

# A numeric comparison between the Swedish LFS and statistics based on administrative sources

## Abstract

In 2022, Statistics Sweden launched new statistics in the form of Population by Labour market status (BAS) with the aim to provide information on the supply of labour in Sweden per month, quarter, and year. BAS provides statistics based on administrative data on the number of employed, unemployed and people outside the labour force for the population aged 15 to 74 years. BAS is available on a monthly basis from January 2020. Producing BAS is mainly made possible by the introduction of monthly administrative data on paid salaries and tax deductions reported from employers to the tax authority.

While there are similarities between the Swedish Labour force survey (LFS) and BAS, important differences also exist concerning target variables and how these are created. Initial analysis has been carried out to gain more information on the numeric differences and similarities between LFS and BAS. These analyses include employed, unemployed and outside the labour force with a focus on employed and unemployed divided by sex, age, and country of birth (born in Sweden/born abroad).

For employed, the levels and developments over time are relatively similar although the level for BAS is lower than the corresponding level for LFS for most of the months. For the population aged 15 to 74 years, the largest difference obtained is 124 000 employed. This corresponds to a difference of 2.4 percent compared to the number of employed according to LFS. For unemployed, the difference between LFS and BAS is larger than for employed where unemployed according to LFS is on a higher level than the corresponding number of unemployed according to BAS. For the population aged 16 to 64 years, the difference varies between 60 000 and 231 000 unemployed, which corresponds to 14.3 and 40.8 percent, respectively, of the number of unemployed in LFS.

By using the estimation in LFS on BAS, comparisons on estimates for LFS and BAS derived by the same estimation procedure can be conducted. According to this analysis, 95 percent of those who are employed according to LFS are also employed according to BAS. For unemployed, approximately half of those who are unemployed according to LFS are also unemployed according to BAS.

The analyses that have been carried out so far indicate that the comparability between LFS and BAS is higher for employed than for unemployed. The lack in comparability for unemployed can mainly be derived to the age group 15 to 24 years.

There is a need for further analyses between LFS and BAS, for example due to providing information to users of statistics on the labour market on how LFS and BAS relate to one another. Another reason is the future computation of LFS and BAS and investigating potential benefits from combining these two sources.

## Introduction

In May 2022, Statistics Sweden (SCB) established a new statistical product in the labour market area – “Population by Labour market status” (“Befolkningens arbetsmarknadsstatus”, BAS), which is mainly based on administrative data. The introduction of BAS is mainly made possible thanks to the introduction of the monthly administrative source Employer Declaration at individual level (AGI), collected by the Swedish Tax Agency since January 2019. BAS is initially categorized as statistics under development and was first published 31 May 2022. The ambition is that BAS will be classified as official statistics in the future.

In BAS, the population is divided into six labour market statuses: employed, unemployed, students, retirement, illness, and others. This is done with a month as the reference period. Since earlier, Statistics Sweden publishes monthly statistics regarding the population's labour market situation within the framework of the Labour Force Surveys (LFS), which form the basis of official statistics and are internationally regulated. It is of interest to analyse how these two sources relate to one another.

This paper presents a limited version of a more comprehensive report (Statistics Sweden, 2022c) which also covers the group outside labour force and more detailed results regarding gender, age and country of birth. There are also several aspects which are not considered in the study, mainly that numerical differences between BAS and LFS related to uncertainty non-response and frame coverage in LFS, are not highlighted. Also, proportions, such as the relative unemployment rate, are not broached. Despite this, it is hoped that this description can provide useful information about how BAS and LFS relate to each other.

## Population by Labour market status

The Population by Labour market status, BAS, show, among other things, estimates of the number of people in each labour market status (employed, unemployed, student, retirement, illness, other) for the age group 15–74 years. The reference period for labour market status is month, with a publication for month, quarter, and year.

The frame population consists of Sweden's registered population as of the last day of the month before the reference month. The classification of the population into the first five labour market statuses is based on administrative data from the Swedish Tax Agency, Companies Registration Office, Public Employment Service, Board of Student Finance, Pensions Agency and Social Insurance Agency, respectively. The group *other* consists of those persons who lack administrative information that could classify them into one of the previous five statuses. The classification is done hierarchically according to employed, unemployed, student, retirement, illness, other. This means that if a person has been classified as employed, the same person cannot be classified to any other status, and correspondingly for the statuses that follow in the hierarchy.

### Some differences in the definitions of labour market status

#### *Employed*

- In the LFS, based on work performed during the reference week
- In BAS, based essentially on whether the individual has received any compensation that is the basis for employer contribution according to AGI
- BAS therefore uses information about when the payment took place rather than when the work has been carried out

#### *Unemployed*

- In the LFS it is required that the individual is without work, has applied for work and can take a job, based on specific requirements for the time periods for this
- In the LFS, there are therefore no requirements that the individual must have been registered with the Employment Service
- In BAS, the starting point is that the individual must have been registered at some point during the month with one of the applicant categories included in openly unemployed or applicants in programs

#### *Not in the workforce*

- In the LFS, subgroups are created by first identifying those who answered that they are full-time students, then the others are grouped according to the answer to the question what they mainly consider themselves to be
- In BAS, subgroups are created hierarchically according to, students, retirement, illness, and others

## **Comparison of number estimates**

In this section, a comparison is made of monthly estimates from BAS and LFS of the number of people in the labour market statuses employed, unemployed respectively for the age group 15–74 years for the period 2020–2021 (linked estimates have been used for 2020). It is a too short period of time to be able to make a complete analysis of the effect of seasonal variations. In addition, there is a risk that the relationships we see between the statistics from BAS and LFS are different in other time periods. This is because the coronavirus pandemic had an impact on the development of the labour market and the pandemic lasted over almost the entire period.

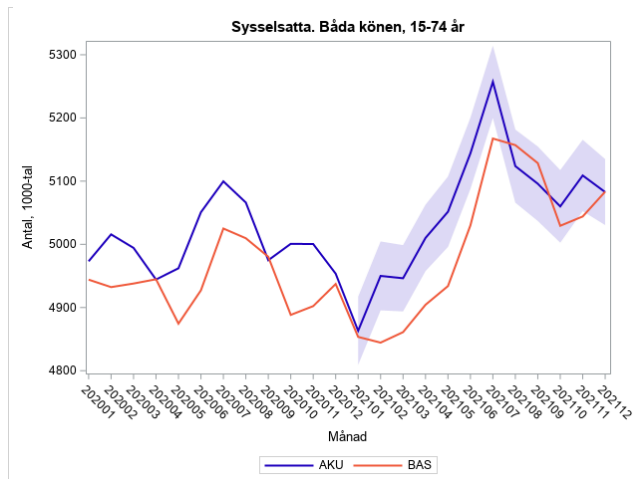
The comparisons are made partly for the entire group, partly for domains by gender, age, and country of birth (born in Sweden/born abroad). The fact that the comparison is made for different domains aims to increase the understanding of how the statistics from BAS and LFS relate to each other.

### **Employed**

Figure 1 shows the number of employed according to BAS (red line) and estimates for LFS (blue line) with corresponding uncertainty intervals, based on a 95 percent confidence level. The estimate of employed for LFS exceeds the estimate from BAS in all but five of the studied months. On the other hand, the estimate for BAS is within the LFS's uncertainty interval for five of the twelve months where there is information about the size of the uncertainty interval. For the studied period, the difference between LFS and BAS is at most 124,000 people with an average difference of 58,000 people.

There appears to be an indication of a time lag for the estimate of employed in BAS in relation to LFS for parts of the studied period. BAS mainly use information about which people have received compensation that is the basis for employer contributions according to AGI to derive the employed. In AGI, there is no information about when the employee has worked, but instead information is used about the month in which the compensation was paid. This could be a contributing factor to the observed lag. This effect is likely to have greater significance for groups where a larger proportion have a weaker connection to the labour market than for the entire age group 15–74 years.

Figure 1: Number of employed, 15-74 years



Apart from the lag in the series for BAS, in comparison with the LFS, the estimate in BAS for the age group 15–24 years exceeds that in the LFS for part of the studied period. This may be an effect of the reference period in the LFS being a week, while in BAS it is a month. A quantity, with the reference period month, expressed as the number of persons employed at any time during the month is greater than a quantity defined in terms of the number of persons employed at any time during the week. Where the latter quantity is expressed as an average over the preceding weeks of the month. This difference is significant for groups that are not employed during all weeks of a month.

There are several factors that contribute to differences in the estimates from BAS and LFS. It concerns, for example, differences in the estimation procedures between BAS and LFS and the extent to which the procedure in LFS compensates for sources of uncertainty such as omissions and coverage.

## Unemployed

Figure 2 shows that the number of unemployed according to BAS (red line) is significantly lower than the corresponding point estimate for LFS (blue line) with an average level difference of 126,000 people. The lowest difference during the period is 60,000 people and the biggest difference is 231,000 people. The number in BAS is also outside the LFS's uncertainty interval for all months in 2021. Even broken down by gender, the estimate for BAS is lower than that in the LFS and also lower than the uncertainty interval for the LFS. The difference is roughly the same order of magnitude for both women and men.

According to the The Swedish Public Employment Service's monthly statistics, the number of registered unemployed varied between 370,000 and 478,000 during the period January 2020–December 2021. It is information on the number of registered people at the Employment Service that forms the basis for the classification as unemployed in BAS. This contrasts with the LFS, where the individual is asked if he or she has tried to find a job, which is expected to be more than registered unemployed. This difference in definition is thus a likely main explanation for the difference in level between BAS and LFS. The lower number of unemployed according to BAS compared to Employment Service can be derived mainly from the fact that the definition of labour market status is done according to a hierarchical principle, where certain people who are registered as unemployed with The Swedish Public Employment Service are classified as employed in BAS.

Figure 2: Number of unemployed, 16-64 years

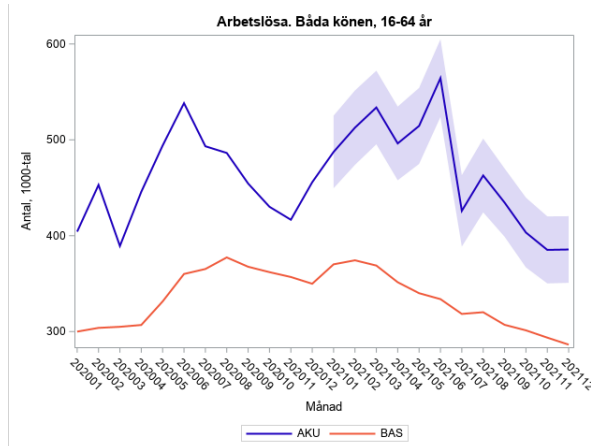
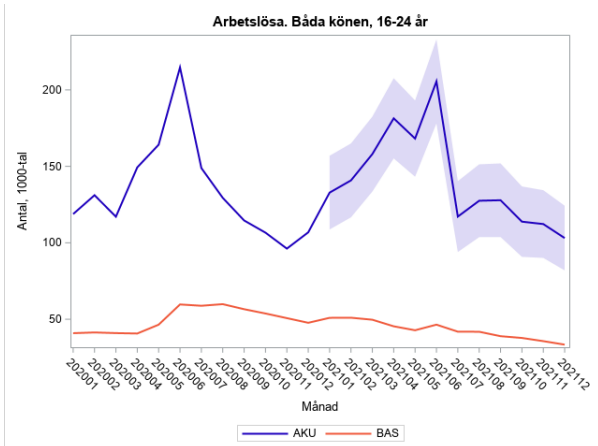


Figure 3: Number of unemployed, 16-24 years



For more information on how estimates of the number of unemployed according to BAS relate to the number of registered unemployed according to the Employment Service's statistics, see Statistics Sweden (2022a). Another explanation for the large number of people who are registered with the Employment Service but who are still employed according to BAS could be that there is a lag in the register data linked to the time for follow-up and updating. If the registered do not notify the Employment Service when they have found a job, this can lead to an overestimation of the unemployment period and the number of unemployed.

The difference in the number of unemployed between BAS and LFS is highest for the younger age groups. The biggest difference is seen for the youngest group, 16–24 years, which can be seen in Figure 3. The difference in this age group constitutes a large part of the total deviation for all unemployed people in the 16–64 age group. Among other age groups, the differences are smaller and mostly not statistically significant. The main explanation for the lower number of unemployed young people in BAS in relation to the corresponding estimates according to the LFS is likely that many in this group have a weaker incentive to register with the Employment Service than people in other age groups. As a rule, students are not entitled to unemployment compensation and among young people, fewer meet the qualification requirements for unemployment insurance. This is also indicated by the fact that the difference in figure 3 increases sharply before the summer months when students look for holiday jobs, without registering with the Employment Service.

## Comparison based on the LFS estimation procedure

In this second part of the analysis, the estimation procedure for the LFS is used to calculate estimates for BAS. This approach eliminates differences in the estimation procedure between BAS and the LFS to focus on similarities and differences in variable values. However, the analysis does not consider differences that can exist due to sources of uncertainty, e.g. nonresponse that are related to the estimation procedure in the LFS. Because of the break in the time series for the LFS from Jan 2021, the analysed period is limited to Jan-Dec 2021.

Estimates on the relation between labour market status in BAS and the LFS is seen in Table 1. These estimates are derived from an individual level where status in BAS is compared to status in the LFS for each individual that has participated in the LFS. The largest difference is found among the unemployed, where only 51 percent of the 489 000 unemployed in the LFS have the same status also in BAS. Among the 350 000 unemployed in BAS, 71 percent have the same status in the LFS. For the employed, the consistency between the measures is considerably higher. Among the 5 058 000 employed in the LFS, 97 percent have the same status in BAS.

**Table 1. Number of people, 15–74 år, per labour force status in LFS and BAS using the LFS estimation procedure, 2021, estimate (margin of error), thousands.**

15–74 years	BAS			
LFS	Employed	Unemployed	Not in l.f.	Total
Employed	4 800,7 (23,6)	25,3 (5,7)	231,9 (15,3)	5 057,9 (25,3)
Unemployed	81,8 (6,5)	247,4 (7,9)	159,6 (10,1)	488,8 (13,7)
Not in l.f.	192,3 (13,5)	76,8 (8,0)	1 694,9 (24,4)	1 963,9 (26,6)
Total	5 074,7 (22,2)	349,5 (7,5)	2 086,3 (22,8)	7 510,6 (7,5)

Regarding the number of unemployed, the lowest consistency between LFS and BAS is found among young people, as shown in corresponding table 2 for the ages 15-24 years. In this group, only 17 percent of the 154 000 unemployed in the LFS have the same status in BAS. Among the 40 000 unemployed in BAS, 66 percent are unemployed also according to the LSF.

**Table 2. Number of people, 15–24 år, per labour force status in LFS and BAS using the LFS estimation procedure, 2021, estimate (margin of error), thousands.**

15–24 years	BAS			
LFS	Employed	Unemployed	Not in l.f.	Total
Employed	407,7 (12,4)	4,0 (2,6)	53,9 (7,3)	465,6 (13,4)
Unemployed	27,8 (4,0)	26,4 (3,6)	99,7 (7,1)	153,9 (8,7)
Not in l.f.	61,1 (7,8)	9,6 (1,7)	457,2 (13,3)	527,9 (14,5)
Total	496,5 (10,9)	40,0 (4,3)	610,8 (11,7)	1 147,4 (5,2)

## Conclusions

When comparing BAS and LFS regarding estimates of the number of employed people, we find that the levels and developments are fairly similar, but where BAS in most of the months gives lower estimates than LFS. For the group aged 15–74, the difference is at most 124,000 people in 2020 and 2021, which corresponds to 2.4 percent in relation to the estimate of the number of employed people in the LFS. When we compare BAS and LFS by also producing estimates for BAS using LFS's estimation procedure, we find that the proportion employed in BAS among those employed in LFS is 95 percent for all months in 2021.

For the unemployed, the comparison shows that the differences between BAS and LFS are greater than for the employed and vary more over time. The estimates according to LFS are numerically larger than those for BAS. For the group aged 16–64, the difference varies between 60,000 and 231,000 people, which corresponds to a difference of 14.3 and 40.8 percent, respectively, in relation to the estimate of the number of unemployed according to the LFS. A large part of the difference between BAS and LFS is made up of young people. This is probably because they have a weaker incentive to register with the Employment Service. This is particularly noticeable in the case of temporary unemployment before the summer holidays, which also largely explains why the seasonal variation is significantly lower in BAS than in LFS. From the comparison with regard to the LFS's estimation procedure, it appears that only about half of the unemployed in the LFS are classified as unemployed in BAS as well.

Overall, this leads to the conclusion that the agreement of the statistics on the number of employed between BAS and LFS is high, while it is significantly lower for the number of unemployed. The low level of agreement for the unemployed primarily applies to people in the age group 15–24, who make up a significantly lower proportion of the total number of unemployed in BAS compared to LFS. In addition, there seems to be a greater variation in the agreement between the sources for the number of unemployed than for the number of employed, which means, for instance, that the statistics during certain periods show different developments.

There may be a need to carry out further analyses by highlighting other aspects. But also to analyse a longer period of time in order to study seasonal effects or a period which is less affected by circumstances that are significant for the development of the labour market, such as the corona pandemic. It is the future use of BAS that will have significance for the content of such analyses. Content and scope for further analyses can be formulated as Statistics Sweden gains more knowledge about the way in which BAS is used.

## References

Statistics Sweden (2014). *Mätfelsstudie i LFS*.

[AM0401\\_2014A01\\_BR\\_AM76BR1402.pdf \(scb.se\)](#), obtained 2023-05-12

Statistics Sweden (2017). *Analys av bortfallsbias avseende Arbetskraftsundersökningarna (LFS)*.

<https://www.scb.se/publikation/32401>, obtained 2023-05-12

Statistics Sweden (2018). *Jämförande studie LFS och Af 2016 – Arbetslöshetstid i*

*Arbetskraftsundersökningarna (LFS) och tid utan arbete på Arbetsförmedlingen (Af)*.

<https://www.scb.se/publikation/36661>, obtained 2023-05-12

Statistics Sweden (2022a). En numerisk beskrivning av relationen mellan BAS och Arbetsförmedlingens verksamhetsstatistik.

[En numerisk beskrivning av relationen mellan BAS och Arbetsförmedlingens verksamhetsstatistik \(pdf\) \(scb.se\)](#) obtained 2023-05-12

Statistiska centralbyrån (2022b). Hur förhåller sig BAS till LFS och Arbetsförmedlingens statistik?

[Hur förhåller sig BAS till LFS och Arbetsförmedlingens statistik? \(pdf\) \(scb.se\)](#) obtained 2022-05-12

Statistics Sweden (2022c). En numerisk beskrivning av relationen mellan BAS och AKU.

[En numerisk beskrivning av relationen mellan BAS och AKU \(pdf\) \(scb.se\)](#) obtained 2023-04-12

The Swedish Public Employment Service

<https://arbetsformedlingen.se/statistik/sok-statistik/tidigare-statistik>, obtained 2022-05-10