_{mm_2}\left\{\frac{\rho_{zm_2}}{\rho_{zz_2}}+\frac{\rho_{mz_2}}{\rho_{mm_2}}-\rho_{z_2m_2}\right\}\right]+$$

$$\left.+\left[2\rho_{yz_1}\rho_{mm_2}\left\{\frac{\rho_{ym_2}}{\rho_{yz_1}}+\frac{\rho_{mz_1}}{\rho_{mm_2}}-\rho_{z_1m_2}\right\}\right]+\left[2\rho_{zz_2}\rho_{xm_1}\left\{\frac{\rho_{zm_1}}{\rho_{zz_2}}+\frac{\rho_{xz_2}}{\rho_{xm_1}}-\rho_{z_2m_1}\right\}\right]\right\}.$$

The correlation matrix is here

$$\rho=\begin{pmatrix} 1 & \rho_{yz} & \rho_{yx} & \rho_{ym} & \rho_{yz_1} & \rho_{yz_2} & \rho_{ym_1} & \rho_{ym_2} \\ \rho_{zy} & 1 & \rho_{zx} & \rho_{zm} & \rho_{zz_1} & \rho_{zz_2} & \rho_{zm_1} & \rho_{zm_2} \\ \rho_{xy} & \rho_{xz} & 1 & \rho_{xm} & \rho_{xz_1} & \rho_{xz_2} & \rho_{xm_1} & \rho_{xm_2} \\ \rho_{my} & \rho_{mz} & \rho_{mx} & 1 & \rho_{mz_1} & \rho_{mz_2} & \rho_{mm_1} & \rho_{mm_2} \\ \rho_{z_1y} & \rho_{z_1z} & \rho_{z_1x} & \rho_{z_1m} & 1 & \rho_{z_1z_2} & \rho_{z_1m_1} & \rho_{z_1m_2} \\ \rho_{z_2y} & \rho_{z_2z} & \rho_{z_2x} & \rho_{z_2m} & \rho_{z_2z_1} & 1 & \rho_{z_2m_1} & \rho_{z_2m_2} \\ \rho_{m_1y} & \rho_{m_1z} & \rho_{m_1x} & \rho_{m_1m} & \rho_{m_1z_1} & \rho_{m_1z_2} & 1 & \rho_{m_1m_2} \\ \rho_{m_2y} & \rho_{m_2z} & \rho_{m_2x} & \rho_{m_2m} & \rho_{m_2z_1} & \rho_{m_2z_2} & \rho_{m_2m_1} & 1 \end{pmatrix}$$

and the allowed $\rho-matrix$ criteria for the pairwise dependent correlations can be calculated according to section 4.1.1, but we will stop the theory chapter here and continue with the practice.

4.2 PRACTICE

Some general estimation problems was described under section 3.2, Practice/Levels. The following points where treated:

-- Non fit-design.

-- Unstable estimates of coefficients of regression.

-- Overcount and undercount objects with respect to the auxiliary information.

-- Non-response.

These practical problems with respect to regression estimators for levels was considered more problematical than for the ordinary blown-up estimator, since the regression estimator with variance is a large sample formula and hence much more complex.

If these practical problems are severe for the regression estimator of levels, they are even more problematical for regression estimators of change, which are even more complex large sample formulas.
The conclusion that the ordinary estimator of levels is more robust with regard to these practical problems, will hold, but not to the same degree for the ordinary estimator of changes; since it also is a large sample formula.

Below the following points will be discussed:

4.2.1 Problems of estimation for complex estimators.

4.2.2 Considerations on estimates of levels and changes for the same design.

4.2.3 How much efficiency gains shall it be to to deal with much more complicated formulas?

4.2.1 PROBLEMS OF ESTIMATION FOR COMPLEX ESTIMATORS

A general problem in the estimation procedure of complex estimators and their variances is that after practical modifications the estimator and the variance get different properties, for example a non-response treatment model, and it is easy to make mistakes with straightforward calculations without much afterthought.

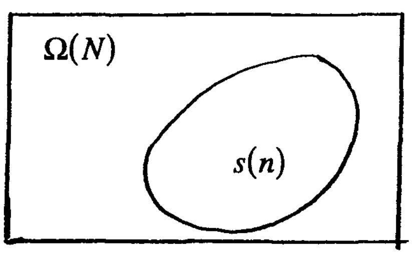
One possible mistake is that the same parameters in different places in the variance are estimated for different subsamples which leads to the variance estimator not being coherent with the variance, which in turn can lead to that the variance estimate is less than zero in extreme cases.
Another possible mistake is that one looks at the original plan of design and estimators and ignores the practical modifications made, for example a non-response treatment model, which leads to non expectation value right variance estimates with respect to the modeling dealing with practical problems.

There are solutions to these problems. Below will be given an example of the estimation problems of an complex estimator and their solutions.

EXAMPLE OF PROBLEMS OF ESTIMATION FOR COMPLEX ESTIMATORS :

We illustrate the problem mentioned above for the regression ratio estimator of change for the case with the same auxiliary information (z = m) at the two consecutive time periods.This is a special case of the formulas given in chapter 4.1 when $z \neq m$. We consider as before a complete panel design for a non-dynamic population and the design STSRS, which we illustrate for the SRS design which is enough to show the points.

Planned design (for consecutive time periods 1 and 2) :



x ; variable of study, time period 1.
y ; variable of study, time period 2.
z ; auxiliary variable for both periods.

A. ESTIMATES FOR THE PLANNED DESIGN WITHOUT MODELING FOR PRACTICAL PROBLEMS

Regression estimator of level at time period 1;

$$\hat{X}_{reg} = \hat{X} + b_{xz}\left(Z - \hat{Z}\right) = \frac{N}{n}\sum_{i \in s_n}^{n} x_i + b_{xz}\left(\sum_{i \in \Omega_N}^{N} z_i - \frac{N}{n}\sum_{i \in s_n}^{n} z_i\right)$$ and for time period 2;

$$\hat{Y}_{reg} = \hat{Y} + b_{yz}\left(Z - \hat{Z}\right) = \frac{N}{n}\sum_{i \in s_n}^{n} y_i + b_{yz}\left(\sum_{i \in \Omega_N}^{N} z_i - \frac{N}{n}\sum_{i \in s_n}^{n} z_i\right)$$ where the estimators of the coefficients of regression are

$$b_{xz} = \frac{\sum_{i \in s_n}^{n} (x_i - \bar{x})(z_i - \bar{z})}{\sum_{i \in s_n}^{n} (z_i - \bar{z})^2} \text{ and } b_{yz} = \frac{\sum_{i \in s_n}^{n} (y_i - \bar{y})(z_i - \bar{z})}{\sum_{i \in s_n}^{n} (z_i - \bar{z})^2} .$$

The variances for the estimators above are

$$V(\hat{X}_{reg}) \cong N^2\left(\frac{1}{n} - \frac{1}{N}\right) S_x^2 \left[1 - \rho_{xz}^2\right] ; \; S_x^2 = \frac{\sum_{i \in \Omega_N}^{N} (x_i - \overline{X})^2}{N - 1} \text{ and}$$

$$V(\hat{Y}_{reg}) \cong N^2\left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 \left[1 - \rho_{yz}^2\right] ; \; S_y^2 = \frac{\sum_{i \in \Omega_N}^{N} (y_i - \overline{Y})^2}{N - 1} \text{ and their estimators are}$$

$$v(\hat{X}_{reg}) \cong N^2\left(\frac{1}{n} - \frac{1}{N}\right) s_x^2 \left[1 - \hat{\rho}_{xz}^2\right] ; \; s_x^2 = \frac{\sum_{i \in s_n}^{n} (x_i - \bar{x})^2}{n - 1} \text{ and}$$

$$v(\hat{Y}_{reg}) \cong N^2\left(\frac{1}{n} - \frac{1}{N}\right) s_y^2 \left[1 - \hat{\rho}_{yz}^2\right] ; \; s_y^2 = \frac{\sum_{i \in s_n}^{n} (y_i - \bar{y})^2}{n - 1} .$$

The estimators of the coefficients of correlation are

$$\hat{\rho}_{xz} = \frac{\sum_{i \in s_n}^{n} (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_{i \in s_n}^{n} (z_i - \bar{z})^2}} \text{ and } \hat{\rho}_{yz} = \frac{\sum_{i \in s_n}^{n} (y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\sum_{i \in s_n}^{n} (z_i - \bar{z})^2}} .$$

The regression estimator of change is then $\hat{R}_{reg} = \hat{Y}_{reg} / \hat{X}_{reg}$ with the variance

$$V(\hat{R}_{reg}) \cong \frac{1}{X^2}\left[V(\hat{Y}_{reg}) + R^2 V(\hat{X}_{reg}) - 2R\,Cov(\hat{Y}_{reg}, \hat{X}_{reg})\right] \text{ which is estimated by}$$

$$v(\hat{R}_{reg}) \cong \frac{1}{\hat{X}^2}\left[v(\hat{Y}_{reg}) + \hat{R}_{reg}^2 v(\hat{X}_{reg}) - 2\hat{R}_{reg} \operatorname{cov}(\hat{Y}_{reg}, \hat{X}_{reg})\right] .$$

The covariance is equal to $Cov(\hat{Y}_{reg}, \hat{X}_{reg}) \cong N^2\left(\frac{1}{n} - \frac{1}{N}\right) S_y S_x \rho_{yx}\left[1 - \frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]$.

Proof: $Cov\left(\hat{Y}_{reg},\hat{X}_{reg}\,\mathrm{I}\,\tilde{b}=\hat{\tilde{\beta}}\right)=Cov\left(\hat{Y},\hat{X}\right)-b_{xz}Cov\left(\hat{Y},\hat{Z}\right)-b_{yz}Cov\left(\hat{X},\hat{Z}\right)+b_{xz}b_{yz}\underbrace{Cov\left(\hat{Z},\hat{Z}\right)}_{V(\hat{Z})}=$

$$=N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left[S_{yx}-b_{xz}S_{yz}-b_{yz}S_{xz}+b_{xz}b_{yz}S_z^2\right]=\left[\text{if }\tilde{b}=\tilde{\beta}\right]=Cov\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\text{ above.}$$

The estimator of this covariance then follows by

$$\mathrm{cov}\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\cong N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left[s_{yx}-b_{xz}s_{yz}-b_{yz}s_{xz}+b_{xz}b_{yz}s_z^2\right]=$$

$$=N^2\left(\frac{1}{n}-\frac{1}{N}\right)s_y s_x\hat{\rho}_{yx}\left[1-\frac{\hat{\rho}_{yz}\hat{\rho}_{xz}}{\hat{\rho}_{yx}}\right]\text{ and}$$

$\hat{\rho}_{yx}$ is based on the n sample objects as $\hat{\rho}_{xz}$ and $\hat{\rho}_{yz}$.

This is the estimation procedure without practical problems.
So far no problems!

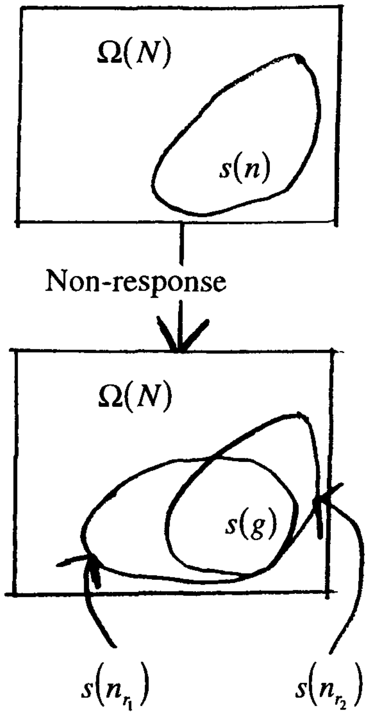
What happens then with the estimation procedure after practical problems?

We just consider non-response problems and assume the non-response treatment straight adjustment ("mean value imputation") with accordingly practical modeling.

First, the design after response objects has changed from the original complete panel design to an overlapping design and second, how does the estimation procedure look like with this practical modeling? This will be described below.

Planned design:

Design after non-response:



for time periods 1 and 2.
x; var. of study, per. 1.
y ; var. of study, per. 2.
z ; aux. var., both periods.

$s(n_{r_1}) = n_{r_1}$ response objects in the sample at time period t.
$s(g) = s(n_{r_1}) \cap s(n_{r_2})$.

B. ESTIMATES FOR THE PLANNED DESIGN WITH MODELING FOR PRACTICAL PROBLEMS

Regression estimator of level at time period 1;

$$\hat{X}_{reg} = \frac{N}{n_{r_1}} \sum_{i \in s(n_{r_1})}^{n_{r_1}} x_i + b_{xz} \left(\sum_{i \in \Omega_N}^{N} z_i - \frac{N}{n_r} \sum_{i \in s(n_{r_1})}^{n_{r_1}} z_i \right)$$ and for time period 2;

$$\hat{Y}_{reg} = \frac{N}{n_{r_2}} \sum_{i \in s(n_{r_2})}^{n_{r_2}} y_i + b_{yz} \left(\sum_{i \in \Omega_N}^{N} z_i - \frac{N}{n_{r_2}} \sum_{i \in s(n_{r_2})}^{n_{r_2}} z_i \right)$$ where the estimators of the

coefficients of regression are

$$b_{xz} = \frac{\sum_{i \in s(n_{r_1})}^{n_{r_1}} (x_i - \bar{x})(z_i - \bar{z})}{\sum_{i \in s(n_{r_1})}^{n_{r_1}} (z_i - \bar{z})^2} \text{ and } b_{yz} = \frac{\sum_{i \in s(n_{r_2})}^{n_{r_2}} (y_i - \bar{y})(z_i - \bar{z})}{\sum_{i \in s(n_{r_2})}^{n_{r_2}} (z_i - \bar{z})^2} .$$

The variances for the estimators above are

$$V\left(\hat{X}_{reg}\right) \cong N^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)S_x^2\left[1-\rho_{xz}^2\right] \; ; \; S_x^2 = \frac{\sum_{i\in\Omega_N}^{N}\left(x_i-\overline{X}\right)^2}{N-1} \quad \text{and}$$

$$V\left(\hat{Y}_{reg}\right) \cong N^2\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)S_y^2\left[1-\rho_{yz}^2\right] \; ; \; S_y^2 = \frac{\sum_{i\in\Omega_N}^{N}\left(y_i-\overline{Y}\right)^2}{N-1} \quad \text{and their estimators are}$$

$$v\left(\hat{X}_{reg}\right) \cong N^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)s_x^2\left[1-\hat{\rho}_{xz}^2\right] \; ; \; s_x^2 = \frac{\sum_{i\in s(n_{r_1})}^{n_r}\left(x_i-\overline{x}\right)^2}{n_{r_1}-1} \quad \text{and}$$

$$v\left(\hat{Y}_{reg}\right) \cong N^2\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)s_y^2\left[1-\hat{\rho}_{yz}^2\right] \; ; \; s_y^2 = \frac{\sum_{i\in s(n_{r_2})}^{n_{r_2}}\left(y_i-\overline{y}\right)^2}{n_{r_2}-1} \; .$$

The estimators of the coefficients of correlation are

$$\hat{\rho}_{xz} = \frac{\sum_{i\in s(n_{r_1})}^{n_{r_1}}\left(x_i-\overline{x}\right)\left(z_i-\overline{z}\right)}{\sqrt{\sum_{i\in s(n_{r_1})}^{n_{r_1}}\left(x_i-\overline{x}\right)^2\sum_{i\in s(n_{r_1})}^{n_{r_1}}\left(z_i-\overline{z}\right)^2}} \quad \text{and} \quad \hat{\rho}_{yz} = \frac{\sum_{i\in s(n_{r_2})}^{n_{r_2}}\left(y_i-\overline{y}\right)\left(z_i-\overline{z}\right)}{\sqrt{\sum_{i\in s(n_{r_2})}^{n_{r_2}}\left(y_i-\overline{y}\right)^2\sum_{i\in s(n_{r_2})}^{n_{r_2}}\left(z_i-\overline{z}\right)^2}} \; .$$

The regression estimator of change is then $\hat{R}_{reg} = \hat{Y}_{reg}/\hat{X}_{reg}$ with the variance

$$V\left(\hat{R}_{reg}\right) \cong \frac{1}{X^2}\left[V\left(\hat{Y}_{reg}\right)+R^2V\left(\hat{X}_{reg}\right)-2RCov\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\right] \quad \text{which is estimated by}$$

$$v\left(\hat{R}_{reg}\right) \cong \frac{1}{\hat{X}^2}\left[v\left(\hat{Y}_{reg}\right)+\hat{R}_{reg}^2 v\left(\hat{X}_{reg}\right)-2\hat{R}_{reg}\operatorname{cov}\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\right] \; .$$

The covariance is equal to

$$Cov\left(\hat{Y}_{reg},\hat{X}_{reg}\right) \cong N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)S_yS_x\rho_{yx}\left[1-\frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right] \; .$$

Proof: $Cov\left(\hat{Y}_{reg},\hat{X}_{reg} \mathrm{I}\, \tilde{b}=\hat{\tilde{\beta}}\right)=Cov\left(\hat{Y},\hat{X}\right)-b_{xz}Cov\left(\hat{Y},\hat{Z}\right)-b_{yz}Cov\left(\hat{X},\hat{Z}\right)+b_{xz}b_{yz}\underbrace{Cov\left(\hat{Z},\hat{Z}\right)}_{v(\hat{z})}=$

$=N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)\left[S_{yx}-b_{xz}S_{yz}-b_{yz}S_{xz}+b_{xz}b_{yz}S_z^2\right]=\left[if\ \tilde{b}=\tilde{\beta}\right]=Cov\left(\hat{Y}_{reg},\hat{X}_{reg}\right)$ above.

See *APPENDIX ON PROOFS* that $Cov\left(\hat{Y},\hat{X}\right)=N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)\underbrace{S_yS_x\rho_{yx}}_{Cov(y_i,x_i)}$.

The estimator of this covariance then follows by

$$\operatorname{cov}\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\cong N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)\left[s_{yx}-b_{xz}s_{yz}-b_{yz}s_{xz}+b_{xz}b_{yz}s_z^2\right]=$$

$=N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)s_ys_x\hat{\rho}_{yx}\left[1-\frac{\hat{\rho}_{yz}\hat{\rho}_{xz}}{\hat{\rho}_{yx}}\right]$ where ρ_{yz} and ρ_{xz} are estimated as previous and

$$\rho_{yx}=\frac{\sum_{i\in s(g)}^{g}(y_i-\bar{y})(x_i-\bar{x})}{\sqrt{\sum_{i\in s(g)}^{g}(y_i-\bar{y})^2\sum_{i\in s(g)}^{g}(x_i-\bar{x})^2}}\ .$$

This is the estimation procedure with practical problems.

Below this estimation procedure will be discussed and possible mistakes that can be made are pointed out.

Mistakes that can be done with straight forward estimation, particulary the estimation of the variances, for complex estimators are listed below. The mistakes mentioned has been done by the author in his survey practice.

-- Ignorance of the overlapping panel design after non-response.

-- Reduction of the variance formulas not finished.

-- The same parameters (for example population-variances and coefficients of correlation) in different places in the estimates of the variance formulas can be estimated by different subsamples and ruin the consistency with the variances; for example negative variance estimates.

SOLUTION: *Look at the conditioned variance after non-response (or other non-measurement error models) and try to get the best variance estimator with the consistency of the variance properties retained.*

With variance consistency we mean $v\left(\hat{\Theta} \text{ I allowed } \hat{\rho}\text{-matrix}\right) \geq 0$
and $E\left(v\left(\hat{\Theta}\right)\right) \cong V\left(\hat{\Theta}\right)$.
We look at the example again.

A. IGNORANCE OF THE OVERLAPPING PANEL DESIGN AFTER NON-RESPONSE

$$Cov\left(\hat{Y}_{reg}, \hat{X}_{reg} \text{ I original complete paneldesign}\right) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S_y S_x \rho_{yx}\left[1 - \frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]$$

and

$$Cov\left(\hat{Y}_{reg}, \hat{X}_{reg} \text{ I panel design after non - response}\right) = N^2\left(\frac{g}{n_{r_1} n_{r_2}} - \frac{1}{N}\right)S_y S_x \rho_{yx}\left[1 - \frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right].$$

Without afterthought the term 1/n may be replaced by 1/g or something else, that either makes the covariance after non-response too large or too small. The solution, above, comes from the conditioned theoretical solution of the covariance after the non-response-treatment-model and the overlapping panel design.

B. REDUCTION OF THE FORMULAS NOT FINISHED

A calculation of the covariance $Cov\left(\hat{Y}_{reg}, \hat{X}_{reg} \text{ I } nonresponse\right)$ leads to

$$Cov\left(\hat{Y}_{reg}, \hat{X}_{reg} \text{ I nonresponse}\right) = N^2\left(\frac{g}{n_{r_1}n_{r_2}} - \frac{1}{N}\right)\left\{S_{yx} - b_{xz}S_{yz} - b_{yz}S_{xz} + b_{xz}b_{yz}S_z^2\right\} \cong$$

$$\cong N^2\left(\frac{g}{n_{r_1}n_{r_2}} - \frac{1}{N}\right)\left\{S_yS_x\rho_{yx}\left[1 - \frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]\right\} \text{ and is estimated by}$$

$$\text{cov}\left(\hat{Y}_{reg}, \hat{X}_{reg} \text{ I } nonresponse\right) \cong N^2\left(\frac{g}{n_{r_1}n_{r_2}} - \frac{1}{N}\right)\left\{s_{yx} - b_{xz}s_{yz} - b_{yz}s_{xz} + b_{xz}b_{yz}s_z^2\right\} =$$

$$= N^2\left(\frac{g}{n_{r_1}n_{r_2}} - \frac{1}{N}\right)\left\{s_ys_x\hat{\rho}_{yx}\left[1 - \frac{\hat{\rho}_{yz}\hat{\rho}_{xz}}{\hat{\rho}_{yx}}\right]\right\} .$$

Say, that we have not finished the reduction of the formula above and now try to estimate it straightforward. Then

$s_{yx} = s_ys_x\hat{\rho}_{yx}$ is based on the sample $s(g)$, and hence here s_y^2, s_x^2 and $\hat{\rho}_{yx}$.
$s_{yz} = s_ys_z\hat{\rho}_{yz}$ is based on the sample $s\left(n_{r_2}\right)$, and hence here s_y^2, s_z^2 and $\hat{\rho}_{yz}$.
$s_{xz} = s_xs_z\hat{\rho}_{xz}$ is based on the sample $s\left(n_{r_1}\right)$, and hence here s_x^2, s_z^2 and $\hat{\rho}_{xz}$.
$b_{yz} = \hat{\rho}_{yz}\, s_y/s_z$ is based on the sample $s(n_{r_2})$, and hence here s_y^2, s_z^2 and $\hat{\rho}_{yz}$.
$b_{xz} = \hat{\rho}_{xz}\, s_x/s_z$ is based on the sample $s(n_{r_1})$, and hence here s_x^2, s_z^2 and $\hat{\rho}_{xz}$.
Should s_z^2 be based on sample $s(n_{r_1})$, $s(n_{r_2})$ or $s(n)$ (the auxiliary variable z is available for all objects in the frame) ?

This gives us two estimators s_y^2 based on samples $s(g)$ and $s(n_{r_2})$. The same is valid for s_x^2 with samples $s(g)$ and $s(n_{r_1})$. We also have two estimators s_z^2 based on samples $s(n_{r_1})$ and $s(n_{r_2})$; and possibly a third based on $s(n)$. For each of the coefficients of correlation there is only one choise;
$\hat{\rho}_{yx}$ is based on sample $s(g)$, $\hat{\rho}_{yz}$ on $s(n_{r_2})$ and $\hat{\rho}_{xz}$ on $s(n_{r_1})$.

This will clearly not be a consistent estimator of the covariance above, where all population-parameters are based on the whole population $\Omega(N)$. If we look at the reduced formulas the estimation problem will be easier. The best estimate for each population parameter will be;
s_y^2 and $\hat{\rho}_{yz}$ based on the sample $s(n_{r_2})$, s_x^2 and $\hat{\rho}_{xz}$ based on $s(n_{r_1})$ and

$\hat{\rho}_{yx}$ based on s(g) (the problem of estimating S_z^2 separatly has disappeared with the formula reduction). *This will probably be the best solution, but we will penetrate similar estimation problems further below.*

C. THE SAME PARAMETERS *(for example population-variances and coefficients of correlation)* **IN DIFFERENT PLACES IN A VARIANCE FORMULA CAN BE ESTIMATED BY DIFFERENT SUBSAMPLES AND RUIN THE CONSISTENCY BETWEEN THE VARIANCE AND THE VARIANCE ESTIMATOR;**
for example negative variance estimates

The variance after nonresponse is according to previous

$$V\left(\hat{R}_{reg} \text{ I nonresponse}\right) \cong \frac{1}{X^2}\left[V\left(\hat{Y}_{reg}\right)+R^2V\left(\hat{X}_{reg}\right)-2RCov\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\right]=$$

$$=\frac{1}{X^2}\left[N^2\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)S_y^2\left[1-\rho_{yz}^2\right]+R^2N^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)S_x^2\left[1-\rho_{xz}^2\right]-\right.$$

$$\left.-2RN^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)\underbrace{S_yS_x\rho_{yx}}_{Cov(y,x)}\left[1-\frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]\right] \text{ and is estimated by}$$

$$v\left(\hat{R}_{reg} \text{ I nonresponse}\right) \cong \frac{1}{\hat{X}_{reg}^2}\left[v\left(\hat{Y}_{reg}\right)+\hat{R}_{reg}^2v\left(\hat{X}_{reg}\right)-2\hat{R}_{reg}\operatorname{cov}\left(\hat{Y}_{reg},\hat{X}_{reg}\right)\right]=$$

$$=\frac{1}{\hat{X}_{reg}^2}\left[N^2\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)s_y^2\left[1-\hat{\rho}_{yz}^2\right]+\hat{R}_{reg}^2N^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)s_x^2\left[1-\hat{\rho}_{xz}^2\right]-\right.$$

$$\left.-2\hat{R}_{reg}N^2\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)\underbrace{s_ys_x\hat{\rho}_{yx}}_{\operatorname{cov}(y,x)}\left[1-\frac{\hat{\rho}_{yz}\hat{\rho}_{xz}}{\hat{\rho}_{yx}}\right]\right].$$

If we for every population parameter takes the best estimator we have

$s_y^2\left[s\left(n_{r_2}\right)\right], s_x^2\left[s\left(n_{r_1}\right)\right], \hat{\rho}_{yx}[s(g)], \hat{\rho}_{yz}\left[s\left(n_{r_2}\right)\right], \hat{\rho}_{xz}\left[s\left(n_{r_1}\right)\right]$ and $b_{xz}\left[s\left(n_{r_1}\right)\right]$ in $\hat{X}_{reg}$.

1. To maintain an allowed estimated correlation structure, since the correlations are pairwise dependent (see section 4.1.1, Correlation matrix), the estimates of the correlations must be based on the same subsample, in our case s(g), which leads to lesser than the best estimates for ρ_{xz} and ρ_{yz} . The correlation matrix in the population is always allowed an based on $\Omega(N)$ and the same is valid for the estimated correlation matrix based on s(n), original design without non-response.

What shall we then chose; an guaranteed allowed estimated correlation matrix based on s(g) or the best possible estimates for each correlation but with a small risk of getting a non-allowed estimated correlation matrix?

Previous in section 3.2.2 we stated that *"The classical problem of regression estimates are the stableness of the estimates of coefficients of regression"* and this is the same for the coefficients of regression. *This makes it easy for the author to chose the second alternative, the best possible estimate for each correlation but with a small risk of getting a non-allowed estimated correlation matrix!*
With a fit design (see section 3.2.1 Non-fit design) and not too large non-response the problem ought to be small, but the risk still exists for negative variance-estimates due to a non-allowed estimated correlation matrix.
A practical point of view is also that the estimators
$v\left(\hat{X}_{reg}\right)$ and $v\left(\hat{Y}_{reg}\right)$ enters $v\left(\hat{R}_{reg}\right)$ without corrections.

2. All population variances must in the variance-formula be estimated by the same subsample, respectivly, to guarantee consistency with the variance $V\left(\hat{R}_{reg}\right)$. The best choise for
s_y^2 is the sample $\mathrm{s}\left(\mathrm{n}_{\mathrm{r}_2}\right)$ and for s_x^2 the sample $\mathrm{s}\left(\mathrm{n}_{\mathrm{r}_1}\right)$.

3. If the population covariance $S_{yx} = Cov(y,x)$ is estimated as $s_{yx} = \mathrm{cov}(y,x)$ by s(g) we get $s_{yx} = \hat{\rho}_{yx}[s(g)]s_y[s(g)]s_x[s(g)]$ and hereby breaks the rule that each population parameter always shall be based on the same sample, and the best possible.

The solution $s_{yx} = \hat{\rho}_{yx}[s(g)]s_y\left[s\left(n_{r_2}\right)\right]s_x\left[s\left(n_{r_1}\right)\right]$ is possible since ρ_{yx} is a non-dimentional measure and independent of the rest.

4. *A further question is what happens with the efficiency formulas after non-response?*

We start with the efficiency formula for the original design (and without the assumption on equal population coefficient of variation for the variables of study, x and y) :

$$V(\hat{R}_{reg}\text{ I original design}) \cong \frac{1}{X^2}N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(S_y^2\left[1-\rho_{yz}^2\right]+R^2S_x^2\left[1-\rho_x^2\right]-2RS_yS_x\rho_{yx}\left[1-\frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]\right)=$$

$$=\frac{1}{X^2}N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left\{\left[S_y^2+R^2S_x^2-2RS_yS_x\rho_{yx}\right]-\left[S_y^2\rho_{yz}^2+R^2S_x^2-2RS_yS_x\rho_{yz}\rho_{xz}\right]\right\}=$$

$$=V(\hat{R})-\frac{1}{X^2}N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(S_y\rho_{yz}-S_x\rho_{xz}\right)^2 \text{ and since } V(\hat{R}_{reg})\geq 0 \text{ and } V(\hat{R})\geq 0$$

it follows that $V(\hat{R}_{reg})\leq V(\hat{R})$, which means that $Eff(\hat{R}_{reg}\text{ I original design})=\dfrac{V(\hat{R}_{reg})}{V(\hat{R})}\leq 1$.

The same calculations after non-response leads to

$$V(\hat{R}_{reg}\text{ I nonresponse})\cong\frac{N^2}{X^2}\left\{\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)S_y^2\left[1-\rho_{yz}^2\right]+R^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)S_x^2\left[1-\rho_{xz}^2\right]-\right.$$

$$\left.-2R\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)S_yS_x\rho_{yx}\left[1-\frac{\rho_{yz}\rho_{xz}}{\rho_{yx}}\right]\right\}=$$

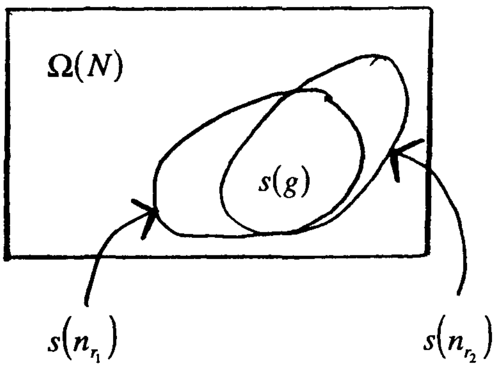
$$=\frac{N^2}{X^2}\left\{\left[\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)S_y^2+R^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)S_x^2-2R\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)S_yS_x\rho_{yx}\right]-\right.$$

$$\left.-\left[\left(\frac{1}{n_{r_2}}-\frac{1}{N}\right)S_y^2\rho_{yz}^2+R^2\left(\frac{1}{n_{r_1}}-\frac{1}{N}\right)S_x^2\rho_{xz}^2-2R\left(\frac{g}{n_{r_1}n_{r_2}}-\frac{1}{N}\right)S_yS_x\rho_{yz}\rho_{xz}\right]\right\}=$$

$$=V(\hat{R})-\frac{N^2}{X^2}\left\{\underbrace{\left[\frac{S_y^2\rho_{yz}^2}{n_{r_2}}+\frac{R^2S_x^2\rho_{xz}^2}{n_{r_1}}-\frac{2RgS_yS_x\rho_{yz}\rho_{xz}}{n_{r_1}n_{r_2}}\right]}_{A}-\underbrace{\left(\frac{S_y\rho_{yz}}{\sqrt{N}}-\frac{RS_x\rho_{xz}}{\sqrt{N}}\right)^2}_{B}\right\}.$$

Let $C=\left(\frac{S_y\rho_{yz}}{\sqrt{n_{r_2}}}-\frac{RS_x\rho_{xz}}{\sqrt{n_{r_1}}}\right)^2=\left[\frac{S_y^2\rho_{yz}^2}{n_{r_2}}+\frac{R^2S_x^2\rho_{xz}^2}{n_{r_1}}-\frac{2R\sqrt{n_{r_1}}\sqrt{n_{r_2}}S_yS_x\rho_{yz}\rho_{xz}}{n_{r_1}n_{r_2}}\right]$.

$\sqrt{n_{r_1}}\sqrt{n_{r_2}}\geq g$ since $\sqrt{n_{r_1}}\geq\sqrt{g}$ and $\sqrt{n_{r_2}}\geq\sqrt{g}$ (Thank you! Monica Rennermalm) .



It is then clear that $C\geq B$ and that $C\leq A$, which means that $A\geq B$.
Since $A\geq B$, $V(\hat{R}_{reg})\geq 0$ and $V(\hat{R})\geq 0$ it is clear that $V(\hat{R}_{reg})\leq V(\hat{R})$;

which means that $Eff(\hat{R}_{reg}\text{ I nonresponse})=\frac{V(\hat{R}_{reg})}{V(\hat{R})}\leq 1$.

For the design and estimators in our example the regression estimator is always equal or more efficient than the ordinary estimator, both for the original design and after non-response.

The estimated efficiency has not strictly this properties since by stochastic reasons $\hat{X}_{reg}\neq\hat{X}$ and $\hat{R}_{reg}\neq\hat{R}$, that is
$\hat{Eff}(\hat{R}_{reg}\text{ I original design})$ and $\hat{Eff}(\hat{R}_{reg}\text{ I nonresponse})$ can be larger than one.

5. The empirical results in chapter 5 have the same regression estimator of change as in the example and corresponding ordinary estimator for the design STSRS and we shall here see estimated efficiences and ratios $\hat{X}_{reg}/\hat{X}$ and $\hat{R}_{reg}/\hat{R}$. Also corresponding results for estimators of levels will be presented here.

4.2.2 CONSIDERATIONS ON ESTIMATES OF LEVELS AND OF CHANGES FOR THE SAME DESIGN

For the sake of simplicity, we assume here as before a non-dynamic population.

For estimates of change it is best with a complete panel design on assumption that all correlations between variables of study are positive, which they almost always are in practice.

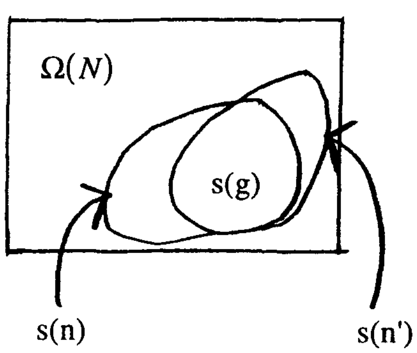
For estimates of levels that do not use auxiliary information from different time periods, besides information in the current frame, it is irrelevant of the degree of overlap in the panel design or independent samples.

If estimates of levels are added to give an estimate of a total for several time periods, it is best with independent samples between the different time periods, since
$\hat{X}+\hat{Y}$ has the variance $V(\hat{X}+\hat{Y}) = V(\hat{X}) + \hat{V}(\hat{Y}) + 2Cov(\hat{X},\hat{Y})$ which is clearly greater than $V(\hat{X}) + V(\hat{Y})$, which is the variance for the case with independent samples between time periods.

For estimators of changes, $\hat{R} = \hat{Y}/\hat{X}$, the variance
$V(\hat{R}) \cong \frac{1}{X^2}\left[V(\hat{Y}) + R^2V(\hat{X}) - 2RCov(\hat{Y},\hat{X})\right]$ is smallest when $Cov(\hat{Y},\hat{X})$ is the greatest possible, which it is with a complete panel design :
$Cov(\hat{Y},\hat{X}\text{ I overlapping panel design and n = n'}) = N^2\left(\frac{g}{n^2} - \frac{1}{N}\right)\underbrace{S_yS_x\rho_{yx}}_{Cov(y,x)}$ and

$\max[g]\ Cov(\hat{Y},\hat{X}) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S_yS_x\rho_{yx}$ when g = n .

See picture below.



That is,

-- *For estimates of change,* $\hat{R} = \hat{Y}/\hat{X}$, *a complete panel design is best.*

-- *For estimates of levels,* $\hat{X}$ and $\hat{Y}$, *it is independent of the degree of overlap in panel designs or independent designs between time periods as far as no auxiliary information are used between time periods, besides current frame information.*

-- *For estimates* $\hat{Y} + \hat{X}$ *it is best with independent sample designs between time periods.*

So, a complete panel design will be best in most cases.

The next question is, how shall the sample be allocated with a panel STSRS design for different estimates of levels and changes?

The optimal sample allocation results for the ordinary estimator of levels with the design STSRS is (see Cochran, reference 3) according to Neyman 1934 (given here without the costfunction) :

$$\hat{X} = \sum_h \hat{X}_h = \sum_h \frac{N_h}{n_h} \sum_{i \in s(n_h)}^{n_h} x_{hi} \text{ with the variance } V(\hat{X}) = \sum_h V(\hat{X}_h) = \sum_h N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_{hx}^2$$

$$\text{where } S_{hx}^2 = \frac{\sum_{i \in \Omega(N_h)}^{N_h} \left(x_{hi} - \overline{X}_h \right)^2}{N_h - 1} \text{ and } \overline{X}_h = \frac{\sum_{i \in \Omega(N_h)}^{N_h} x_{hi}}{N_h} .$$

Planned precision for a 95 % confidence interval : $p = \dfrac{2\sqrt{V(\hat{X})}}{X} \Rightarrow V(\hat{X}) = \left(\dfrac{pX}{2} \right)^2$.

Minimize $n = \sum_h n_h$ for given variance $V(\hat{X}) = \left(\dfrac{pX}{2} \right) = V$:

$$n_{opt} = \frac{\left(\sum_h N_h S_{hx} \right)^2}{V + \sum_h N_h S_{hx}^2} \text{ and } n_h = \frac{n_{opt} N_h S_{hx}}{\sum_h N_h S_{hx}} .$$

These results can be generalized to all estimators $\hat{\Theta}$ that have a variance of the form

$$V(\hat{\Theta}) = \sum_h N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) f_h\left[V(\hat{\Theta}) \right] .$$

Since the only variables to be optimized are the n_h:s and all other terms are constants the proof follows from the $\hat{X}$-case to all $\hat{\Theta}$-cases above by substitution of

$f_h\left[V\left(\hat{\Theta}\right)\right]$ instead of S_{hx}^2, $V\left(\hat{\Theta}\right)$ instead of $V\left(\hat{X}\right)$ and Θ instead of X.

This means that all estimators in this paper can be given an optimal sample allocation since we always have a complete panel design (STSRS) in a non-dynamic population.

Take for example;

A. THE ORDINARY LEVEL ESTIMATOR

$\hat{X}=\sum_h \hat{X}_h$ with the variance $V\left(\hat{X}\right)=\sum_h N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)f_h\left[V\left(\hat{X}\right)\right]$

where $f_h\left[V\left(\hat{X}\right)\right]=S_{hx}^2$.

B. THE ORDINARY COMBINED RATIO ESTIMATOR

$\hat{R}=\frac{\hat{Y}}{\hat{X}}=\frac{\sum_h \hat{Y}_h}{\sum_h \hat{X}_h}$ with the variance $V\left(\hat{R}\right)\cong\sum_h N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)f_h\left[V\left(\hat{R}\right)\right]$

where $f_h\left[V\left(\hat{R}\right)\right]=\frac{1}{X^2}\left[S_{hy}^2+R^2S_{hx}^2-2RS_{hy}S_{hx}\rho_{hyx}\right]$.

C. THE SEPARATE REGRESSION ESTIMATOR

$\hat{X}_{reg}=\sum_h\left[\hat{Y}_h+b_{hxz}\left(Z_h-\hat{Z}_h\right)\right]$ with the variance $V\left(\hat{X}_{reg}\right)\cong\sum_h N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)f_h\left[V\left(\hat{X}_{reg}\right)\right]$

where $f_h\left[V\left(\hat{X}_{reg}\right)\right]=S_{hx}^2\left[1-\rho_{hxz}^2\right]$.

D. THE COMBINED RATIO REGRESSION ESTIMATOR

(with different auxiliary variables m and z for the two time periods)

$$\hat{R}_{reg} = \frac{\hat{Y}_{reg}}{\hat{X}_{reg}} = \frac{\sum_h \left[\hat{Y}_h + b_{hyz}\left(Z_h - \hat{Z}_h\right)\right]}{\sum_h \left[\hat{X}_h + b_{hxz}\left(M_h - \hat{M}_h\right)\right]}$$ with the variance

$$V\left(\hat{R}_{reg}\right) \cong \sum_h \left(\frac{1}{n_h} - \frac{1}{N_h}\right) f_h\left[V\left(\hat{R}_{reg}\right)\right] \text{ where } f_h\left[V\left(\hat{R}_{reg}\right)\right] = \frac{1}{X^2}\left[S_{hy}^2\left(1-\rho_{hyz}^2\right) + R^2 S_{hxz}^2\left(1-\rho_{hxz}^2\right) - \right.$$

$$\left. -2RS_{hy}S_{hx}\rho_{hyx}\left\{1 - \frac{\left[\rho_{hyz}\rho_{hxz} + \rho_{hxm}\rho_{hym} - \rho_{hyz}\rho_{hxm}\rho_{hzm}\right]}{\rho_{hyx}}\right\}\right].$$

Then, how shall the sample be allocated in a multipurpose survey?

The usual way in enterprise statistics is to use an auxiliary variable, number of employees per enterprise or establishment, and to use this for an Neyman sample allocation of levels (the usual design is STSRS and the estimators are the ordinary for levels and changes) and to use this for all estimates in mostly a multipurpose survey. Another good auxiliary variable is available, total sales, but is curiously enough not used either for stratification or sample allocation, although in the survey in chapter 5, with the empirical results, is now used a combined ratio estimator for levels based on this auxiliary variable.

The auxiliary information are usually one to two years old with reference to the time period under study.

Sample allocation with old sample data is presently not used, but is an alternative for allocation for more complex estimators and also with regards to multipurpose aspects.

Auxiliary information in general can be used in several ways; optimal stratification, optimal sample allocation, PPS-sampling, the information used in the estimators (as for example regression estimators); but there is always a limit when the auxiliary information is used up and no more efficiency is possible to gain.
For example, with the same auxiliary information it is possible to use it for optimal stratification and optimal sample allocation. If then the same auxiliary information is used in the estimators, i.e. regression estimators, there is not much gain for variance-reduction since most of the power in the auxiliary information is already used up.

With more than one auxiliary variable much more efficiency (variance-reduction) in the estimates is possible to gain.

In a multipurpose survey an efficient deign must always be a compromise between estimates of levels and changes, between different level estimates and between different estimates of changes and also other aspects of multipurpose (for example domains, timeseries-considerations and need of analysis).

PRIORITY-QUESTIONS ABOUT AN EFFICIENT DESIGN FOR MULTIPURPOSE-USE :

-- *Priority between different survey-estimates; levels/changes and so forth?*

-- *Priority between survey-analysis and survey-estimates?*

-- *Shall sample allocation be with or without a cost-function?*

-- *Shall sample allocation be with considerations of non-response?*

All depends on the most important goal for the survey and how much money is available!

With this we leave these questions to the competent readers!

4.2.3 HOW MUCH EFFICIENCY GAINS SHALL IT BE IN ORDER TO DEAL WITH MUCH MORE COMPLICATED FORMULAS?

a. Simpler formulas are often pedagogically clearer to show and explain to users of statistics!

b. More complicated (complex estimators) and more efficient estimators shall be significantly more efficient (than simpler estimators) since they mostly are pedagogical unclearer (What happens in the survey-process?), they take more data efforts and are more sensitive to nonsampling-errors (mean-square-error considerations).

Let us see if the practical example with the empirical results in next chapter shows significantly efficiency gain for the complicated formulas (regression estimates) versus the simpler ordinary estimates.

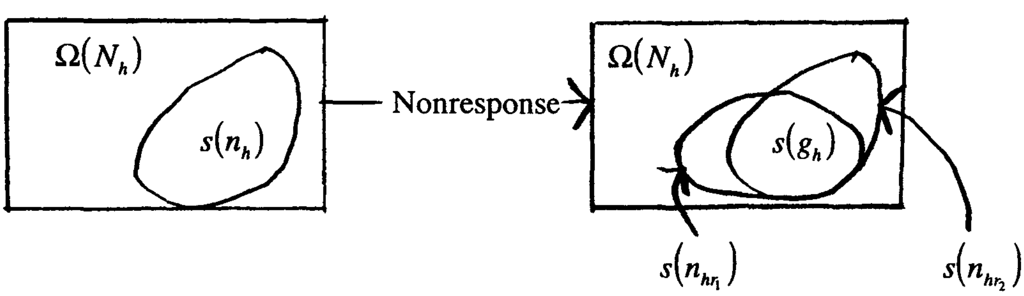
5. PRACTICAL EXAMPLES ON REGRESSION AND ORDINARY ESTIMATORS FROM A SURVEY AT STATISTICS SWEDEN

The practical examples with empirical results of regression and ordinary estimates are from a retail survey at Statistics Sweden, who at the time used the ordinary estimators for levels and changes. At present the combined ratio estimator is used for levels, and hereby the ordinary estimator of change (see Bergdahl, reference 1).

This retail survey is done monthly and quarterly, with according time-periods of study with STSRS and a complete panel design for an approximatly non-dynamic population per calendar year. Stratification variables are number of employees and business groups.The variable of study is turnover in domestic trade. The auxiliary variable used in the regression estimates is also turnover, but one year old information available from a turnover taxation register for almost the whole frame. Enterprise-objects who miss auxiliary information have zero as auxiliary variable value.

The periods of study for the regression versus the ordinary estimators was the first and the second quarter of 1989.
The frame-population consisted of 85.000 enterprises with a sample of about 6500 enterprises. The auxiliary information was the same for both periods of study. The non-response rate was alarmingly high, about 35 % .
This study with regression versus the ordinary estimators is described more in detail in Carlsson & Garås, reference 2.

The design (STSRS and a complete panel) and the estimators are shown below.



x = variable of study, turnover in domestic trade, quarter 1 1989 (period 1).
y = variable of study, turnover in domestic trade, quarter 2 1989 (period 2).
z = auxiliary variable, turnover in domestic trade, both quarters.
r = response objects including known overcount.

A. ORDINARY ESTIMATORS

Estimated total, time period 1 :

$$\hat{X} = \sum_h \frac{N_h}{n_{hr_1}} \sum_{i \in s(n_{hr_1})}^{n_{hr_1}} x_{hi} \text{ with variance - estimator } v(\hat{X}) = \sum_h N_h^2 \left(\frac{1}{n_{hr_1}} - \frac{1}{N_h} \right) s_{hx}^2$$

$$\text{where } s_{hx}^2 = \frac{\sum_{i \in n_{hr_1}}^{n_{hr_1}} (x_{hi} - \bar{x}_h)^2}{n_{hr_1} - 1} \text{ ; } \bar{x}_h = \frac{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} x_{hi}}{n_{hr_1}} \text{ and similarily for time period 2 .}$$

$$\hat{Y} = \sum_h \frac{N_h}{n_{hr_2}} \sum_{i \in s(n_{hr_2})}^{n_{hr_2}} y_{hi} \text{ with variance - estimator } v(\hat{Y}) = \sum_h N_h^2 \left(\frac{1}{n_{hr_2}} - \frac{1}{N_h} \right) s_{hy}^2$$

$$\text{where } s_{hy}^2 = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} (y_{hi} - \bar{y}_h)^2}{n_{hr_2} - 1} \text{ ; } \bar{y}_h = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} y_{hi}}{n_{hr_2}} \text{ .}$$

The estimator of change is then

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} \text{ with variance - estimator } v(\hat{R}) \cong \frac{1}{\hat{X}^2} \left[v(\hat{Y}) + \hat{R}^2 v(\hat{X}) - 2\hat{R} \operatorname{cov}(\hat{Y}, \hat{X}) \right]$$

$$\text{where } \operatorname{cov}(\hat{Y}, \hat{X}) = \sum_h N_h^2 \left(\frac{g_h}{n_{hr_1} n_{hr_2}} - \frac{1}{N_h} \right) \hat{\rho}_{hxy} s_{hy} s_{hx} \text{ and}$$

$$\hat{\rho}_{hxy} = \frac{\sum_{i \in s(g_h)}^{g_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{\sqrt{\sum_{i \in s(g_h)}^{g_h} (x_{hi} - \bar{x}_h)^2 \sum_{i \in s(g_h)}^{g_h} (y_{hi} - \bar{y}_h)^2}} \text{ where } \bar{x}_h = \frac{\sum_{i \in s(g_h)}^{g_h} x_{hi}}{g_h} \text{ and } \bar{y}_h = \frac{\sum_{i \in s(g_h)}^{g_h} y_{hi}}{g_h} \text{ .}$$

All other sub-estimators as above and the implicite non-response treatement model for all estimates and sub-estimates here is mean-value imputation or corresponding.

B. REGRESSION ESTIMATORS

Estimated total, time-period 1 :

$$\hat{X}_{reg} = \sum_h \left[\hat{X}_h + b_{hxz}\left(Z - \hat{Z}_h\right)\right] \text{ with variance - estimator } v\left(\hat{X}_{reg}\right) \cong \sum_h v\left(\hat{X}_h\right)\left[1 - \hat{\rho}^2_{hxz}\right]$$

where $\hat{X}_h$ and $v\left(\hat{X}_h\right)$ according to $\hat{X} = \sum_h \hat{X}_h$ and $v\left(\hat{X}\right) = \sum_h v\left(\hat{X}_h\right)$;

$$Z_h = \sum_{i \in \Omega(N_h)}^{N_h} z_{hi} \text{ , } \hat{Z}_h = \frac{N_h}{n_{hr_1}} \sum_{i \in s(n_{hr_1})}^{n_{hr_1}} z_{hi} \text{ , } b_{hxz} = \frac{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} \left(x_{hi} - \bar{x}_h\right)\left(z_{hi} - \bar{z}_h\right)}{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} \left(z_{hi} - \bar{z}_h\right)^2} \text{ and}$$

$$\hat{\rho}_{hxz} = \frac{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} \left(x_{hi} - \bar{x}_h\right)\left(z_{hi} - \bar{z}_h\right)}{\sqrt{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} \left(x_{hi} - \bar{x}\right)^2 \sum_{i \in s(n_{hr_1})}^{n_{hr_1}} \left(z_{hi} - \bar{z}_h\right)^2}} \text{ where } x_h = \frac{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} x_{hi}}{n_{hr_1}} \text{ and } \bar{z}_h = \frac{\sum_{i \in s(n_{hr_1})}^{n_{hr_1}} z_{hi}}{n_{hr_1}} \text{ .}$$

Estimated total, time-period 2 :

$$\hat{Y}_{reg} = \sum_h \left[\hat{Y}_h + b_{hyz}\left(Z_h - \hat{Z}_h\right)\right] \text{ with variance - estimator } v\left(\hat{Y}_{reg}\right) \cong v\left(\hat{Y}_h\right)\left[1 - \hat{\rho}^2_{hyz}\right]$$

where $\hat{Y}_h$ and $v\left(\hat{Y}_h\right)$ according to $\hat{Y} = \sum_h \hat{Y}_h$ and $v\left(\hat{Y}\right) = \sum_h v\left(\hat{Y}_h\right)$;

$$Z_h = \sum_{i \in \Omega(N_h)}^{N_h} z_{hi} \text{ , } \hat{Z}_h = \frac{N_h}{n_{hr_2}} \sum_{i \in s(n_{hr_2})}^{n_{hr_2}} z_{hi} \text{ , } b_{hyz} = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} \left(y_{hi} - \bar{y}_h\right)\left(z_{hi} - \bar{z}_h\right)}{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} \left(z_{hi} - \bar{z}_h\right)^2} \text{ and}$$

$$\hat{\rho}_{hyz} = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2n}} \left(y_{hi} - \bar{y}_h\right)\left(z_{hi} - \bar{z}_h\right)}{\sqrt{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} \left(y_{hi} - \bar{y}_h\right)^2 \sum_{i \in s(n_{hr_2})}^{n_{hr_2}} \left(z_{hi} - \bar{z}_h\right)^2}} \text{ where } \bar{y}_h = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} y_{hi}}{n_{hr_2}} \text{ and } \bar{z}_h = \frac{\sum_{i \in s(n_{hr_2})}^{n_{hr_2}} z_{hi}}{n_{hr_2}} \text{ .}$$

The estimator of change is then $\hat{R}_{reg} = \hat{Y}_{reg} / \hat{X}_{reg}$ with variance estimator

$$v\left(\hat{R}_{reg}\right) \cong \frac{1}{\hat{X}_{reg}}\left[v\left(\hat{Y}_{reg}\right) + \hat{R}_{reg}^2 v\left(\hat{X}_{reg}\right) - 2\hat{R}_{reg}\operatorname{cov}\left(\hat{Y}_{reg}, \hat{X}_{reg}\right)\right] \text{ where}$$

$$\operatorname{cov}\left(\hat{Y}_{reg}, \hat{X}_{reg}\right) \cong \sum_h N_h^2\left(\frac{g_h}{n_{hr_1} n_{hr_2}} - \frac{1}{N_h}\right) s_{hy} s_{hx} \hat{\rho}_{hxy}\left[1 - \frac{\hat{\rho}_{hxz}\hat{\rho}_{hyz}}{\hat{\rho}_{hxy}}\right] \text{ with } \hat{\rho}_{hxy}$$

according to ordinary estimators and all other sub-estimators according to above.

The implicit non-response treatment model here for all estimates and sub-estimates is mean value imputation or corresponding.

Since the estimators of correlation are not based on the same sample, but the best for each correlation estimator, the possibility exists of a non-allowed estimates correlation matrix, which can lead to negative variance estimates.

The formulas here are motivated in section 4.2.1, Problems of estimation for complex estimators.

Now to the results!

In Table 5.2: Domains of Study some extreme results are marked with a frame.

In Table 5.3: Sample Strata all non-allowed correlation combinations are marked with a frame according to criterias in section 4.1.2.1 on page 16.

Also extreme reults here are marked with a frame.

TABLE 5.1 : GRAND TOTAL

$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
0.91	0.58	0.97	0.51	1.07	0.72
0.80	0.58	0.86	0.51	1.07	0.83

(In line one is reg.est. = ord.est if $n_h = 1$ and in line two not adjusted.)

TABLE 5.2:1 : DOMAINS OF STUDY

Numb.	*Str./domain*	*Domain*	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
01	1	1000	--	--	--	--	--	--
02	6	1902	0.93	0.71	0.98	0.58	1.05	1.01
03	4	1903	0.88	0.84	0.82	0.86	0.93	1.07
04	5	1904	0.91	0.74	$\boxed{1.96}$	0.62	$\boxed{2.15}$	1.15
05	6	1905	0.68	0.98	1.07	0.65	$\boxed{1.58}$	$\boxed{2.32}$
06	5	1906	0.66	0.57	0.72	0.80	1.09	1.15
07	6	1907	1.07	0.63	1.07	0.60	1.00	0.75
08	5	1908	0.78	0.43	0.50	0.45	0.64	1.12
09	4	1909	0.98	0.94	0.99	0.93	1.01	0.99
10	6	1910	0.99	0.75	1.00	0.75	1.01	0.92
11	6	1911	0.93	0.92	0.95	0.89	1.02	0.96
12	5	1912	0.99	0.81	1.00	0.73	1.01	1.01
13	5	1913	0.86	0.52	0.87	0.52	1.01	0.80
14	6	1914	0.95	0.50	0.96	0.59	1.01	0.94
15	6	1915	0.82	0.06	0.88	0.07	1.07	0.12
16	6	1916	0.79	0.07	0.75	0.25	0.95	1.15
17	1	1917	--	--	--	--	--	
18	2	2001	0.99	0.76	0.98	0.53	0.99	0.35
19	6	2002	1.04	0.67	1.06	0.66	1.02	0.81
20	3	2005	0.88	0.64	0.95	0.68	1.08	0.93
21	4	2009	0.76	0.65	0.77	0.72	1.01	1.22
22	6	2010	0.69	0.52	0.74	0.47	1.07	1.00
23	6	2011	1.02	0.98	1.06	0.70	1.04	0.98
24	2	2012	--	--	--	--	--	--
25	1	2013	--	--	--	--	--	--
26	3	2031	0.97	0.25	0.97	0.34	1.00	0.88
27	3	2032	1.05	0.67	1.05	0.67	1.00	0.75
28	4	2041	0.79	0.80	0.79	0.69	1.00	0.86
29	5	2042	0.85	0.48	0.86	0.56	1.01	0.90
30	5	2061	0.35	0.10	1.01	0.90	$\boxed{2.89}$	$\boxed{\sqrt{-1.53}}$
31	5	2062	0.92	0.84	0.91	0.88	0.99	1.05
32	4	2063	0.87	0.68	0.85	0.57	0.98	1.09
33	5	2071	0.44	0.80	0.47	0.78	1.07	2.23
34	4	2072	1.05	0.98	1.06	0.98	1.01	0.94
35	5	2073	0.85	0.37	0.86	0.55	1.01	1.05
36	4	2074	0.85	0.63	0.91	0.56	1.07	1.20
37	4	2075	1.14	0.74	1.17	0.74	1.03	0.83

TABLE 5.2:2 : DOMAINS OF STUDY

Numb.	*Str./domain*	*Domain*	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
38	5	2081	1.01	0.75	1.00	0.70	0.99	0.88
39	3	2082	1.16	0.79	1.09	0.78	0.94	0.70
40	4	2083	1.11	0.84	1.06	0.94	0.95	0.84
41	4	2084	0.90	0.45	0.85	0.43	0.94	0.89
42	5	2101	0.96	0.96	0.96	0.96	1.00	1.03
43	5	2102	1.04	0.85	1.01	0.74	0.97	0.89
44	5	3101	1.01	0.58	1.02	0.49	1.01	0.75
45	4	3201	0.46	0.16	0.45	0.13	0.98	0.45
46	5	3202	1.18	0.76	1.09	0.79	0.92	0.73
47	6	3203	0.99	0.87	0.99	0.76	1.00	1.00
48	5	3204	1.03	0.67	1.03	0.87	1.00	0.76
49	6	3205	0.96	0.34	0.93	0.35	0.97	0.64
50	5	3206	0.99	0.76	0.99	0.84	1.00	0.88
51	5	3207	0.56	0.99	0.58	0.99	1.04	1.87
52	4	3208	0.98	0.73	0.97	0.64	0.99	0.51
53	5	3209	0.96	0.34	0.95	0.40	0.99	0.64
54	6	3210	0.97	0.94	1.01	0.92	1.04	0.98
55	5	3401	1.02	0.84	1.02	0.86	1.00	0.77
56	6	3402	0.80	0.74	0.83	0.82	1.04	1.17
57	5	3901	1.60	0.92	1.58	0.86	0.99	0.61
58	5	4001	1.15	0.50	1.15	0.60	1.00	0.61
59	5	4101	0.98	0.48	0.89	0.52	0.91	0.81
60	5	4201	1.08	0.65	1.07	0.68	0.99	0.79
61	3	4202	1.04	0.92	1.05	0.83	1.01	0.78
62	5	4203	0.38	0.56	0.28	0.52	0.74	1.99
63	4	4204	0.66	0.46	0.62	0.57	0.94	0.71
64	1	9999	--	--	--	--	--	--

(290 Str./domains and 288 exclusive domains 1000 and 1917.)

TABLE 5.3 : 1 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
001	1902 1	180	17	15	0.47	0.64	0.70	0.68	0.76	0.77	0.72	1.14	1.53
002	1902 2	38	10	10	0.93	0.93	0.93	1.10	0.38	1.12	0.38	1.02	1.00
003	1902 2	80	18	17	0.83	0.01	-0.04	1.00	1.00	0.99	1.00	0.99	1.00
004	1902 3	9	5	4	0.59	0.80	0.80	0.82	0.59	0.98	0.60	1.20	1.15
005	1902 3	1	1	1	---	---	---	---	---	---	---	---	---
006	1902 4	10	10	10	0.96	0.98	0.94	1....	---	1....	---	1....	---
007	1903 1	36	12	11	$\boxed{1.00}$	$\boxed{0.55}$	$\boxed{0.51}$	0.72	0.84	0.67	0.86	0.93	1.25
008	1903 2	3	3	3	0.97	-0.60	-0.40	1....	---	1....	---	1....	---
009	1903 2	4	3	3	0.66	-0.57	-0.99	1.11	0.82	1.11	0.12	1.00	0.90
010	1903 3	3	3	3	-0.28	-0.05	0.97	1....	---	1....	---	1....	---
011	1904 1	129	4	3	$\boxed{1.00}$	$\boxed{-0.16}$	$\boxed{0.43}$	$\boxed{0.41}$	0.99	$\boxed{22.05}$	0.91	$\boxed{53.68}$	$\boxed{207.00}$
012	1904 2	19	12	12	0.91	0.77	0.82	0.85	0.64	0.79	0.58	0.92	1.00
013	1904 2	39	5	5	1.00	0.98	0.97	0.79	0.22	0.79	0.26	1.00	1.00
014	1904 3	5	5	5	1.00	1.00	0.99	1....	---	1....	---	1....	---
015	1904 4	3	3	3	$\boxed{1.00}$	$\boxed{-0.20}$	$\boxed{-0.23}$	1....	---	1....	---	1....	---

TABLE 5.3 : 2 : SAMPLE STRATA

Number	*Sample-stratum*	N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
016	1905 1	316	16	15	$\boxed{0.97}$	$\boxed{-0.18}$	$\boxed{0.93}$	0.55	0.98	1.58	0.37	2.85	5.77
017	1905 2	90	18	17	0.97	0.65	0.72	1.12	0.76	1.11	0.70	0.98	1.00
018	1905 2	59	7	7	0.91	-0.15	-0.32	1.05	0.99	1.07	0.95	1.02	0.91
019	1905 3	19	13	12	0.36	0.74	0.71	0.98	0.67	0.94	0.71	0.96	1.00
020	1905 3	4	3	3	1.00	-0.80	-0.84	0.56	0.59	0.56	0.55	1.00	1.00
021	1905 4	2	2	2	1....	1....	1....	1....	---	1....	---	1....	---
022	1906 1	186	12	12	0.90	0.88	0.60	0.60	0.47	0.74	0.80	1.24	0.83
023	1906 2	50	13	13	0.93	0.32	0.44	1.00	0.95	1.00	0.90	1.00	0.98
024	1906 2	53	5	5	1.00	0.99	0.99	0.87	0.12	0.88	0.15	1.01	0.44
025	1906 3	19	16	16	0.88	0.74	0.87	0.97	0.67	0.96	0.50	0.99	0.98
026	1906 4	3	3	3	1.00	$\boxed{-1.08}$	$\boxed{-1.08}$	$\boxed{0.73}$	---	$\boxed{0.08}$	---	$\boxed{1.12}$	$\boxed{1.00}$
027	1907 1	1721	45	41	0.96	0.85	0.88	1.30	0.53	1.30	0.48	1.00	0.58
028	1907 2	356	32	31	0.98	0.77	0.77	1.05	0.64	1.03	0.64	0.98	0.82
029	1907 2	334	23	23	0.97	-0.13	-0.16	1.10	0.99	1.11	0.99	1.01	0.94
030	1907 3	38	15	15	0.97	0.46	0.39	1.00	0.89	1.00	0.92	1.00	0.92
031	1907 3	6	6	6	1.00	-018	-0.18	1....	---	1....	---	1....	---
032	1907 4	13	13	13	0.99	0.63	0.56	1....	---	1....	---	1....	---

TABLE 5.3 : 3 : SAMPLE STRATA

Number	*Sample-stratum*	N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
033	1908 1	247	10	9	0.83	0.84	0.83	0.58	0.54	0.50	0.56	0.85	1.50
034	1908 2	33	8	8	0.98	0.31	0.26	0.87	0.95	0.89	0.97	1.03	1.10
035	1908 2	27	3	3	1.00	1.00	1.00	0.61	0.06	0.62	0.07	1.02	0.54
036	1908 3	5	5	5	0.98	0.35	0.51	1....	---	1....	---	1....	---
037	1908 4	2	2	1	---	1....	---	1....	---	---	---	---	---
038	1909 1	134	5	5	1.00	0.20	0.27	1.13	0.98	1.17	0.96	1.04	0.93
039	1909 2	19	7	7	1.00	-0.41	-0.42	0.88	0.91	0.88	0.91	1.00	1.14
040	1909 2	19	7	7	0.69	0.20	-0.10	0.91	0.98	1.03	1.00	1.14	1.19
041	1909 3	4	4	4	1.00	0.97	0.96	1....	---	1....	---	1....	---
042	1910 1	4110	70	64	0.93	0.59	0.60	0.94	0.81	0.93	0.80	0.78	0.96
043	1910 2	832	80	74	0.87	0.75	0.83	1.07	0.66	1.11	0.55	1.04	0.84
044	1910 2	857	13	12	0.77	0.16	-0.12	1.00	0.99	1.00	0.99	1.01	0.94
045	1910 3	138	44	41	0.81	0.70	0.49	0.99	0.72	0.98	0.87	0.99	0.98
046	1910 3	9	8	8	1.00	-0.15	-0.16	1.03	0.99	1.03	0.99	1.00	0.93
047	1910 4	34	34	33	0.96	0.96	0.96	0.95	---	0.96	0.28	1.01	0.29

TABLE 5.3 : 4 : SAMPLE STRATA

Number	*Sample-stratum*		N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
048	1911	1	842	118	107	0.93	0.63	0.72	1.31	0.77	1.33	0.70	1.01	0.66
049	1911	2	244	28	28	0.98	0.79	0.79	1.02	0.61	1.02	0.62	1.01	0.77
050	1911	2	333	13	12	0.89	0.48	0.37	1.46	0.88	1.37	0.93	0.94	0.68
051	1911	3	35	28	28	1.00	-0.19	-0.15	0.59	0.98	0.66	0.99	1.11	$\boxed{2.81}$
052	1911	3	10	9	9	1.00	0.97	0.97	0.94	0.25	0.93	0.25	0.99	0.79
053	1911	4	20	19	18	1.00	0.41	0.37	1.06	0.91	1.04	0.93	0.98	0.88
054	1912	1	716	50	48	0.58	0.43	0.61	0.90	0.90	0.94	0.79	1.03	1.11
055	1912	2	131	66	66	0.90	0.87	0.81	0.99	0.50	0.99	0.59	1.00	0.98
056	1912	2	206	8	8	0.92	0.57	0.51	$\boxed{1.80}$	0.82	$\boxed{1.63}$	0.86	0.90	0.49
057	1912	3	14	8	8	0.74	0.83	0.67	1.04	0.56	1.03	0.74	1.00	0.96
058	1912	3	4	4	4	0.96	0.56	0.39	1....	---	1....	---	1....	---
059	1913	1	478	27	24	0.98	0.84	0.87	$\boxed{1.51}$	0.55	$\boxed{1.51}$	0.49	1.00	0.41
060	1913	2	101	21	20	0.96	0.85	0.81	0.67	0.53	0.69	0.59	1.03	1.23
061	1913	2	84	8	7	1.00	0.99	0.99	0.67	0.17	0.53	0.17	0.79	0.37
062	1913	3	10	9	9	1.00	0.95	0.97	0.99	0.32	0.99	0.23	1.00	0.99
063	1913	4	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 5 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
064	1914 1	1841	46	42	0.93	0.55	0.53	1.11	0.84	1.06	0.85	0.95	1.00
065	1914 2	379	65	64	0.81	0.86	0.66	0.97	0.52	0.96	0.76	0.99	0.95
066	1914 2	336	18	17	1.00	0.77	0.76	$\boxed{0.44}$	0.64	0.50	0.65	1.14	1.87
067	1914 3	61	39	39	0.89	0.80	0.63	1.09	0.59	1.09	0.78	0.99	0.93
068	1914 3	4	4	4	0.98	0.64	0.78	1....	---	1....	---	1....	---
069	1914 4	6	5	5	1.00	1.00	1.00	0.82	0.04	0.81	0.05	1.00	1.00
070	1915 1	292	4	4	0.98	-0.36	-0.45	1.03	0.93	1.04	0.89	1.01	0.89
071	1915 2	80	4	5	1.00	1.00	1.00	$\boxed{0.37}$	0.05	$\boxed{0.47}$	0.06	1.27	0.25
072	1915 2	76	7	6	0.90	0.96	0.80	$\boxed{2.36}$	0.29	$\boxed{3.20}$	0.60	1.35	0.29
073	1915 3	20	9	8	0.98	0.99	0.99	1.52	0.10	1.33	0.12	0.88	0.29
074	1915 3	3	2	2	1....	1....	1....	0.92	0....	0.69	0....	0.75	0....
075	1915 4	8	8	8	0.98	0.99	0.99	1....	---	1....	---	1....	---
076	1916 1	194	4	4	1.00	1.00	1.00	$\boxed{0.26}$	0.02	$\boxed{0.25}$	0.06	0.96	3.65
077	1916 2	33	13	12	0.98	0.44	0.43	1.21	0.90	1.19	0.90	0.98	1.00
078	1916 2	76	5	6	0.17	0.60	-0.29	$\boxed{2.62}$	0.80	0.59	0.96	$\boxed{0.22}$	0.32
079	1916 3	3	2	2	1....	1....	1....	$\boxed{3.81}$	0....	$\boxed{3.81}$	0....	1.00	0....
080	1916 3	5	5	5	1.00	0.97	0.97	1....	---	1....	---	1....	---
081	1916 4	7	7	6	0.96	-0.07	-0.24	1....	---	1.01	0.97	1.01	1.00

TABLE 5.3 : 6 : SAMPLE STRATA

Number	*Sample-sratum*		N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
082	2001	2	6	4	4	0.95	0.65	0.85	$\boxed{0.48}$	0.76	$\boxed{0.29}$	0.53	0.61	0.57
083	2001	4	3	3	3	0.99	1.00	0.98	1....	---	1....	---	1....	---
084	2002	1	5253	55	54	0.98	0.76	0.80	1.12	0.65	1.16	0.60	1.03	0.68
085	2002	2	1295	65	64	0.95	0.75	0.74	1.01	0.66	1.02	0.68	1.01	0.95
086	2002	2	1708	15	13	0.92	-0.37	-0.33	0.95	0.93	0.96	0.95	1.02	1.00
087	2002	3	88	12	12	0.98	0.59	0.61	1.06	0.81	1.06	0.79	1.00	0.99
088	2002	3	3	3	3	1.00	-0.94	-0.94	1....	---	1....	---	1....	---
089	2002	4	3	3	3	1.00	1.00	1.00	1....	---	1....	---	1....	---
090	2005	1	875	39	35	0.94	0.69	0.69	0.91	0.72	0.93	0.72	1.02	1.11
091	2005	2	23	4	4	0.98	0.96	1.00	0.68	0.28	0.78	0.09	1.16	1.00
092	2005	2	90	10	10	0.46	0.05	0.73	1.02	1.00	1.62	0.69	1.59	0.72
093	2009	1	555	30	30	0.89	0.80	0.80	0.81	0.60	0.80	0.60	1.00	1.00
094	2009	2	20	5	5	0.74	-0.15	-0.08	0.94	0.49	0.96	1.00	1.03	1.00
095	2009	2	153	7	7	0.98	0.08	0.14	0.94	1.00	0.87	0.99	0.92	1.22
096	2009	3	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5:3 : 7 : SAMPLE STRATA

Number	*Sample-stratum*		*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
097	2010	1	1673	34	35	0.97	0.67	0.66	0.73	0.74	0.77	0.75	1.05	1.00
098	2010	2	363	19	19	0.93	0.65	0.73	0.87	0.76	0.84	0.69	0.96	1.12
099	2010	2	757	16	17	1.00	0.88	0.91	$\boxed{-0.01}$	0.47	0.13	0.41	$\boxed{-11.61}$	$\boxed{1982.00}$
100	2010	3	102	25	20	0.90	0.49	0.35	1.01	0.87	1.02	0.94	1.01	0.95
101	2010	3	5	5	5	1.00	-0.41	-0.41	1....	---	1....	---	1....	---
102	2010	4	14	14	14	0.99	-0.05	-0.05	$\boxed{0.99}$	---	$\boxed{0.99}$	---	$\boxed{1.00}$	---
103	2011	1	1577	213	198	0.05	-0.02	0.53	1.00	1.00	1.02	0.85	1.03	1.02
104	2011	2	87	10	9	0.90	0.80	0.76	1.66	0.60	1.67	0.65	1.00	0.53
105	2011	2	547	8	8	0.99	0.74	0.81	1.38	0.67	1.46	0.59	1.06	0.50
106	2011	3	5	5	5	1.00	0.27	0.27	1....	---	1....	---	1....	---
107	2011	3	1	1	1	---	---	---	---	---	---	---	---	---
108	2011	4	9	7	6	0.98	0.79	0.67	0.97	0.61	0.97	0.74	1.00	0.77
109	2012	1	1	$\boxed{0}$	1	---	---	---	---	---	---	---	---	---
110	2012	4	1	1	1	---	---	---	---	---	---	---	---	---
111	2013	4	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 8 : SAMPLE STRATA

Number	*Sample-stratum*		*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
112	2031	1	629	22	22	0.98	0.98	0.95	0.80	0.22	0.81	0.31	1.01	1.22
113	2031	2	60	9	8	0.98	0.83	0.82	$\boxed{1.52}$	0.55	$\boxed{1.43}$	0.57	0.94	0.44
114	2031	2	61	12	12	0.83	-0.18	-0.17	0.95	0.98	0.95	0.98	1.00	1.05
115	2032	1	347	56	53	0.98	0.73	0.74	1.10	0.68	1.11	0.67	1.01	0.75
116	2032	2	13	13	13	0.97	0.98	0.97	1....	---	1....	---	1....	---
117	2032	2	28	8	6	0.96	0.81	0.78	0.79	0.59	0.72	0.63	0.91	0.88
118	2041	1	752	14	12	$\boxed{0.99}$	$\boxed{0.59}$	$\boxed{0.74}$	$\boxed{2.01}$	0.81	$\boxed{2.16}$	0.67	1.08	0.36
119	2041	2	15	3	4	$\boxed{1.00}$	$\boxed{0.99}$	$\boxed{0.99}$	0.71	0.11	0.79	0.11	1.11	0.18
120	2041	2	169	9	9	0.97	0.32	0.34	0.92	0.95	0.92	0.94	0.99	1.07
121	2041	4	1	1	1	---	---	---	---	---	---	---	---	---
122	2042	1	437	46	43	0.93	0.83	0.77	1.10	0.56	1.07	0.64	0.98	0.77
123	2042	2	67	20	19	0.94	0.81	0.81	0.82	0.58	0.78	0.58	0.95	1.01
124	2042	2	53	12	10	0.83	0.10	0.46	1.00	0.99	1.21	0.89	1.20	1.00
125	2042	3	3	2	2	1....	1....	1....	0.90	0....	0.92	0....	1.02	0....
126	2042	4	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 9 : SAMPLE STRATA

Number	*Sample-stratum*		*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
127	2061	1	2460	38	37	0.96	1.00	0.44	1.01	0.10	0.95	0.90	9.43	$\sqrt{-1.35}$
128	2061	2	212	17	17	0.95	0.28	0.27	1.02	0.96	1.02	0.96	1.00	0.98
129	2061	2	308	24	22	0.24	0.52	0.04	1.21	0.85	1.01	1.00	0.83	0.62
130	2061	3	16	8	8	0.87	0.92	0.75	0.99	0.40	1.00	0.66	1.00	0.87
131	2061	4	9	9	9	0.96	0.80	0.90	1....	---	1.05	---	1.05	---
132	2062	1	587	29	28	0.98	0.57	0.50	1.10	0.82	1.07	0.87	0.97	0.89
133	2062	2	65	9	9	0.95	0.46	0.40	1.00	0.89	1.00	0.92	1.00	0.99
134	2062	2	64	10	10	0.69	-0.39	-0.43	0.64	0.92	0.70	0.90	1.10	1.73
135	2062	3	7	6	5	1.00	-0.15	0.01	1.02	0.99	1.01	1.00	0.99	0.98
136	2062	4	1	1	1	---	---	---	---	---	---	---	---	---
137	2063	1	985	36	35	0.93	0.60	0.57	0.94	0.80	0.93	0.82	0.99	1.03
138	2063	2	55	8	8	0.79	0.80	0.99	0.80	0.61	0.71	0.16	0.89	0.98
139	2063	2	91	9	9	0.46	0.29	0.06	1.40	0.96	1.17	1.00	0.84	0.71
140	2063	3	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 10 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\dfrac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\dfrac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\dfrac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\dfrac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\dfrac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\dfrac{v(\hat{R}_{reg})}{v(\hat{R})}}$
141	2071 1	890	45	45	0.72	0.57	0.55	1.13	0.82	1.11	0.83	0.98	0.86
142	2071 2	149	26	25	0.97	0.63	0.76	0.96	0.78	0.91	0.66	1.03	0.66
143	2071 2	162	6	6	0.89	-0.65	-0.58	1.15	0.76	1.15	0.81	1.00	0.87
144	2071 3	4	3	3	1.00	0.18	0.13	1.03	0.98	1.02	0.99	0.99	0.32
145	2071 4	1	1	1	---	---	---	---	---	---	---	---	---
146	2072 1	404	24	24	0.97	0.14	0.02	1.13	0.99	1.02	1.00	0.90	1.00
147	2072 2	114	14	12	1.00	-0.21	-0.22	1.04	0.98	1.06	0.98	1.02	0.95
148	2072 2	63	6	6	0.98	-0.42	-0.37	1.23	0.91	1.23	0.93	1.00	0.81
149	2072 3	8	6	6	0.95	-0.03	0.06	1.00	1.00	1.01	1.00	1.01	0.97
150	2073 1	839	46	45	0.90	0.81	0.69	0.88	0.59	0.88	0.72	1.00	1.09
151	2073 2	60	26	25	0.96	0.96	0.90	0.79	0.30	0.83	0.43	1.05	0.95
152	2073 2	168	14	13	0.90	0.38	0.40	$\boxed{1.76}$	0.93	$\boxed{1.81}$	0.92	1.02	0.56
153	2073 3	3	2	2	1....	1....	1....	1.12	0....	1.12	0....	1.00	0....
154	2073 4	1	1	1	---	---	---	---	---	---	---	---	---
155	2074 1	498	24	23	0.76	0.48	0.61	1.05	0.88	1.33	0.79	1.27	1.02
156	2074 2	32	5	5	0.98	0.99	0.99	0.64	0.16	0.64	0.12	1.00	1.56
157	2074 2	127	6	6	0.96	0.26	0.38	1.01	0.97	1.02	0.93	1.01	0.86
158	2074 4	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 11 : SAMPLE STRATA

Number	*Sample-stratum*		N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
159	2075	1	1096	32	32	0.89	0.53	0.60	1.25	0.85	1.34	0.80	1.07	0.76
160	2075	2	56	8	8	1.00	0.82	0.78	0.83	0.58	0.84	0.63	1.01	0.95
161	2075	2	184	6	6	0.94	0.70	0.72	$\boxed{1.53}$	0.71	$\boxed{1.41}$	0.70	0.92	0.53
162	2075	3	5	5	5	0.91	-0.20	-0.29	1....	---	1....	---	1....	---
163	2081	1	1162	111	106	0.94	0.69	0.76	1.02	0.72	1.01	0.65	0.99	0.85
164	2081	2	60	31	30	0.94	0.49	0.46	0.99	0.87	0.98	0.89	1.00	0.97
165	2081	2	76	2	2	1....	1....	1....	1.36	0....	1.44	0....	1.06	0....
166	2081	3	5	4	4	0.96	0.10	0.38	0.98	0.99	0.93	0.92	0.95	0.24
167	2081	4	2	2	2	1....	1....	1....	1....	---	1....	---	1....	---
168	2082	1	300	26	23	0.98	0.65	0.65	1.14	0.76	1.12	0.76	0.98	0.72
169	2082	2	19	4	4	0.99	0.48	0.50	1.08	0.88	1.08	0.87	1.00	0.92
170	2082	2	47	4	4	-0.17	0.96	-0.39	1.64	0.30	0.75	0.92	0.46	0.33
171	2083	1	1175	55	51	0.75	0.45	0.37	1.08	0.89	1.05	0.93	0.97	0.91
172	2083	2	58	12	12	0.65	0.83	0.31	1.32	0.56	1.11	0.95	0.84	0.59
173	2083	2	312	13	12	0.69	0.70	0.07	0.96	0.71	1.06	1.00	1.10	0.84
174	2083	3	2	2	2	1....	1....	1....	1....	---	---	---	1....	---

TABLE 5.3 : 12 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
175	2084 1	502	39	38	0.97	0.90	0.91	0.81	0.44	0.77	0.41	0.95	0.98
176	2084 2	35	11	11	0.98	0.93	0.86	1.12	0.36	1.12	0.51	1.00	0.89
177	2084 2	100	7	7	0.92	-0.05	-0.41	0.95	1.00	0.68	0.91	0.71	0.42
178	2084 3	2	2	2	-1....	1....	1....	1....	---	1....	---	1....	---
179	2101 1	3836	36	38	0.59	0.31	0.17	0.92	0.95	0.95	0.99	1.04	1.09
180	2101 2	837	33	33	0.84	-0.01	-0.11	1.00	1.00	0.95	0.99	0.95	0.99
181	2101 2	413	26	24	0.81	0.31	0.45	0.86	0.95	0.74	0.89	0.87	1.07
182	2101 3	62	10	10	0.83	0.69	0.82	1.11	0.72	1.18	0.57	1.07	0.74
183	2101 4	14	12	12	1.00	0.86	0.87	0.91	0.51	0.92	0.49	1.00	1.06
184	2102 1	726	21	24	0.60	0.53	0.66	1.01	0.85	0.96	0.75	0.95	0.89
185	2102 2	332	27	28	0.65	0.22	0.46	1.10	0.97	1.08	0.89	0.98	0.87
186	2102 2	91	4	4	0.48	0.94	0.60	2.03	0.33	1.54	0.80	0.76	0.40
187	2102 3	73	20	20	0.56	0.63	0.87	0.95	0.77	0.92	0.49	0.98	1.00
188	2102 4	10	10	10	1.00	-0.29	-0.30	1....	---	1....	---	1....	---

TABLE 5.3 : 13 : SAMPLE STRATA

Number	*Sample-stratum*	N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
189	3101 1	2400	57	56	0.90	0.66	0.53	1.17	0.75	1.13	0.85	0.97	0.80
190	3101 2	302	44	41	0.96	0.86	0.94	0.86	0.52	0.89	0.33	1.03	0.72
191	3101 2	140	12	10	0.99	0.29	0.30	1.03	0.96	0.97	0.95	0.94	0.95
192	3101 3	32	31	31	0.87	0.09	0.00	1.00	1.00	1.00	1.00	1.00	1.00
193	3101 4	14	14	14	0.99	-0.43	-0.41	$\boxed{1.07}$	---	$\boxed{1.07}$	---	$\boxed{1.00}$	---
194	3201 1	888	17	16	0.90	---	---	---	---	---	---	---	---
195	3201 2	60	36	35	0.83	0.38	0.43	0.98	0.92	0.98	0.90	1.00	1.00
196	3201 2	13	10	10	1.00	-0.11	-0.10	0.78	0.99	0.79	0.99	1.00	1.15
197	3201 3	9	9	9	0.81	-0.77	-0.63	1....	---	1....	---	1....	---
198	3201 1	2166	25	25	0.95	0.70	0.69	1.26	0.72	1.25	0.73	1.00	0.79
199	3202 2	187	29	29	0.88	-0.12	-0.16	0.98	0.99	0.97	0.99	0.99	1.01
200	3202 2	58	9	9	0.93	0.85	0.69	1.19	0.53	1.14	0.72	0.96	0.68
201	3202 3	36	25	25	0.71	0.54	0.31	1.41	0.84	1.14	0.95	0.80	0.49
202	3202 4	5	5	5	0.85	0.94	0.98	1....	---	1....	---	1....	---

TABLE 5.3 : 14 : SAMPLE STRATA

Number	*Sample-stratum*	N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
203	3203 1	1695	26	27	0.97	0.45	0.54	1.01	0.89	0.96	0.84	0.95	0.84
204	3203 2	284	17	17	0.68	0.49	0.73	0.87	0.87	0.87	0.68	1.00	1.15
205	3203 2	143	8	8	0.93	0.35	0.39	0.91	0.94	0.90	0.92	0.99	1.09
206	3203 3	69	28	28	0.96	0.49	0.54	1.19	0.87	1.23	0.84	1.03	0.79
207	3203 3	4	4	4	0.74	0.83	0.44	1....	---	1....	---	1....	---
208	3203 4	17	16	16	0.97	0.91	0.85	0.96	0.42	0.96	0.53	1.00	1.00
209	3204 1	1657	37	35	0.01	0.76	0.14	1.01	0.65	1.00	0.99	1.00	0.81
210	3204 2	256	18	17	0.78	0.78	0.54	1.10	0.63	1.12	0.84	1.02	0.80
211	3204 2	87	12	11	0.99	0.80	0.75	1.30	0.60	1.23	0.66	0.94	0.61
212	3204 3	44	12	12	0.95	0.47	0.45	0.96	0.88	0.97	0.89	1.00	1.07
213	3204 4	13	13	13	0.97	0.89	0.83	1....	---	1....	---	1....	---
214	3205 1	2250	37	31	0.95	0.84	0.80	0.98	0.54	0.92	0.59	0.93	0.74
215	3205 2	262	56	53	0.99	0.98	0.98	0.94	0.20	0.91	0.21	0.97	0.41
216	3205 2	236	18	17	0.97	0.40	0.62	1.20	0.92	1.38	0.78	1.15	0.36
217	3205 3	22	21	21	0.99	0.99	0.99	0.95	0.15	0.95	0.16	1.00	1.05
218	3205 3	1	1	1	---	---	---	---	---	---	---	---	---
219	3205 4	4	4	4	0.95	0.98	0.91	1....	---	1....	---	1....	---

TABLE 5.3 : 15 : SAMPLE STRATA

Number	*Sample-stratum*	N	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{yz}$	$\hat{\rho}_{xz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
220	3206 1	167	25	25	0.93	0.72	0.70	1.08	0.69	1.07	0.71	1.00	0.93
221	3206 2	34	14	15	0.97	0.61	0.44	0.93	0.80	0.93	0.90	1.00	0.81
222	3206 2	6	4	4	1.00	0.11	0.11	0.94	0.99	0.95	0.99	1.00	1.00
223	3206 3	4	3	3	1.00	-0.73	-0.73	0.81	0.68	0.81	0.68	1.00	1.00
224	3206 4	3	3	3	0.88	0.89	0.56	1....	---	1....	---	1....	---
225	3207 1	473	27	27	0.99	-0.12	-0.13	$\boxed{-0.04}$	0.99	$\boxed{-0.02}$	0.99	0.60	$\boxed{91.00}$
226	3207 2	48	25	24	0.93	-0.04	-0.03	1.01	1.00	1.01	1.00	1.00	1.00
227	3207 2	17	4	4	1.00	-0.76	-0.76	$\boxed{1.93}$	0.65	$\boxed{1.93}$	0.65	1.00	1.00
228	3207 3	9	8	8	0.83	0.64	0.69	1.02	0.77	1.02	0.72	1.00	1.00
229	3207 4	3	3	3	0.30	0.96	0.03	1....	---	1....	---	1....	---
230	3208 1	56	4	4	0.99	0.23	0.30	0.79	0.97	0.70	0.95	0.88	1.67
231	3208 2	31	5	5	0.99	0.88	0.92	0.93	0.48	0.92	0.40	0.99	0.77
232	3208 3	10	5	5	1.00	0.99	0.98	0.87	0.13	0.85	0.20	0.98	1.00
233	3208 4	8	8	8	0.63	0.13	0.47	1....	---	1....	---	1....	---
234	3209 1	1294	35	34	0.65	0.07	-0.00	1.01	1.00	1.00	1.00	0.99	0.99
235	3209 2	147	18	17	1.00	0.96	0.95	0.80	0.29	0.81	0.31	1.01	0.46
236	3209 2	83	8	7	1.00	0.96	0.96	1.17	0.30	1.01	0.28	0.86	0.25
237	3209 3	21	20	20	0.82	0.64	0.45	1.00	0.77	1.00	0.89	1.00	1.00
238	3209 4	3	2	2	1....	-1....	-1....	1.03	0....	1.15	0....	1.12	0....

TABLE 5.3 : 16 : SAMPLE STRATA

Number	*Sample-stratum*		*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
239	3210	1	603	42	41	0.34	0.72	-0.06	0.86	0.70	1.01	1.00	1.17	0.84
240	3210	2	90	35	33	0.65	0.18	0.37	1.02	0.98	1.02	0.93	1.00	0.95
241	3210	2	404	17	15	0.08	-0.02	0.73	1.00	1.00	1.04	0.68	1.04	0.93
242	3210	3	8	8	7	1.00	-0.32	-0.29	1....	---	1.02	0.96	1.02	0.96
243	3210	3	19	4	4	-0.37	0.21	0.66	1.01	0.98	1.06	0.75	1.04	1.00
244	3210	4	4	4	4	0.95	0.96	0.98	1....	---	1....	---	1....	---
245	3401	1	559	12	12	0.98	0.59	0.62	1.15	0.81	1.15	0.78	1.00	0.74
246	3401	2	164	19	19	0.98	0.43	0.44	1.04	0.90	1.04	0.90	1.00	1.00
247	3401	2	8	2	1	---	-1....	---	1.33	0....	---	---	---	---
248	3401	3	30	11	10	0.94	-0.17	-0.00	0.95	0.99	1.00	1.00	1.06	1.00
249	3401	4	14	14	14	1.00	0.99	0.99	1....	---	1....	---	1....	---
250	3402	1	579	10	11	0.91	0.75	0.82	0.98	0.66	1.00	0.57	1.02	0.91
251	3402	2	172	18	18	0.97	0.29	0.29	0.96	0.96	0.95	0.96	0.99	1.00
252	3402	2	34	6	6	0.42	0.93	0.73	$\boxed{0.06}$	0.37	$\boxed{0.18}$	0.69	$\boxed{2.85}$	$\boxed{25.00}$
253	3402	3	24	17	17	0.92	0.56	0.44	0.97	0.83	0.98	0.90	1.01	0.88
254	3402	3	2	2	2	1....	1....	1....	1....	---	1....	---	1....	---
255	3402	4	1	1	1	---	---	---	---	---	---	---	---	---

TABLE 5.3 : 17 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
256	3901 1	3738	84	82	0.94	-0.03	-0.02	1.01	1.00	1.01	1.00	1.00	0.99
257	3901 2	322	39	39	0.71	0.32	0.56	1.13	0.95	1.17	0.83	1.04	0.91
258	3901 2	55	6	5	0.30	-0.30	-0.45	$\boxed{-3.67}$	0.95	$\boxed{-4.74}$	0.89	1.29	1.00
259	3901 3	66	40	39	0.32	0.57	0.62	1.15	0.82	1.20	0.79	1.04	0.86
260	3901 4	14	13	13	0.98	0.85	0.87	$\boxed{6.46}$	0.52	$\boxed{6.28}$	0.49	0.97	1.00
261	4001 1	2428	38	35	0.90	0.90	0.75	1.36	0.44	1.28	0.66	0.94	0.48
262	4001 2	172	18	18	0.94	0.79	0.81	1.34	0.61	1.40	0.59	1.04	0.71
263	4001 2	282	20	20	0.88	0.58	0.67	1.07	0.82	1.09	0.75	1.02	0.89
264	4001 3	28	11	10	0.95	0.87	0.89	0.86	0.49	0.90	0.46	1.05	0.83
265	4001 4	5	4	4	0.83	-0.54	-0.72	1.02	0.84	1.03	0.69	1.01	0.95
266	4101 1	631	18	16	0.94	0.91	0.95	1.00	0.41	0.95	0.30	0.96	0.68
267	4101 2	30	4	4	0.68	0.50	0.08	0.98	0.86	1.00	1.00	1.02	0.92
268	4101 2	81	6	5	$\boxed{0.93}$	$\boxed{0.12}$	$\boxed{-0.79}$	1.02	0.99	$\boxed{0.08}$	0.61	$\boxed{0.08}$	0.91
269	4101 3	1	1	1	---	---	---	---	---	---	---	---	---
270	4101 4	2	2	2	1....	1....	1....	1....	---	1....	---	1....	---
271	4201 1	325	14	14	0.99	0.76	0.74	1.11	0.66	1.09	0.68	0.99	0.82
272	4201 2	63	7	7	0.97	0.98	0.99	1.17	0.19	1.14	0.17	0.98	0.71
273	4201 2	10	5	5	1.00	-0.96	-0.96	0.74	0.30	0.74	0.29	0.99	1.00
274	4201 3	10	5	5	1.00	0.55	0.52	1.18	0.83	1.16	0.86	0.99	1.00
275	4201 4	4	4	4	1.00	1.00	1.00	1....	---	1....	---	1....	---

TABLE 5.3 : 18 : SAMPLE STRATA

Number	*Sample-stratum*	*N*	n_{r_1}	n_{r_2}	$\hat{\rho}_{yx}$	$\hat{\rho}_{xz}$	$\hat{\rho}_{yz}$	$\frac{\hat{X}_{reg}}{\hat{X}}$	$\sqrt{\frac{v(\hat{X}_{reg})}{v(\hat{X})}}$	$\frac{\hat{Y}_{reg}}{\hat{Y}}$	$\sqrt{\frac{v(\hat{Y}_{reg})}{v(\hat{Y})}}$	$\frac{\hat{R}_{reg}}{\hat{R}}$	$\sqrt{\frac{v(\hat{R}_{reg})}{v(\hat{R})}}$
276	4202 2	2059	34	31	0.97	0.40	0.56	1.05	0.92	1.06	0.83	1.01	0.82
277	4202 2	73	17	17	0.99	0.21	0.18	1.03	0.98	1.03	0.98	1.00	1.00
278	4202 2	17	4	4	1.00	0.80	0.82	0.62	0.59	0.60	0.57	0.98	1.00
279	4203 1	569	11	10	0.88	0.83	0.85	0.83	0.56	0.69	0.52	0.82	0.93
280	4203 2	37	12	13	0.85	0.81	0.49	1.30	0.59	1.20	0.87	0.93	0.62
281	4203 2	47	7	7	0.94	-0.60	-0.65	1.02	0.80	1.02	0.76	1.00	0.98
282	4203 3	9	9	9	0.99	0.96	0.97	1....	---	1....	---	1....	---
283	4203 4	3	3	3	0.99	-3.18	-3.23	-1.57	---	-1.60	---	1.02	---
284	4204 1	301	23	23	0.97	0.86	0.78	1.14	0.51	1.12	0.63	0.97	0.57
285	4204 2	32	9	9	0.95	0.90	0.78	0.70	0.45	0.74	0.62	1.05	1.32
286	4204 2	29	6	6	0.71	0.65	0.99	0.91	0.76	0.59	0.12	0.65	0.16
287	4204 4	1	1	1	---	---	---	---	---	---	---	---	---
288	9999 9	1	1	1	1.00	-4.84	-4.67	---	---	---	---	---	---

COMMENTS ON THE TABLES:

The results here clearly shows on all sorts of problems with designs with small samples. In some cases there are errors due to file error.

The tables also clearly shows great efficiency gains for the regression estimators versus the ordinary estimators with a proper large sample design. This is especially true for estimators of levels. For the estimators of change the efficiency gains are less but significant.

6. CONCLUSIONS OF REGRESSION ESTIMATORS IN THEORY AND IN PRACTICE

6.1 ESTIMATORS OF LEVELS

With a proper large sample design the regression estimators are always more efficient than the ordinary estimators and with a good auxiliary variable very much so, that is the theoretical properties holds.

With small sample designs this is not true and then the ordinary estimators are more robust. In this case the auxiliary variable can be used in different ways but not in the estimator.

6.2 ESTIMATORS OF CHANGE

With a proper large sample design the regression estimators are mostly more efficient than the ordinary estimators, that is the theoretical properties holds. The efficiency gains here are much less than in the case of estimators of levels.

With small sample designs peculiar things can happen and then the more robust ordinary estimator (combined ratio estimator) is to be prefered. But also this estimator is large sample dependent! In this case as above the auxiliary variables can not be used in the estimator but used in different other ways.

So, with a proper large sample design the regression estimators for both levels and change are to be prefered to the ordinary estimators.

With a small sample design the ordinary estimators are to be prefered both for estimators of levels and changes. The auxiliary variables can then be used in oyher ways.

7. REFERENCES

(1) Bergdahl, M. (1992) : Översyn av omsättningsstatistiken. F-metod nr. 32, 1992-12-15. Statistics Sweden.
The empirical results of regression estimation shown in chapter 5 was done as a part of a rewiew of a retail survey at Statistics Sweden.
Here is reported the final methodology choises that was implemented in the reatail survey.

(2a) Carlsson, F. och Garås, T. (1990) : Regressionsestimat med hjälp av MOMS-registret. F-metod nr. 30, 1990-10-18. Statistics Sweden.

(2b) Carlsson, F. and Garås, T. (1992, english version of 1a) : Regression Estimation Using the VAT Register. RA PM nr. 7, Oktober 1992. Statistics Sweden.
Here is reported the test with regression estimation in a retail survey at Statistics Sweden.

(3) Cochran, W. G. (1977) : Sampling Techniques. 3rd ed. Wiley.

(4) Feller, W. (1971) : An Introduction to Probability Theory and Its Applications. Vol. II. Wiley.

(5a) Forsman, G. och Garås, T. (1982) : Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Promemorier från P/STM nr. 7, 1982-11-01 (R&D Report). Statistics Sweden.

(5b) Forsman, G. and Garås, T. (1982, english shorthand version of 5a) : Optimal Estimation of Change in SAmple Surveys. ASA 1982, Proceedings of the Section on Survey Research Methods.
Deals with composite estimators, which includes two-phase regression estimators.

(6) Gadd, A. (1963) : Urval och estimationsteknik vid de kvartalsvisa investeringsundersökningarna. Statistisk Tidskrift 1963:6. Statistics Sweden.
This paper deals with "two-phase combined regression estimators" and presents the formulas that has been used in an investment survey at Statistics Sweden.

(7) Garås, T. (1985) : Optimal Estimation of Levels and Changes. Working paper 1985-10-17, P/STM, Statistics Sweden.
Includes about reference 5b and then deals with further studies concerning composite estimators.

(8) Garås, T. (1987) : Möjligheter att utnyttja hjälpinformation i tjänstenäringarna i samband med urval och estimation.
Working paper 1987-06-26, F/STM (Strukturprojektet), Statistics Sweden.
Rewiew over methods where auxiliary information is used in the estimators.

(9) Garås, T. (1989) : Förändringsestimatorer vid dynamiska populationer. F-metod nr. 21, 1989-02-27. Statistics Sweden.
Deals with estimators of change, overlapping panel designs and dynamic populations; especially when the Jales-technique is used for sampling in the SAMU-system, which includes the enterprise-register (CFAR) as a frame.

(10) Marsden, J. E. and Tromba, A. J. (1981) : Vector Calculus. 2nd ed. W. H. Freeman and Company.

(11) Särndal, C-E, Swensson, B. and Wretman, J. (1992) : Model Assisted Survey Sampling. Springer-Verlag.
Modern sampling book with general approach. Note specificly the general regression estimator with variance and variance-estimator at Result 6.6.1, page 235.

(12) Wolter, Kirk M. (1985) : Introduction to Variance Estimation. Springer-Verlag.
Here methods are given to have approximate variances for complex estimators.

APPENDIX ON PROOFS

What are needed too proofe the variances and variance-estimators of all the estimators in this paper, besides basic definitions of variances and covariances, are :
(Complete panel-design, stratified SRS (STSRS) and a non-dynamic population.)

1. **Taylor expansion methods** to approximate non-linear estimators with linear estimators, whose variances and variance-estimators can be found by conventinal methods and hereby have approximate variances.

See for example Wolter, reference (12), chapter 6; Taylor Series Methods. Taylor expansion formulas can also be found in standard textbooks on Mathematical Calculus.

2. Often the **proof of the SRS-case** will do to have the STSRS-case.

3. **Separate regression estimators for levels and their variances.**
See Cochran, reference (3), chapter 7; Regression Estimators :
Theorem 7.3 on page 194 gives the SRS-case and by applying this to each stratum we have the STSRS-case in (7.57) on page 202.

The variance-estimators in this paper differs slightly from the ones proposed by Cochran, (7.29)-(7.30) (SRS-case) on page 195 and (7.58) (STSRS-case) on page 202.

With Cochran´s divisor n - 2 ($n_h - 2$) replaced by n - 1 ($n_h - 1$) this papers variance-estimator is achieved for the separate regression estimator of levels.

4. *All regression estimators here are*, as customary, ***conditioned*** *on the estimators of of the regression coefficients.*

5. Although we have a complete panel-design we have after non-response, with a straight adjustment model ("mean-value imputation"), *a conditioned sample after non-response* which means **an overlapping panel-design.**

More precisely we need $Cov(\hat{Y},\hat{X}) = \sum_h N_h^2 \left(\frac{g_h}{n_{h_1} n_{h_2}} - \frac{1}{N_h} \right) S_{hx} S_{hy} \rho_{hxy}$

where $s(g_h) = s(n_{h_1}) \cap s(n_{h_2})$ and we then have *the covariance-estimator*

$\mathrm{cov}(\hat{Y},\hat{X}) = \sum_h N_h^2 \left(\frac{g_h}{n_{h_{r_1}} n_{h_{r_2}}} - \frac{1}{N_h} \right) s_{hx} s_{hy} \hat{\rho}_{hxy}$ where $s(g_h) = s(n_{h_{r_1}}) \cap s(n_{h_{r_2}})$ and

r = responding, given time-period.

The covariance and covariance-estimator above can be found as a special case in reference (9), (3.14) on page 10-11.

In reference (5a), there is a proof of the covariance above in the SRS-case on page 20-21. In Sats 1 (in 5a) and in Theorem 1 (in 5b) the covariance above in the SRS-case is given with the restriction $S^2 = S_x^2 = S_y^2$.

Förteckning över utkomna R&D Reports:

R&D Reports är en för U/ADB och U/STM 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 1995-

1995:1 (grön)	Asymptotic Theory for Order Sampling *(Bengt Rosén)*
1995:2 (gul)	PC-programvaror för estimation och statistisk analys *(Dan Hedlin)*
1995:3 (grön)	Vad användarna tycker om SCB - Rapport från två användarstudier med QSP *(Jan Eklöf, Mats Bergdahl)*
1995:4 (grön)	Bortfallsbarometern nr 10 *(Mats Bergdahl, Pär Brundell, Margareta Eriksson, Dan Hedlin, Monica Rennermalm, Per Sandgren)*
1995:5 (gul)	Statistical Metadata - a tutorial *(Bo Sundgren)*
1995:6 (gul)	Making statistical data more available *(Bo Sundgren)*
1995:7 (gul)	Integration of statistical information systems - theory and practice *(Bo Sundgren and Erik Malmborg)*
1995:8 (gul)	Statistical information systems in a modern society - roles, functions, and system designs *(Bo Sundgren)*
1996:1 (grön)	On Sampling with Probability Proportional to Size *(Bengt Rosén)*
1996:2 (grön)	Bortfallsbarometern nr 11 *(Antti Ahtiainen, Stefan Berg, Margareta Eriksson, Åsa Greijer, Dan Hedlin, Monica Rennermalm, Anita Ullberg)*

Tidigare utgivna *R & D Reports* kan beställas genom Ingvar Andersson, SCB, U/SIB, Box 24300, 115 81 STOCKHOLM (telefon 08-783 41 47, fax 08-783 45 99, e-post ingvar andersson@scb.se).

R & D Reports listed above as well as issues from 1988-1994 can - in case they are still in stock - be ordered from Statistics Sweden, att. Ingvar Andersson U/SIB, Box 24300, S-115 81 STOCKHOLM (telephone +46 8 783 41 47, fax +46 8 783 45 99, e-mail ingvar andersson@scb.se).