The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey

Jan Hörngren

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1992:13

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 1992:13. The use of registers as auxiliary information in the Swedish labour force survey / Jan Hörngren.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1992-X101OP9213

The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey

Jan Hörngren

R&D Report
Statistics Sweden
Research - Methods - Development
1992:13

Från trycket November 1992
Producent Statistiska centralbyrån, utvecklingsavdelningen
Ansvarig utgivare Lars Lyberg
Förfrågningar Jan Hörngren, 08/783 42 29

ISSN 0283-8680
Garnisonstryckeriet, Stockholm

The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey

by

Jan Hörngren
Statistics Sweden

Abstract

The possibility to improve the quality in surveys with auxiliary information from administrative registers has increased considerably the last decade in Sweden. Regarding the Swedish Labour Force Survey (LFS) there are, except the Statistics Swedens Register of the Total Population, two registers of special interest; the National Labour Market Board Register of those in search of work (AMSR) and Statistics Swedens newly established Annual Register of Employed (ARE).

The AMSR contains useful information on the unemployed. It is possible to reduce the variance in the estimation of the unemployed by 30% when using information from AMSR to form poststrata.

ARE is based upon six other administrative registers, mainly the Employer's Statements of Income. ARE contains information on the employment situation among individuals. Since 1989 we are using auxiliary information (employed/not employed) in the LFS at the sampling stage. In this report we will study the possibility to use ARE-infomation in the estimation by poststratification. It seems that this method will gain in precision for several domains of employed persons and give better adjustment for nonresponse in comparsion to the present poststratification system.

Key words: Poststratification; register; gain in precision; adjustment for nonresponse.

CONTENTS

1 Introduction

The use of auxiliary information is usually of great importance in sample surveys for the reliability of the estimates. In Sweden the possibilities of using auxiliary information have increased considerably. For the Swedish Labour Force Survey (LFS) there are, besides the Register of the Total Population (RTB), mainly two other registers that contain information that can be used as auxiliary information in the LFS. The registers are Statistics Swedens Annual Register of Employed (ARE) and the Labour Market Boards Register (AMSR).

The information from ARE has been used since 1989 as auxiliary information at the sampling stage. In this report we shall also look into the possibilities of using the information from ARE for poststratification at the estimating stage. Auxiliary information from ARE are especially important when estimating the number of persons that are employed within industry.

AMSR contains important information that can be used when estimating the number of unemployed. We will also analyse the possibilities of using auxiliary information for poststratification when estimating the number of unemployed.

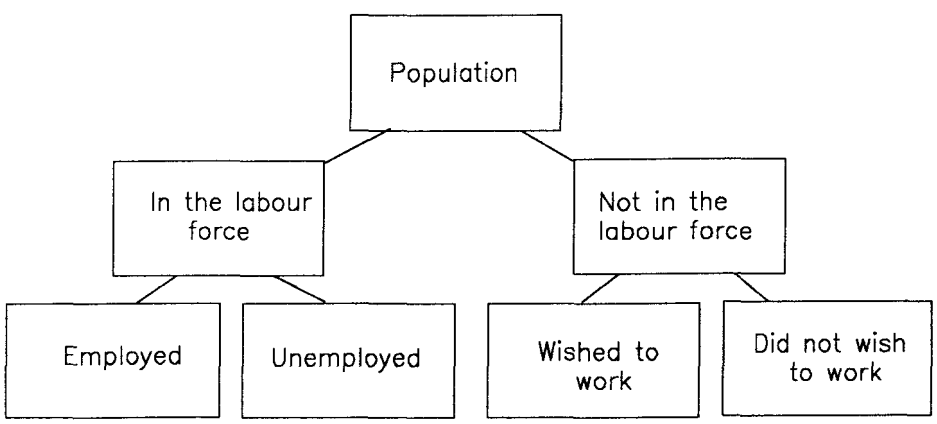
This means that when estimating the total number of unemployed and its different domains we are using a special estimation process which we will not use for the other estimates. Some of the variables will therefore, due to additivity reasons, be differential estimates, formed by components from two different systems of poststratification.

2 General Information on LFS, ARE and AMSR.

2.1 Labour Force Surveys (LFS)

The object of the LFS is to describe the current employment situation and to form a picture of the development on the labour market. LFS is the source of other statistics e.g the National Accounts, labour market analysis and forecasts. The schematic figure below shows the connection between some main concepts in the LFS.

Figure 1 Main concepts in the LFS



Statistics Sweden has carried out the LFS since 1961 and 1970 the LFS changed from being quarterly to becoming monthly surveys. The data is collected by means of telephone interviews. Computer assisted interviews are used since December 1991 (DATI).

The sampled population in LFS consist of persons that are 16 but not 65 years old and who are covered by the civil registration. The age of a person is the actual age in the reference month.

The sample size is at present 18 000 persons per month.

2.2 Statistics Swedens Register of Employed (ARE)

ARE was established in 1985 and is updated every year. The register contains information on the employment situation for the total population in Sweden. ARE is in turn based on six registers. The main source is the statement of income from the employer (KUA). The reference period for ARE is November.

The information of interest for using in the LFS is data on employment status and the employed divided by industry, classified according to the Swedish Standard Industry Classification (SNI).

The employment variable in the ARE is classified as follows:

1= employed
2= not employed
3= no information
4= children 0-15 years

In the ARE, employed are defined as those who are 16 years and over and who did paid work, in average at least one hour per week. Consequently employed are defined by income. In the LFS, employed are defined as those that did paid work during the reference week (more than 1 hour).
A difference between the LFS and the ARE is that in the LFS unpaid family workers are included (e.g in agriculture). The aim is that the definitions used in the ARE will correspond as far as possible to the LFS definitions. This is an advantage when data from registers is used as auxiliary information in LFS.

2.3 The Labour Market Boards Register of Those in Search of Work (AMSR)

The purpose of AMS's register is to describe those looking for work and also to constitute the basis for planning and the follow-up of the activity of employment offices. AMSR is continuously updated with data from employment offices throughout the whole country. Accordingly, the statistics are not made for describing the situation of those in search of work at a specific time, but to constitute information about current data in the register.

The information that is of main interest for the LFS is category data on those in search of work and their need for service. Futhermore in this context data on the date of deregistration of applicants is necessary. The persons that are registered are classified in eight "search categories" and three different "need of service" categories:

"Search category"

1 In search of work without work
2 Part-time unemployed
3 Having temporary work
4 Having a permanent work
5 Doing public relief work
6 Public relief work for youth
7 Special AMS courses
8 Vocational training courses

"Need of service" category

1 Available for work straight away
2 Need for counselling/inquiry
3 Others

The categories that are closely related to the LFS definition of unemployed are search category 1 and need of service category 1 AMS(1:1). We will not cover the differences between the definitions now. The purpose of this report is to study the possibilities of using AMSR as auxiliary information in estimating the number of unemployed.

3 The Present Sampling and Estimation Methods Used in LFS

3.1 The Sampling System

3.1.1 The Sampling Frame and the Rotation System

The sampling frame used in LFS is RTB, sorted by identity numbers and complemented with data on employment status from ARE. The sample consists of three independent samples, one for each month in a quarter. Each sample, 18 000 persons per month, is rotated so that 1/8 is renewed in between two consecutive survey occasions, i.e for each sample it occurs every third month. Each sampled person is in the LFS once every quarter and totally 8 times throughout a period of two years, and then the person is replaced with a new sampled person. The rotation system gives priority to quarterly estimates and changes between consecutive quarters.

The sample is drawn once a year and then it is distributed amongst the following 12 months. The total yearly sample consists of ca 27 000 persons (12x18 000/8). In the middle of each quarter the sampling frame is updated with new immigrants.

3.1.2 The Stratified Sampling

The population is stratified by sex, region, nationality (Swedish/non-Swedish) and employed (employed/not employed) from ARE, which gives 2x24x2x2=192 sample strata, wich are allocated proportional to the stratum sizes. Among those strata systematical samples are drawn. Since the sample frame is sorted by age, we get a more regular distribution by age when using a systematic sample than by using a simple random sample.

The sampling frame is generated by linking and matching the RTB for February year *t* and ARE for year *t-2* (the current value for November). On average the auxiliary information is 23 months old.

3.1.3 The Inclusion Probabilities in the LFS

The inclusion probabilities in the LFS varies. At a certain survey occasion the persons from the latest yearly draw and the complement of immigrants are included. Futher the probability of being included is affected by the matching procedure that is performed towards previous samples to avoid a redraw too soon after being in the LFS. The variation among the inclusion probabilities has a marginal influence on the estimates. This is futher discussed in chapter 3.2.

3.2 The Present Estimation Procedure

The sample adjustment for totals is done with regards to the inclusion probabilities and current monthly data on the number of people from the RTB. At the estimation the sample is devided into 300 poststrata in combination of sex, age (10 groups) and region (2x10x15). A region consists of one or more counties. Totally there are 15 regions. The current poststratification is explained by the classification of poststrata that, to a big extent, correspond to the groups that are important domains in the LFS and also by the variation in nonresponse due to geography. The nonresponse is attended to by so called "straight adjustment" within the poststrata of the estimates. This procedure has the same effect as imputation of means.

An estimate of a total could be denoted as:

$$\hat{Y} = \sum_{g=1}^{300} N_g \frac{\sum_{i=1}^{n_g} \frac{1}{\pi_{gi}}}{\sum_{i=1}^{n_g} \frac{1}{\pi_{gi}}} \qquad (1)$$

Where

$\hat{Y}$ = The estimate of a total, e.g the number of employed people

N_g = The number of persons in the population in poststratum g, g=1,2,...,300.

n_g = The sample size (the number of persons that are not nonrespondents) in poststratum g.

π_{gi} = The inclusion probability of the i:th person in poststratum g.

$$y_{gi} = \begin{cases} 1 \text{ if the person has the characteristic} \\ 0 \text{ otherwise} \end{cases}$$

whithout the finite population correction ($\approx$1) the estimate of the variance can be written:

$$\hat{V}(\hat{Y}) = \sum_{g=1}^{300} N_g^2 \frac{\sum_{i=1}^{n_g} \left(\frac{1}{\pi_{gi}}\right)^2 \hat{p}_g(1-\hat{p}_g)}{\left(\sum_{i=1}^{n_g} \left(\frac{1}{\pi_{gi}}\right)\right)^2 - \sum_{i=1}^{n_g} \left(\frac{1}{\pi_{gi}}\right)^2} \quad \text{where} \quad \hat{p}_g = \frac{\sum_{i=1}^{n_g} \frac{1}{\pi_{gi}} y_{gi}}{\sum_{i=1}^{n_g} \frac{1}{\pi_{gi}}} \qquad (2)$$

Formula 1 and 2 is based on the assumption that the sample, when drawn, could be looked upon as drawn from the poststrata, with adjustments for age, and a simple random sample within the poststrata. The change in-between the time when the sample is drawn and the time of the survey, e.g moving from one region to another and immigration, makes the estimates biased. Dahmström-Kristiansson [1976] have shown that the relative bias in the LFS-estimates are not higher than 0.03. According to Cochran [1977] the confidence interval is relevant if the relative bias is less than 0.10.

As mentioned (3.1.2) the variation among the inclusion probabilities are very small and if we assume that all persons have the same inclusion probability, $\pi_{g1} = \pi_{g2} = \ldots = \pi_{gn}$ for all g, we get the following simple and well known expression for a total estimate:

$$\hat{Y} = \sum_{g=1}^{300} \frac{N_g}{n_g} \sum_{i=1}^{n_g} y_{gi} \qquad (3)$$

and the corresponding estimated variance:

$$\hat{V}(\hat{Y})=\sum_{g=1}^{300}\frac{N_g^2\,\hat{p}_g(1-\hat{p}_g)}{n_g-1} \qquad \text{where} \quad \hat{p}_g=\frac{\sum_{i=1}^{n_g} y_{gi}}{n_g} \tag{4}$$

In appendix 1 the difference between estimates calculated with equal and unequal inclusion probabilities is shown. From there it may be concluded that there is a very marginal difference between the methods. Also the domains of employed that are presented show that the difference is marginal. Due to practical reasons the result that is refered to henceforth has been calculated under the assumption of equal inclusion probabilities. This assumption will most likely be used in the statistical production with the begining in 1993.

4 The Poststratification with Auxiliary Information from a Register

4.1 The Poststratification with Auxiliary Information from ARE

4.1.1 Procedure

Today auxiliary information from ARE is only used when stratifying and not when estimating, i.e we are not using data from ARE in the present poststratification system. This means that the data from ARE, the way it is used today, does not contribute to any important gain in precision. To be able to obtain more effective estimates, we also need to integrate auxiliary information in the estimation process.

Among many users of LFS there is a demand for better precision for the estimate of employment distributed by industry. With the auxiliary information from ARE the prospects of providing those requirements are good.

To be able to use the data from ARE for poststratificaion, there is a need for change in the present system. The number of observations in the poststrata must not be too small, mainly because of the risk of biased estimates. Cochran [1977] states the minimum amount of observations to be 20. This implies that either of the variables region or age, will have to be omitted, or possibly the classification limits could be widened for those variables.

We will study a poststratification system where we replace region by "employed in a certain group of industry according to ARE". To form suitable basis for combinations of poststrata, different samples from the LFS from 1989 have been matched with the ARE. The aim is that the poststrata should coincide with important domains of study presented in the LFS, and also that the number of persons in each poststrata exceeds 25 respondents. Cochran [1977] states 20 as the minimum limit, here we will

intensify the demands a bit, with regards to the variation that might occur over time. The following schematic classification of characteristics from ARE in combination with sex and age (10 groups 5-years intervalls) is an alternative to the current poststratification system. The characteristics from ARE are grouped according to SNI.

Table 1 *Employed Persons by Group of industry (SNI) According to ARE*

Group 1	SNI 1	agriculture, forestry and fishing etc
	SNI 2	mining and quarrying
	SNI 4	electricity, gas and water
	SNI 5	construction
	SNI 7	transport, storage and communication
Group 2	SNI 3 (exkl 38)	manufacturing
Group 3	SNI 38	manufacture of machinery and equipment
Group 4	SNI 6	trade, hotels and restaurants
Group 5	SNI 8	financing and insurance etc
Group 6	SNI 90,91,93	health and education
Group 7	SNI 92,94-96	other community and social services
Group 8	SNI 0	not employed or the SNI-code is missing

The distribution above in combination with sex and age gives 8x2x10=160 poststrata. The sample adjustment for totals with the new poststratification system, means that the monthly data from RTB must be complemented with data from ARE. We obtain adequate population figures in the current production of statistics by matching the monthly data on sex and age from RTB year *t*, with data on employment status and industry from ARE year *t-2*.

The estimates with the corresponding standard deviations can then be calculated according to formula 3 and 4 with the exception of g=1,2,.....160

4.1.2 The Result

We will in this part make a comparsion between the results with the present poststratification system and the new alternative procedure with auxiliary information from ARE. The result in this example is based on data from LFS in May 1989 and ARE 1987. Here the auxiliary information is 18 months old.

Notation:

$\hat{Y}_o$ = An estimate with the present poststratification

$\hat{\sigma}_{\hat{Y}_0} = \sqrt{\hat{V}(\hat{Y}_0)}$ = Standard deviation for $\hat{Y}_0$

$\hat{Y}_1$ = An estimate with auxiliary information from ARE

$\hat{\sigma}_{\hat{Y}_1} = \sqrt{\hat{V}(\hat{Y}_1)}$ = Standard deviation for $\hat{Y}_1$

$$\hat{\upsilon} = 100\left(1 - \left(\frac{\hat{\sigma}_{\hat{Y}_1}/\hat{Y}_1}{\hat{\sigma}_{\hat{Y}_0}/\hat{Y}_0}\right)\right)$$; Estimated gain in precision (%)

The gain in precision is estimated in terms of coefficients of variation (c.v).

Table 2. *Estimates of Employed Persons by Industry when Using Different Poststratifications. LFS May 1989. Thousands.*

Industry (SNI)	$\hat{Y}_0$	$\hat{Y}_1$	$\hat{\sigma}_{\hat{Y}_0}$	$\hat{\sigma}_{\hat{Y}_1}$	$\hat{Y}_1 - \hat{Y}_0$	$\hat{\upsilon}$
SNI 1	161.0	171.2	7.1	7.2	10.2	**4%**
SNI 2-4	1 019.2	1 001.6	16.0	11.0	-17.6	**30%**
of which						
SNI 38	474.3	469.0	11.8	7.9	-5.3	**32%**
SNI 5	281.7	281.1	9.4	8.5	-0.6	**9%**
SNI 6	633.8	628.6	13.8	10.5	-5.2	**23%**
of which						
SNI 61,62	547.7	541.0	13.0	10.2	-6.7	**21%**
SNI 7	315.5	309.7	10.1	8.9	-5.7	**10%**
SNI 8	388.4	362.0	11.1	7.7	-26.4	**25%**
SNI 9	1 621.2	1 587.2	18.2	12.5	-34.1	**30%**

As seen in table 2 the poststratification with auxiliary information from ARE gives important gain in precision for most of the domains of employed that are being studied, distributed by industry. To retain the same precision with the present poststratification system, for example for estimates of the number of employed in manufacturing (SNI 2 3 4), there would have to be a redoubled increase of the sample size. The corresponding calculations with data where the auxiliary information is 20, 22 and 24 months old show the same gain in precision as the one that are shown in table 2.

Another notable difference between the poststratification methods is the change in level of the estimates. We obtain a lower level of employment when we integrate the data from ARE. The most likely explanation to the difference in level that occurs with the two methods, is that the nonresponse causes different effects. In table 3 estimates of totals for the most important variables in the LFS are shown and we are able to evaluate the influence on the level that occurs when we change the poststratification method.

Table 3. *Estimates of Labour Force Participation of the Population When Using Different Poststratification Systems. LFS May 1989. Thousands*

Variabel	$\hat{Y}_0$	$\hat{Y}_1$	$\hat{\sigma}_{\hat{Y}_0}$	$\hat{\sigma}_{\hat{Y}_1}$	$\hat{Y}_1 - \hat{Y}_0$	$\hat{v}$
Employed	4 418.7	4 345.1	14.5	13.6	**-73.6**	4.9%
at work	3 880.3	3 818.1	17.6	17.0	**-62.2**	2.3%
temp. absent	538.4	527.0	12.7	12.5	**-11.4**	0.0%
Unemployed	50.5	54.7	4.1	4.7	**4.2**	-1.5%
In the labour force	4 469.2	4 399.8	14.1	13.2	**-69.4**	-5.1%
Not in the LF	883.4	952.8	14.1	13.2	**69.4**	13.4%

The gain in precision in the estimate of the total number of employed is not as remarkable as for the employed distributed by industry. However, we can see that the precision is not increasing for the most important variables in the LFS by using the new adjustment for totals, with the exception for the estimate of the number of unemployed. We have other auxiliary information that could be used for estimating the unemployed.

What is most interesting in table 3 is the estimates´ change in level that occurs with the establishment of the new poststratification system. One of the estimation methods does not give unbiased estimates. The number of employed is decreased by 74 000 persons or 1.7%. The level, for the number of unemployed and persons not in the labour force, increases considerably. As mentioned before the reason for this is that the non-response causes different effects. The main difference in methods is that

we replace the variable "region" by "employed in a certain industry according to ARE". When we use region for poststratification, the average nonresponse should have the same characteristics as the respondents, regarding the varaibles in the regions. When poststratifying with information from ARE the same should apply for ARE-characteristics. Intuitively, the latter seems to be more probable considering that the employment situation is being analysed.

Previous nonresponse studies indicate that the level of the estimates with auxiliary information from ARE is more correct than with the present method. Employed persons tend to reply to a higher extent than unemployed, which causes an overestimation in the number of employed when using the present method. It has been shown that LFS overestimates the level of employment with 1-2%. With auxiliary information from ARE we can construct poststrata that are more homogeneous with regards to the employment situation. This is not only effective for the precision of the estimates but also to avoid the nonresponse error when straight adjustment is used within the poststrata.

4.1.3 The Effects on the Regional Presentation

To poststratify with data from ARE for regional domains would mean extreme deterioration in precision for estimates on number of employed. For the county of Stockholm for example the variance increases by 450%! Another effect that the removing of "region" as poststratification variable has, is that the population totals within regions are not consistent with the present data from RTB.

Therefore there are more reasons for taking special measures for the regional domains in LFS. One alternative is to keep the present method only for the regional domains and then to correct regional estimates $(\hat{Y}_{reg})$ with $(\hat{Y}_1/\hat{Y}_0)$ afterwards, i.e a proportional correction for all regions to obtain additivity in the domains. The users of LFS demands additivity in the LFS´ results. Another alternative is to use both data from RTB and from ARE for calibrating on known marginal distributions in a two dimensional frequency table.

It is not yet decided upon what the estimates will look like for the regional domains.

4.2 Poststratification with Auxiliary Information from AMSR

4.2.1 Procedure

Even in this case we will study the possibility of using auxiliary information for constructing poststrata. The data in AMSR that is of interest for LFS is persons with the characteristic AMS(1:1), that is persons without work and that were available to start working. A characteristic which is closely related to LFS:s unemployment definition. This is discussed more in detail in (2.3).

Data from LFS and AMSR for May, October and November 1991 have been collected for co-ordination. In contrast to ARE, we have the possibility of using current data from ARE. When constructing poststrata we cannot use a classification as far-reaching as when we are using the present system or information from ARE. Many poststrata would have too few persons. At the co-ordination of LFS and AMSR there were ca 500-550 out of 18 000 persons in the sample that matched. The number varied due to chance and also due to the state of the market.

The following distribution has been found to be suitable when estimating the number of unemployed:

sex
age groups (16-24, 25-34, 35-44, 45-64)
AMS(1:1) / not AMS(1:1)

With this classification we get 16 (2x4x2) poststrata. Estimates and mean errors can be calculated according to formula 3 and 4 with the exception of:

$g = 1,2,...,16.$

where N_g = the number of people in the population that belongs to poststrata g according to AMSR or RTB-AMSR.

4.2.2 The Result

The result that is presented in table 4 is a comparison analogous to the one in part 4.1.2. Estimates and their characteristics, calculated with the present poststratification method and the new method with auxiliary information from AMSR, are being compared.

Notation:

$\hat{Y}_o$ = Estimate with the present poststratification

$\hat{\sigma}_{\hat{Y}_0} = \sqrt{\hat{V}(\hat{Y}_0)}$ = Estimated standard deviation of $\hat{Y}_0$

$\hat{Y}_1$ = Estimate with auxiliary information from the AMS

$\hat{\sigma}_{\hat{Y}_1} = \sqrt{\hat{V}(\hat{Y}_1)}$ = Estimated standard deviation of $\hat{Y}_1$

$$\hat{\upsilon} = 100\left(1 - \left(\frac{\hat{\sigma}_{\hat{Y}_1}/\hat{Y}_1}{\hat{\sigma}_{\hat{Y}_0}/\hat{Y}_0}\right)\right) = \text{Estimated gain in precision (\%)}$$

Tabell 4. *The Estimates of the Total Number of Unemployed when Using Different Poststratification Methods. LFS. Thousands*

Period	$\hat{Y}_0$	$\hat{Y}_1$	$\hat{\sigma}_{\hat{Y}_0}$	$\hat{\sigma}_{\hat{Y}_1}$	$\hat{Y}_1 - \hat{Y}_0$	$\hat{\upsilon}$
May 1991	92.9	92.3	5.7	4.8	-0.6	**16%**
Oct 1991	140.4	149.1	6.8	5.9	8.7	**18%**
Nov 1991	141.3	146.9	6.9	6.0	5.6	**16%**

As seen in table 4 the use of auxiliary information from AMSR means that we can improve the precision by 16-18%. To be able to obtain the same precision, by only increasing the sample, the sample size would have to be changed from the present 18 000 to 25 000 persons, i.e an increase by 7 000 persons per month.

Exactly as when poststratifying with auxiliary information from ARE, we here obtain some notable changes in level for the estimates of the number of unemployed. For October the level of the estimates increases by 8 700 persons or 6% when using auxiliary information. For May there is not a difference of any note in the level of estimates for the different methods. Also in this case the change in level can be explained by the different effects that different poststratification methods have. This is being analysed more in detail in part 4.2.3.

4.2.3 Random Error and Nonresponse Error when Estimating the Number of Unemployed

In part 4.2.2 we saw that the level of the estimate increased considerably in two out of three reference periods and at one point (May 1991) we had the same estimation level for the different poststratification systems. We will try to find the explanation to this phenomenon with help from AMSR. Though AMS(1:1) is a characteristic that is closely related to the definition of unemployment as defined in the LFS, it is of interest to estimate the number of AMS(1:1) based on the LFS sample. Those estimates gives us a rough idea of the effects that the nonresponse have on the estimates of the number of unemployed.

In the present data we have the possibility of estimating the number of AMS (1:1) in the total population, i.e we also have the observational values for the characteristic AMS (1:1) for the persons that are nonrespondents in the LFS. In table 5 the following is presented:

T = The number of AMS(1:1) according to a total count in AMSR, i.e the true value.

$\hat{T}_n$ = The estimate of the number of AMS(1:1) based on the total sample.

$\hat{T}_{n_s}$ = The estimate of the number of AMS(1:1) according to the respondents in the LFS

$\beta = \dfrac{\hat{T}_{n_s} - \hat{T}_n}{\hat{T}_n}$ = The nonresponse error in percentage of $\hat{T}_n$

Table 5 *The Number of Persons with the Characteristic AMS(1:1) According to LFS or AMS Total Count. Thousands.*

Period	$\hat{T}_n$	$\hat{T}_{n_s}$	T	$\hat{T}_{n_s} - \hat{T}_n$	β
May 1991	115.2	106.3	105.5	- 8.9	**7.7%**
Oct 1991	150.8	139.1	156.0	-11.7	**7.8%**
Nov 1991	159.3	147.1	158.0	-12.2	**7.7%**

$\hat{T}_{n_s} - \hat{T}_n$ and β gives a measurement on the bias in the estimates of the number of AMS(1:1) that are caused by the nonresponse. As seen in table 5 the nonresponse underestimates T by 7-8%. The calculations for the different points in time indicate very convincingly the same result. Since AMS(1:1) and unemployed according to LFS are closely related, the results are clear indications of the degree of nonresponse bias in the LFS with the present estimation system.

In table 4 we could establish the fact that the level of estimates for unemployed for October and November increased when we used auxiliary information from AMSR. For May there was no difference in the level of estimates between the different methods. This is due to the fact that the proportion of AMS(1:1) were larger in the sample, than in the corresponding proportion in the population in May, i.e it can be accounted for by the sampling error.

In an analogous way as we created the measurement for the nonresponse error for the *T*-estimate we can construct a measurement for the sampling error. We are able to calculate the sampling error, which we usually denote with the confidence interval, exactly since the true value is known. We can then compare the sampling error with the nonresponse error for AMS(1:1) with the difference in level $\hat{Y}_1 - \hat{Y}_0$. The exact sampling error is given by $\hat{T}_n - T$.

Table 6 *The Nonresponse and Sampling Error for the AMS(1:1) Estimates in Comparison to the Difference in Level of the Unemployment Estimates According to LFS with Different Poststratification Methods. Thousands.*

period	sample error $\hat{T}_n - T$	nonresponse error $\hat{T}_{n_s} - \hat{T}_n$	sample error+ nonresponse error $(\hat{T}_n - T) + (\hat{T}_{n_s} - \hat{T}_n)$	difference in level $\hat{Y}_1 - \hat{Y}_0$
May 1991	9.7	- 8.9	0.8	-0.6
Oct 1991	-5.2	-11.7	-16.9	8.7
Nov 1991	1.3	-12.2	-10.3	5.6

The result for May indicates that the nonresponse error and the sampling error cancel each other out, this should explain why we do not get a remarkable difference in level of estimate of the number of unemployed in the LFS at the same point in time. For October the sampling error and the nonresponse error are going in the same direction, which could explain that difference between the estimate levels is largest between the points in time when we estimate the number of unemployed.

The result in table 4 and 5 indicates that the information from AMSR is not only effective auxiliary information for reducing the sampling error but also very effective for reducing the nonresponse error in the estimates of the number of unemployed in the LFS.

4.3 A New Estimation System

On the assumption that we can use the information from AMSR in the current production of statistics, two different types of auxiliary information will be integrated in the estimation process, data from ARE and AMSR. AMSR data will, as has been shown earlier, be used for estimating the total number of unemployed and their domains. Information from ARE is used for estimating different employment variables.

Due to additivity reasons the estimates of the number of persons "not in the labour force" or a domain of this category will be an estimate of difference constructed from components from the two different poststratification methods that has been described earlier . In this context it is important to make sure that there is no risk for this "residual" to be less than 0 in any domain.

To give a clear description of the new complete estimation system we assume that the total number of persons not in the labour force is formed from an estimate as a difference estimate. From a more global aspect of the most important variables we then get the following estimation process.

1) With auxiliary information from AMSR we estimate the number of unemployed, $(\hat{Y}_{LF1})$

2) With auxiliary information from the ARE we estimate the number of employed, $(\hat{Y}_{LF2})$

3) The number of persons in the labour force, $(\hat{Y}_{LF3} = \hat{Y}_{LF1} + \hat{Y}_{LF2})$

4) The number of persons not in the labour force,

$(\hat{Y}_{LF4} = Population - \hat{Y}_{LF1} - \hat{Y}_{LF2})$

The system guarantees consistency if the $Population \geq (\hat{Y}_{LF1} + \hat{Y}_{LF2})$ for all groups that are presented.

The new system implies that the estimates of the number of persons not in the labour force (estimates of difference) get a larger variance in comparison to the corresponding estimate with the present system. But naturally, the estimates of the number of unemployed and employed have the highest priority in the LFS, and those estimates will be, as established, considerably more reliable with the new estimation process.

5 Conclusions

There are very good reasons for using information from both AMSR and ARE, as auxiliary information in the estimation process in the LFS. When we introduce auxiliary information for constructing poststrata, the gain in precision is estimated to be 16-18% when estimating the number of unemployed. When estimating the number of employed disributed by industry, for some domains we gain 25-30% in precision when using auxiliary information from ARE. Also the estimates of the total number of employed becomes more effective.

Futher it can be considered established that poststratification with auxiliary information from ARE and AMSR is an effective method for reducing serious nonresponse error in the LFS estimates. The estimates get a more correct level in comparison with the present system, when we integrate register information in the estimation. Nonresponse studies in the LFS show a bias that coincide with the difference in level that occurs between the present and the new poststratification system. Here we have a good and relatively simple method for overcoming one of the most serious causes of error in sample surveys, i.e the nonresponse.

Statistics Sweden is planning to integrate the auxiliary information from AMSR and ARE in the estimation beginnig in 1993.

6 References

Cochran, W.G. [1977]: Sampling Techniques, Third edition, Wiley, NY

Dahmström, P. and Hörngren, J. [1992]: The Labour Market Board Register as Auxiliary Information in the Swedish Labour Force Survey, (in Swedish), Statistics Sweden.

Dahmström, P. and Kristiansson, K-E. [1976]: Sampling and Estimation in the Swedish LFS, (in Swedish), Statistics Sweden.

Hörngren, J [1992]: Two Nonresponse Studies in the Swedish Labour Force Survey, (in Swedish), Statistics Sweden.

Larsson, M [1992]: A Nonresponse Study in the Swedish Labour Force Survey with the Statistics Swedens Annual Register of Employed, (in Swedish), Statistics Sweden.

Appendix 1.

LFS Estimates with Corresponding Estimated Standard Deviations Computed with Equal and Unequal Inclusion Probabilities (π) May 1989. Thousands.

Variable	***Estimates***			***Standard deviations***		
	equal π *a*	*unequal π* *b*	*ratio* *a/b*	*equal π* *c*	*unequal π* *d*	*ratio* *c/d*
Employed	4 418.7	4 421.9	0.999	14.5	14.5	0.999
Unemployed	50.5	50.2	1.006	4.1	4.1	1.002
In the Labour force	4 469.2	4 472.1	0.999	14.1	14.0	1.008
Not in the Labour force	883.4	880.6	1.003	14.1	14.0	1.008
Employed Persons by Industry						
Agriculture, fishing etc	161.0	163.9	0.982	7.1	6.9	1.025
Manufacturing, mining	1 019.2	1 037.3	0.983	16.0	15.8	1.010
Construction	281.7	278.2	1.013	9.4	9.4	0.998
Trade	547.7	537.5	1.019	13.0	13.0	1.003
Financing, insurance	388.4	379.2	1.024	11.1	10.7	1.039
community, soc. services	1 621.2	1 629.2	0.995	18.2	18.2	1.002

Acknowledgements:

This paper was presented at the International Workshop on Uses of Auxiliary Information in Surveys, in Örebro, Sweden, October 1992.

I would like to thank Per Dahmström, University of Stockholm; Dep. of Statistics, who participated in the development of this topic. I am also grateful to K-E Kristiansson, Statistics Sweden, for valuable comments during this work, and Lilli Japec, Statistics Sweden, for helping me with the English.

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

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

Reports published during 1992:

1992:1 (grön)	Industrins konkurrenskraft och produktivitet i fokus - en utvärdering av statistiken **(Margareta Ringquist)**
1992:2 (grön)	Automated Coding of Survey Responses: An International Review **(Lars Lyberg and Pat Dean)**
1992:3 (grön)	TABELLER ,... TABELLER ,... TABELLER ,... - Variation och Förnyelse **(Per Nilsson)**
1992:4 (grön)	Basurval vid SCB? Studier av reskostnadseffekter vid övergång till basurval **(Elisabet Berglund)**
1992:5 (beige)	Abstracts I - sammanfattning av metodrapporter från SCB
1992:6 (grön)	Utvärdering av framskrivningsförfarande för UVAV-statistik **(Kerstin Forssén & Bengt Rosén)**
1992:7 (grön)	Cross-Classified Sampling for the Consumer Price Index **(Esbjörn Ohlsson)**
1992:8 (grön)	Bortfallsbarometern nr 7 **(Mats Bergdahl, Pär Brundell, Anders Lindberg, Håkan Lindén, Peter Lundquist, Monica Rennermalm)**
1992:9 (beige)	Abstracts II - sammanfattning av metodrapporter från SCB
1992:10 (gul)	Organizing the Metainformation Systems of a Statistical Office **(Bo Sundgren)**
1992:11 (grön)	CLAN - ett SAS-program för skattningar av medelfel **(Claes Andersson, Lennart Nordberg)**
1992:12 (grön)	KVALITETSRAPPORTEN - Utveckling av kvaliteten för SCBs statistikproduktion **(Jan Eklöf, Per Nilsson)**

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Inga-Lill Pettersson, U/LEDN, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 49 56.

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