Research and Development – Methodology reports from Statistics Sweden

Statistiska centralbyrån Statistics Sweden

Linking a system of time series

Recalculation of Labour Force Surveys 1987–1992

Britt Wallgren
Anders Wallgren

The series entitled "**Research and Development** – **Methodology Reports from Statistics Sweden**" presents results from research activities within Statistics Sweden. The focus of the series is on development of methods and techniques for statistics production. Contributions from all departments of Statistics Sweden are published and papers can deal with a wide variety of methodological issues.

Previous publication:

2006:1 Quantifying the quality of macroeconomic variables

2006:2 Stochastic population projections for Sweden

2007:1 Jämförelse av röjanderiskmått för tabeller

2007:2 Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator.

2007:3 Kartläggning av felkällor för bättre aktualitet

2008:1 Optimalt antal kontaktförsök i en telefonundersökning

2009:1 Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias

2009:2 Demographic methods for the statistical office

2011:1 Three Factors to Signal Nonresponse Bias – With applications to Categorical Auxiliary Variables

2011:2 Quality Assessment of Administrative Data

2012:1 Aspects of Responsive Design for the Swedish Living Conditions Survey

Linking a system of time series

Recalculation of Labour Force Surveys

1987–1992

Britt Wallgren
Anders Wallgren

Statistiska centralbyrån
2012

Linking a system of time series

Recalculation of Labour Force Surveys 1987–1992

Statistics Sweden
2012

Producer Statistics Sweden, Research and Development Department
SE-701 89 ÖREBRO
+ 46 19 17 60 00

Enquiries Hassan Mirza, +46 8 506 946 72
Hassan.mirza@scb.se

Elisabet Andersson, +46 8 506 946 45
elisabet.andersson@scb.se

Cover Ateljén, SCB

ISSN 1653-7149 (Online)
URN:NBN:SE:SCB-2012-X103BR1202_pdf

This publication is only published electronically on Statistics Sweden's website www.scb.se

Preface to the English report

Improvements in statistical products often lead to estimation shifts and can therefore make comparability over time more difficult. Understandably, there are strong calls from statistics users for such time series breaks to be linked. Regarding the Statistics Sweden Labour Force Surveys, it was obvious that we should try to link the series to rectify the breaks that occurred.

The Swedish Labour Force Surveys (LFS) have undergone several major changes during their 50-year history creating the need to link time series. Such an occasion was at the end of 1992/beginning of 1993, and another was between the first and second quarters of 2005. The report on the 1992/1993 linking has been published in Swedish and the report on the 2005 linking is due to be published in English in the near future.

The authors of this report, Anders and Britt Wallgren, took on the task of linking the LFS series on both these occasions. The method used to link time series can differ from one case to the next, depending on the causes of the time series breaks. The work done by Anders and Britt on the 1992/1993 linking does however contribute to general linking methodology, which can also be of use in other contexts. Their work contains two new approaches of a general nature:

i) linking a set of time series where good consistency is called for
ii) linking a system of series by performing the linking on the micro level using calibration.

These contributions from 1992/1993 have also been used in connection with the 2005 linking. For these reasons, the report (Background Facts regarding the Labour and Education Statistics 1998:2) into the 1992/1993 linking is being published in English.

Statistics Sweden, November 2012

Inger Eklund

Lilli Japec

Preface to the Swedish report

Statistics Sweden has conducted Labour Force Surveys (LFS) every month since the beginning of 1970. Up until the mid-1980s, the surveys were performed using basically the same definitions and interview form. As a result of major changes in working life during this period, it became increasingly difficult to capture new phenomena on the labour market.

A comprehensive overhaul of LFS was therefore performed in 1987 involving changes to both content and definitions. This overhaul caused a time series break. Due to the fact that data were collected during 1986 pertaining to both the old and the new LFS, linkage between 1986 and 1987 could be achieved.

Further major changes were carried out in LFS in 1993. A new survey week system was introduced and the estimation procedure was altered. In addition, the definition of 'unemployed' was adapted to bring it more into line with the ILO's international recommendations. The alterations in 1993 led to time series breaks, which manifested themselves both as changes in levels and as modified seasonal patterns.

In order to be able to make comparisons both prior to and after 1993, a great deal of work has been done both by LFS department staff and by Britt Wallgren and Anders Wallgren, who developed the method of linking a system of time series used for the recalculation of the Labour Force Surveys 1987-1992.

All LFS basic tables 1987-1992 have been recalculated using the method described in this report. The recalculated series are not comparable with the corresponding series prior to 1987.

This report describes the various steps in the linking process, consisting of linking work on the time series level, calibration of weighting factors on the individual level and the evaluation of linked series. The appendix also presents the linking done as a result of the new industrial classification in LFS in 1995.

The report has been compiled by Britt Wallgren U/STM and Anders Wallgren U/STM. Hassan Mirza and others at AM/AKU have contributed valuable comments. The authors have discussed the calibration methodology with Professor Carl Erik Särndal at Montreal University, who presented this calibration application at the 1995 annual meeting of the Finnish Statistical Society.

Statistics Sweden, February 1997

Department of Labour and Education Statistics

Berndt Öhman

Contents

Background and aim

Between December 1992 and January 1993, the Swedish Labour Force Surveys (LFS) underwent major changes regarding *definitions, estimation methods* and *measurement periods*. The aim was to improve the quality of the surveys. As a result of these alterations, time series breaks occurred manifested both as changes in levels as well as changed seasonal patterns.

In order to increase comparability with old data, the monthly LFS for the period 1987-1992 have been recalculated. Greater use can therefore be made of data collected at a considerable cost. Between 1986 and 1987, the RIDA changeover was carried out[1], resulting in level discrepancies in many series. The converted series are not comparable with the corresponding series prior to 1987, however.

About 30 000 time series are reported in the LFS system every month. All these series have been recalculated maintaining internal consistency and we have developed new methodology for this purpose. The aim of the recalculations is both to improve the comparability and quality of old data and to facilitate seasonal adjustment of the new series. To be able to seasonally adjust a time series, we need about 5 years of monthly data – without the conversion of old values, we would have to wait under 1998 before being able to seasonally adjust the new series.

The new industrial classification SNI92 has been used in LFS since the beginning of 1995. A description of the SNI-linking for the period 1987-1994 can be found in the appendix.

---

[1] Review of content and definitions in the Labour Force Surveys, see Statistics Sweden (1988)

Overview of the method

The new method for linking time series systems consists of two parts and we have also done extra correction for holiday absence:

1) On the *time series level*, about one hundred key series are linked and the linkages are done so that the key series remain consistent. For each of these series, we then obtain 72 recalculated monthly values for the period 1987-1992 which are comparable to corresponding new series from 1993 onwards.
2) On the *individual level*, the weighting factor of every respondent is recomputed using the recalculated key series as calibration conditions. This means that the weighting factors for each of the 72 surveys from 1987 to 1992 are calibrated. As a result, all 30 000 series are linked with preserved internal consistency.
3) Prior to 1993, absence as a result of public holidays was not studied. Since there wasn't anyone with holiday absence previously, calibration cannot solve the problem – there are no categories that can be given increased weight so that data will contain people with holiday absence. We have therefore created synthetic holiday absence for some people.

Consistency in the linked series means for example that *the number of employed persons + the number of unemployed persons + the number of persons not in the labour force = the number of persons in the population* and this is true for all sub-divisions by age and sex.

About 1/350 of the population are interviewed each month as part of LFS. When population values are estimated, the values of each individual will be multiplied by his/her weighting factor, which is in the region of 350. The individual weights vary between different strata.

Assume that we have established from the time series work in Point 1 above that the number of employed persons in a particular month is to be reduced by 2 percent. Since the population is to remain unchanged, the number of persons that are not employed must increase, we assume by 8 percent. Explained in very simple terms, we can imagine the calibration being done in the following way, illustrated using data for two respondents:

Personal ID number	Labour force status	Original weight	New calibrated weight
1	employed	350	0.98 • 350 = 343
2	not employed	360	1.08 • 360 = 389

By calibrating on the individual level, all time series produced using the new weights will be consistent. The key series we have used will continue to be the same as those calculated under Point 1. All other time series that can be formed will be linked in concordance with the key series. The natural relationships between the variables on the microdata level govern how these other series are linked.

Step 1: Linking of key series

A total of more than 100 time series have been selected and linked. This linking work has been done in several steps. We have made changes based on new definitions and also corrected for non-response error and periodization error. Finally, we have made the corrections needed due to the fact that the entire month is now measured. Up to and including December 1992, only one or two weeks out of every month were studied and the weeks selected did not contain any public holidays. As from January 1993, all weeks of the year are measured.

Recalculation based on new definitions

From 1993 onwards, the definition of unemployment was changed in LFS to bring it more into line with the ILO[2] recommendation. Those who sought work 1-4 weeks prior to the survey week and received a negative response were previously counted as *not in the labour force* but are now counted as *unemployed.* Full-time students are still counted as not in the labour force.

Using the questions included in LFS during the period 1987-1992, we can classify the respondents according to the new definition. This reclassification has been performed for all LFS data sets from this period, i.e. the definition change has been applied retroactively as from January 1987 in the data sets that form the basis of the linked time series.

Selection of key series as calibration conditions

We have chosen key series based on two criteria. Firstly, *important main series* must be included, i.e. series that are often divided into sub-groups. The main series have to be well linked for the sub-groups to be similarly well linked. Secondly, *series that are changed substantially* must be included, i.e. series affected by public holidays.

The key series we chose are as follows, as shown in Tables 1A, 2A, 3, 19 and 20 in Basic tables LFS. For each combination of sex and seven age groups[3], series were linked with regard to number of employed persons by degree of attachment[4], persons at work, absent the entire week, unemployed and not in the labour force.

Furthermore, we also chose to link the number of persons at work and hours worked broken down by 21 industries[5]. Two series were also linked with regard to vacation (number of absent persons the entire week and part of the week respectively).

---

[2] International Labour Organization

[3] Age groups: 16-19, 20-24, 25-34, 35-44, 45-54, 55-59 and 60-64 years

[4] Degree of attachment: Permanently employed, temporarily employed and self-employed/unpaid family worker

[5] SNI69: 11, 1 apart from 11, 31, 32, 33, 341, 35, 37, 38, 2-4 apart from these, 5, 61, 62, 63, 7, 8, 91, 931, 933, 934 and 9 apart from these

Correction for non-response error

Better estimation methods have been used in LFS since January 1993. The use of register data can reduce sampling error and better correction for non-response error can be done. The new estimation method is described by Hörngren (1992). As a result of the new better correction for non-response, the employment level for 1992 is reduced by about 1 percent and the number of unemployed persons increases by about 5 percent.

The new and old estimation methods have been compared for all series in LFS for 1992. The estimated level discrepancies are presented in SCB (1993), where the given multiplicative coefficients for unemployed persons and persons not in the labour force also cover the definition change. Since we have already corrected data on the individual level for the definition change, other coefficients that only cover the new estimation procedure are therefore used.

During the period 1987-92, non-response has increased from 11 to 14 percent. We assume that the non-response error depends on the size of the non-response and therefore correct the level of all key series using a function of the size of the non-response. If, for example, the level discrepancy in 1992 is 5 percent for a certain series, we have corrected the month t by $5 \bullet nonresponse_t / 0.14$ ($nonresponse_t$ is a moving average of actual non-response for month t).

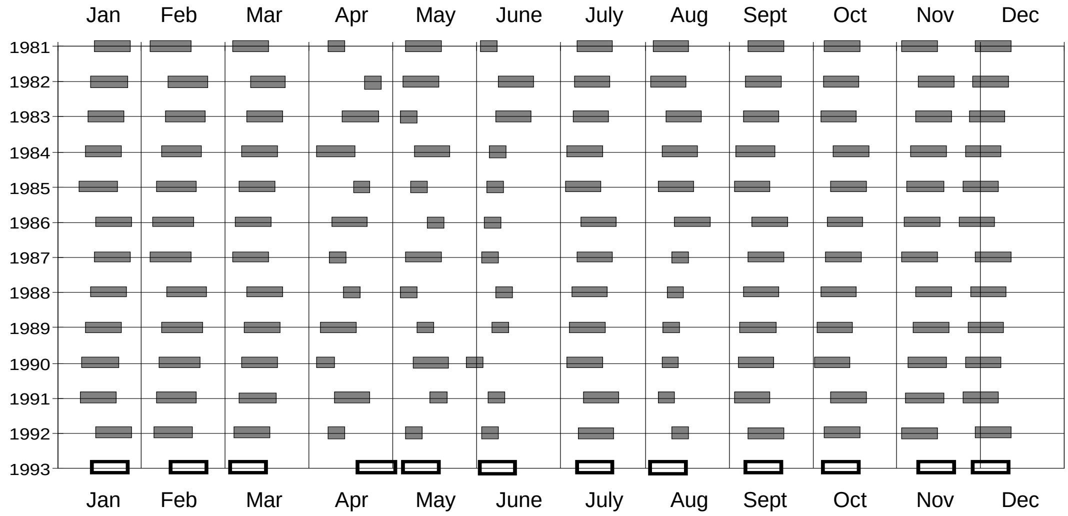
Measurement periods and time series quality

Up to and including 1992, the measurement period for a monthly survey consisted of one or two adjacent reference weeks. These weeks have been chosen because they have no public holidays in them. This way of choosing measurement periods is common for labour force surveys in many countries. Due to the fact that week divisions and public holidays vary for the same month from one year to the next, the positioning of the measurement period has also varied.

As from 1993, all weeks of the year have been measured. The monthly surveys are now based on the four or five-week periods that tally most closely with each month respectively.

The diagram below shows how the measurement periods have varied. The shaded rectangles depict measurement periods. We can see that the positioning of the measurement period has varied considerably and that it has occurred early in each month respectively. Measurements were performed for entire months in 1993, but measurements relating to the old method of two-week periods have been processed separately to facilitate comparability with earlier surveys and will be used in the section *Correction for holiday effects* later on.

Chart 1. Measurement periods in LFS, 1981 - 1993



The significance of the measurement period for the equidistance
An equidistant time series is one where the time between the mid-points of the various measurement periods is of the same length. As is clear from Chart 1, LFS time series were not equidistant prior to 1993. Employing the previous method of selecting measurement periods, the time between the mid-points of the periods could vary between ½ to 1½ months, making the extent of the changes difficult to interpret. All time series analysis requires equidistant series. The measurement period corrections we present in the next section involve changing over to measurement periods that are in the middle of each month.

The significance of the measurement periods for the seasonal pattern
The old way of selecting measurement periods destroyed the seasonal pattern, especially for series such as hours worked, number of persons at work and absence in hours or persons. The seasonal pattern for number of employed persons is also affected by moving measurement periods.

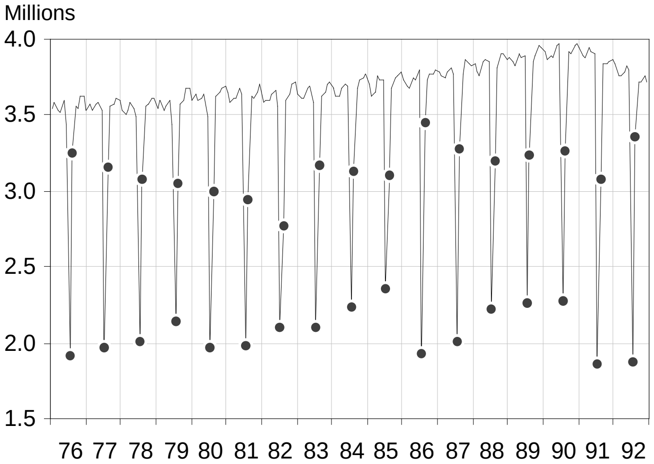
Chart 2 below shows the published original series of the number of persons at work. The July and August values have been marked. The series of July values (bottom) has three clear interruptions in the seasonal pattern: 79/80, 85/86 and 90/91.

Chart 3 below provides the explanation – the measurement period has been moved from early in July (fewer people on vacation) to later in July (more people on vacation), which means that the number of persons at work decreases. The series of August values also includes sizeable breaks in the seasonal pattern, which are explained by the fact that the measurement period was abruptly changed.

The ideal situation had been for the measurement periods to always be positioned the same in the middle of each month for every year. Variation in the positioning of the measurement period leads to periodization error:

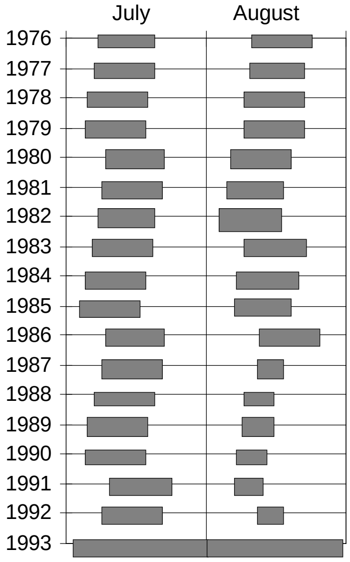
Periodization error = value of the series during the actual measurement period - value of the series during an ideal measurement period

Chart 2
Number persons at work, uncorrected series



Note: The RIDA break 86/87 is about 18 000 persons for this series, a negligible amount in the context.

Chart 3
Measurement periods in LFS



Correction for periodization error

Modern time series analysis[6] software includes methods for corrections of different types of disturbances. A model for the relationship between the disturbing factor and the time series is firstly estimated and the series is then corrected using this estimated relationship. Such corrections must be done prior to seasonal adjustment.

To be able to correct for periodization error, we have estimated the relationship between where the mid-point of the measurement period is in relation to the mid-point of the month and the value of each time series respectively (different *transfer function models* for different series in LFS). With the help of the estimated relationships, each series is then recalculated to correspond to a series in which the measurement period has always been in the middle of the month. This gives us series with stable seasonal patterns.

To estimate e.g. the relationship between the positioning of the measurement period *(x)* and the number of persons at work *(y)*, a function with the following format is estimated; see Liu, Hudak et al. (1992):

$$y_t = \beta_1 \bullet x_t \bullet d_{1t} + \beta_2 \bullet x_t \bullet d_{2t} + \ldots + \beta_{12} \bullet x_t \bullet d_{12t} + noise$$

where d_{1t} is 1 for January, 0 otherwise, d_{2t} to d_{12t} are the equivalent for February to December. The coefficient estimates β_1 to β_{12} have been made using *transfer function analysis*, a method of estimating relationships between time series, where the noise follows an ARIMA model. The noise then shows stochastic trends and seasonal patterns not caused by the model's explanatory variables. The estimated model for the number of persons at work was, after conventional model checking according to Box and Jenkins (1970):

---

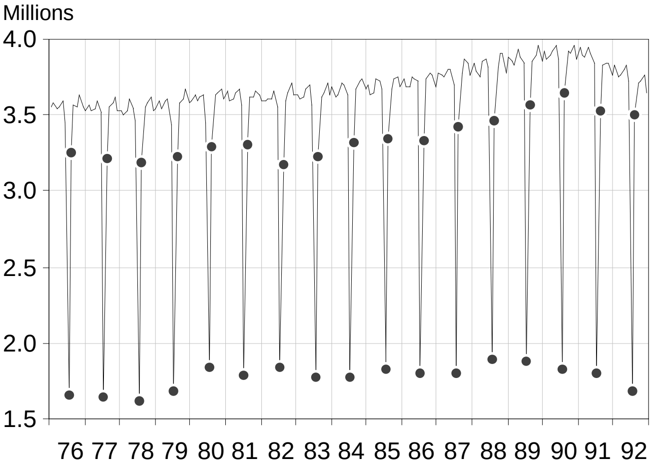
[6] The Bureau of the Census uses X-12-ARIMA that contains REGARIMA for model-based pre-corrections. Eurostat uses TRAMO/SEATS where the TRAMO part contains similar methods. The SCA software package has contained these methods for a long time. The SCA software package has been used for the analyses that we have performed.

Coefficient	Month	Estimated coefficient 1000s of persons	t-ratio	Residual standard deviation
β_4	April	137	2.8	*for this model:* 36 thousand persons
β_6	June	-208	-3.2	
β_7	July	-2016	-17.2	*for an ARIMA-model without explanatory*
β_8	August	1824	18.1	*variables*: 86 thousand persons
β_{12}	December	-228	-1.8	
noise	**ARIMA**	**0 1 1 0 1 1**		*standard error for the estimate of*
θ_1		0.69	11.1	*a monthly value:* 18 thousand persons
Θ_{12}		0.34	3.9	

The estimated model is used to correct for periodization error. The table below shows how data for certain months in 1990 are corrected. The variable x_t is the distance between the mid-date of the measurement period and the 15th of the month, expressed in months. Corrected y_t corresponds to $x_t = 0$, i.e. a measurement period that is in the middle of the month.

Year	*Month*	x_t	$x_t \bullet d_{4t}$	$x_t \bullet d_{6t}$	$x_t \bullet d_{7t}$	$x_t \bullet d_{8t}$	$x_t \bullet d_{12t}$	y_t		y_t *corrected*
90	4	-0.33	-0 33	0.00	0 00	0 00	0 00	3873	-137• (-0.33)=	3919
90	5	-0.06	0 00	0 00	0 00	0 00	0 00	3955		3955
90	6	-0 53	0 00	-0 53	0 00	0 00	0 00	3980		3869
90	7	-0 23	0 00	0 00	-0 23	0 00	0 00	2281		1826
90	8	-0 21	0 00	0 00	0 00	-0 21	0 00	3265		3647
90	12	-0 45	0 00	0 00	0 00	0 00	-0 45	3975		3872

Chart 4
Number of persons at work, corrected series



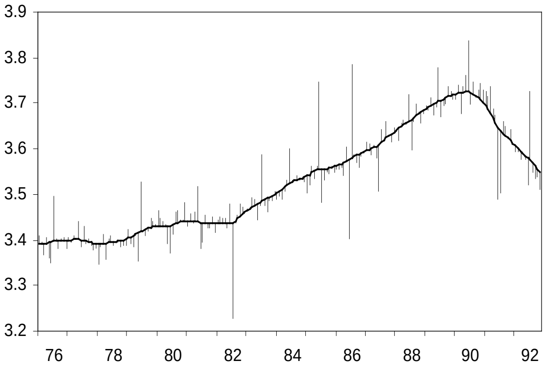
The correction reduces time series noise, from 86 to 36 thousand per month. The interruptions in the seasonal pattern have disappeared, making seasonal adjustment a meaningful exercise. The values for July are on a lower level and on a higher level for August compared to an uncorrected series, as the values are measured early in the month – more people are on vacation in the middle of July compared to the beginning of July and the converse is true for August.

The way of correcting for periodization error follows international practice and the model above is rather like a traditional calendar adjustment model. See references in Footnote 6 above.

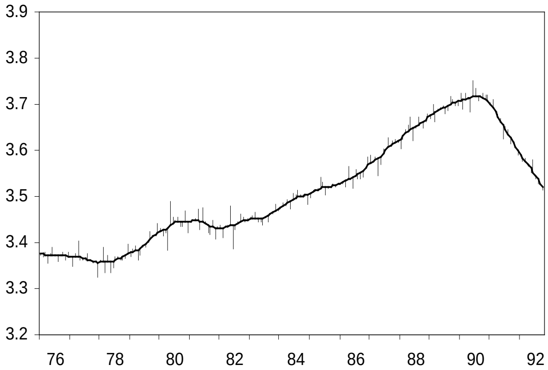
Correcting for periodization error increases the quality of the seasonally adjusted series, as is illustrated in Charts 5 and 6. The estimated trend is also affected; the economic boom around 1980 is only visible in Chart 6 with the corrected series.

Charts 5 and 6
Number of persons at work – seasonally adjusted series and estimated trend

5. Uncorrected series
Millions



6.Series corrected for measurement period
Millions



Note: The SEATS software package has been used for seasonal adjustment and trend estimation

Time series with annual data are also affected by the positioning of the measurement periods, which can be seen in the diagram below.

Chart 7. Number of persons at work
Millions, annual mean values

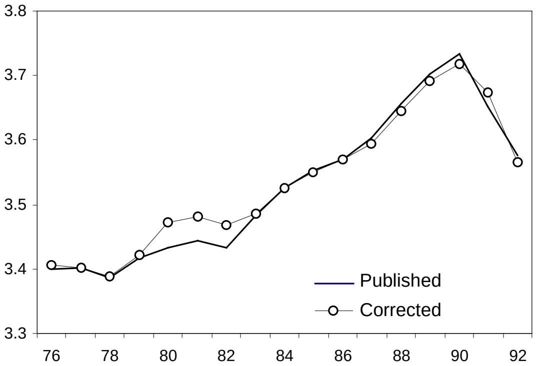
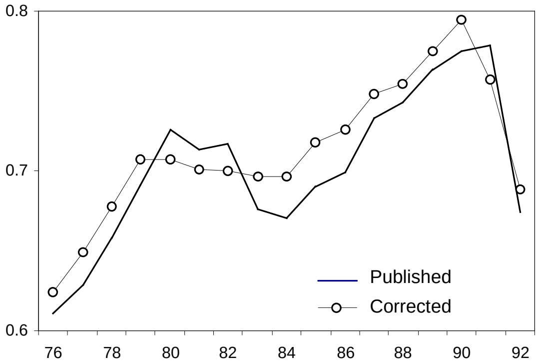


Chart 8. Number of absent persons
Millions, annual mean values



For these series, Charts 7 and 8 show that the periodization error is a very serious source of error, as even the annual data provide an entirely inaccurate picture of the trend. The fact that periodization error must not be ignored is an important conclusion to be drawn from this example. We discovered the magnitude of this error in LFS during the linking work. Several important and advanced users had previously misinterpreted the trend – Chart 8 is shown to have a shocking content due to the fact that the trend in the corrected series is so different to what was believed before.

The linked time series that have now become available cover the period 1987-1992 and these series have been corrected for periodization error. We recommend that data prior to 1987 also be corrected, if series from this time period are to be of good use. It is necessary, however, to take the time series break in 1986/87, the missing value for April 1980 and the extreme value for May 1980 into account in connection with the labour market conflict. We have considered these problems in the transfer function model above.

Correction for periodization error in a system of series
After having first corrected individual data for definition changes and corrected the series for non-response error, we examined how a large number of series are affected by the positioning of the measurement period. For employed persons who are temporarily employed (especially young

people), for persons at work, absent persons and persons not in the labour force, the relationships were so strong that we felt that correction was a suitable course of action. Series for number of unemployed persons have not been corrected, as the relationships were insignificant. We have adapted different models for women and men by age (7 age groups) using data for the period 1987-1993[7]. For persons at work and hours worked respectively, we have adapted different models for each individual industry.

The estimated β-coefficients in all models for the number of persons do not exhibit total internal consistency as a result of estimation errors. Since the measurement period-corrected values must also fulfil consistency requirements such as:

number of persons at work + absent persons = number of employed persons and

number of persons in the labour force + number of persons not in the labour force = the population,

then the β-coefficients must fulfil the following conditions:

For a certain month (e.g. August) β for employed persons should be equal to the sum of β for persons at work and β for number of absent persons. Furthermore, β for persons in the labour force should be equal to $-\beta$ for persons not in the labour force, since the number in the population must remain unchanged. Furthermore, the sum of the 14 β-coefficients for the number of persons at work by age and sex should be equal to the β-coefficient in the model for both sexes and all ages. Finally, the sum of the 21 β-coefficients for the various industries should also tally with the β-coefficient for the total number of persons at work.

We have therefore adjusted all estimated β-coefficients so that they are consistent. We have done this by using the generalized least squares method where the consistency terms give the x-variables and the variances of the estimated coefficients give the weights. We do this partly to improve the estimated coefficients and partly to maintain consistency in the corrected series. The linked series in Step 1 must be completely consistent in order for them to be used as calibration conditions in Step 2 of the linking work.

A brief example based on real data can illustrate the working method. From a larger system of estimated equations, we show the equations for number of employed persons, number of persons at work and number of persons absent from work (1000s).

[7] For 1993, we have used the results of the two-weekly measurements according to the old system.

We have obtained the following estimated β-coefficients:

	estimated β-coefficient	s(β-estimate)	estimate corrected for consistency
Employed	-191	37	-190.68
At work	1824	101	1821.64
Absent	-2010	100	-2012.32

Since 1824 - 2010 = -186 and not -191, these estimates must be corrected. If the x-variables $x_1 = 1$ for at work, 0 otherwise and $x_2 = 1$ for absent, 0 otherwise, the following applies:

$$-191 = b_1 * x_1 + b_2 * x_2 + e_1 = b_1 * 1 + b_2 * 1 + e_1$$
$$1824 = b_1 * x_1 + b_2 * x_2 + e_2 = b_1 * 1 + b_2 * 0 + e_2$$
$$-2010 = b_1 * x_1 + b_2 * x_2 + e_3 = b_1 * 0 + b_2 * 1 + e_3,$$

where b_1 and b_2 are determined using the generalized least squares method[8] for a purely heteroscedastic model with the weights $1/s^2$(β-estimate). Using the estimated b_1 (1821.64) and b_2 (-2012.32) the consistency-corrected estimates of the β-coefficients are calculated. These are consistent since 1821.64 - 2012.32 = -190.68

Correction for holiday effects

The correction in the previous section was performed because the positioning of the measurement periods *varied*, but it was always a question of measurement periods *without* public holidays. As a result of the correction, we have obtained values that have been corrected for both non-response error and periodization error. After these recalculations, we have values that refer to measurement periods of one or two weeks that are in the middle of each month. These periods do not contain any holidays.

In this section, we describe the correction we have performed to convert to values that refer to measurement periods of four or five weeks. This means that even weeks that include public holidays are now measured. For 1993, we have compared the old measurement period system of two-week periods and the new system of four-week periods for all key series. For e.g. number of persons at work, this means:

Persons at work in 1993, millions. Old and new measurement periods
Two weeks - the old system (corrected for periodization error) Whole month = the new system

Month	Two weeks	Whole month	Difference	Ratio
Jan	3.567	3.483	-0.084	0.976
Feb	3.535	3.523	-0.012	0.997
Mar	3.497	3.513	0.016	1.005
Apr	3.589	3.465	-0.124	0.965
May	3.575	3.533	-0.042	0.988
June	3.501	3.310	-0.192	0.945
July	1.723	1.899	0.176	1.102
Aug	3.208	3.061	-0.147	0.954
Sept	3.523	3.536	0.013	1.004
Oct	3.511	3.531	0.020	1.006
Nov	3.498	3.508	0.011	1.003
Dec	3.563	3.307	-0.256	0.928

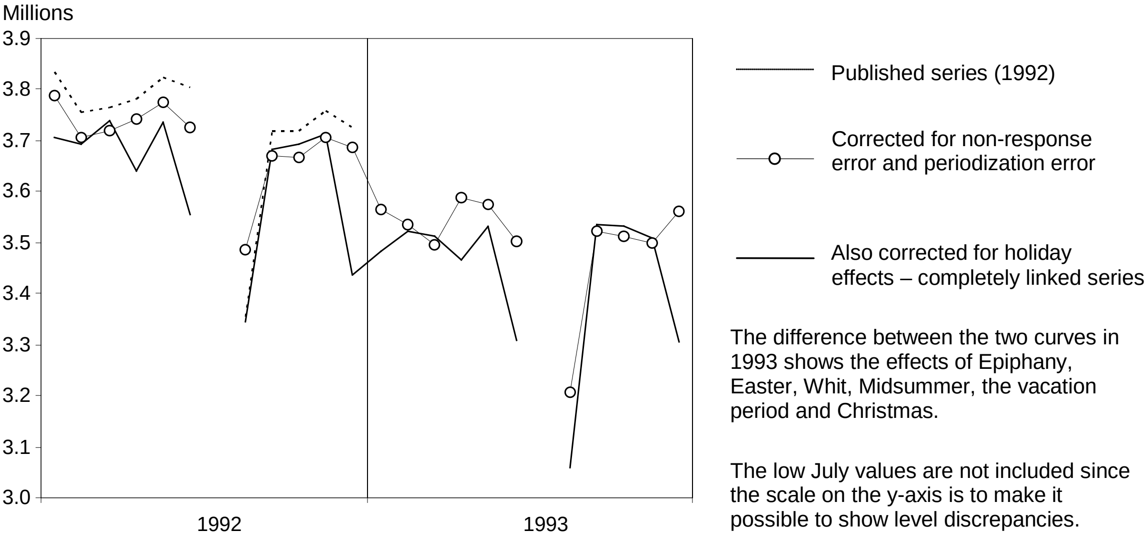
[8] See for example Kmenta (1971)

The two-week measurements are based on half samples, making them less accurate than whole-month estimates. This means that Jan, Apr, May, June, July, Aug and Dec display significant discrepancies, *s(whole month)=0.018*

All key series apart from those referring to unemployed persons (minor differences for these) have been corrected for holiday effects. The correction factors calculated are based on the *1993 calendar*, where e.g. Easter was in the middle of April.

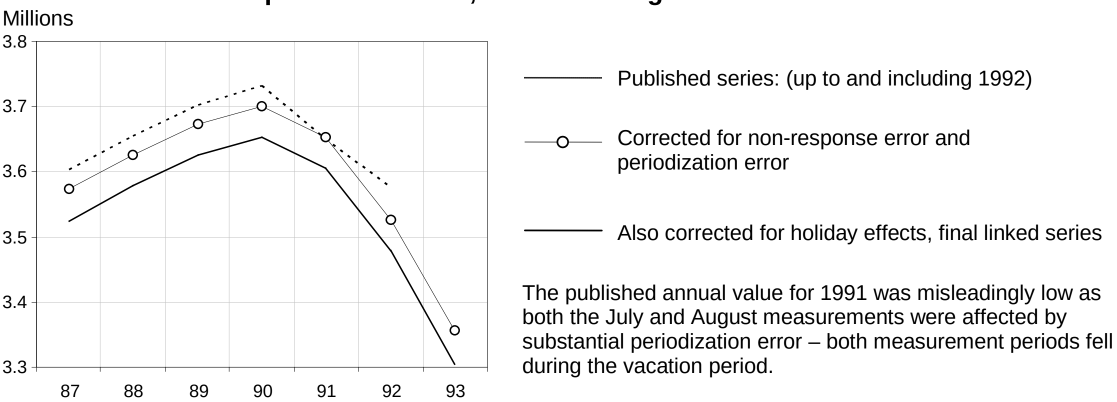
The correction factors we finally use are differences. The reasons why we have chosen an additive model are given in the next section. Chart 9 below shows the results of the final linking.

Chart 9. Number of persons at work, monthly data excluding July values



After correction, the *1993 calendar* will be valid for the entire period 1987-1992. 1993 is the only year we can compare two-week estimates with whole-month estimates. It is therefore unavoidable that the 1993 calendar effects will affect the linked series. Chart 9 with monthly data shows how the level for different months is affected by the change in measurement periods. This means that we obtain entirely new seasonal patterns in the series of monthly values. How the *annual levels* are affected is shown in Chart 10.

Chart 10. Number of persons at work, annual averages



Consistency adjustment of key series

The 100 or so linked key series will be used as calibration conditions when calibrating the sampling weights of the respondents. The conditions must be completely consistent if they are to be used in Step 2 of the linking work.

When correcting for non-response error, periodization error and holiday effects, we can choose between an additive model (correcting using differences) and a multiplicative model (correcting using ratios). We have chosen a multiplicative model for the non-response error and corrected all series, but we have chosen an additive model for the corrections for periodization error and holiday effects. It was however not necessary to correct the unemployment series for these two errors. If we use an additive model, no inconsistencies will emerge.

When correcting for *non-response error*, the key series are multiplied by different factors and despite the factors being close to one, inconsistencies still occur. The table below shows this.

Population aged 16-64 by labour force status, 1000s of persons, annual mean values

Year	Original data corrected for new definitions				Corrected for non-response error					
	Employed	Unem-ployed	Not in labour force	The popu-lation	Employed	Unem-ployed	Not in labour force	The popu-lation	Population-Employed - Not in labour force	Inconsi-stencies
87	4337	87	874	5299	4294	91	906	5299	5299-4294-906 = 99	91-99 = -8
88	4399	75	848	5323	4355	78	879	5323	89	-11
89	4466	65	825	5356	4417	67	857	5356	82	-14
90	4508	73	817	5397	4454	76	850	5397	92	-17
91	4430	128	866	5424	4372	134	905	5424	147	-13
92	4250	222	974	5447	4195	233	1019	5447	233	0

Original data are consistent:

employed + unemployed + not in the labour force = the population.

When these series are multiplied by different factors, this similarity does not apply. The correction for non-response error has been calculated using data for 1992. The model for non-response error that was correct at that time is not entirely perfect prior to 1992. For example, the number of unemployed persons can be calculated in two ways that differ by 8 thousand people in 1987. We have consistency-adjusted the series proportionally[9] so that there is only negligible change in the 14 unemployment series[10] in terms of number of unemployed.

During the period 1987-1992, it is only the unemployment series that have changed level significantly – from 65 thousand in 1989 to 222 thousand in 1992, an increase of 242 percent. The choice between additive and multiplicative model is therefore of considerable significance for the unemployment series. It was our assessment that the multiplicative model is the only appropriate model in this case.

When we corrected for *periodization error*, we found that the unemployment series didn't need to be corrected. Since the other series don't have such

---

[9] For 1991, the number of employed persons, unemployed persons and economically inactive persons would decrease proportionally by 13/5424.

[10] Women and men broken down into seven age classes.

level discrepancies relatively speaking, the choice between additive and multiplicative model is not so pivotal – the corrections will be approximately of the same size. Here we chose an additive model since it is easiest to adjust the coefficients in the transfer function models for consistency. After correction, the series are still consistent.

Even when we correct for *holiday effects*, we made a choice between an additive and a multiplicative model. The structure of the LFS tables is additive – the totals and their parts are the main content. Using a multiplicative model for an additive structure always causes problems, as illustrated in the table below. The number of employed persons should tally with the sum of the number of persons at work and the number of absent persons. When we use the multiplicative model that applied in 1993 for previous years, the inconsistencies in certain months are quite substantial. These inconsistencies vary from one year to the next. If we had chosen a multiplicative model, we would have been forced to adjust for these inconsistencies and this would have made the seasonal patterns more unstable. We therefore chose an additive model partly to avoid having to adjust for inconsistencies and partly so as not to impair the seasonal patterns.

The effects of using ratios when correcting for holiday effects

Ratios:	Whole month divided by two-weekly measurement 1993			Inconsistencies: Employed persons - (at work + absent persons) 1000s of persons						
	Employed	At work	Absent	1987	1988	1989	1990	1991	1992	1993
Jan	0.998	0.976	1.179	-8	-5	-9	-11	-5	3	0
Feb	0 998	0.997	1.006	0	-1	0	-1	-1	0	0
Mar	0 998	1.005	0.955	0	2	1	2	1	0	0
Apr	0.996	0.965	1.287	-37	-31	-33	-53	-26	-27	0
May	1.000	0.988	1.113	-11	-9	-15	-13	-12	-3	0
June	1.010	0.945	1.450	-43	-20	-29	-33	-30	-16	0
July	1.003	1.102	0.931	21	9	18	29	24	33	0
Aug	1.002	0.954	1.211	-19	-25	-8	11	9	23	0
Sept	1 006	1.004	1.027	-3	-3	-4	-3	-3	-2	0
Oct	0.999	1 006	0.942	5	5	8	9	6	2	0
Nov	1 004	1 003	1.015	-1	-1	-1	-1	-1	0	0
Dec	0 997	0.928	1.670	-76	-94	-92	-79	-59	-36	0

If the unemployment series had needed to be corrected, we would have preferred to use a multiplicative model, but since these series are only negligibly affected by holiday effects, we decided *not* to correct them[11]. The remaining series were corrected using an additive model in a consistent manner. In addition to avoiding retroactive consistency adjustment, the seasonal patterns are also less affected.

A fundamental principle has been to perform the recalculations carefully. The structure of the original time series is to be adhered to as much as possible. No unnecessary or unjustified corrections are to be made. This applies in particular to the unemployment series.

After this, the linking work on the time series level is now complete. As a result, we have obtained about one hundred key series that have been corrected for changes in definition, non-response error, periodization error and holiday effects. Every series consists of 72 monthly values that are

[11] This is discussed in the section describing Step 3: Evaluation of linked series.

comparable to the series that started in January 1993. Please note that calendar effects for 1993 have been transferred to the period 1987-1992. If calendar correction is necessary in the future, we must remember that the 1993 calendar is valid for the entire period 1987-1993.

Summary of Step 1

Linking work on the time series level (Step 1) consists of a number of different stages. Before we go on to describing the linking work on the individual level, here is a summary of the stages in Step 1.

Recalculation based on new definitions
The definition of *unemployed* has been changed on the individual level for the period 1987-1992 so that it tallies with the new definition that applies from 1993 onwards. Likewise, the definition of *not in the labour force* has been changed. It was possible to do these recalculations with information on the individual level that existed in the microdata for 1987-1992.

Selection of key series
The time series for different labour force statuses (At work, Absent, Unemployed, Not in the labour force) and employed broken down by degree of attachment (Permanently employees, Temporary employees, Self-employed) are classified by sex and age (7 age groups). All these series have been linked on the time series level.

Persons at work and hours worked (both sexes, 16-64 years) broken down into 21 industries and number of absent persons (both sexes, 16-64 years) whole week and part of week respectively due to vacation have also been linked on the time series level.

Correction for non-response error
All selected monthly series in 1987-1992 have been corrected for the level changes measured for 1992 as a result of the new estimation method. The level changes are due to the fact that the new estimation method is deemed to reduce the non-response error. Since the correction is made multiplicatively, the series must also be adjusted for consistency. New further corrections are made to the series that refer to unemployed persons,

Comparable series 1987-1992 and 1993 according to the old measurement period
As a result of the work above, we now have monthly data for the period 1987-1992 that are comparable to monthly data from 1993 that refer to the old measurement period, i.e. 1-2 weeks without public holidays. The recalculated series for 1987-1992 are extended using estimates for 1993 with regard to the old type of measurement period.

Correction for periodization error
The measurement periods according to the old system are sometimes earlier and sometimes later in each month respectively. This significantly disrupts the seasonal pattern in certain series. This disruption is similar to calendar variation and has been corrected for using methodology that is common in this context.

The series that have been corrected for each sex and age group respectively are: temporary employees, at work, absent and not in the labour force for the period 1987-1993 (old type of measurement period). Number of absent persons due to vacation the whole week or part thereof has also been corrected.

Correction for holiday effects
For 1993, we have compared new four-week measurements and measurements according to the old measurement period system[12] and corrected for periodization error. The differences between these provide an estimate of the holiday effects. All key series (apart from the unemployment series) are corrected using these estimated holiday effects.

Comparable key series 1987-1993
As a result, we have now obtained corrected key series that are comparable over time.

[12] For 1993, two-week periods were also chosen for April, June and August, as the material would otherwise have been too scant.

Step 2: Calibration of weights

Within the sampling theory, there are methods that make use of auxiliary information to improve estimates. In recent years, a special method has been developed, in which the weighting factors are *calibrated* using known population totals $\mathbf{t}_x$ on the auxiliary variables $\mathbf{x}_k$. The aim is to improve the estimates by reducing their standard deviation. At the same time, we obtain consistency between the estimates and the known population totals.

Before adding auxiliary information, the weighting factors d_k provided by the sample design are applied. According to the method described in Deville & Särndal (1992), new weighting factors w_k are formed that should be as close to the original weights d_k as possible. By minimizing the distance between d_k and w_k subject to the conditions provided by the known population totals, we can determine the new weights:

– for the known totals $\mathbf{t}_x$ the weights w_k must conform to $\Sigma_s w_k \mathbf{x}_k = \mathbf{t}_x$

– the minimization of $\Sigma_s(w_k - d_k)^2 / d_k$ gives $w_k = d_k (1 + \mathbf{x}'_k \lambda)$, where $\lambda = \mathbf{T}_s^{-1}(\mathbf{t}_x - \Sigma_s d_k \mathbf{x}_k)$ and $\mathbf{T}_s = \Sigma_s d_k \mathbf{x}_k \mathbf{x}'_k$, where *s* stands for sample.

How do we use calibration to link time series?

We use this method but with another aim[13]. Our aim is not to reduce standard deviations for estimates but to link time series. The known population totals $\mathbf{t}_x$ in Devill & Särndal are equivalent here to the linked values of the key series for a given month. In the same way as the method in Deville & Särndal provides consistency between estimates and known population totals, we will obtain consistency between all new linked series and the key series we have as calibration conditions.

For each month, we have a number of different time series values that are estimates of various population totals such as the number of employed persons or the number of hours worked. Some of these estimated totals (for the key series) have been adjusted as we described in the section on Step 1. We now want the respondents from each month to have new weighting factors that "tally" with these approximately 100 new population total estimates. What value is the weighting factor of each of the approximately 16 000 respondents to have in order for us to obtain the new adjusted values of the key series? We perform this type of calibration of the respondents' weighting factors for each of the 72 months in the period 1987-1992.

[13] We have discussed this application with Professor Särndal who used the example at a conference in Helsinki in 1995.

The calibration for a given monthly survey is performed as follows:

1) We start from the weighting factors that were originally used. These correspond to the weights d_k
2) The values of the linked key series for this month give us 106 calibration conditions $\mathbf{t}_x$. Example: The number of permanently employed women aged 16-19 years should be 38 957 for this month.
3) An x-variable is defined for each calibration condition. Example: x = 1 if the respondent is a permanently employed woman aged 16-19 years, otherwise x = 0.
4) We then calculate the matrix $\mathbf{T}_s = \Sigma_s d_k \mathbf{x}_k \mathbf{x}'_k$ and its inverse $\mathbf{T}_s^{-1}$
5) The vector λ is calculated: $\lambda = \mathbf{T}_s^{-1}(\mathbf{t}_x - \Sigma_s d_k \mathbf{x}_k)$
6) Finally, we calculate $w_k = d_k (1 + \mathbf{x}'_k \lambda)$

Choosing calibration conditions

In our first calibration attempt, we used calibration conditions on *number of persons* by sex, age, degree of attachment and labour force status. Furthermore, we had calibration conditions on *number of hours worked* by industry and two calibration conditions that relate to number of persons on vacation the whole week or part thereof. We chose these calibration conditions because we wanted both series of number of persons and series of hours worked to be well linked.

The results of this calibration were as follows: During the month of December, about 70 negative weights[14] w_k occurred, which may be a sign that the calibration conditions and structure of microdata are not in complete concordance with each other. Most of the tables with new estimates obtained reasonable results, although the results were unreasonable for some categories. For months with public holidays, the number of persons who had worked only a few hours, e.g. women with infants, was unreasonably high.

We gradually realized that these unreasonable results were due to the fact that the linked key series for number of hours worked and number of persons at work contradicted each other somewhat. Public holiday absence was an entirely new reason for absence from 1993 onwards. Public holidays lead to persons at work being absent for part of the week. This reduces the number of hours worked despite the number of persons at work remaining unaffected. When we calibrated with both number of hours worked and number of persons at work as the calibration conditions, persons who only worked a few hours were given greater weight leading to the estimates of the number of employed women with infants being unreasonable.

We therefore performed a new calibration in which the calibration conditions for hours worked were replaced by calibration conditions for number of persons at work broken down into different industries. This calibration produced much fewer negative weights w_k than the first calibration. Out of 72 months, we obtained no negative weights for 59 months and between 1 and 6 for the others. After producing and studying

[14] Negative weights are not unreasonable. On the contrary, the calibration estimators provide good estimates despite this. They may however confuse users who work with individual data and sometimes give negative estimates in small study domains.

new tables of recalculated estimates, we assessed all tables containing number of persons to be reasonable and could be used as linked time series. However, the calibration produced estimates that were too high for number of hours worked during months that contained public holidays (January, April, May, June and December). We performed a special correction for this, as described in the section on synthetic absence.

The table below shows the results for a few respondents involved in the survey in December 1992. The first person is an example of how the number of persons at work should be adjusted downwards. The next is a person not in the labour force, which represents a category only negligibly affected by the calibration. Unemployed persons have only been corrected for non-response error – the key series for unemployed persons was multiplied by about 1.05. The weight for the unemployed person has increased by a corresponding degree. Finally, we have two examples of persons who have been given increased weighting factors. The 30-year-old man is an example of a person who is absent for a part of the week due to vacation. When linking this key series, it became apparent that this time series needed to be multiplied by about 3 for the December months. The last person is a woman who works in the health service and who was absent during the survey week. As the actual measurement was performed at the beginning of December with low absence, her weighting factor will increase considerably.

LFS for December 1992, results for a few respondents

Sex	Age	Labour force category	Degree of attachment	SNI Swedish NACE	Original weight	New weight	Ratio w_k / d_k
Female	44	At work	Self-employed	3560	300.7	166.7	0.55
Female	23	Not in the labour force	Out of work	0	314.5	311.0	0.99
Female	26	Unemployed	Out of work	0	327.7	346.0	1.06
Male	30	At work part of week	Permanently employed	7200	344.4	1006.7	2.92
Female	42	Absent	Permanently employed	9330	352.3	2718.9	7.72

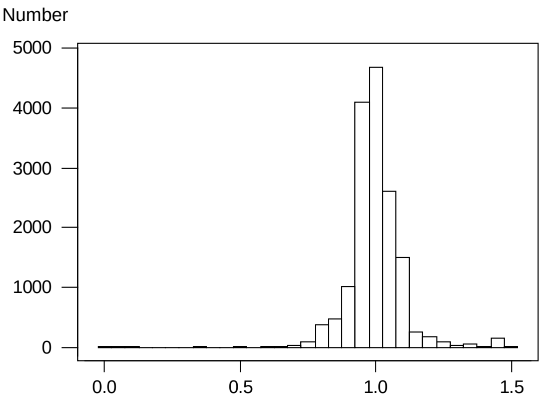
The new weighting factors

To compare the original and the calibrated weights, you normally calculate w_k / d_k which is the ratio between the new and the original weight. These ratios are on average one, but the distribution around this average value gives us an idea of the size of recalculations needed to fulfil the calibration conditions.

Chart 11 shows the ratios for a month with minor recalculations whereas Chart 12 shows that the absence over the Christmas holidays results in major recalculations.

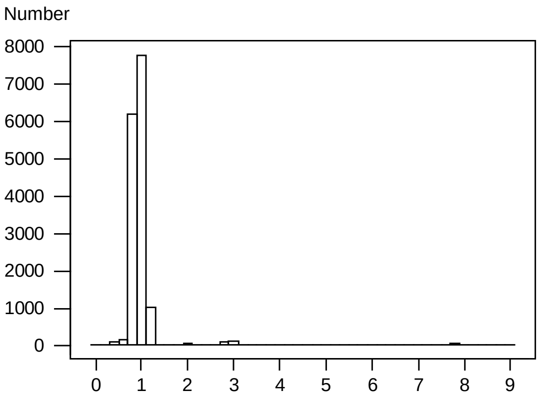
The fact that some persons have been given much higher weights when calibrating for December 1992 could lead to a risk that the calibration had provided "strange" values. This risk is unavoidable – major recalculations are necessary for some series.

Chart 11
Distribution for w_k / d_k November 92



Minimum	Lower quartile	Upper quartile	Maximum
0.002	0.957	1.037	1.514

Chart 12
Distribution for wk / dk December 92



Minimum	Lower quartile	Upper quartile	Maximum
0.018	0.858	0 999	8.987

Synthetic absence

In the final calibration, we had, among others, calibration conditions on number of persons at work, number of persons absent the whole week and number of persons absent part of the week due to vacation. During months containing public holidays, the number of persons at work fell while the number of persons absent the whole week or who were on vacation for part of the week increased. All this leads to the number of actual hours worked decreases. Unfortunately, the decrease was shown to be insufficient, i.e. the number of hours worked in months containing public holidays was still too high. The reason for this is that holiday absence was not studied prior to 1993 due to the fact that only periods without public holidays in them were chosen as measurement periods.

The first calibration attempt demonstrated that it is not possible to rectify this using calibration since there is no suitable category of respondents that could obtain increased weight. We must therefore perform a different type of correction for how holiday absence should affect the number of hours worked.

The table below shows the reductions achieved by the calibration and what further reductions need to be done in accordance with the linked key series for the number of hours worked.

Total number of hours worked, millions of hours per week

	January 1992	April 1992	May 1992	June 1992	December 1992
Original values	143.76	142.59	145.50	142.71	140.35
After final calibration	137.90	137.36	141.71	132.97	127.77
Should be according to linked time series	130.23	127.93	133.64	124.10	118.37

The additional reduction needed for January (137.90-130.23) is equivalent to a quarter of 3.7 million persons at work being absent for about 8 hours during the survey week. Epiphany is included in the first survey week in January 1993 and should therefore produce such holiday absence for a quarter of the sample.

We performed the correction in the following way:

1) For January and May 1987-1992, we chose[15] a quarter of the sample for which absence and hours worked were adjusted so that it would correspond to the effect of Epiphany and Ascension Day respectively.
2) For April, June and December, half the sample was chosen since different public holidays affect about half of the sample during these months.
3) The holiday corrections we perform are based on the 1993 surveys, in which we have studied the holiday effects. Please note that it is the 1993 calendar effects that are valid for the linked series throughout the entire period 1987-1992.
4) Holiday absence is different depending on the industry. In the health service and in other industries where operations are continuous, holiday absence is low. We therefore performed different corrections for different industries.

[15] Microdata was sorted according to SNI code. A systematic sample was then selected and given a synthetic holiday absence.

Step 3: Evaluation of linked series

How successful has the linking been? To give an idea of how successful the linking has been, we have studied a number of linked series and compared them with each original series for 1987-1993. The opinion within the Department for Labour Market Statistics has been that a total of 216 time series should be studied. This section presents the evaluation results in the following order:

Number of persons by age (16-24, 25-34, 35-44, 45-54, 55-64 and 16-64 years) and sex (men, women and both sexes):

1) Unemployed
2) Employed
3) At work
4) Temporarily employed
5) Permanently employed

Number of persons by sex and industry SNI69 (1, 2-4, 38, 5, 6, 61-62, 7, 8, 9, 91 and 931-934):

6) Number of persons at work

Broken down by sex and detailed industry SNI69 (1, 2-4, 38, 5, 6, 61, 62, 7, 71, 72, 8, 81, 82, 9, 91, 92, 93, 931, 933, 934, 932+935-939, 942, 941+949):

7) Number of hours worked
8) Number of employed persons

Evaluation method

In order to evaluate linked series, we use not only graphical methods but also analysis of the time series quality with the help of ARIMA models. By comparing the ARIMA model residuals for the original series and the corresponding linked series, we obtain a measure of whether the linking has increased or decreased the time series noise.

The ARIMA residual for a given point in time *t* describes the deviation between the actual value and the estimated value according to the model. The model value describes the stochastic trend of the series and stochastic seasonal patterns for the point in time *t*. As a measure of the noise, we use the standard deviations for the ARIMA model's residuals throughout. For all 216 series, ARIMA(310011)-models were adapted both for the original series and the linked series. This model proved to be the best for most of the series when comparing the four models: 011011, 210011, 310011 and 013011. Since the selected model was not unsatisfactory in any of the cases, it was used for all series.

We analyse ARIMA residuals in the following way: The noise for the period 1987-1992 for the original series and corresponding linked series are compared. The linked series must not have stronger noise than the original series, but this is not in itself a reason for declaring the linking a success. It shows that the monthly time series pattern is good for the period 1987-1992. If, on the other hand, the noise of the linked series is stronger than the original series, this is an indication that the linking has led to a more unstable seasonal pattern. But we must remember that many series are affected by definition changes and that the new definitions can give series with higher noise, which must also be visible in the linked series.

The residual analysis above is supplemented by analysis of time series diagrams. By graphically analysing monthly data, we can compare the seasonal pattern in the original series with the pattern in the linked series. These diagrams show whether or not we have succeeded in obtaining correct seasonal patterns. Finally, by analysing diagrams with annual data, we can see whether the linking has successfully obtained the right level.

The results of the residual analysis are described in the diagram below. For each series, the noise before and after linking is compared by forming ratios between residual standard deviations. Since the linking affects the level of the series, we have divided each residual standard deviation by the average of the series for the period 1987-1992. The ratios are presented for each of the eight groups of series.

Comparison of noise before and after linking for the period 1987-1992
Ratios s(residual linked series) / s(residual original series), denoted with a point for each series

The scatter diagram above shows that the linking has reduced the noise for 115 out of 216 series. This is the result of the work done to correct periodization error. In only one of the groups (employed persons by sex and detailed SNI) are there series where the linking has increased the noise by more than 15 percent. This group also contains an extremely high ratio.

For each group, we will present time series diagrams (where the ratio above is given in the heading) for the group's total series and for one with a low ratio and a series with a high ratio respectively. We state whether the linking covers definition changes on the individual level, correction for non-response error, for periodization error and for holiday effects. We also state whether the linking has been done on the time series level under Step 1 or after calibration of weighting factors in Step 2.

Group 1: Unemployed persons by age and sex

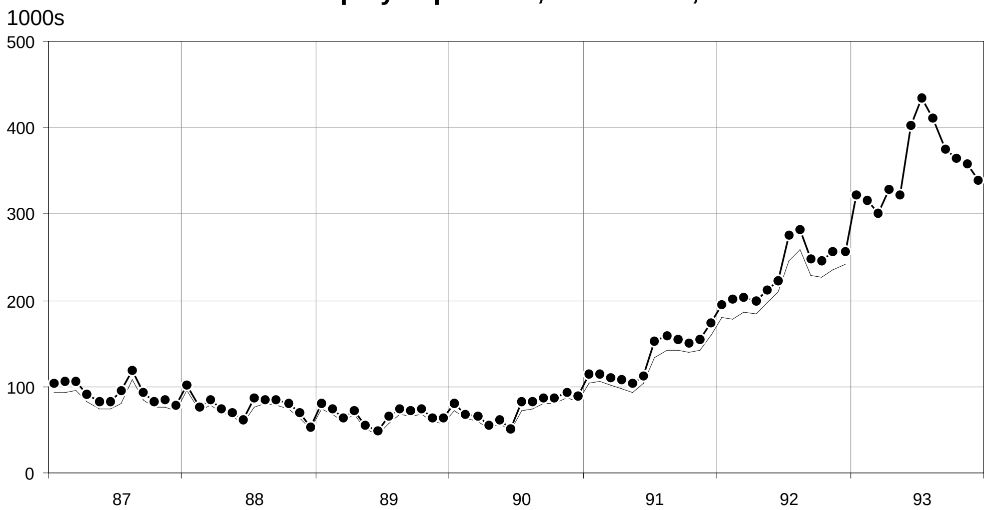
The series of unemployed persons have *only been corrected for non-response error.* The discrepancies between original series and linked series are hence only moderate, which is apparent from Charts 13-14[16].

Chart 13 below shows that the seasonal pattern for June 1993 differs from previous years. In 1993, unemployment was high in June, July and August. In 1992 and earlier, only July and August have high seasonal values. For the unemployment series, we have as a rule no statistically significant differences between the measurement period according to the old and the new system.

There is one exception, however: In June 1993, the estimate for the first two weeks was 357 thousand (both sexes, 16-64 years), while the estimate for the whole of June was 402 thousand. The difference of 45 thousand can be explained by a reduction of 87 thousand persons in government-supported training programmes, which occurred in June of that year. The value of 87 thousand refers to the change between the last day of May and the last day of June. If this reduction occurs gradually, it would be equivalent to a difference of about 44 000 persons between the first and second half of June.

Since the difference for June 1993 between the old and the new measurement period could be a one-off as a result of the extreme economic recession during 1993, it was our assessment that it was less appropriate to assume that there were similar discrepancies during previous years. We chose therefore not to make any corrections for holiday effects for unemployment series. Since the deviating pattern for June 1993 depends mostly on the behaviour of young age groups, we will comment more on this in connection with Chart 19.

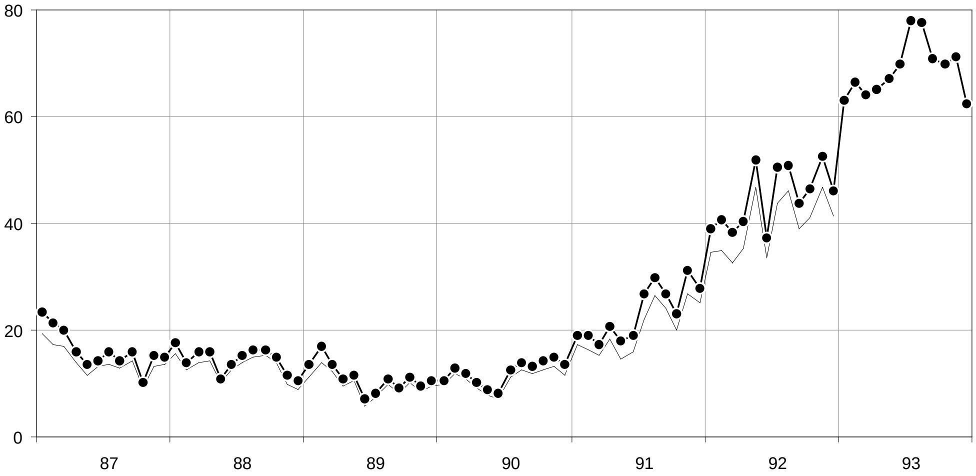
Chart 13. Number of unemployed persons, both sexes, 16-64



In all diagrams, the original series is denoted by a thin unbroken line and the linked series by a thick line where the monthly values are depicted as filled circles.

[16] Some charts have been excluded in the English version of the report. The numbers are however, the same as in the Swedish report.

Chart 14. Number of unemployed persons, both sexes, 35-44 years
1000s



Charts 13 and 14 indicate level discrepancies between December 1992 and January 1993. These level discrepancies are still there after linking to describe real changes on the labour market. Many persons lost their jobs from January 1[st] 1993 onwards. The aim of the linking is not to remove all types of time series breaks, only breaks that are caused by the LFS changeover.

The linking covers:	
Definition change	Yes
Correction for nonresponse error	Yes
Correction for periodization error	No
Correction for holiday effects	No
Linking on time series level or after calibration	Time series level

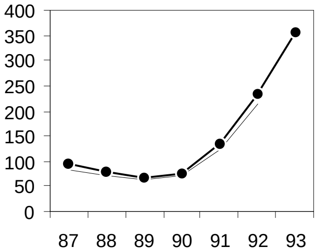
Ratios between noise for linked series and noise for original series

	16-24 years	25-34 years	35-44 years	45-54 years	55-64 years	16-64 years
Men	0.99	1.07	0.97	1.03	1.05	1.02
Women	0.98	0.95	0.93	0.97	1.00	0.94
Both sexes	1.02	1.05	0.94	0.89	1.04	1.00

The diagrams containing annual values below describe the small level changes caused by the linking.

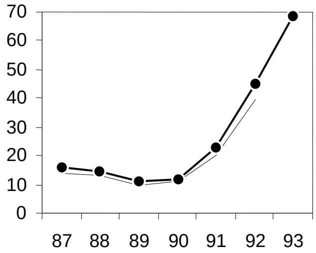
Number of unemployed persons, annual values, 1000s

Chart 16. Both sexes, 16-64 years



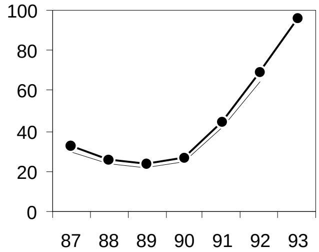
Level shift 1992: 18.8 thousands

Chart 17. Both sexes, 45-54 years



Level shift 1992: 1.4 thousands

Chart 18. Men, 25-34 years



Level shift 1992: 4.4 thousands

Chart 19. Number of unemployed persons, both sexes, 16-24 years

1000s

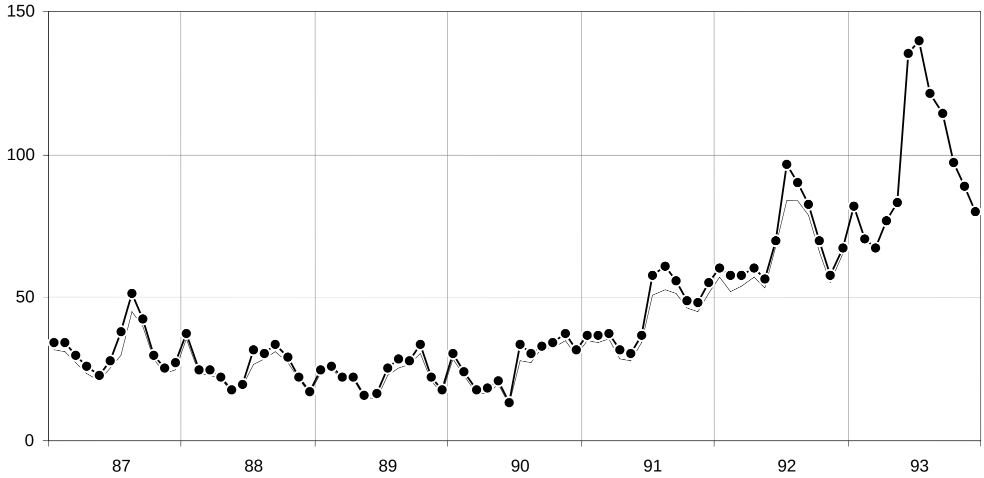


Chart 19 shows that the change in definition has a particular effect on the values during the summer months of 1991 and 1992. This shows that it would have been inappropriate to correct for new definitions by multiplying by the same coefficient for all months of all the years.

As we pointed out on page 33, the time series pattern for June 1993 deviates from previous years. At the end of the spring term, students from upper secondary school and higher education start looking for jobs on the labour market. Dependent on the economic situation, they either become unemployed or employed during the latter half of June.

In the period 1987-1993, LFS was performed in early June, which is why there are no obvious upswings between May and June in the original series in Chart 19. In June 1982-83 LFS was performed between the 7 and 20 June and between 6 and 19 June respectively.

In the table below, we can compare the upswings between May and June for these years with corresponding upswings in 1993 when the whole of June was studied:

Both sexes, 16-24 years	Unemployed persons, 1000s	Employed persons, 1000s	
June 82 - May 82	16	85	Values for 82 and 83 have been corrected for time series breaks 86/87 and 92/93
June 83 - May 83	21	68	
June 93 - May 93	52	75	

The increase in the number of unemployed persons in 1993 is affected by a reduction of 35 500 in the number of youth work experience places in June.

If the whole of June had been studied in the 1987-1992 period, the unemployment Charts would probably have been higher. The question is then: how much higher? This is very difficult to tell with the data we have available and we have therefore consciously refrained from correcting for this June upswing since the economic situation and labour market policy measures vary from one year to the next.

Group 2: Employed persons by age and sex

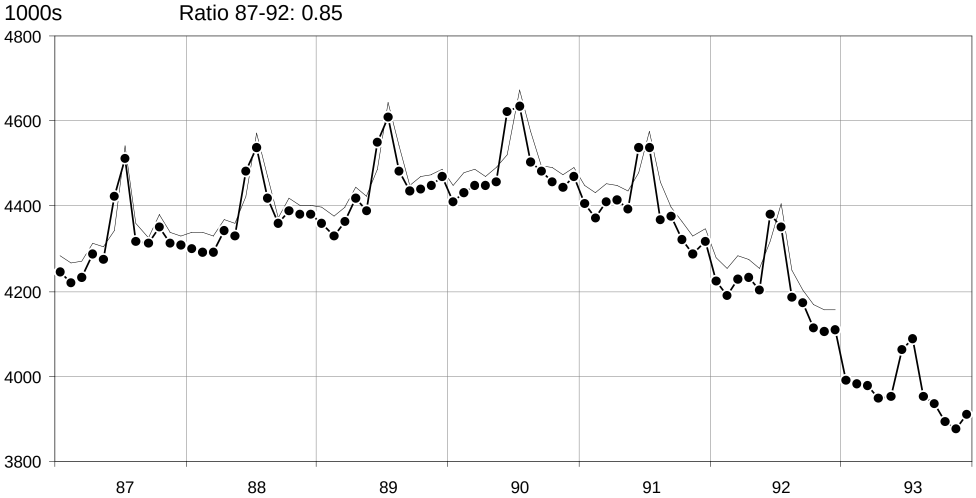
The series of employed persons are not affected by definition changes but have been linked to correct for all the other effects of the changeover.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	Time series level

Ratios between noise for linked series and noise for original series

	16-24 years	25-34 years	35-44 years	45-54 years	55-64 years	16-64 years
Men	0.84	1.11	1.15	1.11	1.02	0.99
Women	0.88	1.05	1.07	1.02	1.00	0.95
Both sexes	0.85	1.15	1.12	1.07	0.98	0.85

Chart 20. Number of employed persons, both sexes, 16-64 years



The linking has led to the series being given a lower level and to a change in the seasonal pattern, especially for June. The linked series has a more stable seasonal pattern than the original series due to the correction for periodization error. The level break between December 1992 and January 1993 depicts a real downturn in employment.

Group 3: Persons at work by age and sex

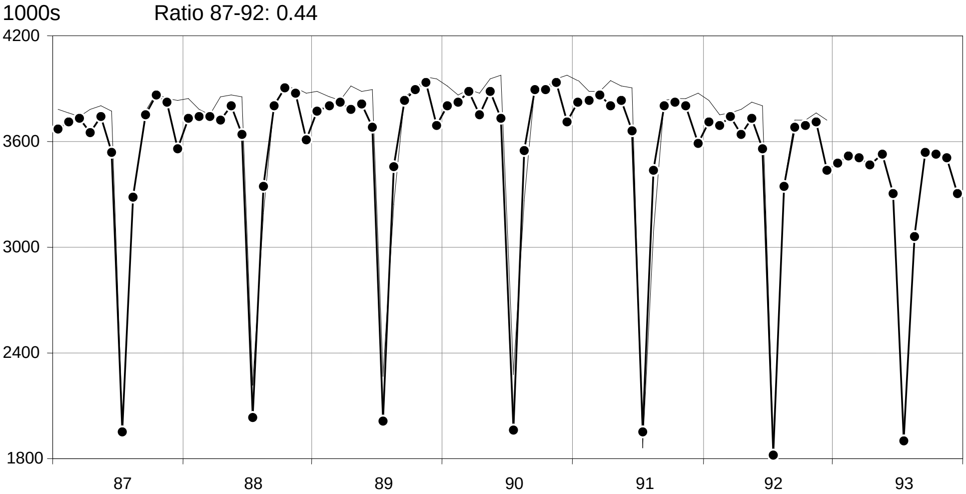
The corrections for measurement period and holiday effects are relatively major for series in this group, which creates major level discrepancies between unlinked and linked series. Correction for periodization error leads to a significant reduction in noise as a result of the linking.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	Time series level

Ratios between noise for linked series and noise for original series

	16-24 years	25-34 years	35-44 years	45-54 years	55-64 years	16-64 years
Men	0.88	0.70	0.59	0.58	0.84	0.47
Women	0.96	0.81	0.72	0.64	0.73	0.58
Both sexes	0.89	0.66	0.57	0.55	0.72	0.44

Chart 24. Number of persons at work, both sexes, 16-64 years



The linking has given the series a new seasonal pattern and we see that this pattern tallies with the pattern of the new series in 1993.

Group 4: Temporary employees by age and sex

It is our assessment that the linking for this group of time series has also been successful – the ratios between the noise after and prior to the linking are mostly under one. This is due to the corrections done for periodization error.

Apart from the total series for all temporarily employees, the group consists of series for small categories of persons, which is why sample errors and panel effects can be seen in the diagrams. During every quarter, three different samples were studied which were then examined in the same order during the next quarter. This can produce panel effects – a positive correlation between values with an interval of three months. These effects are not affected by the linking.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	Time series level

Ratios between noise for linked series and noise for original series

	16-24 years	25-34 years	35-44 years	45-54 years	55-64 years	16-64 years
Men	0.76	1.03	1.06	0.88	0.96	0.84
Women	0.87	0.93	1.05	0.94	0.89	0.86
Both sexes	0.84	0.98	1.06	0.92	0.92	0.85

Group 5: Permanent employees by age and sex

This group of series depicts large and stable categories on the labour market. The linking mainly relates to correction for non-response error.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	Time series level

Ratios between noise for linked series and noise for original series

	16-24 years	25-34 years	35-44 years	45-54 years	55-64 years	16-64 years
Men	1.04	1.07	1.05	1.06	1.03	1.05
Women	1.02	1.05	1.04	1.04	1.02	1.04
Both sexes	1.03	1.06	1.05	1.06	1.03	1.04

Group 6: Employees at work by sex and industry

This group of series has been linked indirectly in connection with the calibration of the weighting factors during Step 2. Despite this, the linking has resulted in the seasonal pattern becoming more stable since all ratios between the noise after and prior to the linking are less than one. During Step 1, the series for persons at work (broken down by industry) were linked. In connection with this linking, the series were corrected for the periodization error caused by the variation in survey periods. As these series were calibration conditions, the correction for periodization error has been transferred to the series of employees at work by sex and industry. The low noise ratios for Group 6 indicate this.

Chart 43. Number of employees at work, both sexes in public administration, teaching, health and social care (SNI69=91, 931-934)

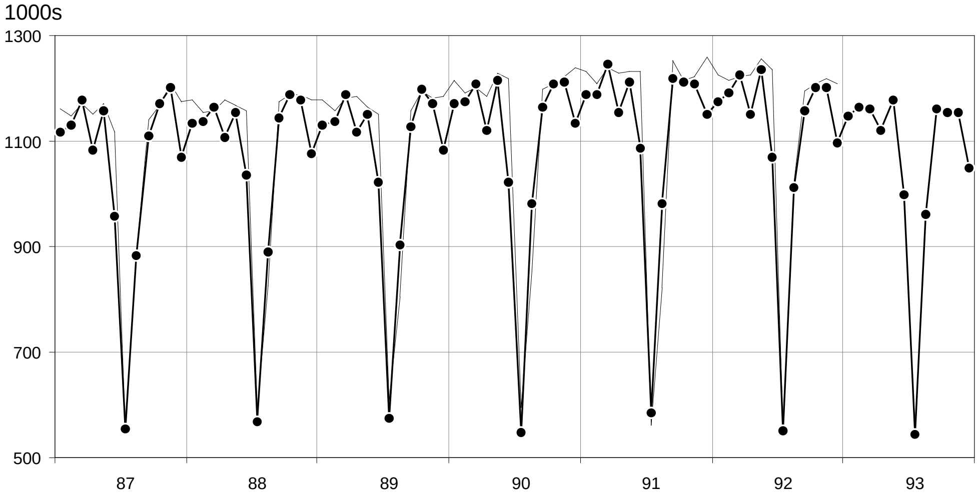
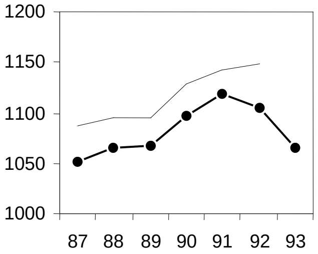


Chart 46. Both sexes SNI 91, 931-934



The linking has resulted in a changed seasonal pattern. The linked series has a more stable seasonal pattern than the original which is clear from the ratio 0.67. This series contain education (931) that was difficult to link for *employed* persons. This is discussed in the section on Group 8. For *employees at work* in Chart 43 the linking was successful.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	After calibration

Ratios between noise for linked series and noise for original series

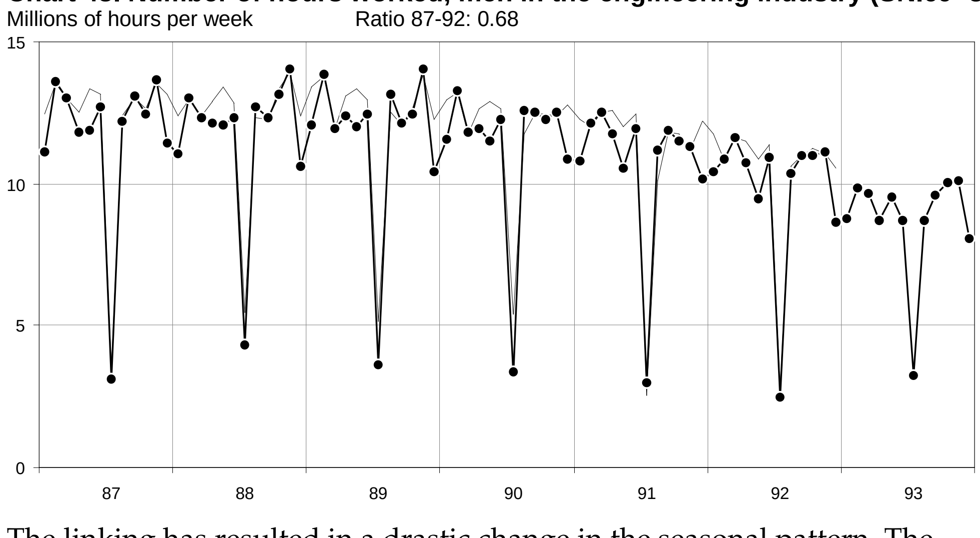
SNI69:	SNI 1	SNI 2-4	SNI 38	SNI 5	SNI 6	SNI 61-62	SNI 7	SNI 8	SNI 9	SNI 91, 931-934
Men	0.99	0.70	0.66	0.77	0.84	0.84	0.86	0.89	0.68	0.78
Women	0.92	0.88	0.99	0.99	0.87	0.85	0.96	0.94	0.67	0.69
Both sexes	0.97	0.80	0.68	0.79	0.81	0.79	0.78	0.93	0.66	0.67

Group 7: Hours worked by sex and industry

This group of series has not been linked directly during "Step 1 Linking of key series", but indirectly via the calibration in "Step 2: Calibration of weighting factors". Number of persons at work for different industries were linked directly and then we corrected the number of hours worked by giving certain persons at work what is known as "synthetic holiday absence".

Despite the linking occurring indirectly, the noise in the linked series is generally lower than in the original series – this is due to the correction for periodization error, which was done on the series of persons at work. The effect of this correction has been transferred to the hourly series through the calibration. The series over hours worked have also been checked carefully in the unit working with the LFS.

Chart 48. Number of hours worked, men in the engineering industry (SNI69=38)



The linking has resulted in a drastic change in the seasonal pattern. The linked series has a more stable seasonal pattern than the original since the ratio between the noise after and prior to the linking is less than one (0.68). The seasonal pattern for the summer months is different for 1993 compared to previous years. This depicts a real change in the seasonal pattern – the industrial vacation is less concentrated to July in 1993.

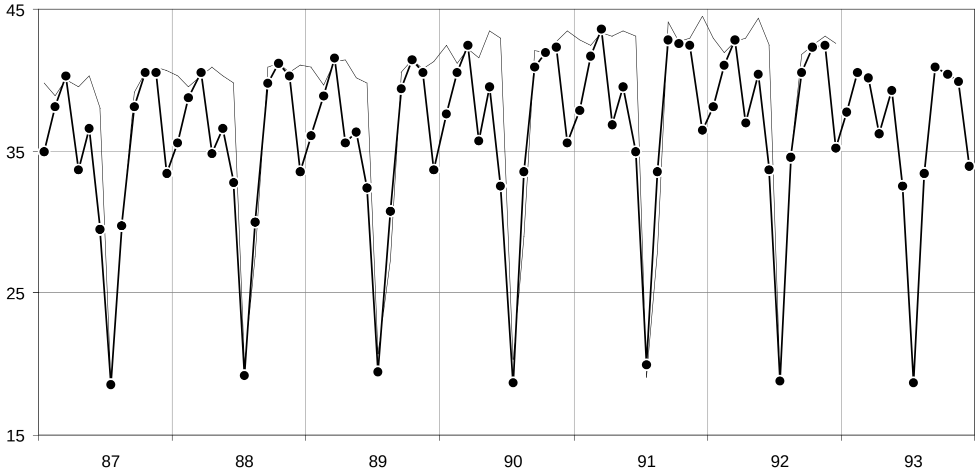
The linking covers: (Group 7)	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	After calibration

Ratios between noise for linked series and noise for original series

SNI69:	SNI 1	SNI 2-4	SNI 38	SNI 5	SNI 6	SNI 7	SNI 8	SNI 9	SNI 91, 931-934
Men	1.09	0.68	0.68	0.76	0.97	0.86	0.89	0.69	0.79
Women	1.10	0.91	1.06	0.95	0.86	1.01	0.98	0.63	0.67
Both sexes	1.12	0.74	0.70	0.77	0.85	0.82	0.92	0.64	0.63

Chart 49. Number of hours worked, both sexes in public administration, teaching, health and social care (SNI69=91, 931-934)

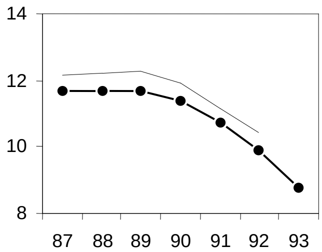
Millions (m) of hours per week Ratio 87-92: 0.63



The linking has resulted in the series having a radically changed seasonal pattern that tallies well with the pattern during 1993.

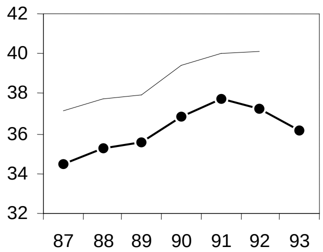
Number of hours worked, annual values, millions of hours per week

Chart 51. Men in SNI 38



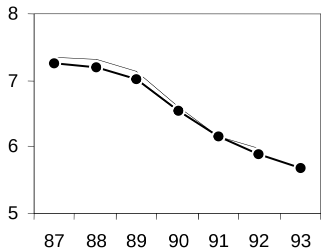
Level shift 1992: 0.54 millions

Chart 52. Both sexes SNI 91, 931-934



Level shift 1992: 2.84 millions

Chart 53. Both sexes in SNI 1



Level shift 1992: 0.08 millions

The annual series have lower levels as a result of the calibration changing the seasonal pattern. As a result of the linking, the public holiday and vacation effects reduce certain monthly values so that the annual average value is also lower. SNI 1, mainly Agriculture, is not affected by holydays.

Group 8: Employed by sex and detailed industry

For the seven previous groups of series, the linking worked so that no series had a ratio of over 1.15, i.e. the noise after the linking was at most 15 percent stronger than the noise prior to the linking. For this eighth group, however, there are several series in which the linking has increased the noise by more than 15 percent. We have therefore chosen to study this group in more detail.

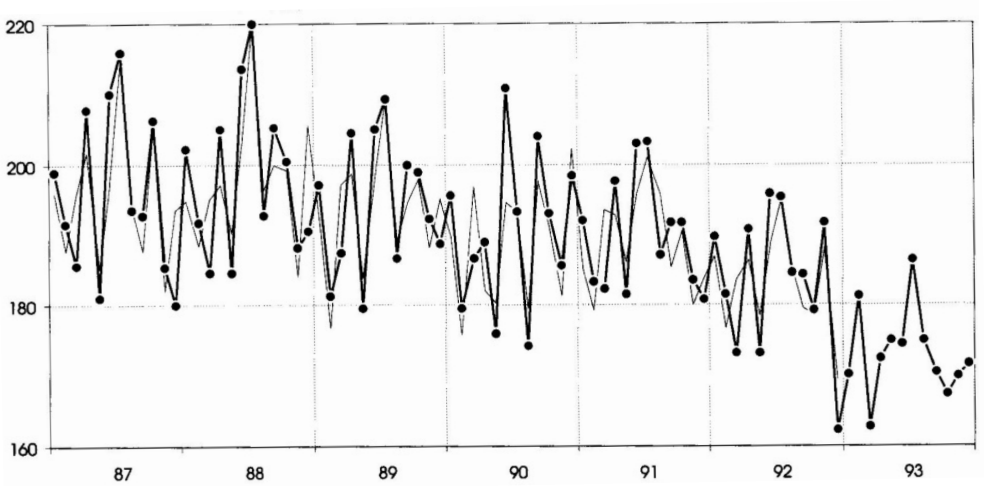
Group 2, employed persons by sex and age, could be linked in a satisfactory manner. This linking was performed on the time series level in Step 1. We therefore have complete control over the changes that occur in the linking. We have also linked persons at work by industry in Step 1. It was no problem to link this group either. For Group 8, employed persons by sex and industry, the linking is done by calibrating the weighting factors in Step 2. Since *employed persons = persons at work + absent persons* , the linking result for group 8 is dependent on how successfully the calibration managed to distribute absenteeism by industry.

The linking covers:	
Definition change	No
Correction for nonresponse error	Yes
Correction for periodization error	Yes
Correction for holiday effects	Yes
Linking on time series level or after calibration	After calibration

Ratios between noise for linked series and noise for original series

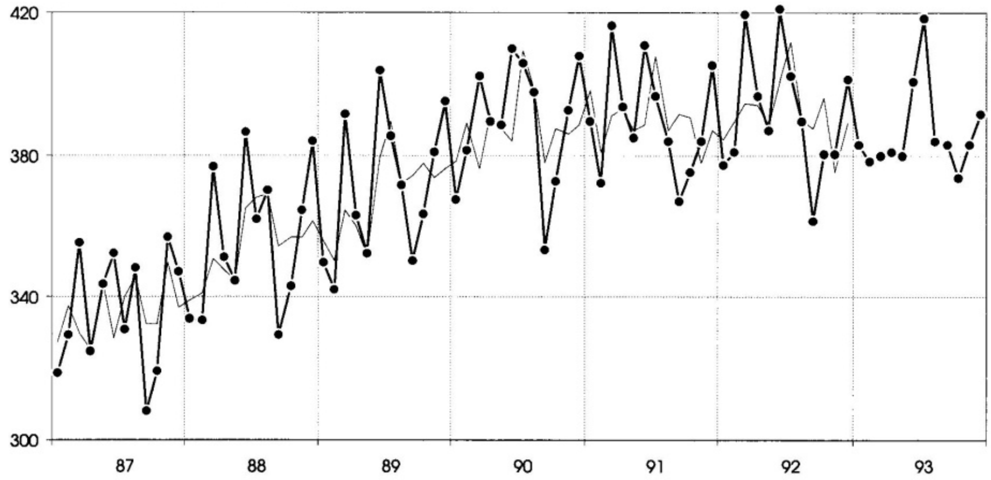
SNI69:	SNI 1	SNI 2-4	SNI 38	SNI 5	SNI 6	SNI 61	SNI 62	SNI 7	SNI 71	SNI 72
Men	1.06	1.15	1.09	1.16	1.16	1.28	1.11	1.17	1.19	1.20
Women	1.03	1.16	1.18	1.32	1.03	1.19	0.99	1.25	1.08	1.38
Both sexes	1.07	1.17	1.16	1.25	1.18	1.36	1.11	1.16	1.08	1.07
SNI69:	SNI 8	SNI 81	SNI 82	SNI 9	SNI 91	SNI 92	SNI 93	SNI 931	SNI 933	SNI 934
Men	1.20	1.27	1.10	1.31	1.29	1.16	1.24	1.48	1.11	1.04
Women	1.31	1.36	1.45	1.14	1.24	1.24	1.23	1.49	1.24	1.07
Both sexes	1.24	1.47	1.16	1.35	1.16	1.19	1.38	1.74	1.19	1.09

Chart 54. Number of employed persons, women in the retail trade (SNI69=62) 1000s



This series has an irregular pattern that is probably due to panel effects. The linking makes these panel effects slightly more obvious, probably as a result of the corrections for periodization that have been carried out. The panel effects can be seen e.g. in the autocorrelations in the series between values at three-monthly time intervals. The linking is to retain any panel effects, which has occurred in this case.

Chart 56. Number of employed persons, women in social care (SNI69-934)
1000s Ratio 87-92: 1.07



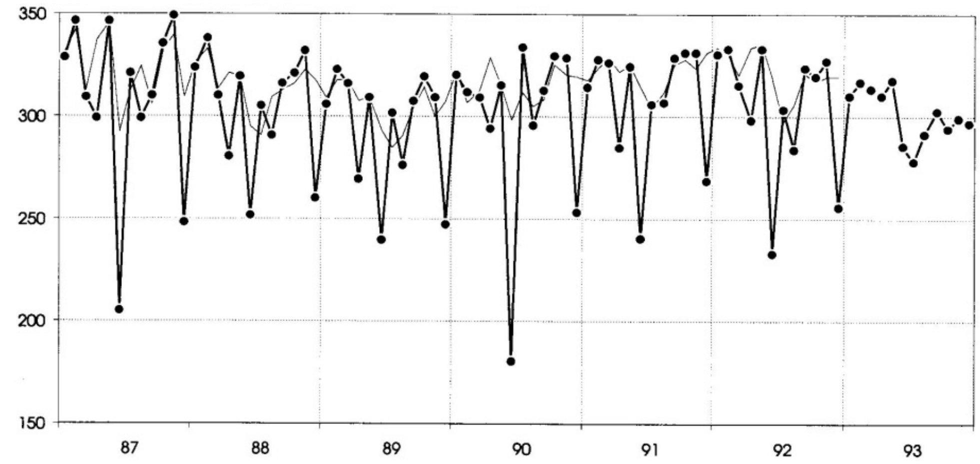
The linking has resulted in a changed seasonal pattern, but the new seasonal pattern does not tally well with 1993. This is not due to the linking method not having worked, but is due to the fact that the linking is based on the estimated discrepancies between new and old survey periods.

The adjacent table here shows these differences and that the linking should increase the March, June, November and December values and reduce the September and October values. Despite the linking having occurred indirectly via calibration, the seasonal pattern has been changed as intended. This is clear from Chart 56.

1993	Whole month, new measurement period	Two weeks, old measurement period	Difference
Jan	383.1	383.6	-0.5
Feb	378.1	381.1	-3.0
Mar	379.9	348.1	31.8
Apr	380.7	372.6	8.1
May	379 9	385.5	-5.6
June	400.5	363.6	36.9
July	418.3	428.1	-9.8
Aug	383.8	393.5	-9.7
Sept	382.8	402.0	-19.2
Oct	373.6	386.5	-12.9
Nov	382.7	371.6	11.1
Dec	391.7	372.2	19.5

Since the estimates of the discrepancies between new and old survey periods can be unreliable, the series patterns, as in this case, can be changed in a less desirable way, but for want of better estimates, we have to use these estimated discrepancies, which hold true for 1993. Finally, the series with the highest ratio between the noise after and prior to the linking is analysed.

Chart 63. Number of employed persons, both sexes in teaching (SNI69=931)
1000s Ration 87-92: 1.74



The linking of this series has failed. The corresponding series for men and women respectively also show the same incorrect pattern. It can be seen from Chart 63 that the linking of the monthly values for April, June and December has produced values that are too low. According to the table below, April 92 should, for example, be written down by about 16 thousand in order to correct for the new measurement period.

Number of employed persons, both sexes in teaching (SNI69=931), 1000s

	Whole month, new measurement period 1993	Two weeks, old measurement period 1993	Difference: New – Old 1993	Difference: Linked – Original 1992
Jan	310.0	304.3	5.7	-2.9
Feb	316.8	300.9	15.9	3.0
Mar	313.4	325.4	-12.0	-4.8
Apr	310 0	325.6	-15.6	-36.1
May	317.7	313.3	4.4	-2.0
June	285.9	313 4	-27.5	-85.4
July	277.8	266.0	11.8	6.6
Aug	291.5	289.4	2.1	-21.8
Sept	302.2	302.9	-0.7	2.3
Oct	294.0	289.9	4.1	3.0
Nov	299.4	284.3	15.1	7.1
Dec	296.6	317.3	-20.7	-63.8

To correct for non-response error, we should also reduce it by a further 1 percent, i.e. by about 3 thousand persons. The actual change according to the linking was a reduction of 36 thousand, i.e. 17 thousand too many. For the whole of 1992, the linking has given a value that can be estimated to be about 12 thousand persons or 3.7 percent too low.

We can thus ascertain that the series of employed persons in SNI 931 has been linked so that the level for certain months and for the whole year has come out too low. This is also true of the aggregate SNI 93 and 9 that contain series for SNI 931.

After checking the linking result for this group of series for employed persons by sex and industry, we can ascertain that it is series for SNI 931 teaching that have been poorly linked. As this failure was caused by low levels in these linked series, we can ask the question: Has this led to major level error in series in other industries? We have performed a calculation in the table below to illustrate this.

Analysis of levels in annual values for 1992

	Original value 1000s	Linked value 1000s	Desired linked value, 1000s	Wrong level of linked value 1000s	Wrong level of linked value Percent
All industries	4 250	4 208	4 208	0	0 %
Education, 931	321	305	317	-12	-3.7 %
All, except 931	3 929	3 905	3 891	12	+0.3 %

If the compensation for the fact that SNI 931 turned out too low affects all other industries, the level error in these should be insignificant. To check that certain industries don't have to compensate for the level error in SNI 931, we have compared the desired and the actual level change resulting from the linking for 1992 and all sub-series. Our check indicates that there are mainly four service industries that may have been given values that are too high during the linking. The linked annual values for 1992, relating to number of employed persons, both sexes, in SNI 6, 7, 8 and 933 are about 3 thousand too high for each industry respectively. All these errors are less than 1 percent of the level.

The problem with the series of employed persons in SNI 933 is due to the fact that the linking has not been successful with the weight calibration for absentees in this industry. The reason why the teaching sector should be so difficult to link may be because this sector has a special pattern of absenteeism. The linking had perhaps been more successful if we had had more calibration conditions for different SNIs, in addition to number at work, we may also have needed number of employed persons or number of absent persons for different industries as calibration conditions.

Summary

Sample surveys and the time dimension
Before the changeover in 1993, LFS was a sample survey in which probability samples of people were drawn in order to avoid systematic errors. Reference periods were not selected in this way, however. This leads to systematic errors in the time dimension. If we want to estimate *levels* without systematic errors, we must study all the weeks. *Change estimates* will also be affected by systematic errors if reference periods are selected using the old method.

The changeover done in LFS is therefore significant and represents a major improvement in quality. The same change should also be considered for other surveys.

Periodization error
A consequence of the survey period not being exactly the same for the same month in different years was considerable disruption in certain important time series. It was a surprise to us that this was such a significant error. Vi have previously observed interruptions in the seasonal pattern in time series of hours worked, but thought that this was due to changes in vacation agreements.

Periodization error should be given the same attention as sampling error and non-response error. The fact that this source of error is not dealt with sufficiently in many surveys is a serious matter.

Linking systems of series
Surveys must sometimes be revised. Both the RIDA changeover in 1986/87 and the changeover dealt with here in 1992/93 led to a significant improvement in quality of LFS. The downside of such changeovers is however the breaks in thousands of time series they cause. All these time series breaks make things extremely complicated for the users and our awareness of such complications slows down the introduction of necessary methodological changes.

The new method of linking systems of series that we have presented in this report could therefore be used when revising surveys. If we know that it is relatively easy to link all series, the reluctance to making methodological changes should disappear.

User-friendly statistics producers should take responsibility for the time series breaks they themselves have caused. Rectifying time series breaks is far too complicated for most users. The risk is that the user misinterprets data or doesn't dare to use the series.

Solving problems
When we took on the task of linking LFS for the 1992/93 break, we did not know the extent of it. We started with a straightforward time series approach and linked about one hundred series – those we call key series in previous sections of this report. But how can these results be used when producing other series? We managed to make consistent changes in a further 1000 or so series. But what about the other 29 000 series?

But once we became aware of how the individual data sets in LFS are used to produce optional series for different users, we saw the opportunity to work on the individual level. Thanks to being made aware of the new calibration technology through discussions with our colleagues, we realized that this methodology could also be applied to our purpose.

References

Box, G. E. P, Jenkins, G. M. (1970): *Time Series Analysis: Forecasting and Control.* Holden-Day.

Deville, J. C., Särndal, C. E. (1992): *Calibration Estimators in Survey Sampling.* Journal of the American Statistical Association.

Gómes,V., Maravall, M. (1994) Program TRAMO: *Time Series Regression with ARIMA Noise, Missing Observations and Outliers*. *Instructions for the User*. European University Institute.

Grundtabeller Arbetskraftsundersökningen December 1992, Statistics Sweden

Hörngren, J. (1992): *The Use of Data sets as Auxiliary Information in the Swedish Labour Force Survey*. R&D Report 1992:13, Statistics Sweden

Kmenta, J. (1971): *Elements of Econometrics.* Macmillan.

Liu, L., Hudak, G. et al. (1992): *Forecasting and Time Series Analysis Using the SCA Statistical System* Volume 1. Scientific Computing Associates Corp.

Maravall, M., Gómes,V. (1994) Program SEATS: *Signal Extraction in ARIMA Time Series*. *Instructions for the User*. European University Institute.

Statistics Sweden (1988): *Revision av innehåll och definitioner i arbetskraftsundersökningarna. RIDA Slutrapport.* Bakgrundsfakta till arbetsmarknadsstatistiken 1988:5.

Statistics Sweden (1993): *Länkningskoefficienter för AKU-resultat 1987-1992 avseende förändringar i uppräkningsförfarande och arbetslöshetsdefinitioner.* AM/AKU 1993

Standardavvikelser AKU Månad 1992 respektive 1993. Statistics Sweden

Särndal, C. E. (1995): Lecture in Helsinki

X-12-ARIMA Reference Manual. Bureau of the Census. Washington DC 1994

Appendix: Industrial Classification linking

From 1995 onwards, the new Swedish industrial classification SNI92 has been used in LFS. During the fourth quarter in 1994, both the old SNI69 and the new SNI92 were used in parallel. Using the information from this quarter, all individual data for surveys during the period 1987-94 have been supplemented with a code that corresponds to the new SNI92. By working in the individual level, all series relating to industry will be linked consistently.

Codes for SNI69 and SNI92
A total of 70 codes were used in SNI69 for the various industries, whereas in SNI92 the corresponding Chart is 64 codes. As a basis for the linking, the matrix for Quarter 4, 1994 is used with 70 * 64 cells.

It was clear from the matrix that it would be difficult for certain small industries to do a simple translation between SNI69 and SNI92. We therefore decided to simplify things by differentiating only 48 industries according to SNI92 for the period 1987-94. As an example of such a simplification, we can mention the food and drink industry and tobacco products. In LFS for the period 1987-94, there are three SNI69 codes:

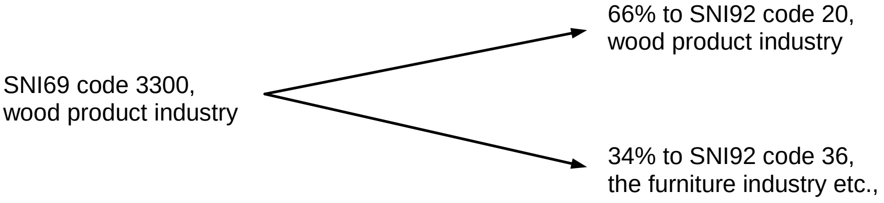
3111 Protected food industry
3113 Deregulated food industry
3130 Beverage and tobacco product manufacture

As from 1995, there are two SNI92 codes in LFS that correspond to these:

15 Food and beverages industry
16 Tobacco product manufacture

Our simplification in this case means that all individuals with the SNI69 codes 3111-3130 are given the SNI92 codes 15-16, I.e. the linked series will not differentiate between code 15 and code 16. By do similar simplifications, the 64 SNI92 codes were reduced to only 48.

Recoding models
Of the 70 SNI69 codes, 45 could be recoded to *one* specific SNI92 code. For the 25 remaining codes, models along the following lines were developed:



where the percentages 66% and 34% were retrieved from the matrix for Quarter 4, 1994. For each month during the period 1987-1994, those with the SNI69 code 3330 are randomly distributed between the SNI92 codes 20 and 36. These are distributed so that the new code 20 gets 66 percent of the estimated population volume for the old code 3300. For every month, a new randomization takes place, so that the same persons who reoccur can then be allocated different SNI92 codes in different months. Longitudinal studies of changes between industrial sectors should hence be conducted using SNI69 codes.

The quality of the recoding

Of the 48 codes for SNI92, we achieved good concordance between estimated and actual codes for the fourth quarter of 1994 apart from for 2 codes where the net errors were about 12 respondents. The concordance17 is presented in detail on the next page. For employed persons in agriculture, we have allowed SNI69 code 1100 to equal SNI92 code 1. The table below compares estimated and actual SNI92. The net error is 98.5-98.0 and the gross error is 2.0+1.5 thousand persons.

Agriculture, 1000s of persons	SNI92: 1	SNI92: not 1	Total
SNI69: 1100 = estimated sni92: 1	96.5	2.0	98.5
SNI69: not 1100	1.5	3 831.9	3 833.4
Total	98 0	3 833.9	3 931.9

[17] We had originally intended to perform another weight calibration to obtain linked series without breaks, but since the concordances were so good, we felt this work would be unnecessary.

Recoding of SNI69 with the model compared to SNI92 according to LFS for qtr 4 1994

Recoded SNI69 according to model		*SNI92 according to measurements in LFS*	
new SNI92 code	*number of persons*	***measured SNI92 code***	*number of persons*
01	98541	***01***	97992
02	27153	***02***	27338
05	3570	***05***	3288
Total	**129264**	***Total***	**128618**
10-14	9383	***10-14***	9316
15-16	65550	***15-16***	66488
17-19	21730	***17-19***	20553
20	39518	***20***	39143
21	45857	***21***	47274
22	66452	***22***	68516
23-24	31397	***23-24***	30793
25	25076	***25***	24965
26	19665	***26***	19810
27	37157	***27***	36511
28	75363	***28***	74381
29	98200	***29***	95796
30-33	83084	***30-33***	84325
34-35	92415	***34-35***	93347
36-37	31576	***36-37***	31032
40-41	32065	***40-41***	31908
Total	**774488**	***Total***	**774158**
Total 28-35	**349062**	***Total 28-35***	**347849**
45.1, 2, 5	138392	***45.1, 2, 5***	139118
45.3-4	87774	***45.3-4***	89010
Total	**226166**	***Total***	**228128**
50	73826	***50***	74560
51	185931	***51***	186676
52.1	86699	***52.1***	86422
52.2-5	146910	***52.2-5***	147312
52.6	5668	***52.6***	5628
52.7	6438	***52.7***	6550
Total	**505472**	***Total***	**507148**
60-63	177255	***60-63***	177191
64	85992	***64***	85121
Total	**263248**	***Total***	**262312**
65, 67.1	56446	***65, 67.1***	56536
66, 67.2	23790	***66, 67.2***	24558
70	66901	***70***	69277
71	7376	***71***	8111
72	39174	***72***	39474
74	202061	***74***	201977
Total	**395747**	***Total***	**399933**
73	30406	***73***	29565
80	297576	***80***	293305
Total	**327982**	***Total***	**322870**
85.1-2	369334	***85.1-2***	373129
85.321-322	168230	***85.321-322***	168039
85.311, 323	180892	***85.311, 323***	178141
85.312-315, 324-325	94563	***85.312-315, 324-325***	94898
Total	**813020**	***Total***	**814207**
55	93026	***55***	92722
90	9782	***90***	10116
91	57693	***91***	59140
92	94893	***92***	94596
93, 95	36262	***93, 95***	34580
Total	**291656**	***Total***	**291154**
75, 99	**201623**	***75, 99***	**201598**
missing data	**3275**	***missing data***	**1814**

Linking a system of time series

Between December 1992 and January 1993, the Swedish Labour Force Surveys (LFS) underwent major changes regarding definitions, estimation methods and measurement periods. The aim was to improve the quality of the surveys. As a result of these alterations, time series breaks occurred manifested both as changes in levels as well as changed seasonal patterns.

In order to increase comparability with old data, the monthly LFS for the period 1987–1992 have been recalculated.

About 30 000 time series are reported in the LFS system every month. All these series have been recalculated maintaining internal consistency and we have developed new methodology for this purpose. The new method uses the calibration technique that was introduced by Deville and Särndal 1992. However, the aim was not to reduce sampling and nonresponse errors – instead the aim was to link a system of time series so that all series remain consistent.

All official statistics can be found at: **www.scb.se**
Statistics Service, phone +46 8 506 948 01