

Model estimation of number of hours worked

Daniel Lennartsson, Statistics Sweden, daniel.lennartsson@scb.se

Susanne Gullberg Brännström, Statistics Sweden,
susanne.gullbergbrannstrom@scb.se

Abstract

There is a general interest in the number of hours worked in the Swedish economy. Hours worked are needed to be able to calculate work effort and productivity in the economy. But also, to analyze and evaluate the economy.

The purpose of the present work is to present two models that have been developed to estimate the number of hours worked in the Swedish economy; an estimate that is not based on the Labor force survey (LFS), to describe the hours worked in the economy using data that covers the whole working population. The work has been made possible by Statistics Sweden receiving monthly employer declarations at the individual level (PAYE) from the Swedish Tax Agency (SKV) since 2019. PAYE together with information from Statistics Sweden's business register (BR) and short-term statistics on wages and salaries (KL) have been used in the estimation of each model. An estimate like this with good quality are also planned to be presented in the new register-based labor market statistics (BAS).

To validate the models and their estimates the paper includes an analysis section in which a comparison of estimated hours worked is made based on the two models developed against hours worked according to the National Accounts (NA).

Keywords: model based, hours worked

1. Introduction

At Statistics Sweden (SCB), the number of hours worked is measured mainly by the LFS. Other surveys that collect hours worked are the wage statistics¹, although hours worked do not constitute a target quantity in these surveys. In LFS, the individual person is the object, which itself states how many hours he or she has worked during

¹ With wages and salary statistics we mean short-term wages and salary statistics for private sector (KLP), government sector (KLS), municipalities (KLK) and regions (KLR).

the reference week, while in the salary surveys it is the company or organization that is the object and which is responsible to report how many hours the employee has worked during the reference period. The object differs between the LFS and the wage statistics, as well as the purpose of the surveys. The LFS describes the Swedish labor market regarding unemployment, employment, hours worked, etc. The wage statistics have the function of describing wage levels, wage structures and wage changes, etc., in the Swedish labor market.

2. Description of used data

2.1 PAYE:s target population

Three main sources are used. These are monthly freezes by BR, PAYE and KL respectively. Information is obtained from BR regarding the enterprise, here legal unit unless otherwise stated, such as industry (NACE) and sector, according to NACE2007 and SECTOR2014, respectively, and whether the company is active or not. When information regarding industry affiliation is missing in the monthly freezes for the reference period 201901 - 201909, the information is imputed from later reference periods when the information is available. Only active companies are included in the continued work².

Excluded from the number of individuals in PAYE are those who do not have a social security number, who are on board or working at special workplace units. Individuals with a total salary sum that is equal to or less than SEK 1,000 for a given reference period are also excluded³. The compensation wage that these remaining individuals received according to PAYE is an important component in the continued work. For each person, there is information about which company the individual is employed at. This information, together with information regarding the reference period (year and month) is used to link the two sources, BR and PAYE. Only enterprises that match,

² We are interested in the national accounts Swedish operating units and employees at these.

³ This definition is in good agreement with the definition of employees in BR for legal entities with a local unit. The difference is that here we also apply it to all individuals and all enterprises, regardless of whether the enterprise has one or more local units. The reason we can do that is that we have not passed on the information to local units (LU). Furthermore, the reason why this definition of number of employees has been used in this work is that KLP has used BR's approach (which is close to the one used here) in its selection design.

and employees at these enterprises are included in the further analysis⁴. This set of objects, enterprises and employees, is called PAYE's target population. The Swedish economy data is divided into different institutional sectors. Table 1 shows a summary of PAYE's target population for two different reference periods, 201901 and 202001 respectively. These reference periods are not chosen for reasons other than that 201901 is the first period for which PAYE exists and 202001 is 12 months later. The table shows the number of enterprises, the number of employees, total remuneration and average monthly salary per employee and its development over 12 months for the companies included in the target population for this work, per sector⁵.

Table 1 Description av PAYE:s target population

Reference period	Sector	Number of enterprises (PAYE)	Number of employees (PAYE)	Wages (PAYE, SEK)	Wages per employee (SEK)	Yearly change of wages per employee (%)
201901	Business	273 370	3 334 323	102 356 648 846	30 698	
201901	Non profit org.	12 722	129 259	2 966 408 035	22 949	
201901	State	306	264 680	9 143 229 104	34 544	
201901	Regions	49	290 076	9 526 390 266	32 841	
201901	Municipalities	481	946 940	23 692 199 737	25 020	
202001	Business	279 254	3 351 224	106 408 986 542	31 752	3,4
202001	Non profit org.	12 371	127 645	3 016 011 370	23 628	3,0

⁴ There are active enterprises in BR (monthly freezes) that are not included in the PAYE data and there are enterprises in PAYE that are not active in BR or that are not found in BR

⁵ With regard to industry, it should be noted that there are no industry boundaries.

202001	State	290	269 735	9 678 163 010	35 880	3,9
202001	Regions	51	295 941	10 071 923 188	34 034	3,6
202001	Municipalities	484	930 827	24 202 523 288	26 001	3,9

2.2 Short term statistics on wages and salaries (KL) target population

The theoretical target population for KL is not the same as the target population for PAYE. PAYE includes the departments of agriculture and forestry (NACE A) and gainful employment in households (NACE T). KL does not. The reason for this is that the remuneration for work performed (salary) is not assumed to be market-based, which is required in the wages and salary statistics, as the work is often performed for own use or by self-employed persons and its family members. When NACE divisions A and T are excluded from PAYE's target population, what we choose to call is obtained KL's theoretical target population. See Table 2 for a description.

Table 2 Description of KL:s theoretical target population

Reference period	Sector	Number of enterprises (PAYE)	Number of employees (PAYE)	Wages (PAYE, SEK)	Wages per employee (SEK)	Yearly change of wages per employee (%)
201901	Business	264 363	3 298 845	101 520 016 658	30 774	
201901	Non profit org.	12 659	129 004	2 962 247 296	22 962	
201901	State	298	264 487	9 137 618 938	34 548	
201901	Regions	49	290 076	9 526 390 266	32 841	
201901	Municipalities	481	946 940	23 692 199 737	25 020	
202001	Business	270 205	3 314 934	105 536 421 328	31 837	3,5

202001	Non profit org.	12 315	127 384	3 011 843 437	23 644	3,0
202001	State	289	269 721	9 677 828 425	35 881	3,9
202001	Regions	51	295 941	10 071 923 188	34 034	3,6
202001	Municipalities	484	930 827	24 202 523 288	26 001	3,9

By excluding division NACE A and T from PAYE's target population, the relative relationship between the sectors is not affected. On the other hand, the number of companies is reduced by around 9,000, the number of employees is reduced by around 35,000 and the total remuneration is reduced by around SEK 850 million. It is above all sectors N and H that are affected by the industry delimitation. The changes in average remuneration per employee are nearly unchanged. Why is this not KL's actual target population? The reason for this is that the surveys that make up the official wage statistics are not designed on the basis of the sectors but on the basis of the main objects within each sector.

The survey on short term statistics for wages in the government sector (KLS) has authorities and their employees as objects. In the same way, the wage survey for regions (KLR) and wage survey for municipalities (KLK) has regions and municipalities and their employees, respectively, as objects. As can be seen from Table 1 or 2, there are, for example, more objects in sector M than the 290 municipalities that exist. The same applies to sectors S and R.

KLP, on the other hand, includes, in addition to sectors B and H, also limited companies in public administration, i.e. sectors S, R and K⁶. Nevertheless, there are objects in public administration that are not covered by the KL surveys.

2.3 Variables of interest

In this work, we are interested in hours worked in the Swedish economy per reference period (years and quarters) and the change in these. Let us see the total

⁶ In the report "Introduction of INSEKT2014 in the statistics" which forms a basis for a decision on "The introduction of INSEKT2014 in the statistics" (no. 2015/383), is a probable explanation for why KLP has these limited enterprises today in its target population.

number of hours worked as a sum of the hours worked in each institutional sector, i.e.

$$T_y = \sum_{i=1}^I T_{y_i}$$

We are interested in T_{y_i} , for the institutional sectors, but for business, the information per quarter industry used in the national accounts is also requested. This allows us to see the total for each sector as the sum of hours worked across all quarterly industries, i.e.

$$T_{y_i} = \sum_{d=1}^D T_{y_{id}}$$

There are two sets of quarterly industries in the national accounts at quarterly level, partly at calculation level and partly at accounting level. In this work, the quarterly industries at the calculation level have been used and include 95 industry groups (96 with the total) across the entire NACE nomenclature. Wage statistics, on the other hand, KLP and SLP more specifically, are designed according to the accounting industries, which are 64 in number (65 including the total), when departments A, T and O have been excluded.

3. Models for estimating the number of hours worked

The purpose of this work is to present how a model estimate of the number of hours worked in the Swedish economy can be produced, per reference period, institutional sector (i) and quarterly industry (d). That is, we are looking for an estimate of

$$T_{y_{id}}$$

In classical randomization theory, unknown population totals, here for example $T_{y_{id}}$, are replaced with design-based estimates, which means that we expand observed values with a weight that takes into account the probability for the company to be included in the sample and dropout compensation and more. See Figure 1 for an illustration.

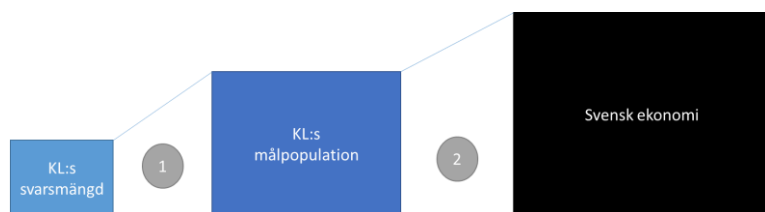


Figure 1 Relation between micro and macro

However, we have (at least) two problems. Due to the cut-off sample (that KLP does not survey enterprises with fewer than five employees), the weights cannot expand the response rate so that an underestimation of totals can be avoided. On the other hand, here, just as in the official statistics from KLP, it can be assumed that the averages (here compensation per hour or working hours) for the employees in enterprises with less than five employees are the same as for the employees in enterprises with five or more employees. Given this assumption holds, the population of inference is KL's target population. One way of assessing the reasonableness of the assumption is by comparing the average remuneration per employee in PAYE's target population, with the corresponding size for KL's target population and selection framework. Regardless of the assumption's validity, however, totals are underestimated.

The second problem is the difference between the target population for PAYE and KL, see Tables 1 and 3 and Figure 1. Above all, the fact that there are quarterly industries in PAYE that are not covered by KL. Just as we lack information about enterprises with less than five employees, especially in sectors B and H, we also lack information about companies in division A and T. By assuming that the companies in these industries are like the companies in general, we can extrapolate the results for KL's target population to the target population for PAYE. See Figure 1 for illustration.

If we let $T_{x_{1id}}$ denote the total remuneration the employees receive in sector i and quarterly industry d for a given reference period, we can express the sought (unknown) quantity, the number of hours worked

$$T_{y_{id}} = T_{x_{1id}} \left(\frac{T_{x_{1id}}}{T_{y_{id}}} \right)^{-1},$$

which can also be expressed as

$$T_{y_{id}} = T_{x_{1id}} R_{x_{1id}, y_{id}}^{-1},$$

that is, the total remuneration is divided by the average remuneration per hour (or multiplied by the inverse of the average remuneration per hour). We call this model 1.

If, on the other hand, we let $T_{x_{2id}}$ denote the total number of employees in sector i and quarterly industry d for a given reference period, we can express the sought (unknown) quantity as

$$T_{y_{id}} = T_{x_{2id}} \left(\frac{T_{y_{id}}}{T_{x_{2id}}} \right),$$

which can also be expressed as

$$T_{y_{id}} = T_{x_{2id}} R_{y_{id},x_{2id}},$$

that is, the total number of employees with the average working time. We call this model 2.

We know $T_{x_{2id}}$ and $T_{x_{1id}}$ for PAYE's target population, but not $R_{x_{1id},y_{id}}^{-1}$ and $R_{y_{id},x_{2id}}$. However, we also know x_1 and x_2 for each enterprise in KL's response rate. Since the KL surveys also collect information about hours worked, we can estimate average remuneration per hour and average working hours per employee, as the hours available in KL should be the hours that PAYE would have had if hours worked had been included among in the PAYE data. This allows us to produce an estimate for $R_{x_{1id},y_{id}}$ and $R_{y_{id},x_{2id}}$, respectively, by producing estimates of $T_{y_{id}}$, $T_{x_{1id}}$ and $T_{x_{2id}}$ and then calculate the respective ratio.

To produce the estimates, we need to treat the KL surveys as one survey. This is because KLP is a sample survey with enterprise not only in sectors B and H but also in public administration. KLS, KLR and KLK are included in the KL survey as separate strata with weight 1. Since KLP is divided by personnel category (workers and salaried employees), a new, common weight scheme was adopted, i.e. a new, different, set of weights than what is used- in regular statistics production, for KLP⁷.

⁷ Each enterprise, given the category of staff, is the respondent (4), belongs to the coverage (6) or belongs to the dropout (9). If a given company has at least one four for one of the staff categories, the enterprise is considered in the common weight scheme to be equivalent, if it does not have a four, but at least one six, then the enterprise is considered to belong to the coverage. Other enterprises belong to the dropout.

For the enterprises from KLP, they are placed in their usual strata and the weights come from the new weight scheme.

A description of the response rate for KL is given in Table 3. The response rate comprises almost 100 percent of the information in the target population for enterprises in public administration (sectors S, R and M). The exception is the number of enterprises, where just over 90 per cent is represented in the response rate.

Table 3 KL:s responding population

Reference period	Sector	Number of enterprises (KL)	Number of employees (PAYE)	Wages (PAYE, SEK)
201901	Business	5211	1 481 318	49 023 756 827
201901	Non profit org.	470	40 281	1 107 737 803
201901	State	220	261 123	9 011 054 829
201901	Regions	28	289 141	9 504 523 228
201901	Municipalities	296	927 817	23 185 564 569
202001	Business	5218	1 490 283	50 841 215 856
202001	Non profit org.	473	42 149	1 181 331 684
202001	State	219	266 663	9 557 626 545
202001	Regions	30	295 115	10 051 480 923
202001	Municipalities	297	911 698	23 686 965 399

In the case of sectors B and H, the situation is different. The response rate covers about 50 percent of the employees and their compensation for the employees at enterprises in the target population in sector B, but only 1 out of 50 enterprises, which is explained by the cut-off sample. This proportion has increased slightly, in terms of the selection framework, as the response rate includes 1 of 16 of these enterprises.

When it comes to the number of employees and their remuneration, it is for sector H that the response rate has the worst coverage, with about 40 percent. In terms of the coverage of the number of enterprises, it is about twice as large as the coverage for enterprises in sector B. This can possibly be explained by the fact that the propensity to respond is higher among the small companies in sector H. Another, more probable explanation is that sector H is small in relation to sector H and neither the KL nor the KLP survey is designed specifically for sector H.

Design-based estimates $T_{y_{id}}$, $T_{x_{1id}}$ and $T_{x_{2id}}$ and $R_{x_{1id},y_{id}}$ and $R_{y_{id},x_{2id}}$, respectively, are produced in the same way as in the regular production of official statistics belonging to the KL surveys for respective reference period (24 pieces).

In order to be able to evaluate these estimates, interval estimates (95% uncertainty intervals) have been produced for number of companies, number of employees ($T_{x_{2id}}$), the total remuneration for these employees ($T_{x_{1id}}$) and number of hours worked ($T_{y_{id}}$). Table 4 shows interval estimates for these unknown quantities for the total per sector, for the two reference periods used in this work, 201901 and 202001, respectively.

Table 4 Interval estimates for different variables

Reference period	Sector	CI number of enterprises (KL)		CI number of employees (PAYE)		CI wages(PAYE, SEK)		CI number of hours worked (KL)	
		Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit	Lower limit	Upper limit
201901	Business	83 350	84 259	2 813 833	2 883 751	88 622 864	90 549 281	323 619 762	330 819 046
201901	Non profit org.	3 620	4 519	98 044	113 734	2 391 473	2 722 810	9 451 241	10 926 973
201901	State	192	312	260 918	262 277	9 003 569	9 052 813	33 510 093	33 678 724

201901	Regions	27	29	288 918	289 474	9 498 966	9 512 825	35 861 322	35 907 230
201901	Municipalities	273	373	927 298	930 522	23 172 133	23 255 242	102 257 827	102 655 527
202001	Business	83 611	84 625	2 824 684	2 905 869	92 530 324	95 225 589	314 974 569	321 537 808
202001	Non profit org.	3 833	4 833	104 544	123 569	2 611 292	3 043 310	10 076 716	11 943 704
202001	State	187	321	266 454	267 893	9 549 693	9 603 326	34 033 344	34 213 697
202001	Regions	29	31	295 056	295 196	10 049 964	10 053 541	35 450 058	35 464 679
202001	Municipalities	276	369	911 243	913 880	23 673 350	23 753 795	98 363 690	98 729 845

Normally we do not know the values of the totals for which estimates are sought. The only total that is unknown is the number of hours worked ($T_{y_{id}}$). Normally, only for the response rate the information is known (Table 4) at the object level, given that no auxiliary information has been used. Help information may be available at the population level, or at least the sample level.

The fact that we know the totals for certain quantities enables a different analysis than what is normally possible. As can be seen, the totals are underestimated, as expected. What is important is that we obtain good estimates of the totals of KL's selection framework (see Table 3) and their functions⁸. Estimates of average remuneration per hour ($R_{x_{1id},y_{id}}$) and average working hours per employee ($R_{y_{id},x_{2id}}$) are needed to obtain estimates of the number of hours worked according to model 1

⁸ Good estimates refer to estimates that have a small minimum square error.

and 2. Estimates according to model 1 are obtained from the following way, for a given reference period:

$$\hat{T}_{yid} = T_{x_{1id}} \hat{R}_{x_{1id},yid}^{-1}$$

where $T_{x_{1id}}$ is, known, total remuneration per sector in and quarterly industry d for PAYE's target population and $\hat{R}_{x_{1id},yid}$ is the estimate of (unknown) average remuneration per hour based on KL's response rate. In those cases, for example for quarterly industries that belong to sections A and T, there is no basis for producing estimates of $R_{x_{1id},yid}$, whereby we instead use the estimate of $R_{x_{1id},yid}$ which refers to the total, for these quarterly industries⁹.

Similarly, model 2 estimates are obtained as follows, for a given reference period:

$$\hat{T}_{yid} = T_{x_{2id}} \hat{R}_{x_{2id},yid}$$

where $T_{x_{2id}}$ is, known, the total number of employees per sector and industry d for PAYE's target population and $\hat{R}_{x_{2id},yid}$ is the estimate of (unknown) average working hours per employee based on KL's response rate. Imputations are performed in the same way as for model 1.

Finally, to obtain estimates for a given sector in, the estimates for the included quarterly industries are summed up, i.e.

$$\hat{T}_{yi} = \sum_{d=1}^D \hat{T}_{yid}$$

to obtain sum consistency.

Table 5 shows the estimates of each model for the total number of hours worked per sector for two reference periods, 201901 and 202001, respectively, for PAYE's target population.

Table 5 Estimates of total hours worked for PAYE:s target population, for model 1 and model 2

Reference period	Sector	Hours worked (model 1)	Hours worked (model 2)

⁹

Models 1 and 2 are linear models, which means that it is to produce estimates not only at domain level per reference period, but also object level. It should be noted, however, that such should be interpreted with caution if they are produced.

201901	Business	376 795 658	384 757 273
201901	Non profit org.	12 130 622	12 791 429
201901	State	34 000 314	33 958 842
201901	Regions	35 957 977	35 968 864
201901	Municipalities	104 548 727	104 438 317
202001	Business	362 365 491	373 481 374
202001	Non profit org.	11 771 300	12 335 894
202001	State	34 471 638	34 419 949
202001	Regions	35 530 886	35 536 444
202001	Municipalities	100 569 172	100 506 417

All estimates, except for the estimate for sector H 201901 according to model 1, are outside the interval estimate of hours worked based on KL's response rate. This is according to expectations, because KL's target population is smaller than PAYE's target population, and partly because KL's sample is a cut-off sample. Another thing to note is that the estimates of each model, at this level, are very similar to each other. However, it should be noted that there are differences across the quarterly industries.

KL's cut-off sample is also the reason why the estimates of each model for the total number of hours worked for KL's theoretical target population, see Table 6, and outside the interval estimates for hours worked based on KL's response rate. The exception is again sector H 201901 for model 1, which is within the corresponding interval estimate.

Table 6 Estimates of total hours worked for KL:s theoretical target population, for model 1 and model 2

Reference period	Sector	Hours worked (model 1)	Hours worked (model 2)
------------------	--------	------------------------	------------------------

201901	Business	373 739 801	380 682 180
201901	Non profit org.	12 114 044	12 766 892
201901	State	33 979 438	33 934 057
201901	Regions	35 957 977	35 968 864
201901	Municipalities	104 548 727	104 438 317
202001	Business	359 407 403	369 450 518
202001	Non profit org.	11 755 069	12 310 698
202001	State	34 470 445	34 418 161
202001	Regions	35 530 886	35 536 444
202001	Municipalities	100 569 172	100 506 417

With regard to the estimates for KL's actual target population of the total number of hours worked, according to each model, see Table 7, the estimates for the various sectors of public administration are within the interval estimates based on KL's response rate, for both models, for both reference periods. One explanation for this may be that the number of companies with less than five employees is small.

For sector B, the model estimates of the total number of hours worked for both models and both reference periods are outside the interval estimates based on KL's response rate. It is in line with expectations.

The same applies to sector H for 201901. For the second reference period, 202001, however, they are within the interval estimates. The explanation should be as before.

Table 7 Estimates of total hours worked for KL:s actual target population, for model 1 and model 2

Reference period	Sector	Hours worked (model 1)	Hours worked (model 2)
------------------	--------	------------------------	------------------------

201901	Business	360 567 918	364 164 511
201901	Non profit org.	11 219 349	11 521 160
201901	State	33 628 205	33 595 682
201901	Regions	35 886 033	35 885 467
201901	Municipalities	102 343 304	102 348 616
202001	Business	347 394 961	354 330 738
202001	Non profit org.	11 187 704	11 460 565
202001	State	34 179 259	34 146 569
202001	Regions	35 457 846	35 457 214
202001	Municipalities	98 449 380	98 455 108

Finally, Table 8 shows the estimates according to models 1 and 2, respectively, for KL's selection framework. That all estimates of model 1 and model 2, respectively, are in the corresponding interval estimate of the number of hours worked based on KL's response rate is a positive result and indicates that the estimates of model 1 and 2 give reasonable results.

Tabell 8 Estimates of total hours worked for KL:s sample frame, for model 1 and model 2

Reference period	Sector	Hours worked (model 1)	Hours worked (model 2)
201901	Business	332 172 845	330 164 490
201901	Non profit org.	10 260 526	10 151 590
201901	State	33 627 406	33 595 313
201901	Regions	35 886 033	35 885 467
201901	Municipalities	102 341 692	102 345 748

202001	Business	319 927 582	321 041 424
202001	Non profit org.	10 304 778	10 186 251
202001	State	34 176 833	34 145 144
202001	Regions	35 457 846	35 457 214
202001	Municipalities	98 447 841	98 452 949

As usual, there is an exception. The estimate of model 1 in sector B for 201901 is higher than the interval estimate of the number of hours worked. What explains this remains to be seen. Finally, since the model estimates of hours worked are a function of, among other things, KL (through estimates of average hourly wages and average working hours), these model estimates are also subject to selection uncertainty. However, estimating this and other uncertainties has been outside the scope of the work.

4. Results

In this part of the report, a comparison is made of estimated hours worked based on the two models developed against hours worked according to the National Accounts (NA).

Hours worked in NA's annual calculations are based on the LFS for the entire economy. The only difference is that NA also makes a supplement for black hours. For industries and sectors, other sources are used. In NA's quarterly calculations, the LFS's development figures for the entire economy are used, while for industry and sectors, KL are used together with the short-term employment statistics (KS).

4.1 Model 1 and model 2 compared to national accounts

In this section, the calculation of hours worked is evaluated according to model 1 and 2. Model 1 and model 2 will reach 6.6 respectively 6,7 billion hours for the entire economy by 2020. The corresponding figure for NA is 7.6 billion hours. Model 1 underestimates the hours worked by 14 percent and model 2 by 12 per cent. It is an

underestimation for all sectors. The underestimation is stable for the years 2019 and 2020 and varies marginally. NA's hours include black hours. The black hours accounted for 2.5 percent of the total number of hours for employees in 2019.

Table 9 Hours worked (millions) for model 1 and NA

Sector	Model 1		NA		Share		Model 2		Share	
	2019	2020	2019	2020	2019	2020	2019	2020	Share	Share
Business	4 641	4 442	5 484	5 255	0,85	0,85	4 778	4 608	0,87	0,88
Non profit org.	148	145	167	163	0,88	0,89	158	153	0,94	0,94
State	396	408	415	427	0,96	0,96	396	407	0,95	0,95
Regions	407	417	452	459	0,90	0,91	408	417	0,90	0,91
Municipalities	1 167	1 146	1 349	1 318	0,87	0,87	1 166	1 146	0,86	0,87
Total	6 759	6 557	7 867	7 622	0,86	0,86	6 906	6 731	0,88	0,88

Is the underestimation of hours worked stable over time? In Table 10, we can see the underestimation for the four quarters of 2020. The picture looks broadly the same in 2019, which is why only 2020 is reported here. The underestimation is greatest in Q2 and amounts to 17 percent for model 1. For quarter 3, the corresponding figure is only 9 respectively 7 percent. For quarter 3, the regions are even more overestimated. At present, we have no hypotheses as to why the underestimation varies between different quarters. It may just as well be the case that NA's quarterly pattern is incorrect, such as that model's pattern is incorrect.

Table 10 Shares per quarter 2020

Sector	Model 1				Model 2			
	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4

Business	0,81	0,82	0,88	0,87	0,84	0,86	0,91	0,90
Non profit org.	0,78	0,94	0,89	0,96	0,82	0,99	0,94	1,03
State	0,97	0,89	1,00	0,96	0,97	0,89	1,00	0,96
Regions	0,91	0,86	1,02	0,86	0,91	0,86	1,02	0,86
Municipalities	0,86	0,83	0,96	0,85	0,86	0,83	0,96	0,85
Total	0,84	0,83	0,91	0,87	0,86	0,86	0,93	0,90

In NA's quarterly calculations, they work to a large extent with development figures that write out levels set in the annual calculations. Although the levels are underestimated, the development figures from the models in this report could be used. In the next analysis, we therefore look at the development figures for hours worked, where 2020 is compared with 2019 at sector level. As we can see from Table 11, the model's estimates meet NA's estimates to a large extent. For business and the economy as a whole, it is a difference of 0.1 percentage points between the model 1 and NA. For some industries, there are relatively large differences

Table 11 Comparison of growth rates 2020/2019 for model 1 and NA

Sector	Model 1	Model 2	NA
Business	-4,3	-3,6	-4,2
Non profit org.	-2,2	-2,8	-2,6
State	2,8	2,8	2,8
Regions	2,4	2,2	1,5
Municipalities	-1,8	-1,8	-2,3
Total	-3,0	-2,5	-3,1

Table 12 shows the development figures for all quarters 2020 for the entire economy. As we can see, there are relatively small differences between the models and NA.

Tabell 12 Comparison of growth rates for the whole economy by quarter according to models and NA

Quarter	Model 1	Model 2	NA
2020Q1	-0,9	-0,6	-0,8
2020Q2	-5,1	-4,6	-5,1
2020Q3	-3,8	-3,4	-4,6
2020Q4	-2,3	-1,7	-2,3
2020	-3,0	-2,5	-3,1

4.2 A comparison of the number of employees according to PAYE and NA

For model 2 to work well, it is important that the measure of the number of employees is good. In the work of producing model-estimated hours worked, which was described in the section on PAYE's target population, the number of employees, in this work, has been calculated according to the business register approach where a person can be counted in both his main and second job. In LFS and NA, you want a person to only be counted once and then in their main job. In the present work, the number of employees in the economy has been estimated at 5.1 million employees by 2020. The corresponding figure in NA is 4.9 million employees. Although a person can be counted twice in model 2, the number of employees is approximately the same as NA. If model 2 had counted absentees (parental leave, sick leave) as employed, PAYE would have ended up higher than NA.

Table 13 Number of employees (thousand) for model 2 and NA

Sector	PAYE		NA data		Share	
	2019	2020	2019	2020	2019	2020
Business	3 422	3 320	3 333	3 271	1,027	1,015

Non profit org.	143	132	127	127	1,126	1,039
State	272	275	265	268	1,026	1,026
Regions	297	299	304	306	0,977	0,977
Municipalities	958	938	904	893	1,060	1,050
Total	5 092	4 964	4 933	4 866	1,032	1,020

Another important key figure for users is average working hours. It is calculated by dividing the hours worked by the number of employees. Both models 1 and 2 underestimate the average working hours when comparing with NA. This is logical because the underestimation of the hours is greater than the number of employees.

Table 14 Number of hours worked per employee 2020

Sector	Model 1	Model 2	NA
Business	1 338	1 388	1 607
Non profit org.	1 092	1 157	1 285
State	1 484	1 481	1 590
Regions	1 394	1 395	1 498
Municipalities	1 222	1 221	1 475
Total	1 321	1 356	1 566

5. Conclusion

The purpose of this work has been to produce estimates of hours worked from a source other than the LFS. The approach has been relatively free in the sense that different approaches can be tested and evaluated. The working group sees great

potential in the model and can hopefully finalize a version with relatively small corrections of models and population delimitations.

The working group managed in this first stage to produce estimates of hours worked based on two different models, where one model uses average remuneration per hour (whose inverse is multiplied by the total remuneration) and the other model uses average hours worked per employee (multiplied by the number employees).

The work has used three different sources, BR, PAYE and KL surveys, three sources with differences, which has been tried to be clarified in the introductory section of the work. There are differences in the target population of the different sources and there is a difference in the topicality of the sources. Possibly also periodicity.

It is a positive result that the estimates of model 1 and model 2 for KL's sample frame are covered by the interval estimates of the number of hours worked based on KL's response rate as this indicates that the model estimates are reasonable based on the setup and information used in this work.

The fact that there is then an exception, business sector 201901, indicates that there are questions left to consider.

5.1 Conclusion for national accounts

Both models underestimate the number of hours worked if you compare with the hours available in NA even if you adjust for black hours. The underestimation is slightly greater in model 1 than model 2. Further investigation of the levels over hours worked is required both in the LFS and according to the models. When you look at the development figures, model 1 ends up somewhat closer to NA than model 2. However, there are small differences at the rough sub-sector level. The estimates of the development figures look promising, even though there are major discrepancies with NA when looking at detailed industry.

6. International contacts

Before the work in Sweden with developing the models, some interviews were made with our Nordic colleagues and Ireland to collect their experiences in the matter.

Norway and Denmark have information on hours worked on register. Ireland and Iceland are more in the same situation as Sweden and need, or have already, developed models.