Statistiska centralbyrån Statistics Sweden

Revising Surveys – Linking Old and New Data

2013:2

The series Background facts presents background material for statistics produced by the Department of Labour and Education Statistics at Statistics Sweden. Product descriptions, methodology reports and various statistic compilations are examples of background material that give an overview and facilitate the use of statistics.

Publications in the series Background facts on Labour and Education Statistics

2000:1	Övergång till yrkeskodning på fyrsiffernivå (SSYK) och införande av jobbstatus-kod i SCB:s lönestatistik
2000:2	The Information System for Occupational Injuries and the Work-related Health Problems Survey – A comparative study
2000:3	Konferens om utbildningsstatistik den 23 mars 2000
2001:1	Avvikelser i lönesummestatistiken – en jämförelse mellan LAPS och LSUM
2001:2	En longitudinell databas kring utbildning, inkomst och sysselsättning 1990–1998
2001:3	Staff training costs 1994–1999
2001:4	Studieresultat i högskolan i form av avklarade poäng
2001:5	Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)
2001:6	Svar, bortfall och representativitet i Arbetsmiljöundersökningen 1999
2001:7	Individ- och företagsbaserad sysselsättningsstatistik – en jämförelse mellan AKU och KS
2002:1	Tidsseriebrott i utbildningsregistret 2001-01-01
2002:2	En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–1999
2003:1	Exempel på hur EU:s ”Quality Reports” kan skrivas – avser Labour Cost Survey (LSC) 2000
2003:2	Förändrad redovisning av högskolans personal
2003:3	Individ- och företagsbaserad sysselsättningsstatistik – en fortsatt jämförelse mellan AKU och KS
2003:4	Sjukfrånvarande enligt SCB och sjukskrivna enligt RFV
2003:5	Informationssystemet om arbetsskador och undersökningen om arbetsorsakade besvär. En jämförande studie
2004:1	Samlad statistik från SCB avseende ohälsa
2004:2	Översyn av forskarutbildningsstatistiken. Bedömning av kvaliteten
2004.3	Sjukfrånvaro och ohälsa i Sverige – en belysning utifrån SCB:s statistik
2005:1	En longitudinell databas kring utbildning, inkomst och sysselsättning (LOUISE) 1990–2002
2005:2	Nordisk pendlingskarta. Huvudrapport
2005:3	Nordisk pendlingskarta. Delrapport 1–4.
2005:4	Flödesstatistik från AKU
2005:5	Flow statistics from the Swedish Labour Force Survey
2006:1	Sysselsättningsavgränsning i RAMS – Metodöversyn 2005

Continued on inside of the back cover!

Background Facts

Revising Surveys – Linking Old and New Data

Labour and Education Statistics 2013:2

Statistics Sweden
2013

Background Facts

Labour and Education Statistics 2013:2

Revising Surveys – Linking Old and New Data

Statistics Sweden
2013

Producer	Statistics Sweden, Population and Welfare Department BOX 24300 SE-104 51 STOCKHOLM
Inquires	Elisabet Andersson +46 8 506 946 45 elisabet.andersson@scb.se Hassan Mirza +46 8 506 946 72 hassan.mirza@scb.se

ISSN 1654-465X (Online)

URN:NBN:SE:SCB-2013-AM76BR1302_pdf

This publication is only available in electronic form on www.scb.se.

Preface

In April 2005, changes were made in the LFS questionnaire and in the derivation of the labour status variable. The reason for these changes was that the Swedish LFS should fully comply with the EU regulations. The changes led to an increase in both employment and unemployment, while the number of persons not in the labour force decreased compared to before.

In addition, the seasonal pattern in the series over the number of unemployed and in the unemployment rate was altered. Beginning October 2007, the Swedish LFS presented the number of full-time students, who met the conditions for being classified as unemployed, as unemployed. The changes from 2005 and 2007 caused a break in the time series, and thus a lack in comparability in almost all the series.

The work conducted to adjust data from January 1987 to March 2005 to attain comparable data with the period April 2005 and onwards was based on the nature of the changes. As a last step in the linking process, the respondents were given a new weight in each month of the period, whereby the entire LFS system has been linked on the microdata level.

The new classification standard, SNI 2007, which was based on NACE Rev 2, was introduced in January 2009. This change caused breaks in time series where the data are presented by branch of industry. Therefore, data for branch of industry for the period January 1987 – December 2007 have also been adjusted.

Anders Wallgren and Britt Wallgren from Örebro University were responsible for the methodology and are the authors of this report. The employees in the LFS unit conducted an extensive audit work and updated the LFS series.

The authors have also conducted a previous linking project in the LFS in relation to the changes of 1992. This work is documented in Wallgren and Wallgren (2012).

Statistics Sweden, April 2013

Inger Eklund

Hassan Mirza

Contents

1 The changes in 2005 and 2007

During April 2005 a new questionnaire was introduced in the Swedish Labour Force Survey. The intention was to harmonise the measurement methods with the standard that is used within the European Union. All labour status categories were affected by the change. As a result, it became difficult to compare data produced with the old and new measurement methods.

During October 2007 another change took place when the definition of unemployed was changed to be similar to the ILO definition. People studying full-time but looking for a job were previously defined as persons not in the labour force, but now they were defined as unemployed.

1.1 User demands

The Swedish Labour Force Survey plays a very important role in Sweden. It is used by many analysts in central government. As different analysts can use different variables, the number of time series that are important is very large. These users requested thousands of long monthly and quarterly time series, at least from 1987 and onwards.

The users also request that linked data are *consistent*, e.g. that linked (or backcasted) series for unemployed males + linked unemployed females = linked unemployed of both sexes.

1.2 Our aim

The manager of the Labour Force Survey gave us the following task:

Link all time series for the level shifts and new seasonal patterns that were created by the changes 2005 and 2007. All linked series should be consistent. Also link for revised ISIC and NACE during 2003 and 2009.

1.3 The seven changes in 2005 and 2007

We have adjusted old LSF data for the following seven changes:

- January 2003, Industry classification SNI 1992 was replaced by SNI 2002.
- April 2005, persons living in Sweden but working abroad are now classified as employed, earlier they were classified as not in the labour force.
- April 2005, persons waiting to start a new job within three months are now classified as unemployed, earlier persons waiting between five weeks and three months were classified as not in the labour force.
- April 2005, measurement method (questions in the questionnaire) used for classifying persons that are not students as unemployed was changed.
- April 2005, measurement method (questions in the questionnaire) used for classifying students as unemployed was changed.
- October 2007, new definition of unemployed. Persons studying full-time but looking for a job are now classified as unemployed; earlier they were classified as not in the labour force.
- January 2009, Industry classification SNI 2002 was replaced by SNI 2007.

1.3.1 Changes in 2005

In 2005, changes were made in the LFS in the questions in the questionnaire or in the derivation of labour force status. The aim of the changes was to comply fully with the EU regulations. The changes were in reference to:

1. Employed persons abroad: Up to March 2005, these persons were classified in the category not in the labour force. This practice did not comply with EU regulations. Beginning in April 2005, these people were classified as employed persons and thus in the labour force.

2. Unemployed persons: Up to and including March 2005, a person was classified as unemployed if the person was without work during the reference week, but had *applied* for work during the last four weeks (reference week and three weeks back in time) and was *willing* and *able* to work during the reference week or begin within 14 days from the end of the reference week. Unemployed persons also include persons who have work that will begin within four weeks (awaiting work).

In order to comply with EU regulations, the definition of unemployed persons was changed to: Persons who were without work during the reference week but who had *applied* for work during the last four weeks (reference week and three weeks back in time) and were *able* to work during the reference week or begin within 14 days from the end of the reference week. Unemployed persons also include persons who have work that will begin within three months. Thus for unemployed persons, the condition of *willing* to work was removed and included persons who have work that begins within three months, instead of four weeks as previously.

3. The structural variables, which are regulated by the EU, were collected through March 2005 at the end of the interview as a supplementary survey. Beginning April 2005, these variables were integrated into the basic interview and the questions were put in appropriate places in the questionnaire.

Points 1-2 and the changes related to full-time students resulted in level shifts and changed seasonal patterns in the time series.

1.3.2 Changes in 2007

Up to October 2007, persons who met all the criteria for unemployment were classified as not in the labour force if the person at the same time was engaged in full-time studies. This was contrary to ILO's recommendations and EU regulations.

1.4 The methodological challenge

Surveys must be revised now and then. How to handle these changes is a neglected part of survey methodology. Comparability over time is an important factor of survey quality that is not sufficiently discussed. A general methodology for survey revisions is needed and we discuss in this paper a number of different methods that can be used. This report gives a detailed description of the methods used for linking all time series in the LFS in taking account of the seven changes listed in Section 1.3.

The users need many series due to many subdivisions by labour status category, sex, age, region etc. Consistency is required between a large number of linked series that must be additive in many dimensions.

2 Different methods for linking

In this chapter we give a general background and compare different methods. When surveys are revised, a plan for the revision should be made and should include a plan for the work with linking old and new data. What kind of revision will be made? What kind of data must be available to be used as a basis for the linking? What kind of linking method will be used? These questions should be answered before the revision; otherwise, problems may arise and make the linking unnecessarily difficult. In sections 8.3 and 9.3 in Wallgren and Wallgren (2007), some methods for linking time series are discussed.

2.1 Different kinds of revisions

All surveys can be improved, and with better methods different kinds of survey errors can be reduced. Preconditions may change and as a result surveys must be revised. We list here different kinds of revisions that can lead to old and new data becoming not comparable:

1) *Different frame*
   Administrative systems change and this can lead to changes in the administrative registers used by the National Statistical Institute (NSI). Both the administrative population and administrative variables can change. The NSI can also improve the statistical registers that are used as frames. E.g. Statistics Sweden's Business Register is now being improved and the change will produce smaller coverage errors, but this will also give rise to time series level shifts in many economic surveys. Statistics Sweden's Population Register can also be improved if more sources are used, which will give level shifts in the Swedish LFS.
2) *New survey design*
   New ways of collecting data can be introduced, e.g. web questionnaires may be introduced. A new way of defining measurement periods is also a fundamental change of design. E.g. the Swedish LFS changed 1993 from one or two weeks each month into a system of measuring all weeks during the year. This gave rise to both new seasonal patterns and times series level shifts. This is discussed in Wallgren and Wallgren (2012).
3) *New definitions, new questionnaire and measurement methods*
   A questionnaire may be improved to reduce measurement errors; new definitions coming from new international regulations will also result in a revised questionnaire. This will generate time series breaks. In Section 3 we describe the effects of the changed measurement methods and definitions that were introduced in the Swedish LFS during 2005.
4) *Revised classifications*
   International classifications are changed. As society develops the old classifications are getting less relevant and are therefore revised. However, this generates times series breaks that are problems for the users. In Section 4 we describe how time series can be linked to compensate for the changed economic classification in the Swedish LFS 2003 and 2009.

5) *New estimation methods*
   New administrative sources can be used for calibration purposes and this will lead to a reduction in nonresponse errors. This will generate level shifts in many time series. This kind of improved estimation methods was implemented in the Swedish LFS 1993. This is discussed in Wallgren and Wallgren (2012).

2.2 Why should we link old and new data?

According to Statistics Sweden's model for quality declarations (2001), comparability over time is one of the quality indicators. However, this indicator is given a very short description without clear definitions of what one should do when a survey has been revised. The description in ESS (2012) is also short: *"Changes in methods are clearly identified and their impact measured to facilitate reconciliation... Breaks in the series are explained and methods for ensuring reconciliation over a period of time made publicly available."* Our interpretation of this statement is that it is sufficient to explain that the revision at point t in time produced a level shift of p% in the variable x.

These vague guidelines differ from the clear demands made by the users of LFS data in Sweden. The Ministry of Finance required that a large number of monthly and quarterly time series from January 1987- March 2005 should be linked with the new time series published from April 2005 and onwards. This means that it was not sufficient to tell the users in central government that the level shift for March 2005 was p%; the users wanted to know the linked time series values for each month during the period January 1987-March 2005. It took almost five years to produce the linked series, which created considerable irritation among the users.

Data from the Swedish LFS for about 18 years have been linked and these data represent production costs equivalent to about EUR 200 million. The linking contributes to a more frequent and improved use of these data.

LFS data are additive in many dimensions
LFS data are additive, e.g. *men* + *women* = *both sexes,* etc. This means that linked data also must be additive, otherwise there will be conflicting descriptions of the same reality; the sum of two series *men* + *women* is one description of both sexes and the series for *both sexes* is another. The users therefore not only wanted linked series, they also wanted *consistent* linked series without conflicting information.

The table below illustrates the problems that arise when we try to link old and new data that should be additive so that *employed* + *unemployed* = *persons in the labour force*. For March 2005 we have both old data and linked data. Assume that we only publish linked data for March 2005 and explain what happened 2005 and how the linked data for March were derived.

Chart 1. Linking data that are additive. Both sexes 16-19 years, the Swedish LFS.

1000s of persons		Original not linked data			New linked data			Ratios new/original			Linked with ratios		
		Em-ployed	Unem-ployed	In labour force	Em-ployed	Unem-ployed	In labour force	Em-ployed	Unem-ployed	In labour force	Em-ployed	Unem-ployed	In labour force
1987	Jan	**199**	**7**	**206**							**201**	**29**	**270**
2005	Mar	98	11	109	99	44	142	**1.01**	**4.00**	**1.31**	99	44	142

What shall the users do if they want linked data from January 1987 and onwards? If they use the ratios for March 2005 they will get conflicting estimates. For January 1987 *employed* + *unemployed* = 201 + 29 = 230

thousand persons in the labour force. But according to the direct estimate there are 270 thousand in the labour force.

This example shows that it is not sufficient to explain the difference between old and new data for one period of time – the users need statistics for all periods. The ratios for March 2005 are not relevant for earlier months as they give unreasonable linked values for e.g. January 1987.

The example in Chart 1 shows also that LFS data with many additivity conditions are difficult to link – how to get consistency in such a system of time series? A similar issue is discussed in Wallgren and Wallgren (2013) in relation to seasonal adjustment of LFS data. The structure of LFS data is the same with the same additivity conditions when you want to seasonally adjust data or when you want to link old and new data.

Different attitudes towards linking
We have earlier met managers and methodologists that were negative towards linking time series. They said that the producer should only give the difference between old and new data for one period of time and nothing more. If the users want to link backwards, it is their problem.

In our opinion, the producer is responsible for the revision and for linking old and new data. Linking time series requires competence regarding model-based estimation methods and good subject matter competence. This competence should exist at the NSI, where the knowledge about the survey in question is located together with the microdata from the survey.

The professional independence of the NSI should imply that the NSI should do the linking and that this linking will be accepted by different users. A political controversy on earlier unemployment figures based on linking done by different users would be the worst alternative.

Lack of linking methods stops improvements
There is one attitude we have met many times: "If we improve our survey there will be time series breaks and our users don't want that!"

Improvements cost money and can be difficult. The trouble with time series breaks is a seemingly excellent excuse to let improvements wait. Necessary improvements can be delayed for many years due to this excuse. If there is a lack of staff with the competence to link old and new time series data, then this would be a serious hindrance for improvements. Revisions do not occur so often, which could lead to competence not being available when a new revision is done.

2.3 Different linking methods

In this section we give an overview of four different methods that can be used when linking old and new time series. These methods are examples of model based estimation. The models used should be described and discussed as they determine the quality of linked data.

The main distinction between different groups of methods is if the work is done with *aggregated time series data* or if the linking work is done with *microdata.* The work with microdata consists of changing the individual weights so that new linked estimates can be produced with these new weights. There are two methods for this kind of linking work with microdata – if a set of linked time series are used as calibration conditions

or if relations with other variables in the survey are used when generating new individual weights. Finally, the work with a revised classification constitutes a methodological issue of its own.

2.3.1 Managing the break

How to manage the revision and interpret the new data? If a monthly survey is revised so that both levels and seasonal patterns change, the effects of the revision must be evaluated so that information on how important time series change can be reported even after the revision of the survey. If this is not done there will be loss of important information for at least three years – that is the shortest time needed to estimate the new seasonal patterns so that the new data can be interpreted. Before the revision is done it should be decided how the short-term comparisons will be made after the revision. And it should also be decided what kind of extra information is necessary, e.g. if a part of the sample should be according to the old survey design and the other part according to the new design.

The revision 1992/1993

This revision of the Swedish LFS was carefully planned. The two main parts of the revision were:

- *New estimation method*
  A new post stratification estimation was introduced in the LFS where a number of auxiliary variables from different registers were used. As a result, nonresponse bias was reduced. Microdata for 1992 were used both for generating estimates with the old and the new estimators. The basic output from the LFS consists of tables with about 15 000 cells; and both estimates were produced for all these cells. The following ratios between new estimates and old estimates were observed:

  Ratio new/old estimate of number of persons in each category for 1992

Employed at work	Employed absent	Unemployed	Not in labour force
0.987	0.987	1.050	1.046

- *New measurement periods*
  Instead of measuring two weeks without holidays every month a four or five week period as close to the calendar month was measured from 1993 and onwards. For each month during 1993, estimates were obtained for the old measurement period and the new four or five week period for all 15 000 cells in the basic tables. The following ratios between old and new measurement periods were observed:

  Ratio new/old measurement period of number of persons in each category

	Employed at work	Employed absent
January 1993	0.976	1.179
...	...	...
December 1993	0.928	1.670

For each month during 1993 it was thus possible to compare estimates with the old time series and when 1993 was complete, all information necessary to link old and new series was available.

2.3.2 Linking aggregated time series data

To link old and new time series data the new seasonal pattern and the level shift must be estimated. The linking will be easier if it can be assumed that

the seasonal pattern is not affected by the revision. It is an advantage if the new seasonal patterns and level shifts can be estimated as in Section 2.3.2.

It is common to use ratios to estimate the level shifts and backcast with these ratios. This will give inconsistent backcasts with LFS data that are additive in many dimensions, which is illustrated in Chart 1 in Section 2.2. These inconsistent backcasts can be made consistent with the regression technique that is explained in Wallgren and Wallgren (2013).

This method builds on the assumption that the estimated ratios are relevant not only at the time of the revision but also during the whole period with old data.

2.3.3 Linking with calibration conditions

This method was developed for the revision of the Swedish LFS 1992/1993. The method is described in detail in Wallgren and Wallgren (2012) which is based on an earlier report in Swedish. This method is suitable for revisions where the changes are *general*, e.g. a new survey design or new estimation methods. In these cases there are no defined groups that are changed in a specific way. The method consists of the following steps:

a) On the *time series level*, a number of key series are linked and the linkages are done so that the key series remain consistent. Consistency means that the total population remains unchanged and labour status categories, age classes and sexes and also 16 industries remain consistent with respective totals. The key series should be chosen so that all important changes are included. For each of these series, we obtained recalculated monthly values for the period 1987-1992, which are comparable to the corresponding new series from 1993 onwards.
b) 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 series defined by the 15 000 cells in the basic tables are linked with preserved internal consistency.
c) The results were carefully checked.

The new seasonal patterns and level shifts were estimated for the key series. Backcasting was done of the nonresponse errors observed for 1992 and the differences between two week and four week measurement periods observed in 1993. It was assumed that these observed errors and differences were relevant for the whole period 1987-1992.

The new time series for the 15 000 cells defined by the basic tables were accepted as revised official statistics for the period 1987-1992. All these series are consistent and follow the same additivity conditions as the original data. The checking revealed only one error: the calibration had missed the vacations for teachers; we should have had one more key series for this.

2.3.4 Linking with relations with other variables

This method was developed for the revision of the Swedish LFS 2005 and is described in detail in Section 3.

This method is suitable for revisions where the changes are *specific*; there are defined groups that are changed in specific ways, e.g. due to new

definitions or new questionnaire and measurement methods. The weighting factor of every respondent is also recomputed with this method. As a result, all series are linked with preserved internal consistency.

This method does not start with estimation of level shifts or new seasonal patterns; instead, it starts by finding models that explain the relations between the old and the new concepts. It is necessary to develop models that explain all important changes due to the revision. If these models are good, then levels and the seasonal patterns will be adjusted through the models. As a result of applying the models, old and new time series will be comparable.

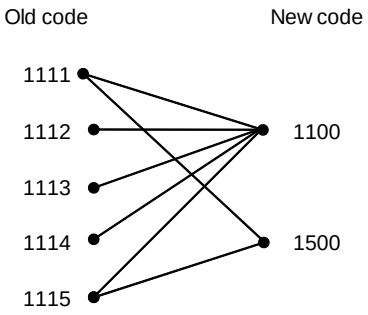
2.3.5 Linking a revised classification

This method was developed for the revision of the classification of economic activity and is described in detail in Section 4. The same method can be used for other kinds of classifications.

For one period of time, the statistical units in a register or a sample are classified according to both the old and new classifications. The old and new classifications are related by *one-to-many* or *many-to-many* relations, i.e. one old category can be replaced by a number of new categories.

Chart 2. Relation between old and new classifications

Old code	New code	Weight
01111	01100	0.94804
01111	01500	0.05196
01112	01100	1
01113	01100	1
01114	01100	1
01115	01100	0.92687
01115	01500	0.07313



In this example, five categories of the old classification are linked to the corresponding new categories.

The first and last are examples of *one-to-many* relations.

The whole table is an example of *many-to-many* relations.

In the LFS for 2008, all persons were coded according to both the old and new classifications and the weights were derived from this double coding.

The method that is described in Section 4 uses weights of the kind that are shown in Chart 2 for linking old and new classifications. To simplify calculations, database modelling is used.

3 Linking for a new questionnaire

At the end of 2009 we were requested to link for the revision of the Swedish LFS during 2005. The linking for the revision 1992/1993 had been successful and the same method had been tried for the revision 2005. However, the method that worked for 1993 did not work for 2005 – the linked series were in many cases unreasonable. We started the work with two questions:

– Why did the old method not work?

– What kind of method was suitable for the break for 2005?

The definition of unemployed was changed in 2007 to comply with the ILO definition. Persons studying full-time but looking for a job were now classified as *unemployed*. Earlier, they were classified as *not in the labour force*. This change was not a problem, as all information necessary to recode old data was available in all LFS surveys since 1987.

After recoding for this new definition, we started to analyse microdata to compare how people answered when interviewed with the old and the new questionnaire. We compared data with old questionnaire for the four quarters 2004Q2-2005Q1 with the new questionnaire for the period 2005Q2-2006Q1. We would have preferred to compare data from old and new questionnaires for the same panels of persons, but unfortunately it had been decided to start with new panels after the change in April 2005.

After analysing microdata we developed four models that were used to link old and new data. The final linking for a small subpopulation is shown in Chart 3. The ratios between new linked time series values and old values are quite different for January 1987 and March 2005 and this explains why the method developed for the revision 1992/1993 did not work for 2005. The linked series that were used as calibration conditions were reasonable for 2005 but not for 1987.

When the revision does not produce general changes to all, but instead specific changes that are different for different categories, it is impossible to find out how times series should be linked by only estimating level shifts and new seasonal patterns. The shifts and the new seasonal patterns are in accordance with the structure of the data near the change, but not necessarily with data from earlier periods. This means that a different kind of analysis is necessary for the revision 2005.

Chart 3. The final linking for the break 2005. Both sexes 16-19 years

1000s of persons		Original not linked data			New linked data			Ratios new/original		
		Em-ployed	Unem-ployed	In labour force	Em-ployed	Unem-ployed	In labour force	Em-ployed	Unem-ployed	In labour force
1987	Jan	199	7	206	199	21	220	**1.00**	**2.91**	**1.07**
2005	Mar	98	11	109	99	44	142	**1.01**	**4.00**	**1.31**

The following variables and questions in the LFS-questionnaire were used when microdata were analysed:

– Labour status category: employed/unemployed/not in labour force/ill

– Age, sex, main activity among persons not employed

– "Would you have wanted to work during the measured week?"

– Studying full-time or not

– Persons with no job but awaiting new job within 4 weeks/3 months

3.1 A model for employed abroad

Persons living in Sweden but working abroad are now classified as *employed*; before April 2005 they were classified as *not in the labour force*. This category is the simplest to link and we use this example to illustrate the general method. The variable *"Main activity among persons not employed"* contains categories defining persons working abroad. With this variable we can find persons employed abroad in all surveys January 1987-March 2005.

Chart 4a illustrates how the recalculations are done. There are different models for each combination of sex and age class. The reason is that we do not want to change the sex and age distributions in the surveys. In each subpopulation defined by sex and age class, there is one category with *givers* of weights and another category with *receivers* of weights.

The persons with *labour force status 5* (= not in the labour force) that are *employed abroad* are now going to be counted as employed (labour force status 2 or 3). But as we have no data describing these persons activities (no information of e.g. hours worked or NACE) we just cannot recode labour force status for these persons from 5 to 2 or 3. Instead, we create new individual weights and put their new weights = 0. The number of persons not in the labour force will then be reduced just in the way we want. This category of persons is in this case the *givers*.

As we want to increase the number of *employed persons,* we increase the individual weights for the persons that are employed according to the old classification. This category of persons is in this case the *receivers* of weights. In Chart 4a the time series values describing employed women 20-24 years during March 2005 is increased by 3.8%.

Chart 4a. Linking at the micro data level. Women 20-24 years, employed, March 2005

Age	Sex	Month	LF Status	Employed abroad	Hours worked	NACE	Original weight	New weight	Ratio new/original
20	2	200503	5	1	null	null	483.70	0.00	0
23	2	200503	5	1	null	null	483.70	0.00	0
24	2	200503	5	1	null	null	483.70	0.00	0
...									
21	2	200503	2	0	24	45110	490.92	509.55	1.038
23	2	200503	2	0	45	49410	579.76	601.76	1.038
23	2	200503	3	0	24	53100	371.98	386.10	1.038
...									

The sum of the weights given should be equal to the increased weights among the receivers. Chart 4b shows that there were 9 women 20-24 years employed abroad that gave weights summing to 4 353 to the 230 receivers. The new time series value increased from 114 730 employed women to 119 083 employed women according to the new definition of employed.

Chart 4b. Sums of individual weights	Original weights	New weights	Number in the sample
Employed, labour status category 2 or 3	114 730	119 083	230
Employed abroad, labour status category 5	4 353	0	9
Employed women 20-24 years	119 083	119 083	239

In this case the time series value for March 2005 was increased by 3.8% due to the linking. Each month in the series for women 20-24 years increased similarly.

Below are the increments for employed women 20-24 years during 1987 and 2004:

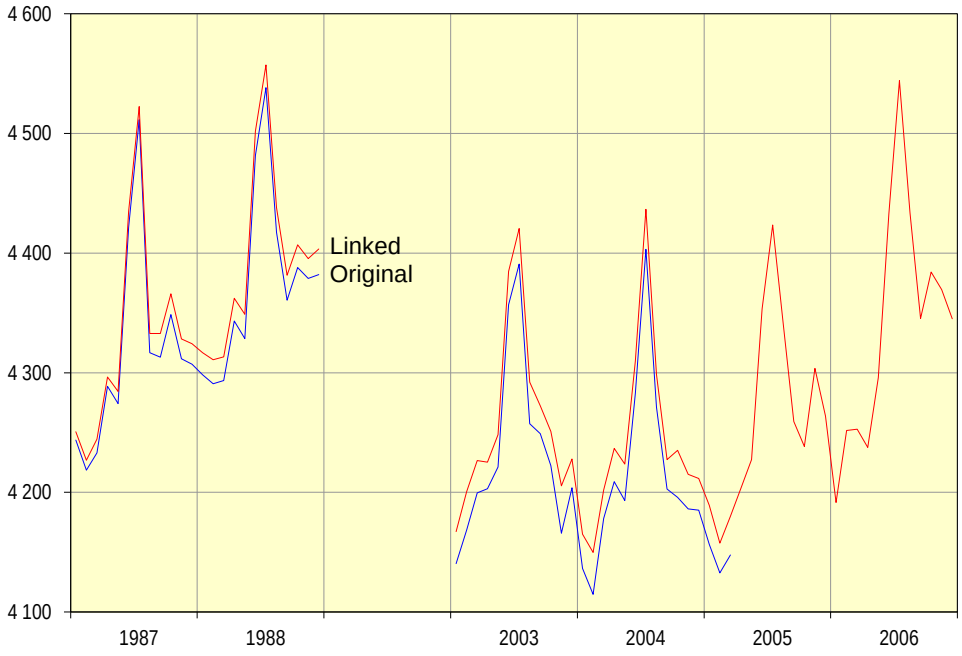
Chart 4c	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1987	1.005	1.008	1.013	1.006	1.007	1.010	1.001	1.002	1.018	1.011	1.003	1.015
2004	1.025	1.020	1.033	1.020	1.021	1.032	1.025	1.010	1.033	1.029	1.008	1.014

The aim of the recalculations in Chart 4 is to link time series so that the linked series are consistent. The most efficient way of doing this linking is to create revised individual weights in the set of microdata from each survey. In Chart 4 the recalculations for women 20-24 years are illustrated.

The individual weights in all subpopulations defined by sex and age class were recalculated in the same manner. With the original weights the original time series can be produced; with the new weights the linked series can be produced. In the following sections we use the same kind of recalculations of individual weights for different categories in the LFS.

In Chart 5 the original and linked series for all employed 16-64 years are shown for the years 1987-1988 and 2003-2006. It should be remembered that this linking of persons employed is perfectly correct as we had information on the number of persons employed abroad for the whole period. Other variables describing employed, such as hours worked, are linked in accordance with the variable employment. The model is not exact for such variables, the liked series are model-based and approximate. The model assumes that persons working abroad have the same patterns regarding e.g. hours worked by NACE as person working in Sweden. As persons working abroad is a very small category, this simple assumption can be used without generating disturbing errors.

Chart 5. Employed both sexes 16-64 years, original and linked series, 1000s



3.2 Models for persons waiting to start a new job

From April 2005, persons waiting to start a new job within three months are classified as *unemployed*. Earlier, persons waiting between five weeks and three months were classified as *not in the labour force*.

We must define one group of persons that are givers of individual weights and another group of persons that are receivers of weights. As we must reduce persons *not in the labour force*, this category is the *givers* in this model divided in subpopulations by sex and age. We decided that persons classified as *unemployed but waiting to start a new job within four weeks* are the *receivers* of individual weights in this model. We expect that this is the

group that is most similar to persons waiting to start a new job within five weeks-three months. The aim of this part of the work with linking time series is to increase the number of unemployed for this category.

To obtain a good model we must have data that show the relation between the old category (persons waiting to start a new job within four weeks) and the new category (persons waiting to start a new job within three months). This relation is estimated with data from March 2006-February 2007. During that period an extra question was included in the LFS interview. Those who answered that they are waiting to start a new job within three months got the question: *"Are you going to start that job within four weeks?"*

This illustrates that it is important to plan a revision beforehand and that there must be relevant data available to do the linking. In this case it was necessary to add an extra question to make the linking possible.

For the period January 1987-March 2005 the category *unemployed waiting for a new job within 4 weeks* are the receivers. These receivers are divided in two groups:

– Full-time students received an increased individual weight of 52%.
– All others received an increased individual weight with 33%.

Persons not in the labour force are the givers. Some categories not likely to take a job, such as persons with long time illness, etc., are excluded from the givers. The individual weights for the givers are reduced so that total weights are the same.

3.3 Models for unemployed not willing to take a job

From April 2005, questions in the questionnaire used for classifying persons as unemployed were changed. In Chart 6 the structure of the responses to the two questionnaires are compared. Three variables are tabulated: labour status category, student full-time yes/no and willingness to take a job. *Unemployment* and *not in the labour force* are measured in different ways. Among those who are not willing to take a job, we find new categories of unemployed in the white cells with 21.5 and 23.5 thousand.

Chart 6. Givers and receivers of individual weights to link unemployed, 1000s

Studying full-time						Others, not studying full-time					
Old questionnarie 2004Q2-20055Q1			**New questionnarie** 2005Q2-2006Q1			**Old questionnarie** 2004Q2-20055Q1			**New questionnarie** 2005Q2-2006Q1		
Category 1: "Would you have wanted to work during the reference week?" Did not get the question											
Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force
152.9	0.0	0.0	151.3	0.1	0.8	4 090.5	0.0	402.2	4 124.5	0.1	1.6
Category 2: "Would you have wanted to work during the reference week?" "Yes"											
Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force
0.0	**56.4**	55.1	0.0	66.0	59.4	0.0	**242.6**	52.9	0.0	243.4	100.8
Category 3: "Would you have wanted to work during the reference week?" "No"											
Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force
0.0	0.0	**416.0**	0.0	**21.5**	380.7	0.0	0.0	**274.8**	0.0	**23.5**	607.0
All categories 1+2+3			**All categories 1+2+3**			**All categories 1+2+3**			**All categories 1+2+3**		
Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force	Em-ployed	Unem-ployed	Not in labour force
152.9	56.4	471.1	151.3	87.6	440.9	4090.5	242.6	729.9	4 124.5	267.0	709.4

Category 1
The employed belong to category 1 both before and after the revision. Persons not in the labour force are handled differently. With the old questionnaire, all those who were not full-time students belonged to category 1; but with the new questionnaire they belong to category 2 or 3.

Category 2
The unemployed are almost the same both before and after the revision. After the revision, persons waiting for a new job within five weeks-three months are included.

Category 3
Due to the new categories that from April 2005 are defined as unemployed, unemployment among full-time students increased 32.6% (= 21.5/66.0) and among the others, unemployment increased 9.6% (= 23.5/243.4). How should time series regarding unemployment be linked for these changes?

Again we must define givers and receivers of individual weights (the arrows in Chart 6). The category not in the labour force should be reduced and unemployed should be increased. The *givers* are thus persons not in the labour force, but otherwise as similar as possible to the new category unemployed. The *receivers* are those already classified as unemployed. Givers and receivers should have the same distributions by sex and age.

Chart 7 shows that for persons studying full-time, almost all (= 15.9+2.4 of 21.5) in the new category with unemployed come from the categories *Studies* and *Job-seeker* of the variable *"Main activity among persons not employed"*. Among other persons, not studying full-time, 22.8 of 23.5 thousand of the new kind of unemployed come from the category *Job-seeker*.

Chart 7. Givers of individual weights specified

Studying full-time

Old questionnarie 2004Q2-20055Q1			**New questionnarie** 2005Q2-2006Q1		
Category 3: "Would you have wanted to work during the reference week?" "No"					
	Unem-ployed	Not in labour force		Unem-ployed	Not in labour force
Studies	0.0	369.7	Studies	15.9	380.7
Job-seeker	0.0	18.6	Job-seeker	2.4	0.0
Category 3	0.0	**416.0**	**Category 3**	**21.5**	380.7

Others, not studying full-time

Old questionnarie 2004Q2-20055Q1			**New questionnarie** 2005Q2-2006Q1		
	Unem-ployed	Not in labour force		Unem-ployed	Not in labour force
Job-seeker	0.0	41.0	Job-seeker	22.8	12.2
Category 3	0.0	**274.8**	**Category 3**	**23.5**	607.0

The final linking by adjusting individual weights can now be done. The 56.4 thousand full-time students that are unemployed according to the old questionnaire in Chart 6 should be increased by 32.6%, i.e. by 18.8 thousand. The individual weights should be taken from the 369.7+18.6 thousand with the main activity *Studies* or *Job-seeker.*

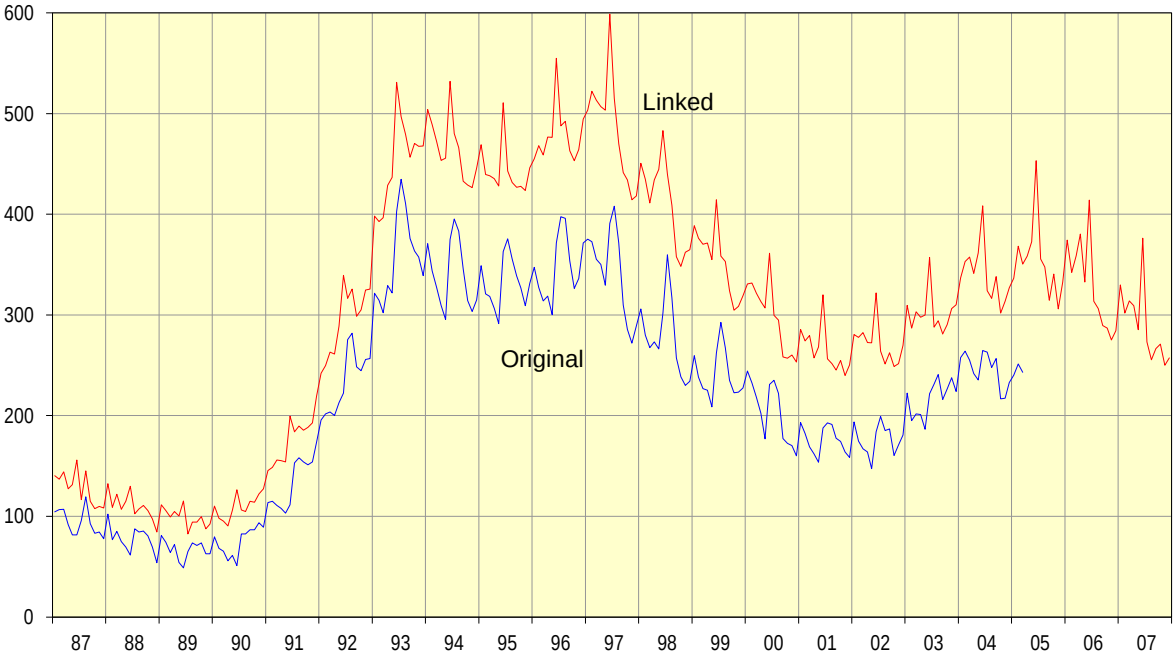
In a similar way, the 242.6 thousand of other persons (not full-time students) that are unemployed according to the old questionnaire should be increased by 9.6%, i.e. by 23.3 thousand. The individual weights should be taken from the 41.0 thousand with the main activity *Job-seeker.*

The ratios used to adjust individual weights should be monthly as we must link to get revised seasonal patterns. The table below is an example for one sex-age category. Note that summer months differ from the other months.

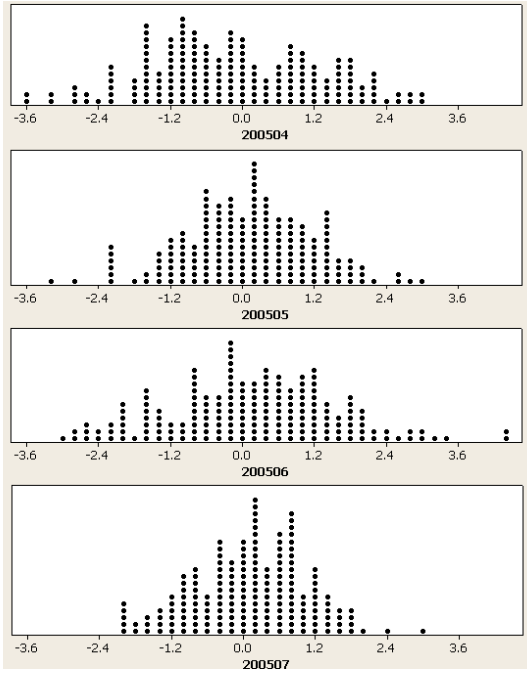
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0.021	0.024	0.044	0.030	0.065	0.294	0.102	0.025	0.011	0.013	0.023	0.039

The original and linked series are illustrated in Chart 8. Notice that both level and seasonal pattern have been changed.

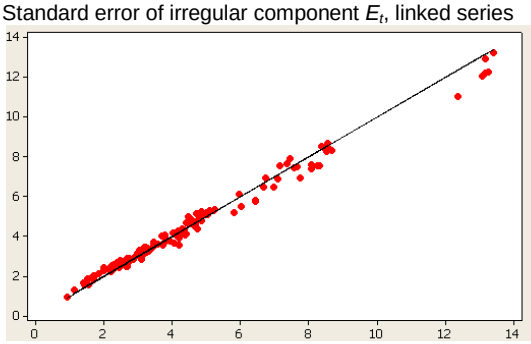
Chart 8. Unemployed both sexes 16-64 years, original and linked series, 1000s



Standardised ARIMA-residuals of 180 linked LFS-series



Standard error of irregular component E_t, linked series



Standard error of irregular component E_t, original series

The quality of the linked series

The quality of the linked series can be measured in different ways.

If levels and/or seasonal patterns are not well linked, then the standardised ARIMA-residuals after the break will be extreme. In the charts to the left, these ARIMA-residuals are plotted for 180 series with *employed*, *unemployed* and persons *not in the labour force* by sex and age class. We judge these residual plots to be acceptable and that levels and seasonal patterns seem to be linked so that they agree with the revised series that started April 2005.

In the last chart, we have plotted the standard errors of the X12-ARIMA estimates of the irregular components of these 180 series. In the chart we can compare the standard error of the original series with the linked series. The chart shows that the linking did not increase the standard error of the irregular component.

4 Linking for new classification

From January 2009 economic activity according to SNI 2007 based on NACE Rev 2 was introduced in the Swedish LFS. Industry classification SNI 2002 was then replaced by SNI 2007 and a similar revision took place January 2003 when Industry classification SNI 1992 was replaced by SNI 2002.

Economic activity or industry classification is important for all economic statistics and the variable is revised repeatedly. When the revision is extensive, as the revision 2009, this generates level shifts in all economic time series. How should these time series be linked? That is the issue of this section. In Wallgren and Wallgren (2012) we linked industry classification SNI 1969 and SNI 1992 in the Swedish LFS with a simpler method than the method we use here. The present method uses data more efficiently. The present method was introduced in Wallgren and Wallgren (2007).

4.1 What data should be used?

For one period of time, most suitably a year, the statistical units in the survey are classified both according to the old and the new classification. With this information the statistical units in earlier surveys can be classified according to the new classification and linked time series can be produced.

When a new version of industry classification is introduced in the Swedish LFS, all employed persons in the samples during one year are classified according to both the old and the new classifications. During 2008, SNI 2002 and SNI 2007 can be compared. The comparison can be based on number of employed person by industry or number of hours worked by industry. These frequency distributions are rather similar and we have chosen to base the linkage on employed persons by industry.

Chart 9. Employed for each old group by the new classification, percent 2008

SNI 2007	SNI 2002 A-B	C-E	Kpart+M	F	G+I	H+O+P	J+Kpart	L	N
A	**95.19**	0.02	0.00	0.10	0.03	0.06	0.05	0.00	0.00
B-E	0.11	**94.61**	0.32	0.32	0.12	3.41	0.24	0.00	0.00
F	0.00	0.12	0.00	**98.49**	0.30	0.03	0.16	0.02	0.00
G	0.23	0.23	0.03	0.08	**66.22**	0.05	0.25	0.00	0.00
H	0.00	0.06	0.00	0.00	28.18	0.03	0.04	0.00	0.00
I	0.00	0.00	0.00	0.00	0.01	37.22	0.03	0.00	0.00
J	0.14	3.97	0.02	0.06	2.69	5.07	15.10	0.02	0.00
K-N	4.10	0.70	9.26	0.76	1.92	0.68	**83.51**	0.21	0.67
O-U	0.00	0.00	0.01	0.01	0.00	0.10	0.00	**99.69**	0.07
P	0.00	0.00	**90.32**	0.09	0.02	0.71	0.04	0.03	0.04
Q	0.00	0.00	0.01	0.00	0.00	0.06	0.03	0.02	**98.95**
R-T	0.16	0.27	0.00	0.00	0.50	**52.58**	0.48	0.00	0.00
Unknown	0.07	0.02	0.04	0.08	0.00	0.02	0.06	0.00	0.27
All	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Chart 9 illustrates how the time series reported with SNI 2002 differ from the series reported with SNI 2007 in the Swedish LFS. The level shifts are large for some series. The linkage we have done is based on similar weights describing the relation between old and new classification on as detailed level as possible.

4.2 The method illustrated with a small example

Let us assume that we have the following microdata from Labour Force Surveys from three years for a small region. Each dataset contains the employed persons and three variables: *identity numbers*, the *individual weights* and *economic activity*. During year 1, economic activity is classified according to SNI 01; during the second year a new industry classification is introduced, SNI 02, and all persons in the LFS are classified according to both the new and old systems. During the third year, a new industry classification is again introduced, SNI 03, and also here all employed persons in the LFS for year 3 are classified according to both the old (SNI 02) and new (SNI 03) systems.

Chart 10a. Microdata for all employed in the surveys for year 1, 2 and 3

Year 1		
PIN	Weights	SNI 01
10	263.4	2
11	315.4	1
12	465.4	1
13	425.4	2
14	399.6	2
15	331.4	1
16	424.4	2
17	399.9	2
18	363.9	1
19	370.1	2

Year 2			
PIN	Weights	SNI 01	SNI 02
20	372.2	2	B
21	352	1	A
22	498.9	2	A
23	474.4	2	B
24	354.3	2	B
25	345.6	1	A
26	431.7	2	A
27	341.7	1	A
28	376.5	2	B

Year 3			
PIN	Weights	SNI 02	SNI 03
30	449.5	B	II
31	407.2	B	III
32	414.9	B	II
33	242.7	A	I
34	521.9	B	II
35	277.1	B	III
36	397.1	A	II
37	380.4	B	III
38	220.3	A	I
39	353.8	A	I

In Chart 10b the statistics describing employed persons by industry are shown. The estimates are derived as the sums of the individual weights in Chart 10a by industry. How can the time series for employed by industry be linked for the revised industry classifications?

Chart 10b. Employed persons by industry, year 1-3

Year 1	
SNI 01	Employed
1	1 476
2	2 283
All	3 759

Year 2			
SNI 01	Employed	SNI 02	Employed
1	1 039	A	1 970
2	2 508	B	1 577
All	3 547	All	3 547

Year 3			
SNI 02	Employed	SNI 03	Employed
A	1 214	I	817
B	2 451	II	1 783
All	3 665	III	1 065
		All	3 665

When there are *one-to-one* relations as between code 1 in SNI 01 and code A in SNI 02, then we have no problems regarding linking. But when there are *one-to-many* relations as between code 2 in SNI 01 and codes A and B in SNI 02 then we must have weights (e.g. 931/2508=0.371) that tell how often code 2 becomes A and how often code 2 becomes B. The cross tabulations in Chart 10c of the two industry classifications give the following weights that can be used to link SNI 01 and SNI 02 and also the weights for linking SNI 02 and SNI 03.

Chart 10c. The weights used to link old and new industry classifications

Year 2				
SNI 01	SNI 02	Employed	Sum	Weight w_1
1	A	1 039	1 039	1
2	A	931	2 508	0.371
2	B	1 577	2 508	0.629

Year 3				
SNI 02	SNI 03	Employed	Sum	Weight w_2
A	I	817	1 214	0.673
A	II	397	1 214	0.327
B	II	1 386	2 451	0.565
B	III	1 065	2 451	0.435

The two tables in Chart 10c contain all combinations of old and new codes. The weights sum to 1 for each old code, e.g. the weights for the two

combinations year 2 for the old code 2 in SNI 01 sum to 0.371+0.629=1 and the two weights year 3 for the old code A in SNI 02 sum to 0.673+0.327=1 .

The next step is to match the dataset with microdata regarding year 1 in Chart 10a with the table with weights based on data for year 2 in Chart 10c. The *linking variable* should be the *old classification* SNI 01. After this matching we get the dataset regarding year 1 in Chart 10d. The statistical units in this dataset are not persons; the units are *combinations* of old and new codes. The first person with identity number PIN10 is represented with two records for the two combinations SNI 01 = 2 and SNI 02 = A and SNI 01 = 2 and SNI 02 = B. These two combinations have weights 0.371 and 0.629 that sum to 1.

Chart 10d. Linking classifications, SNI 01 and SNI 02 for the LFS year 1

Year 1 from Chart 10a			Year 2 from Chart 10c			
PIN	Individual w	SNI 01	SNI 01	SNI 02	Weight w_1	Individual w · w_1
10	263.4	2	2	A	0.371	97.8
10	263.4	2	2	B	0.629	165.6
11	315.4	1	1	A	1	315.4
12	465.4	1	1	A	1	465.4
13	425.4	2	2	A	0.371	157.9
13	425.4	2	2	B	0.629	267.5
14	399.6	2	2	A	0.371	148.3
14	399.6	2	2	B	0.629	251.3
15	331.4	1	1	A	1	331.4
16	424.4	2	2	A	0.371	157.5
16	424.4	2	2	B	0.629	266.9
17	399.9	2	2	A	0.371	148.4
17	399.9	2	2	B	0.629	251.5
18	363.9	1	1	A	1	363.9
19	370.1	2	2	A	0.371	137.4
19	370.1	2	2	B	0.629	232.7

The first three columns in Chart 10d come from the original dataset for the LFS for year 1 in Chart 10a. The next three columns come from the table with weights in Chart 10c. The last column with the original individual weights multiplied with the weights w_1 for linking SNI 01 and SNI 02 has been created after the linking of these two sets of data. This column is used for the final estimates of employed persons by SNI 02 in Chart 10e.

It should be noted that the database table in Chart 10d has been created by linking two database tables. The linking has not been done by the usual matching with identity numbers PIN. As the aim of the linking is to link time series with different industry classifications, the database tables have instead been linked with the old classification SNI 01. The matched dataset then gives comparisons between SNI 01 and SNI 02.

If the last column in Chart 10d is summed up by industry SNI 02, we obtain employed by SNI 02 for year 1 in Chart 10e.

Chart 10e. Employed persons by industry, year 1

Year 1		Year 1 SNI 02 linked	
SNI 01	Employed	SNI 02	Employed
1	1 476	A	2 323
2	2 283	B	1 436
All	3 759	All	3 759

As industry classifications are revised repeatedly, we must also have methods to link *linked* time series with a new classification. The method is

the same but now we link the database table in Chart 10d with the table with weights for linking the code for SNI 02 and the new codes for SNI 03 in Chart 10c.

The next step is to match the dataset with microdata regarding year 1 in Chart 10d with the table with weights based on data for year 2 in Chart 10c. The linking variable should be the old classification *SNI 02*. After this matching we get the dataset regarding year 1 in Chart 10f below. The first person with identity number PIN10 is now represented with four records for the four combinations

- SNI 01 = 2, SNI 02 = A, SNI 03 = I
- SNI 01 = 2, SNI 02 = A, SNI 03 = II
- SNI 01 = 2, SNI 02 = B, SNI 03 = II
- SNI 01 = 2, SNI 02 = B, SNI 03 = III

If the weights w_1 and w_2 are multiplied, the four weights for the person with identity number PIN 10 will sum up to 1. If all three weights are multiplied the four weights *Individual w* $\cdot w_1 \cdot w_2$ will sum up to the original individual weight 263.4 as they should.

Chart 10f. Linking classifications, SNI 02 and SNI 03 for the LFS year 1

The first step – linking SNI 01 and SNI 02						Linking SNI 02 and SNI 03			
PIN	Individual w	SNI 01	SNI 02	Weight w_1	Individual w $\cdot$ w_1	SNI 02	SNI 03	Weight w_2	Individual w $\cdot$ w_1 $\cdot$ w_2
10	263.4	2	A	0.371	97.8	A	I	0.673	65.8
10	263.4	2	A	0.371	97.8	A	II	0.327	32.0
10	263.4	2	B	0.629	165.6	B	II	0.565	93.7
10	263.4	2	B	0.629	165.6	B	III	0.435	72.0
11	315.4	1	A	1	315.4	A	I	0.673	212.3
11	315.4	1	A	1	315.4	A	II	0.327	103.1
12	465.4	1	A	1	465.4	A	I	0.673	313.2
12	465.4	1	A	1	465.4	A	II	0.327	152.2
13	425.4	2	A	0.371	157.9	A	I	0.673	106.3
13	425.4	2	A	0.371	157.9	A	II	0.327	51.6
13	425.4	2	B	0.629	267.5	B	II	0.565	151.3
13	425.4	2	B	0.629	267.5	B	III	0.435	116.2
14	399.6	2	A	0.371	148.3	A	I	0.673	99.8
14	399.6	2	A	0.371	148.3	A	II	0.327	48.5
14	399.6	2	B	0.629	251.3	B	II	0.565	142.1
14	399.6	2	B	0.629	251.3	B	III	0.435	109.2
15	331.4	1	A	1	331.4	A	I	0.673	223.0
15	331.4	1	A	1	331.4	A	II	0.327	108.4
16	424.4	2	A	0.371	157.5	A	I	0.673	106.0
16	424.4	2	A	0.371	157.5	A	II	0.327	51.5
16	424.4	2	B	0.629	266.9	B	II	0.565	150.9
16	424.4	2	B	0.629	266.9	B	III	0.435	116.0
17	399.9	2	A	0.371	148.4	A	I	0.673	99.9
17	399.9	2	A	0.371	148.4	A	II	0.327	48.5
17	399.9	2	B	0.629	251.5	B	II	0.565	142.2
17	399.9	2	B	0.629	251.5	B	III	0.435	109.3
18	363.9	1	A	1	363.9	A	I	0.673	244.9
18	363.9	1	A	1	363.9	A	II	0.327	119.0
19	370.1	2	A	0.371	137.4	A	I	0.673	92.5
19	370.1	2	A	0.371	137.4	A	II	0.327	44.9
19	370.1	2	B	0.629	232.7	B	II	0.565	131.6
19	370.1	2	B	0.629	232.7	B	III	0.435	101.1

If the last column in Chart 10f is summed up by industry SNI 03, we get employed by SNI 03 for year 1 in Chart 10g below.

Chart 10g. Employed persons by industry, year 1

Year 1 SNI 01	
SNI 01	Employed
1	1 476
2	2 283
All	3 759

Year 1 SNI 02 linked	
SNI 02	Employed
A	2 323
B	1 436
All	3 759

Year 1 SNI 03 linked	
SNI 03	Employed
I	1 563
II	1 571
III	625
All	3 759

With this small example we have illustrated how revised classifications can be linked by using individual weights and weights for combinations of old and new categories. Statistical registers can also be linked in this way; here the individual weights are equal to one for all statistical units in the register.

There are many advantages with this method; all series are linked so that internal consistency is preserved. Not only different economic activities are additive; all series that are reported by industry remain consistent – hours worked, number of persons employed, number of persons at work, persons absent from work, hours worked due to overtime, etc.

The computations to do this linking are also easy – in a simple database model the original dataset is combined with a small database table with the weights for different combinations of the old and the new codes. These weights can also be modified to adjust the linking. If e.g. one industry did not exist in earlier years, the weights can be adjusted so that this industry is not linked backwards.

4.3 Linking industry classification in the Swedish LFS 1987-2009

During the period 1987-2009 the industry classification in the Swedish LFS was changed four times:

– 1987-1994 one system of codes based on SNI 1992
– 1994-2002 one system based on more detailed SNI 1992
– 2003-March 2005 based on SNI 2002
– April 2005-2008 based on SNI 2002 at the five-digit level

During 2008 all employed were coded according to both SNI 2002 at the five-digit level and SNI 2007 also at the five-digit level.

For each of these changes, we created tables with the relations between old and new codes based on microdata from the LFS where both codes were used for the same persons. As these tables have sampling errors, we tried to use weights for combinations of codes for SNI 2002 and SNI 2007 that had been created by the Swedish Employment Register.

The Employment Register used the industry codes from the Swedish Business Register where all enterprises were coded according to both SNI 2002 and SNI 2007. Based on these SNI codes, weights based on number of employed persons were produced with the Employment Register.

However, when we used these weights there appeared level shifts between 2007 and 2008 in the LFS series. The explanation to this is that the industry codes in the Business Register and the LFS are not coordinated.

We therefore decided to use weights based on double coded data from the LFS.

We did the linking in four steps starting with the oldest data:

1. Linking of data based on SNI 1992 with SNI 2002 for 1987-1994. The chart below shows the *one-to-many* links between old and new codes.

2. Linking of data for the period 1987-2002. This linking had to be done to link the special codes used in the LFS with the correct codes for SNI 2002. The grouping of economic activity was quite different during the period 1987-1994 and the period 1994-2002.

3. Linking of data for the period 1987-March 2005 to adjust for the change from SNI 1992 to SNI 2002. Some examples of the weights used are in the chart below.

4. The final linking for the period 1987-2007 to adjust for the change from SNI 2002 to SNI 2007. The linking was done to SNI 2007 at the three-digit level. Some examples of the weights used are in the chart below.

1987-1994

LFS 1987	SNI 2002	Weight
82	85320	0.12
82	80101	0.88
83	85310	0.664
83	85320	0.336
84	85310	0.686
84	85320	0.314

2003-March 2005

LFS 2003	SNI 2002	Weight
92	59600	0.14053
92	62630	0.01318
92	74000	0.00413
92	78100	0.0025
92	85560	0.01939
92	90000	0.21708
92	91000	0.1713
92	92930	0.43189

April 2005-2007

SNI 2002	SNI 2007	Weight
92621	85500	0.04231
92621	93100	0.95769
92622	93100	1
92623	85500	0.24712
92623	93100	0.64716
92623	93200	0.10572

References

ESS (2012): *Quality Assurance Framework*.

Statistics Canada (2009): *Quality Guidelines*. Fifth Edition – October 2009.

Statistics Sweden (2001): *Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik (in Swedish).*

Wallgren, A., Wallgren, B. (2007): *Register-based Statistics – Administrative Data for Statistical Purposes.* John Wiley & Sons Ltd, Chichester, England.

Wallgren, A., Wallgren, B. (2012): *Linking a system of time series – Recalculation of Labour Force Surveys 1987-1992.* Research and Development – Methodology reports from Statistics Sweden 2012:2.

Wallgren, A., Wallgren, B. (2013): *Consistent Seasonal Adjustment and Trend-cycle Estimation.* Background facts on Labour and Education Statistics 2013:3, Statistics Sweden.

2007:1	Några väsentliga sysselsättningsbegrepp i den officiella statistiken
2007:2	Registerbaserad aktivitetsstatistik
2008:1	Ungdomar utan fullföljd gymnasieutbildning – en undersökning med många utmaningar
2009:1	Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2007
2010:1	Principiella grunder för Arbetskraftsundersökningarna (AKU) och arbetsmarknadsstatistiken
2010:2	Rekryteringsstatistik från AKU
2011:1	Basic principles for Labour Force Surveys (LFS) and labour market statistics
2011:2	Recruitment Statistics for the Swedish Labour Force Survey
2011:3	Arbetskraftsundersökningarna (AKU) 50 år. Fyra forskarperspektiv på arbetsmarknaden
2011:4	Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2009
2011:5	Yrkesregistret med yrkesstatistik. En beskrivning av innehåll och kvalitet
2011:6	Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)
2011:7	Konjunkturberoende i inflödet till och utflödet från högre studier
2012:1	Actual Hours Worked in the Swedish LFS. Four articles
2013:1	Ungdomsarbetslöshet – jämförbarhet i statistiken mellan ett antal europeiska länder

ISSN 1654-465X (Online)

All officiell statistik finns på: **www.scb.se**
Statistikservice: tfn 08-506 948 01

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