Incomes Across the Distribution Database

Stefan Thewissen*, Brian Nolan†, and Max Roser‡

May 2016

1 Introduction

How widely are the benefits of economic growth shared in advanced societies? Are the gains only going to the rich so that the middle of the distribution sees little or no improvement in living standards? Is growth raising the incomes of the poor so that they keep pace with or even narrow the gap to the middle?

To answer such questions one must be able to track real incomes at different points of the income distribution over time (Thewissen et al., 2015; Nolan et al., 2016b,a).

This paper describes the Incomes Across the Distribution Database (Thewissen et al., 2016), which aims to provide the information required for a large number of OECD countries for recent decades. It is based primarily on the microdata brought together by the Luxembourg Income Study (LIS), and includes observations for 180 country-years. This document provides a description of the database, including the sample selection and methodological choices employed.

2 LIS database and country coverage

The aim of the LIS database is to bring together microdata on household income and to standardise these insofar as possible – see `http://www.lisdatacenter.org` (LIS, 2015; Ravallion, 2015). For the Incomes Across the Distribution Database we restrict ourselves to developed countries for which at least 2 waves of data are available. In total we include 27 developed countries between 1978 and 2013, though the panel is strongly unbalanced in that the period covered is much longer for some countries than others.

The data are structured in waves rather than annual, often with a gap between observations of about five years. Appendix 1 provides an overview of the exact country coverage and the national surveys on which LIS relies. This also

*Institute for New Economic Thinking at the Oxford Martin School, Department of Social Policy and Intervention, and Nuffield College, University of Oxford. Corresponding author: stefan.thewissen@spi.ox.ac.uk

†Institute for New Economic Thinking at the Oxford Martin School, Department of Social Policy and Intervention, and Nuffield College, University of Oxford

‡Institute for New Economic Thinking at the Oxford Martin School, Oxford Martin School, and Nuffield College, University of Oxford

lists a small number of observations that we left out due to data breaks that gave rise to substantial changes in definitions or coverage, based on information about the sources (or its absence) and/or implausible patterns in the data. Appendix 1 lists all countries and years included in the database and lists for each country and year the underlying national micro data set. Appendix 2 lists all the variables included, and their construction is described below.

3 Population definition, decile cut-offs and means

In studying trends in living standards across the distribution, we gather information on decile cut-offs and decile means. We also include mean income (the means of the entire distribution). The 5th decile cut-off is the median of the distribution, which is often used as the point of reference in deriving relative income poverty thresholds (most often as 50% or 60% of the median) and now advocated as a key indicator of trends in living standards to complement GDP (Aaberge & Atkinson, 2013). Decile means capture developments at the very top and bottom better than cut-offs in theory, although household surveys tend not be particularly strong in measuring the tails of the distribution (Burkhauser et al., 2016). All information is gathered and presented both for the entire population and for working age households, defined for this purpose as households headed by someone aged between 18-65 (using the predefined variable for the household head available in the LIS micro data). In order not to affect the decile cut-offs or means in some form, we choose not to apply top and bottom coding. We set negative reported household incomes to zero (we will more extensively discuss our income definitions in Section 5) but we keep in all households with zero income.

4 Inequality measures

In addition to providing levels of incomes across the distribution, we include a set of summary inequality measures. Based on our income data by decile it is straightforward to calculate decile ratios such as the P90/P10, P90/P50, and P50/P10. We also calculate and include Gini indices based on the same sample, taking into account zero incomes (using the `ineqdec0` routine in Stata). We provide bootstrapped standard errors for these Gini indices as well.

Generally Gini indices are calculated on the basis of top- and bottom-coded data. We have compared our non-top and bottom coded Gini figures to the ones provided on the LIS website (as "key figures") which are based on the same income definition and equivalence scale but employing top- and bottom-coding. As expected, the correlation is very high (0.996), and our figures are never lower which is due to the fact that we do not 'reduce' the measured inequality by applying top and bottom coding. For some waves the difference between the two is larger than 0.01 Gini points (shown here for the entire population):

- Belgium 2000 (our Gini index is 0.318 vs. LIS key figures of 0.279)
- Norway 2000 (0.261 vs. 0.25)

- Norway 2004 (0.283 vs. 0.256)
- UK 2004 (0.354 vs. 0.344)
- USA 1994 (0.371 vs. 0.361)
- USA 1997 (0.374 vs. 0.36)
- USA 2000 (0.372 vs. 0.357)
- USA 2004 (0.377 vs. 0.364)
- USA 2007 (0.383 vs. 0.371)

For the USA the differences between our figures and the key figures provided by LIS are relatively constant over time, so that the change over time is not much affected. This is different for Belgium and in particular for Norway. As we show in Figure 1, for Norway top and bottom coding has a significant effect, which should be kept in mind when using the Gini index in our database (the one in the database is shown as "Including zero incomes"). We show the Gini for the entire population, equivalised disposable household income here. The discrepancy in the case of Norway might be explained by the high volatility of top incomes driven by tax reforms (Aaberge & Atkinson, 2010).

Figure 1: The Gini index for Norway

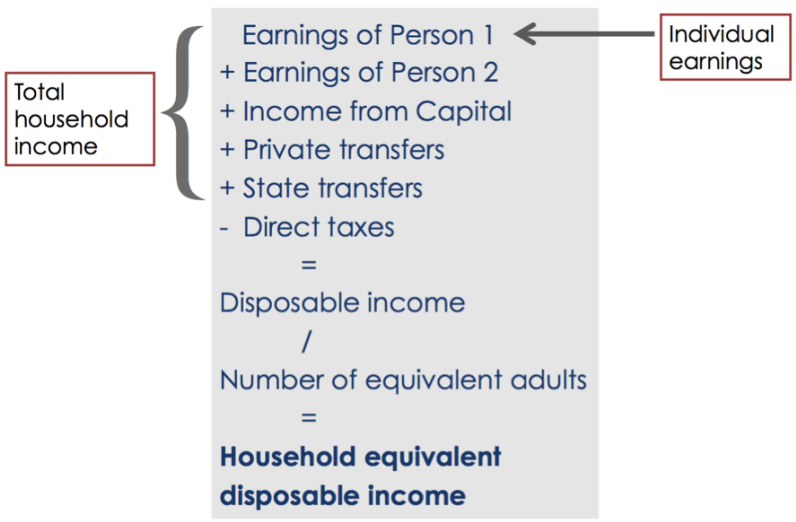
5 Income definitions

The decile cut-offs and means are calculated on the basis of disposable household income. The measure of disposable household income employed in LIS is paid employment and self-employment income, capital income, transfer income, which includes social security transfers (work-related insurance transfers, universal benefits, and assistance benefits) and private transfers, minus income

taxes and social security contributions. This follows the definitions of the Canberra Group, see also Figure 2.

The Gini index is available both for disposable and for market income, for the equivalised income concept (see Section 6). Market income is directly available in the LIS database and is defined as the sum of labour income (paid employment and self-employment income), and capital income (in the LIS variable list it is called "factor income"). The inclusion of both disposable and market income Gini indices allows us to calculate the measure of (absolute) redistribution defined as the Gini for market income minus the Gini for disposable income.

Figure 2: The composition of equivalised disposable household income



For 9 LIS waves, market income is not available (Estonia 2000; Ireland 1987, Poland 1999, 2004, Slovak Republic 1996, Spain 1985, Switzerland 2000, 2002, 2004). Moreover, the population sample on which market income is calculated might differ from the sample for the calculation of disposable income. This can happen if information for market income for a household is missing, but disposable income information is available. To check whether this affects our estimate, we conducted a sensitivity test where we calculated the Gini for disposable income on the sample for which market income is available (not included in the database). The correlation is almost perfect (0.9998) with the disposable income Gini for the original (full) sample. Only for a couple of waves the difference between the two is larger than 0.002 Gini points (here shown for the entire population, for the working age population very comparable differences):

- France 1978 (full sample Gini is 0.319, vs. market income sample 0.312)
- Greece 2007: 0.322 vs. 0.317
- Greece 2010: 0.338 vs. 0.331

- Hungary 2007: 0.277 vs. 0.287
- Hungary 2010: 0.279 vs. 0.276
- For Hungary 2012 the difference is only noticeable for the working age population: 0.294 vs. 0.301.

6 The equivalence scale

Two households on the same income but one comprising a single individual and another a couple with two children will have differing living standards, because a household with several individuals benefits from economies of scale in consumption. We include both equivalised and per capita decile information in the database. For per capita income, we calculate household income and divide it by the number of household members. For equivalised income, we assume economies of scale by applying the square root of the household size as the equivalence scale. Both per capita and equivalised income assume equal sharing of income across individuals within a household.

7 Data on CPI and deflators

To use income levels to measure living standards, we correct for differences in price levels over time and in purchasing power across countries. We use the consumer price index (CPI, from OECD Consumer Prices (MEI), all items) to deflate household income. To convert to a common currency we apply Purchasing Power Parities for actual individual consumption to household incomes sourced from OECD National Accounts. We express all PPP-adjusted figures in 2011 international dollars. We provide all incomes in both nominal terms and in inflation-adjusted 2011 international dollars.

8 Calculating growth rates

All data are provided in levels. Due to the gaps in the time series for each country, growth rates should be calculated as a compound annual growth rate. In Stata this can be done using this command for variable VAR:

```
foreach v of varlist VAR {
      generate CAGR`v' = ((`v'/`v'[_n-1])^(1/(year-year[_n-1])) - 1)
            * 100 if country==country[_n-1]
}
```

Another way is to index the data to for instance the first year available for each country.

9 Citation

When using the database, please cite:

Nolan, B., Roser, M., & Thewissen, S. (2016). GDP per capita versus median household income: What gives rise to divergence over time? *INET Working Paper Series no. 2016-03*.

Thewissen, S., Nolan, B., & Roser, M. (2016). Incomes across the distribution database. `https://ourworldindata.org/incomes-across-the-distribution`.

References

Aaberge, R. & Atkinson, T. (2010). *Top Incomes: A Global Perspective*, chapter Top Incomes in Norway. Oxford University Press.

Aaberge, R. & Atkinson, T. (2013). The median as watershed,. *Statistics Norway Discussion papers no. 749*.

Burkhauser, R., Herault, N., Jenkins, S., & Wilkins, R. (2016). What has been happening to uk income inequality since the mid-1990s? answers from reconciled and combined household survey and tax return data. *Melbourne Institute Working Paper Series Working Paper No. 5/16*.

LIS (2015). Micro data runs for multiple countries completed in september 2015.

Nolan, B., Roser, M., & Thewissen, S. (2016a). GDP per capita versus median household income: What gives rise to divergence over time? *INET Working Paper Series no. 2016-03*.

Nolan, B., Roser, M., & Thewissen, S. (2016b). Models, regimes, and the evolution of middle incomes in oecd countries. *LIS Working Paper Series no. 660*.

Ravallion, M. (2015). The Luxembourg Income Study. *The Journal of Economic Inequality*, 13(4), 527–547.

Thewissen, S., Kenworthy, L., Nolan, B., Roser, M., & Smeeding, T. (2015). Rising income inequality and living standards in oecd countries: How does the middle fare? *LIS Working Paper Series no. 656*.

Thewissen, S., Nolan, B., & Roser, M. (2016). Incomes across the distribution database. *https://ourworldindata.org/incomes-across-the-distribution*.

Appendix 1

Table A1: List of included waves

Country	Year	Source that LIS relies on
Australia	81, 85, 89, 95, 01, 03, 08	Survey of Income and Housing Costs (SIHC)
	10 (designed to be comparable)	Household Expenditure Survey (HES) and Survey of Income and Housing (SIH)
Austria	*87, 95 (incomparable, left out)*	*Austrian Microconsensus*
	94, 97, 00	European Household Panel / AT ECHP
	4	EU-SILC
Belgium	85, 88, 92, 97	Socio-Economic Panel
	95, 00	Panel Study of Belgian Households (PSBH) / BE ECHP
Canada	*71, 75 (historical data, left out)*	*Survey of Consumer Finances (SCF)*
	81, 87, 91, 94, 97	Survey of Consumer Finances (SCF)
	98, 00, 04, 07, 10 (designed to be comparable)	Survey of Labour and Income Dynamics (SLID)
Czech	92, 96	Czech Microconsensus
	04, 07, 10	EU-SILC
Denmark	87, 92, 95, 00, 04, 07, 10	Law model
Estonia	0	Household Budget Survey
	04, 07, 10	Estonian Social Survey / EU-SILC
Finland	87, 91, 95, 00, 04	Income Distribution Survey (IDS)
	07, 10 (designed to be comparable)	SILC formerly known as IDS
France	78, 84, 89, 94, 00, 05, 10	Family Budget Survey (BdF)
Germany	*73, 78, 83 (incomparable, left out)*	*Income and Consumer Survey (EVS)*
	81 (incomparable, left out)	*German Transfer Survey*
	84, 89, 94, 00, 04, 07, 10	GSOEP
Greece	95, 00	Household Income and Living Conditions Survey / ECHP
	04, 07, 10	EU-SILC
Hungary	91, 94, 99, 05, 07, 09, 12	Household Monitor Survey
Iceland	04, 07, 10	EU-SILC
Ireland	87	Survey of Income Distribution, Poverty, and Usage of State Services
	94, 95, 96, 00	Living in Ireland Survey / IE ECHP
	04, 07, 10	EU-SILC
Israel	*79 (left out as no CPI/PPP data available)*	*Household Expenditure Survey*
	86, 92, 97, 01, 05, 07, 10	Household Expenditure Survey
Italy	86, 87, 89, 91, 93, 95, 98, 00, 04, 08, 10	Survey on Household Income and Wealth (SHIW)
Luxembourg	85, 91	Socio-Economic Panel (PSELL)
	94, 97, 00	ECHP
	4	SILC
	07, 10	Panel socio-economique "Liewen zu Letzebuerg" (PSELL III) / EU-SILC
Netherlands	*83, 87, 90 (incomparable, left out)*	*Additional Enquiry on the Use of (Public) Services (AVO)*
	93, 99	Socio-Economic Panel Survey
	04, 07, 10	EU-SILC
Norway	79, 86, 91, 95, 00, 04	Income Distribution Survey (IF)
	07, 10 (designed to be comparable)	Household Income Statistics (formerly based on the Income Distribution Survey)
Poland	*86 (left out as no CPI/PPP data available)*	*Household Budget Survey*
Poland	92, 95, 99, 04, 07, 10	Household Budget Survey
Slovak R	92, 96	Slovak Microconsensus
	04, 07, 10	EU-SILC
Slovenia	97, 99, 04, 07, 10	Household Budget Survey

Country	Year	Source that LIS relies on
Spain	80	Family Expenditure Survey
	85	Household Budget Continuous Survey (Encuesta Continua de Presupuestos Familiares – ECPF)
	90	Family Expenditure Survey
	95, 00	Spanish ECHP
	04, 07, 10	Encuesta de Condiciones de Vida (ECV) / EU-SILC
Sweden	*67 (incomparable, left out)*	*Income from Register Data, Demographics from the Level of Living Survey*
	75 (historical data, left out)	*Income Distribution Survey (HINK)*
	81, 87, 92, 95, 00, 05	Income Distribution Survey (HINK)
Switzerland	*82 (incomparable, left out)*	*Swiss Income and Wealth Survey*
	92 (incomparable, left out)	*Swiss Poverty Study*
	00, 02, 04	Income and Consumption Survey (EVE/ERC)
United Kingdom	*69, 74 (historical data, left out)*	*Family Expenditure Survey (FES)*
	79, 86, 91, 94, 95	Family Expenditure Survey (FES)
	99, 04, 07, 10 (designed to be comparable)	Family Resources Survey (FRS)
United States	*74 (historical data, left out)*	*Current Population Survey (CPS) – March supplement*
	79, 86, 91, 94, 97, 00	Current Population Survey (CPS) – March supplement
	04, 07, 10, 13 (designed to be comparable)	Current Population Survey (CPS) – Annual Social and Economic Supplement (ASEC)

Appendix 2

Table A2: List of variables

Variable name	Variable definition
country	Country
year	year
copentire10	Nominal cut-off 10 entire pop, equivalised disposable household income
copentire20	Nominal cut-off 20 entire pop, equivalised disposable household income
...	...
copentire90	Nominal cut-off 90 entire pop, equivalised disposable household income
copworking10	Nominal cut-off 10 working age pop, equivalised disposable household income
copworking20	Nominal cut-off 20 working age pop, equivalised disposable household income
...	...
copworking90	Nominal cut-off 90 working age pop, equivalised disposable household income
decentire10	Nominal decile mean 10 entire pop, equivalised disposable household income
decentire20	Nominal decile mean 20 entire pop, equivalised disposable household income
...	...
decentire100	Nominal decile mean 100 entire pop, equivalised disposable household income
decworking10	Nominal decile mean 10 working age pop, equivalised disposable household income
decworking20	Nominal decile mean 20 working age pop, equivalised disposable household income
...	...
decworking100	Nominal decile mean 100 working age pop, equivalised disposable household income
realPPP11_copentire10	Real PPP cut-off 10 entire pop, equivalised disposable household income
realPPP11_copentire20	Real PPP cut-off 20 entire pop, equivalised disposable household income
...	...
realPPP11_copentire90	Real PPP cut-off 90 entire pop, equivalised disposable household income
realPPP11_copworking10	Real PPP cut-off 10 working age pop, equivalised disposable household income
realPPP11_copworking20	Real PPP cut-off 20 working age pop, equivalised disposable household income
...	...
realPPP11_copworking90	Real PPP cut-off 90 working age pop, equivalised disposable household income
realPPP11_decentire10	Real PPP decile mean 10 entire pop, equivalised disposable household income
realPPP11_decentire20	Real PPP decile mean 20 entire pop, equivalised disposable household income
...	...
realPPP11_decentire100	Real PPP decile mean 100 entire pop, equivalised disposable household income
realPPP11_decworking10	Real PPP decile mean 10 working age pop, equivalised disposable household income
realPPP11_decworking20	Real PPP decile mean 20 working age pop, equivalised disposable household income
...	...
realPPP11_decworking100	Real PPP decile mean 100 working age pop, equivalised disposable household incom
percapentire10	Nominal cut-off 10 entire pop, per capita disposable household income
percapentire20	Nominal cut-off 20 entire pop, per capita disposable household income
...	...
percapentire90	Nominal cut-off 90 entire pop, per capita disposable household income
percapworking10	Nominal cut-off 10 working age pop, per capita disposable household income
percapworking20	Nominal cut-off 20 working age pop, per capita disposable household income
...	...
percapworking90	Nominal cut-off 90 working age pop, per capita disposable household income

Variable name	Variable definition
realPPP11_percapentire10	Real PPP cut-off 10 entire pop, per capita disposable household income
realPPP11_percapentire20	Real PPP cut-off 20 entire pop, per capita disposable household income
...	...
realPPP11_percapentire90	Real PPP cut-off 90 entire pop, per capita disposable household income
realPPP11_percapworking10	Real PPP cut-off 10 working age pop, per capita disposable household income
realPPP11_percapworking20	Real PPP cut-off 20 working age pop, per capita disposable household income
...	...
realPPP11_percapworking90	Real PPP cut-off 90 working age pop, per capita disposable household income
entiremean	Nominal mean equivalised disposable household income entire pop
workingmean	Nominal mean equivalised disposable household income working age pop
realPPP11_entiremean	Real PPP mean equivalised disposable household income entire pop
realPPP11_workingmean	Real PPP mean equivalised disposable household income working age pop
percapentiremean	Nominal mean per capita disposable household income entire pop
percapworkingmean	Nominal mean per capita disposable household income working age pop
realPPP11_percapentiremean	Real PPP mean per capita disposable household income entire pop
realPPP11_percapworkingmean	Real PPP mean per capita disposable household income working age pop
eydhientiregini	Gini equivalised disposable household income entire pop
eydhiworkinggini	Gini equivalised disposable household income working age pop
stdeydhientiregini	Standard error Gini equivalised disposable household income entire pop
stdeydhiworkinggini	Standard error Gini equivalised disposable household income working age pop
eymarketentiregini	Gini equivalised market household income entire pop
eymarketworkinggini	Gini equivalised market household income working age pop
absredisentiregini	Absolute redistribution equivalised disposable household income entire pop
absredisworkinggini	Absolute redistribution equivalised disposable household income working age pop
copentirep90p10	P90/p10 cut-off entire pop, equivalised disposable household income
copentirep90p50	P90/p50 cut-off entire pop, equivalised disposable household income
copentirep50p10	P50/p10 cut-off entire pop, equivalised disposable household income
percapentirep90p10	P90/p10 cut-off entire pop, per capita disposable household income
percapentirep90p50	P90/p50 cut-off entire pop, per capita disposable household income
percapentirep50p10	P50/p10 cut-off entire pop, per capita disposable household income
copworkingp90p10	P90/p10 cut-off working age pop, equivalised disposable household income
copworkingp90p50	P90/p50 cut-off working age pop, equivalised disposable household income
copworkingp50p10	P50/p10 cut-off working age pop, equivalised disposable household income
percapworkingp90p10	P90/p10 cut-off working age pop, per capita disposable household income
percapworkingp90p50	P90/p50 cut-off working age pop, per capita disposable household income
percapworkingp50p10	P50/p10 cut-off working age pop, per capita disposable household income
Consumerpricesallitems	Consumer prices - all items (OECD Consumer Prices (MEI) dataset)
pppp41	Purchasing Power Parities for actual individual consumption (OECD PPPs dataset)
pppp41_in2011	Purchasing Power Parities for actual individual consumption in 2011 (OECD PPPs dataset)