# LEIDEN LIS SECTORAL INCOME INEQUALITY DATASET

Chen Wang, Stefan Thewissen, and Olaf van Vliet

Leiden University, Department of Economics

Version 1.1, March 2014

## 1. Introduction

The Leiden LIS Sectoral Income Inequality Dataset contains information on multiple indicators of earnings inequality and employment within 9 sectors and 12 subsectors. Compared to version 1.0 of the dataset, version 1.1 presents updated data for the main part of the first version, namely, for 8 developed countries and 31 LIS waves between 1984 and 2005. Combined with the original data from the first edition, a total of 49 LIS waves providing data for 12 developed countries between 1969 and 2005 are available. We provide additional information of earnings and employment at the country level.

This dataset draws upon data from the Luxembourg Income Study (LIS) micro dataset, which is a time series of household survey data containing information on earnings and employment, standardised across countries. The Leiden LIS Sectoral Income Inequality Dataset allows researchers and public policy analysts to compare sectoral earnings inequality and employment levels across developed countries over the last three decades, based on a standardised classification of sectors across countries and periods. The data can be linked to other sectoral databases, for instance to the OECD Structural Analysis (STAN) database. The database extends the work of Mahler, Jesuit, and Roscoe (1999) who calculate sectoral earnings inequality in 10 countries around the years 1985 and 1990. The full list of variable definitions can be found in Appendix 1.

### 1.1 Difference with Version 1.0

Version 1.1 of the Leiden LIS Sectoral Income Inequality Dataset updates data for the main part of version 1.0. This main part of the data consists of the waves between 1984 and 2005 for the set of countries for which comparable earnings information is available. This part of the dataset can be used for panel data analysis; the exact countries and waves are presented in Table 2.

Version 1.1 uses a slightly broader and more consistent sample definition for all variables.[1] In this second version the number of individuals included is higher, leading to more accurate inequality and earnings estimations.[2] Nevertheless, correlations between the variables from old and new data are very high.[3] The new sample definition is also consistent with the definition for the indicators based on household information. Hence, these are not updated.

For the variables based on individual information the new LIS data are used (version 7; October 2013 update). The variables based on household information are, as in version 1.0 of the Leiden LIS Sectoral Income Inequality Dataset, based on LIS data version 7; March 2012 update. The differences between the LIS March 2012 and October 2013 updates are very small. Thus, variables from versions 1.0 and 1.1 can be used interchangeably.

## 2. Calculating sectoral earnings inequality and employment

### 2.1 Labour earnings and sample definition

We calculate annual earnings both at the household and individual level. We follow the earnings and sample definitions of Mahler *et al*. (1999), that is, we only include income from wages and salaries or self-employment. Income from other sources, such as interest and rent, is excluded. Also excluded are public benefits and income taxes. For all

[1] In addition, information for all variables is now available for Denmark, wave 1992, industry 71.
[2] The average number of all individuals aged 25-54 with nonzero earnings within a sector increases from 1078 to 1213.
[3] For example, for the calculations based on individual information, correlations are 0.996 for the first order corrected Gini index (GINIALLFOC), 0.999 for the sectoral relative median earnings (BETWEENALL), and 0.999 for the sectoral relative employment size of a sector (RELFREQALL).

calculations we apply standard LIS top- and bottom coding conventions, with 1 per cent of (for household inequality: equivalised) mean earnings as our bottom, and ten times the median (for household inequality: non-equivalised) earnings as our top boundary.

We restrict our sample to 'prime age workers', people aged between 25 and 54 with nonzero earnings following Mahler *et al*. (1999). This is the part of the population that is for its income most dependent on earnings from labour. In addition, this group probably has the strongest labour market attachment as their earnings are less affected by retirement and schooling decisions (Atkinson *et al*., 1995; Mahler *et al*., 1999). Based on this sample, we calculate the earnings inequality using household earnings (following Mahler *et al*., 1999) and using individual earnings for multiple sample definitions.

For the calculations based on household earnings, we correct for differences in household size using the square root equivalence scale. We apply household weights as standard in LIS.[4] We follow Mahler *et al*. (1999) by defining households as working in a particular sector if the household head is working in this sector.[5]

Yet, a problem with using household earnings is that the members of a household might work in different sectors, so that earnings are attributed to sectors in which they were not necessarily made. Therefore, we also calculate inequality based on individual earnings. We use the personal weights[6] and we distinguish between three groups of individuals where we again only include people aged between 25 and 54 with nonzero earnings:

1. Household heads;
2. Household heads and spouses;
3. All individuals.

Here, we attribute the individual earnings to the sector in which the specific individual is working.[7]

We also show the absolute and relative number of households and individuals classified in a sector, both weighted and unweighted, and their weighted absolute and median earnings. The LIS weightings are used to transpose the sample indicators to the

[4] HWEIGHT in LIS.
[5] D16 in LIS.
[6] PPOPWGT in LIS.
[7] IND1_C in LIS.

population level. In this case, the population is the total number of households or individuals with the age and earnings restriction.

Contrary to Mahler *et al*. (1999), we do not include sectoral information for disposable income and the amount of redistribution, as taxes and transfers are set at the national level so that these regulations do not differ between sectors.

### 2.2 Sectoral classification and country sample

We standardise the classification of sectors based on the International Standard of Industrial Classification (ISIC) rev. 3.0 at the two digit level. The manufacturing and transport and telecommunication sector are further broken down using the ISIC 3.0 three digit level, as can be seen in Table 1.

**Table 1 Sectoral definitions based on the ISIC 3.0 codes**

| No. | Sector | ISIC rev. 3.0 code |
|---|---|---|
| 1 | Agriculture, hunting, forestry and fishing | C01T05 |
| 2 | Mining and quarrying | C10T14 |
| 3 | Manufacturing | C15T37 |
| 31 | Food products, beverages and tobacco | C15T16 |
| 32 | Textiles, textile products, leather and footwear | C17T19 |
| 33 | Wood and products of wood and cork | C20 |
| 34 | Pulp, paper, paper products, printing and publishing | C21T22 |
| 35 | Chemical, rubber, plastics and fuel products | C23T25 |
| 36 | Other non-metallic mineral products | C26 |
| 37 | Basic metals and fabricated metal products | C27T28 |
| 38 | Machinery and equipment | C29T33 |
| 39 | Transport equipment | C34T35 |
| 30 | Manufacturing n.e.c. and recycling | C36T37 |
| 4 | Electricity, gas and water supply | C40T41 |
| 5 | Construction | C45 |
| 6 | Wholesale and retail trade – restaurants and hotels | C50T55 |
| 7 | Transport and telecommunications | C60T64 |
| 71 | Transport and storage | C60T63 |
| 72 | Post and telecommunications | C64 |
| 8 | Finance, insurance, real estate and business services | C65T74 |
| 9 | Community, social and personal services | C75T99 |

In the LIS database multiple sectoral definitions are used across countries and waves, such as ISIC 2.0, or NAICS for the US. To consistently classify industries, we recompute all classification schemes to the ISIC 3.0 definitions. In general this did not require much interpretation, although sometimes some sectors needed to be excluded (mainly when no

distinction was made between C34T35 Manufacturing of transport equipment and C36T37 Manufacturing n.e.c. and recycling). Seven classification dummies are included. The classification scheme is included as a separate worksheet in the dataset.

As displayed in Table 2, this second version of our dataset presents updated data for eight developed countries data at the sectoral level. Combined with the original data from the first edition, data are available for twelve developed countries. Contrary to Mahler *et al*. (1999), we include Czech Republic and Ireland, and we have data from our previous version for Austria, Belgium, Poland, and Spain. Yet, we leave out Australia, Canada, Italy, and the Netherlands for which the data does not have enough detail to calculate inequality for a sufficient number of sectors.

Three waves in our original dataset, namely, the information for the year 2000 for Belgium, Ireland, and Spain, are based on net earnings. They are shown in italics in Table 2. The calculations for Germany in 1984 and 1989 are based on West-Germany. For Ireland, three consecutive waves with only few observations, 1994-1996, have been combined (with YEAR=1994-1996) where earnings information has been recalculated to 1995 levels using information on inflation from the World Bank. Due to the higher number of surveyed people, we recommend to use this combined observation for 1994-1996, which also has been updated, instead of the observations for the separate years from the original data. The inclusion of Spain, and to a lesser extent Belgium, requires caution as the number of surveyed people is low, leading to possibly inaccurate inequality estimations.

**Table 2 Country and wave sample**

| Country | Available waves |
|---|---|
| **Version 1.1 (updated data)** | |
| Czech Republic | 1996, 2004 |
| Denmark | 1987, 1992, 1995, 2000, 2004 |
| Finland | 1987, 1991, 1995, 2000, 2004 |
| Germany | 1984, 1989, 1994, 2000, 2004 |
| Ireland | 1994-1996, 2004 |
| Sweden | 1987, 1992, 2000, 2005 |
| UK | 1986, 1999, 2004 |
| US | 1986, 1991, 1994, 2000, 2004 |
| **Version 1 (data not updated in version 1.1)** | |
| Austria | 2004 |
| Belgium | 1995, *2000* |
| Ireland | 1994, 1995, 1996, 2000 |
| Poland | 1986, 1992, 1995, 1999, 2004 |
| Spain | 1995, 2000 |
| Sweden | 1981 |
| UK | 1969, 1979 |
| US | 1979, 1997 |

For a number of variables in the original dataset information is missing for Austria, Belgium, Poland, Spain, the waves with net earnings, the individual waves 1994-1996 for Ireland, Sweden 1981, UK 1969 and 1979, and the US 1979 and 1997. Thus, for these variables a total of 31 waves and 639 observations are available.

One possible application of this dataset is to use the data in panel data analysis. The waves can be included in an unbalanced panel dataset of five year periods, for instance from around 1985 to around 2005, leaving out the calculations based on net earnings. This leads to the exclusion of the UK 1969 and 1979, Sweden 1981, and the US 1979 and 1997, and the three individual years 1994, 1995, and 1996, and the year 2000 for Ireland. This is also the sample that has been updated for this version.

## 3. Codebook and descriptives

In this section we show definitions of our main variables. In addition, we show descriptives for the updated dataset.

### 3.1 Country-level data based on household information

The dataset contains a number of indicators at the country level based on household information. The first two columns in Table 3 show the sum of unweighted (SUM) and weighted (SUMW) number of individuals within included households in the calculations at the sectoral level in the respective wave. The total sum of (weighted) households is potentially underestimated within a wave if information for a single industry is missing. We do not replace the sums of (weighted) households as missing. This problem could occur for the UK 1986, where information for transport and telecommunications is missing, and for Sweden 1987 and 1992 and Denmark 1987 and 1992, where information is missing for two subsectors within the manufacturing industry. Next, GINIC shows the level of equivalised earnings inequality as measured by the Gini index, pooled for all households part of our sectoral sample. P50C gives us the weighted median household earnings. Its summary statistics are not shown here as the indicator is expressed in national currency and current prices, making it not internationally comparable.

**Table 3 Country-level indicators based on household information**

| Variable name | SUM | SUMW | GINIC |
|---|---|---|---|
| Mean | 12,298 | 30,751,966 | 0.317 |
| Standard dev. | 12,109 | 53,294,823 | 0.042 |
| Minimum | 1,967 | 4,606 | 0.259 |
| Maximum | 39,944 | 176,450,466 | 0.404 |
| No. waves | 31 | 31 | 31 |

Thus, the average country-wave observation Gini index is 0.317 based on household information.

### 3.2 Country-level data based on individual information

The following country-level indicators are constructed on the basis of individual information. Table 4 shows the sum of the unweighted number of individuals using the three groups of individuals (SUMALL for all individuals, SUMHS for household heads and spouses, and SUMH for household heads only). Table 5 shows the same information but then for the weighted frequencies. As with the sums of households, the total sum of (weighted) individuals is potentially underestimated in the waves for which information

for a single industry is missing (UK 1986, Sweden 1987 and 1992, and Denmark 1987 and 1992).

**Table 4 Unweighted individual frequencies at the country level**

| Variable name | SUMALL | SUMHS | SUMH |
|---|---|---|---|
| Mean | 19,956 | 18,360 | 11,420 |
| Standard dev. | 20,392 | 18,455 | 11,500 |
| Minimum | 3,418 | 3,108 | 1,963 |
| Maximum | 70,288 | 61,857 | 38,980 |
| No. waves | 31 | 31 | 31 |

**Table 5 Weighted individual frequencies at the country level**

| Variable name | SUMWALL | SUMWHS | SUMWH |
|---|---|---|---|
| Mean | 20,194,416 | 1,7925,895 | 11,846,810 |
| Standard dev. | 29,840,568 | 25,821,000 | 16,886,444 |
| Minimum | 1,176,274 | 970,426 | 589,259 |
| Maximum | 96,144,183 | 82,184,773 | 53,097,115 |
| No. waves | 31 | 31 | 31 |

Next, Table 6 summarises the level of inequality for the sample based on all individuals, using a number of indicators. We report the Gini index (GINIALLC), the mean log deviation (GE0ALLC), the Theil index (GE1ALLC), and the Atkinson index with inequality aversion parameter $\varepsilon = 0.5$ (AT05ALLC). The dataset also contains information on the median individual earnings for the three sample definitions (P50ALLC, P50HSC, and P50HC), for which summary statistics are not shown here as they are expressed in national currency and current prices.

**Table 6 Individual earnings inequality at the country level**

| Variable name | GINIALLC | GE0ALLC | GE1ALLC | AT05ALLC |
|---|---|---|---|---|
| Mean | 0.322 | 0.236 | 0.194 | 0.098 |
| Standard dev. | 0.051 | 0.064 | 0.060 | 0.027 |
| Minimum | 0.257 | 0.145 | 0.125 | 0.066 |
| Maximum | 0.421 | 0.359 | 0.335 | 0.155 |
| No. waves | 31 | 31 | 31 | 31 |

### 3.3 Sectoral data based on household information

Now we move to sectoral data based on household information. In Table 7 we show descriptives for the unweighted and weighted number of households per sector (FREQ and WFREQ respectively). Next, RELFREQ shows the weighted relative employment size of a sector, defined as the number of households classified in a sector divided by the total number of households (WFREQ/SUMW * 100%). The relative employment size maps sectoral employment shifts relative to the total labour market per country, sector, and over time. As an example, the average sector contains 639 households.

The relative employment size of sectors is potentially overestimated for the waves for which information on the number of households for a single industry within that wave is missing, as this causes the denominator to be underestimated (as stated above, this problem could occur for UK 1986, Sweden 1987 and 1992, and Denmark 1987 and 1992).

**Table 7 Frequencies and relative employment size based on household information**

| Variable name | FREQ | WFREQ | RELFREQ |
|---|---|---|---|
| Mean | 764 | 1,924,483 | 6.413 |
| Standard dev. | 1,528 | 5,608,224 | 7.562 |
| Minimum | 5 | 22 | 0.139 |
| Maximum | 13,115 | 50,300,000 | 40.373 |
| No. observ. | 639 | 639 | 639 |

Next, Table 8 summarises descriptives for sectoral earnings inequality for multiple indicators based on equivalised household information. We report the Gini index (GINI), the P90/P10 ratio (P90P10), the mean log deviation (GE0), the Theil index (GE1), and the Atkinson index with inequality aversion parameter $\varepsilon = 0.5$ (AT05). The Gini coefficient is to a certain extent sensitive to the sample size for which the Gini index is calculated. For the Gini index bootstrapped standard errors with 250 repeats are calculated (BSSE250) to provide a confidence interval of the level of inequality.

Deltas (2003) shows this for different cumulative distributions, using Monte Carlo simulations. The Gini index underestimates the 'true' inequality level when the sample size is low. Deltas calculates that by multiplying the Gini index by $N / (N - 1)$, which he calls the first order correction, the underestimation bias is significantly reduced. As for some industries, in particular mining and wood manufacturing, the number of people

interviewed is often low, we include his first order procedure by calculating GINIFOC as the GINI * FREQ / (FREQ – 1). We use the unweighted frequencies here as the bias arises from the number of people interviewed (the sample).

Last, we show the relative median wage (BETWEEN), a measure of inequality between rather than within industries, calculated as the sectoral median wage divided by its counterpart at the national level (P50/P50C). Again, summary statistics for the P50 are not shown here as the indicator is expressed in national currency and current prices, so that it is not internationally comparable.

**Table 8 Earnings inequality based on household information**

| Variable name | GINI | GINIFOC | BSSE250 | P90P10 | GE0 | GE1 | AT05 | BETWEEN |
|---|---|---|---|---|---|---|---|---|
| Mean | 0.286 | 0.289 | 0.020 | 4.270 | 0.171 | 0.149 | 0.075 | 89,942 |
| Standard dev. | 0.058 | 0.058 | 0.015 | 1.709 | 0.076 | 0.066 | 0.031 | 86,108 |
| Minimum | 0.131 | 0.138 | 0.002 | 1.813 | 0.027 | 0.026 | 0.013 | 864 |
| Maximum | 0.560 | 0.578 | 0.121 | 16.552 | 0.598 | 0.626 | 0.268 | 340,069 |
| No. observ. | 639 | 639 | 639 | 639 | 639 | 639 | 639 | 639 |

### 3.4 Sectoral data based on individual information

The first three columns of Table 9 summarise the unweighted number of persons classified in a sector for the three groups of individuals. Columns 4-6 provide the same information for the weighted number of persons.

**Table 9 Frequencies based on individual information**

| Variable name | FREQALL | FREQHS | FREQH | WFREQALL | WFREQHS | WFREQH |
|---|---|---|---|---|---|---|
| Mean | 1,213 | 1,116 | 716 | 1,235,864 | 1,098,555 | 745,679 |
| Standard dev. | 2,694 | 2,473 | 1,424 | 3,428,399 | 3,004,466 | 1,892,857 |
| Minimum | 6 | 6 | 3 | 1,201 | 1,201 | 974 |
| Maximum | 23,340 | 22,133 | 12,604 | 30,800,000 | 27,100,000 | 16,700,000 |
| No. observ. | 639 | 639 | 639 | 639 | 639 | 639 |

Table 10 describes the weighted relative employment size of a sector, defined as the number of individuals classified in a sector divided by the total number of individuals for the three groups of individuals. As was the case for household information, the sectoral relative employment sizes of the UK 1986, Sweden 1987 and 1992, and Denmark 1987 and 1992 are potentially overestimated.

**Table 10 Relative employment size based on individual information**

| Variable name | RELFREQALL | RELFREQHS | RELFREQH |
|---|---|---|---|
| Mean | 0.062 | 0.062 | 0.064 |
| Standard dev. | 0.084 | 0.085 | 0.078 |
| Minimum | 0.001 | 0.001 | 0.001 |
| Maximum | 0.438 | 0.440 | 0.386 |
| No. observ. | 639 | 639 | 639 |

Next, we show in Table 11 the Gini index for the three groups of individuals, both the 'normal' one and the first-order corrected version, which decreases bias due to small sample size.

**Table 11 Gini based on individual information**

| Variable name | GINIALL | GINIHS | GINIH | GINIALLFOC | GINIHSFOC | GINIHFOC |
|---|---|---|---|---|---|---|
| Mean | 0.186 | 0.186 | 0.166 | 0.285 | 0.284 | 0.269 |
| Standard dev. | 0.114 | 0.117 | 0.116 | 0.073 | 0.075 | 0.075 |
| Minimum | 0.010 | 0.010 | 0.010 | 0.083 | 0.083 | 0.082 |
| Maximum | 1.099 | 1.143 | 1.130 | 0.597 | 0.602 | 0.595 |
| No. observ. | 639 | 639 | 639 | 639 | 639 | 639 |

Table 12 provides information on the mean log deviation or GE(0) and the weighted relative median wage for the three groups of individuals. These are again calculated by dividing the sectoral median wage by its counterpart at the country level. Again, summary statistics for the sectoral median wage themselves are not shown here because they are expressed in national currency and current prices and therefore not internationally comparable.

**Table 12 GE(0) and relative median wage based on individual earnings**

| Variable name | GE0ALL | GE0HS | GE0H | BETWEENALL | BETWEENHS | BETWEENH |
|---|---|---|---|---|---|---|
| Mean | 0.186 | 0.186 | 0.166 | 1.050 | 1.049 | 1.018 |
| Standard dev. | 0.114 | 0.117 | 0.116 | 0.194 | 0.194 | 0.185 |
| Minimum | 0.010 | 0.010 | 0.010 | 0.159 | 0.151 | 0.145 |
| Maximum | 1.099 | 1.143 | 1.130 | 1.897 | 1.820 | 2.772 |
| No. observ. | 639 | 639 | 639 | 639 | 639 | 639 |

## 4. Comparison to Mahler *et al*. (1999)

Comparing to Mahler et al. (1999), we extend the dataset of sectoral earnings inequality in three ways as can be seen in Table 13. Firstly, we calculate earnings inequality for 12

countries and between 1969 and 2005, while Mahler et al. (1999) provide data for 10 countries and between 1984 and 1992. Secondly, we include more inequality measures. Mahler *et al*. only calculate the P90/P10 whilst we also include the Gini index, the Atkinson index ($\varepsilon$ =0.5), the mean log deviation (GE(0)), and the Theil index (GE(1)) for household earnings. Next to calculations based on household information, we calculate the GE(0) and Gini coefficient using individual information, which allows us to more carefully attribute earnings to sectors. In addition, we follow the first order correction to reduce the underestimation bias by low sample size.

## Table 13 Comparison to Mahler *et al*. (1999)

| | **Mahler *et al*.** | **Leiden LIS Sectoral Income Inequality Dataset** |
|---|---|---|
| Launched | 1999 | June 2013 |
| Last update | 1999 | Version 1.1: May 2014 |
| No. of countries | 10 | 12 (updated in second version: 8) |
| Countries | Australia, Canada, Denmark, Finland, Germany, Italy, the Netherlands, Sweden, UK, and US | Version 1.1: Czech Republic, Denmark, Finland, Germany, Ireland, Poland, Spain, Sweden, UK, and US. Version 1 also includes Austria, Belgium, Poland, and Spain. |
| LIS waves | II and III | Version 1.1: II, III, IV, V, and VI. Version 1 also includes 0 for UK and I for a set of countries |
| Time series | 1984-1992 | Version 1.1: 1984-2005. Version 1: 1969-2005 |
| Total no. of included LIS waves | 18 | Version 1.1: 31. Version 1: 49 |
| ISIC scheme | ISIC 2.0 | ISIC 3.0 |
| Income unit | Equivalised household earnings | - Equivalised household earnings<br>- Individual income using three individual definitions |
| Earnings definition | - income from wages and salaries or self-employment<br>- disposable income<br>- redistribution | - income from wages and salaries or self-employment |
| Within sector earnings inequality indicators | For household earnings:<br>- P90/P10<br><br>For individual earnings:<br>None | For household earnings:<br>- Gini index<br>- Gini index with first order correction (Deltas, 2003)<br>- Atkinson index ($\varepsilon$ =0.5)<br>- P90/P10<br>- Mean log deviation (GE(0))<br>- Theil index (GE(1))<br><br>For individual earnings:<br>- Gini index<br>- Gini index with first order correction (Deltas, 2003)<br>- Mean log deviation (GE(0)) |
| Between sector inequality | - Sectoral median earnings / country median earnings | - Sectoral median earnings / country median earnings<br>- Sectoral employment size / country employment size |

**Contact**

Any questions about the Leiden LIS Sectoral Income Inequality Dataset may be addressed to:

Chen Wang, Economics Department, Leiden University, PO Box 9520, 2300 RA Leiden, The Netherlands. E-mail: c.wang@law.leidenuniv.nl

Stefan Thewissen, Economics Department, Leiden University, PO Box 9520, 2300 RA Leiden, The Netherlands. E-mail: s.h.thewissen@law.leidenuniv.nl

Olaf van Vliet, Economics Department, Leiden University, PO Box 9520, 2300 RA Leiden, The Netherlands. E-mail: o.p.van.vliet@law.leidenuniv.nl

## Appendix 1. Full variable list

| Variable name | Definition |
|---|---|
| **Identifiers** | |
| COU | Country abbreviation |
| CNTRY | Country code |
| YEAR | LIS survey year |
| PERIOD | Period number {1, …6} (for panel data analysis, with six periods of each around five years between 1980-2005) |
| INDUS | Sectoral code based on ISIC rev. 3.0 |
| CLASSIFICATION | Full sectoral name based on ISIC rev. 3.0 |
| **Industry classification scheme** | |
| ISIC 3 | Dummy for ISIC 3.0 (ISIC 3=1 if the LIS micro data (variable D16) of that wave is classified based on ISIC 3.0) |
| ISIC 2 | Dummy for ISIC 2.0 (ISIC 2=1 if the LIS micro data (variable D16) of that wave is classified based on ISIC 2.0) |
| SIC | Dummy for SIC (SIC=1 if the LIS micro data (variable D16) of that wave is classified based on SIC) |
| OLD NAICS | Dummy for old NAICS classification (OLD NAICS=1 if the LIS micro data (variable D16) of that wave is classified based on an older version of NAICS classification) |
| NEW NAICS | Dummy for new NAICS classification (NEW NAICS=1 if the LIS micro data (variable D16) of that wave is classified based on the new version of NAICS classification) |
| Other | Dummy for other classification schemes (OTHER=1 if the LIS micro data (variable D16) of that wave is classified based on none of the aforementioned classification schemes |
| NET EARN | Dummy indicating waves for which net earnings are used |
| WGD | Dummy for West Germany (WGD=1 for Germany 1989, 1994) |
| **Country level data based on household information** | |
| SUM | Total number of individuals within households with household head aged 25-54 with nonzero household earnings classified in a sector |
| SUMW | Total weighted number of individuals within households with household head aged 25-54 with nonzero household earnings classified in a sector |
| GINIC | Gini index for households with household head aged 25-54 with nonzero household earnings classified in a sector |
| P50C | Weighted median household earnings with household head aged 25-54 with nonzero household earnings classified in a sector in national currency, current prices |
| **Country level data based on individual information** | |
| SUMALL | Total number of all individuals aged 25-54 with nonzero earnings classified in a sector |
| SUMHS | Total number of household heads and spouses aged 25-54 with nonzero earnings classified in a sector |
| SUMH | Total number of household heads aged 25-54 with nonzero earnings classified in a sector |
| SUMWALL | Total weighted number of individuals aged 25-54 with nonzero earnings classified in a sector |
| SUMWHS | Total weighted number of household heads and spouses aged 25-54 with nonzero earnings classified in a sector |
| SUMWH | Total weighted number of household heads aged 25-54 with nonzero earnings classified in a sector |
| GINIALLC | Gini index for all individuals aged 25-54 with nonzero earnings classified in a sector |
| GE0ALLC | Mean log deviation for all individuals aged 25-54 with nonzero earnings classified in a sector |
| GE1ALLC | Theil index for all individuals aged 25-54 with nonzero earnings classified in a sector |
| AT05ALLC | Atkinson's index (0.5) for all individuals aged 25-54 with nonzero earnings classified in a sector |

| | |
|---|---|
| P50ALLC | Weighted median earnings for all individuals aged 25-54 with nonzero earnings classified in a sector in national currency, current prices |
| P50HSC | Weighted median earnings for household heads and spouses aged 25-54 with nonzero earnings classified in a sector in national currency, current prices |
| P50HC | Weighted median earnings for household heads aged 25-54 with nonzero earnings classified in a sector in national currency, current prices |
| **Sectoral data based on household information** | |
| FREQ | Number of households with household head aged 25-54 with nonzero earnings |
| WFREQ | Weighted number of households with household head aged 25-54 with nonzero earnings |
| RELFREQ | Weighted relative sectoral employment size: WFREQ / SUMW * 100% |
| GINI | Sectoral Gini index (without first order correction) of households with household head 25-54 with nonzero household earnings |
| GINIFOC | Sectoral first order corrected Gini index based on Deltas (2003): FREQ / (FREQ – 1) * GINI, of households with household head 25-54 with nonzero household earnings |
| BSSE250 | Bootstrapped standard errors of the sectoral Gini (without first order correction) with 250 repeats |
| P90P10 | Sectoral P90/P10 ratio of households with household head 25-54 with nonzero household earnings |
| GE0 | Sectoral mean log deviation of households with household head 25-54 with nonzero household earnings |
| GE1 | Sectoral Theil index of households with household head 25-54 with nonzero household earnings |
| AT05 | Sectoral Atkinson index with parameter $\varepsilon = 0.5$ of households with household head 25-54 with nonzero household earnings |
| P50 | Weighted sectoral median earnings with household head aged 25-54 with nonzero household earnings, national currency, current prices |
| BETWEEN | Weighted relative sectoral median wage or inequality between sectors: P50 / P50C |
| **Sectoral data based on individual information** | |
| FREQALL | Number of all individuals aged 25-54 with nonzero earnings |
| FREQHS | Number of household heads and spouses aged 25-54 with nonzero earnings |
| FREQH | Number of household heads aged 25-54 with nonzero earnings |
| WFREQALL | Weighted number of all individuals aged 25-54 with nonzero earnings |
| WFREQHS | Weighted number of household heads and spouses aged 25-54 with nonzero earnings |
| WFREQH | Weighted number of household heads aged 25-54 with nonzero earnings |
| RELFREQALL | Weighted relative sectoral employment size, all individuals: WFREQALL / SUMWALL * 100% |
| RELFREQHS | Weighted relative sectoral employment size, household heads and spouses: WFREQHS / SUMWHS * 100% |
| RELFREQH | Weighted relative sectoral employment size, household heads: WFREQH / SUMWH * 100% |
| GINIALL | Sectoral Gini index (without first order correction) for all individuals aged 25-54 with nonzero earnings |
| GINIHS | Sectoral Gini index (without first order correction) for household heads and spouses aged 25-54 with nonzero earnings |
| GINIH | Sectoral Gini index (without first order correction) for household heads aged 25-54 with nonzero earnings |
| GINIALLFOC | Sectoral first order corrected Gini index based on Deltas (2003): FREQALL / (FREQALL – 1) * GINIALL, all individuals aged 25-54 with nonzero earnings |
| GINIHSFOC | Sectoral first order corrected Gini index based on Deltas (2003): FREQHS / (FREQHS – 1) * GINIHS, household heads and spouses aged 25-54 with nonzero earnings |
| GINIHFOC | Sectoral first order corrected Gini index based on Deltas (2003): FREQH / (FREQH – 1) * GINIH, household heads aged 25-54 with nonzero earnings |
| GE0ALL | Sectoral mean log deviation for all individuals aged 25-54 with nonzero earnings |
| GE0HS | Sectoral mean log deviation for household heads and spouses aged 25-54 with nonzero earnings |
| GE0H | Sectoral mean log deviation for household heads aged 25-54 with nonzero earnings |

| | |
|---|---|
| P50ALL | Weighted median individual earnings for all individuals aged 25-54 with nonzero earnings, national currency, current prices |
| P50HS | Weighted median individual earnings for household heads and spouses aged 25-54 with nonzero earnings, national currency, current prices |
| P50H | Weighted median individual earnings for household heads aged 25-54 with nonzero earnings, national currency, current prices |
| BETWEENALL | Weighted relative sectoral median wage or inequality between sectors, all individuals: P50ALL / P50ALLC |
| BETWEENHS | Weighted relative sectoral median wage or inequality between sectors, household heads and spouses: P50HS / P50HSC |
| BETWEENH | Weighted relative sectoral median wage or inequality between sectors, household heads: P50H / P50HC |