

Income from main job

- Expanding the use of register information in the Norwegian Labour Force Survey

Ingvild Johansen and Susie Jentoft, Statistics Norway

1. Introduction

The Norwegian Labour Force Survey (LFS) is a quarterly sample survey¹ based on telephone interviews. Register information is used to some extent, but mainly as auxiliary information. Statistics Norway is currently exploring the possibility of expanding the use of register information in the LFS by using it as a direct source for measuring income, a work financed through a Eurostat grant project for the period 2017-2019.

During recent years, there has been an increasing focus within Eurostat on improving the quality of income from the main job in the LFS. This variable is not included in the Norwegian LFS today. However, a new data source for the register-based employment and wage statistics provides Statistics Norway with more accurate and frequent information on employment and income, in addition to a broader coverage of employees from all registered firms.

There are some clear advantages for using register information to measure income in the LFS rather than adding more questions to an already extensive questionnaire. The register information is more accurate when it comes to obtaining the exact amount in gross terms, compared to a telephone interview where people might be reluctant to answer questions on income or uncertain about the exact gross values. There is also a risk of people over-stating or under-reporting their income. However, the quality of the end results depends on the coherence between the two sources, for example establishing the statistical unit to match and harmonizing the time periods, as well as creating a satisfying linking procedure.

The paper is organized as follows; the new data source is described in chapter 2, methodological issues and some results are outlined in chapter 3, while chapter 4 makes some concluding remarks.

2. The new data source

As of 2015, the register-based employment statistics and the wage statistics are based on data from a new data source called A-ordningen. A-ordningen is a coordinated digital service, which gathers the reporting on employment, income and tax deductions from employers to the Tax Administration, the Norwegian Labour and Welfare Organization and Statistics Norway.

The information is submitted electronically via the employer's payroll system² to the Tax Administration Shared Services Agency, which administers A-ordningen. The deadline for reporting is the 5th of the following month. Anyone who has employees or who pays salary, pension or other benefits must submit an electronic form³. Employers are required to report information on

¹ The sample in the Norwegian LFS consists of about 12 000 family units (24 000 persons) each quarter (Bø and Håland, 2015). Each family participates in the survey 8 times during a period of 8 quarters. Data are collected weekly.

² The data submitted via the payroll systems are automatically checked for errors and omissions. 90 percent of employers have a payroll and HR system, which account for over 98 percent of the information that is submitted via the A-ordning. Small enterprises that do not have a payroll system, can submit the information electronically via the governmental dialogue system (Altinn).

³ Regulated by law (A-opplysningsloven).

employment, i.e. the type of employment, start and end data for the job, full-time equivalent (FTE) percentage, leave etc., in addition to salary and benefits, i.e. cash benefit, expense allowances, payments in kind, pension etc., and information on different deductions, employer's national insurance contribution and financial activity tax. Salaries are reported in gross values before tax.

3. Method

The target population for *incgross* consists of employed persons aged 15 to 74 years and the target concept being their income from the main job. The unit in the register-based wage statistics is jobs based on a cross-section of the population, as they cover jobs which are active during the third week of each month. The integrated data set is produced by combining the sample of employed persons from the LFS with wage information from the register at a micro level. The quality of the integrated data set may be affected by differences in the coverage and the method for combining the two data sources. A quality assessment of different issues concerning micro data integration is based on Zhang's (2012) two-phase life cycle. The first phase deals with each data source separately, while the second phase, which is the focus in this paper, deals with the actual integration of micro data.

3.1. Definition of Incgross

Income from main job (*incgross*) is defined as gross pay which refers to the monetary component of the remuneration of employees in cash payable by an employer to an employee before deduction of income tax and National Insurance Contributions. It can, however, be collected in either net or gross terms, but the transmission to Eurostat in gross value is compulsory. Since Statistics Norway have access to salaries in gross terms, no conversion is needed.

According to the Eurostat guidelines, the income variable has a broader coverage compared to the wage statistics⁴. In addition to basic salaries, variable additional allowances and bonuses, it also includes regular overtime, extra compensation for shift work, seniority bonuses, regular travel allowances and per diem allowances, tips and commissions, and compensation for meals⁵. Although most of the components mentioned above are available through A-ordningen, not all the data are processed through the production system due to the large amount of data.

Some of the main advantages of using the same definition as the wage statistics is that the wage components are checked every month and that the data are easily accessible, ensuring good quality data and a cost-efficient production system in addition to consistency. By adding overtime and holiday payments to the definition in the wage statistics, about 98 percent of the salary is covered. The remaining components seem to have little effect on the average salary.

Payments based on a period longer than a month, primarily bonus- and holiday payments, should according to the Eurostat guidelines, be proportionally included in the gross monthly pay. Holiday- and bonus payments are usually received once a year, although the frequency of bonus payment varies between different groups of the population. Furthermore, the share of expected future amounts for variable payments should also be included, which might be possible through an interview process, but not when using register information, at least not in an ordinary production process (only retrospectively). In order to avoid an overly complicated production system we suggest that bonus- and holiday payments are treated in the same manner for all employees and that one-twelfth of the payments over the last 12 months is included in the monthly pay as a proxy.

⁴ The definition of earnings in the wage statistics is regulated through council and commission regulations (reg. 530/1999 and 1916/2000).

⁵ The same definition as PY200G in EU-SILC.

3.2. Statistical unit

The statistical unit in the LFS is employed **persons** whom are employees in the main job. The unit in the register is **jobs**, i.e. a person working in a given establishment⁶, which may differ from the LFS response. For instance, a person working part time in two different schools within the same municipality, will appear as two jobs in the register, however, the respondent may report only one job in the LFS. Therefore, we have explored an alternative transformation of the object to unit by re-organizing the register data. The transformation of the data involves aggregating the jobs from the register in cases where a person has more than one job in different establishments belonging to the same enterprise⁷ and is registered with the same occupational code. In these cases, the agreed FTE percentage and the wages are summed across the establishments to one job. When comparing agreed working hours between the two sources, the aggregated figures seem to correspond somewhat better. On average we find that the aggregated wages are about 25-30 Euro higher for women and 5-10 Euro for men (September 2017).

An alternative approach could be to adjust the wages according to the working hours in the LFS. However, the working hours in the LFS and register do not coincide as well as expected. An ongoing project to improve the measure of contractual working hours in the register might result in better consistency. Some further investigation is needed in order to conclude.

3.3. Linking the LFS and the register

The main purpose of implementing income from main job in the LFS is to measure the effects of individual characteristics and labour market characteristics on monthly income. The aim is not to provide an additional measure of the wage levels. Focus has therefore been on the quality at micro level.

A personal identification number is the primary key when combining the LFS and the register. 95 per cent of the employed persons in the LFS are also classified as employed in the register. Since the register contains information on **all** jobs a person might have, the correct job must be identified as well. The *Incgross* variable shall represent the income from the **main** job and identifying this is a potential error when combining the two sources. We compared two alternative procedures to identify the main job: 1) Using the register definition and 2) using auxiliary information from the interview.

In linkage option 1, the register definition of main job is used, which is classified according to a set of predefined criteria including the type of job and the agreed FTE percentage. The second option for identifying the main job was to use a combination of information about the enterprise, industry and occupational code. We do not have a unique identifier for either the establishment or enterprise available in the LFS. However, since register information is used as auxiliary information in the interview process, we have an indicator variable on whether the main job in the LFS and the register belongs to the same establishment/enterprise or a different unit. The two approaches identify the same jobs in 97 percent of the cases. This confirms that there is a good correspondence between the classification of main job in the register and the LFS. Using the register definition is less complicated and can also be used by EU-SILC.

According to the Eurostat guidelines, the *incgross* variable should contain payments received in the calendar month either preceding or containing the reference week. Using register information from the calendar month containing the reference week gives better consistency between the two sources

⁶ Establishment is defined as a locally delimited functional unit.

⁷ Enterprise is defined as a legal unit.

since payment often are made in arrears, or a combination of that and prepayments. It also increases the probability for capturing wages for new jobs, which in some cases might not be reported until the salaries are paid.

Statistics Norway uses a wave approach for yearly variables, a subsample consisting of 2 of the 8 waves, an approach that also will be applied for linking *incgross*.

3.4. Imputation of earnings in the Norwegian LFS

Employees in the LFS who lack wage information after the linking procedure, consist of persons registered unemployed or outside the labour force in the register, and some employees that are temporarily absent from work. During 2017, there was between 10.1 and 12.4 percent in the group containing outlier values (either missing values or extreme values), see table 1. The 2nd quarter contained the highest share of observations to impute while the 4th quarter the lowest.

The identification of extreme values is performed as a part of the production process for the wage statistics. Extreme values are identified through a pre-defined lower boundary⁸ and a quartile method for the upper boundary⁹ within strata consisting of a combination of occupation and type of industry. Wages outside these boundaries are considered reporting errors. Previous studies have identified around 4 percent of values in the wage statistics as either incorrect or implausible (Jentoft and Holgersen, 2017).

Table 1. Total number of units and percent to impute in each quarter, 2017

	Q1		Q2		Q3		Q4	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Valid wages	2799	89.7	2885	87.6	2905	89.1	2995	89.9
Missing in register	249	8.0	311	9.4	273	8.4	229	7.0
Extreme value - high	19	0.6	21	0.6	18	0.6	30	0.9
Extreme value - low	55	1.8	78	2.4	66	2.0	73	2.2
Total to impute	323	10.3	410	12.4	357	10.9	332	10.1

An investigation of the imputation pattern shows variation in the required imputation rates among groups. The youngest and oldest age-groups are overrepresented in the group missing valid wages. Those absent from work during the reference week, with shorter contractual working hours, or with lower/missing education in Statistics Norway's education database are also more likely to be missing/outlier. We also see significant differences in the average wages. This pattern is likely to bias our estimates and we therefore see a need to impute, to reduce the effect.

Different imputation methods were tested, but overall, a Random forest nearest neighbour imputation seems to provide the best imputation method in this case. It has reasonable accuracy and maintains the distribution of the data to allow sensible percentile estimates. The imputed variable is *incgross*, including distributed bonus and holiday pay. The imputation method is based on quarterly survey data.

3.5. Comparison with Statistics Norway's publication of earnings

The quality of the gross wages obtained for *incgross* is assessed by comparing weighted LFS figures with Statistics Norway's wage statistics based on the definitions, linkage and imputation method described in this paper. The earnings information from the register is based on the same population as the LFS. Since the wage statistics does not include overtime nor holiday payments, these

⁸ The lower boundary is based on a fixed contractual minimum wage and minimum wage rates; NOK 6 000 for apprentices, NOK 12 000 for people younger than 20 years and NOK 18 000 for people 20 years and older.

⁹ The upper boundary is set to 4 times the distance from the 3rd quartile.

components are excluded from *incgross* for these comparisons. Table 2 gives *incgross* as a full-time equivalent (FTE) based on information about contractual hours and fulltime position from the LFS.

Table 2. Comparison of full-time equivalents from weighted LFS excluding overtime and holiday pay (Q3) and register based wage statistics (September 2017). Mean earnings and difference in percent. Euro

		LFS – Incgross ex. overtime and holiday pay	Wage statistics – LFS population	Difference in percent
Total		4 781	4 761	-0.4
Age	15-24	2 977	3 061	2.8
	25-54	4 876	4 862	-0.3
	55-74	5 205	5 198	-0.1
Sex	Men	5 129	5 085	-0.9
	Women	4 368	4 373	0.1
Sector	Private	4 903	4 835	-1.4
	Municipality	4 221	4 310	2.1
	State	5 074	5 098	0.5

At an aggregated level, the difference between the full-time equivalents from the LFS and the register is only 20 Euros. The differences are larger at more detailed levels, especially for the youngest age group. The comparison will be affected by deviations in contractual working hours between the sources, due to differences in the reference period, the statistical unit, or the linking procedure, or due to reporting errors in the register or errors connected to collection and processing of data in the LFS. The use of proxy interviews in the LFS, especially for the youngest age group, tends to give less accurate answers. Furthermore, the LFS is not specifically designed for measuring average wages. Due to reasons explained above, we cannot expect the two sources to provide identical figures.

4. Concluding remarks

The main purpose of implementing income from the main job in the LFS is to measure the effects of individual and labour market characteristics on monthly income. The aim is not to provide an additional measure of the wage levels in Norway, but rather to provide a valuable supplement to the LFS dataset, allowing for new analysis. As shown in this paper, the figures at macro level are not necessarily consistent with the wage statistics as the average earnings based on LFS is subject to sampling and data integration errors. It is therefore important to interpret the results cautiously.

References

Bø, T. P. and Håland, I. (2015). Dokumentasjon av Arbeidskraftsundersøkelsen (AKU) etter omleggingen i 2006, https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/_attachment/218011?_ts=14bdf041650

Commision Regulation (EC) no. 1916/2000 of 8 September 2000 Concerning Structural Statistics on Earnings and on Labour Costs as regards the definition and transmission of information on structure of earnings

Council Regulation (EC) no. 530/1999 of 9 March 1999 Concerning Structural Statistics for Wages and Labour Costs

Jentoft, S. and Holgersen, H. (2017). Imputering av lønn. Dokumentasjon av imputeringsmetode i lønnsstatistikk, <http://www.byranettet.ssb.no/Filer/ID2017-14-1.pdf>

The a-melding guide: <https://www.skatteetaten.no/en/business-and-organisation/employer/the-a-melding/>

Unece (2007). Register-based statistics in the Nordic countries – Review of best practices with focus on population and social statistics

Zhang, Li-Chun (2012). Topics of statistical theory for register-based statistics and data integration