LIS
Technical Working Paper

No. 6

Comparative Research with Net and Gross Income Data: An Evaluation of Two Netting Down Procedures for the LIS

Rense Nieuwenhuis, Teresa Munzi, and Janet C. Gornick

August 2013

Revised in February 2016

Luxembourg Income Study (LIS), asbl

Comparative Research with Net and Gross Income Data: An Evaluation of Two Netting Down Procedures for the LIS Database

Rense Nieuwenhuis[1], Teresa Munzi[2], Janet Gornick[3]

February 25, 2016

Paper accepted for publication Review of Income and Wealth.

Abstract

Researchers seeking to perform country-comparative and trend analyses using income data have to account for the fact that income surveys differ in whether income is measured gross or net of taxes and contributions. We discuss, develop, and evaluate two 'netting down procedures' for data in the LIS Database. Evaluations of these netting down procedures indicate that comparisons across gross and net datasets can be greatly improved when netting down procedures are applied. In several cases, however, substantial amounts of bias remain.

JEL Codes: D3, P5, C8

Keywords: netting down, income, data harmonization, comparative research, LIS

[1]Swedish Institute for Social Research (SOFI), Stockholm University. rense.nieuwenhuis@sofi.su.se
[2]LIS. munzi@lisdatacenter.org
[3]LIS & Graduate Center, City University of New York. jgornick@gc.cuny.edu

1 Introduction

A common challenge in country-comparative and trend analyses of income using microdata, is that income surveys differ in whether income is measured gross or net of taxes and contributions. The issue of comparability between gross and net datasets is common in comparative datasets on income, including the data contained in the LIS Database.

LIS acquires existing income surveys, and harmonises them into a pre-defined template for comparative analysis. All LIS datasets provide fully comparable measures of disposable household income.[4] However, comparability problems arise with other income variables, because LIS provides income data that are *net* of (income) taxes in some countries or years, while providing *gross* income data in others. For users of LIS who seek to perform country-comparative analyses and/or analyses of trends within countries, this results in the challenge that their selected income variables refer to different (net vs. gross) concepts across datasets and therefore in most applications should not be compared directly. Of the 262 LIS datasets available at the time of writing, 74 (28%) were classified net, 175 (67%) as gross, and 13 (5%) as 'mixed'.[5] Datasets on the United States or the United Kingdom have always been gross, while Austria has been net. Other countries, such as Ireland and Luxembourg were covered by both net and gross datasets, at different points in time. Mixed datasets are a special case in which income variables are, for instance, as in France, net of mandatory contributions but gross of income tax. Such 'mixed' datasets are beyond the scope of this paper.

Researchers working with the LIS data have applied at least four different strategies for comparing gross and net datasets. The first is to include both types of datasets in the same (comparative) analysis, acknowledging incomparabilities that could lead to biased results. The second strategy is to restrict all analyses to either gross or net datasets. This results in accurate findings, but clearly limits the scope of the research. Third, LIS users sometimes present separate analyses using gross and net datasets. The limitation of this strategy is that differences in the results based on gross and net datasets could originate from the different earnings concepts, or from real differences across countries, or both. The fourth strategy is to modify the gross income data to approximate net income data. This process is referred to as netting down, and entails subtracting observed or estimated taxes from the gross income amounts. Such netting down procedures, however, have not been evaluated empirically for their capacity to produce measurements of income that are comparable across datasets.

This technical note presents and evaluates two netting down procedures. We present background information on the comparison of income in gross and net LIS datasets.[6] We then present the two

[4] Disposable household income is the income concept usually adopted for poverty and distribution analysis. It is also used for calculation of the LIS Key Figures: http://www.lisdatacenter.org/data-access/key-figures/disposable-household-income/

[5] For a continuously updated overview, see: http://www.lisdatacenter.org/our-data/lis-database/datasets-information/

[6] Although the authors are associated with LIS, the presentation of this method does not represent official LIS work nor an official LIS recommendation; LIS data users should feel free to utilise the method presented here or any other.

procedures for netting down, and evaluate their performance.

We introduce the rationale of netting down with reference to comparing *income*. In the empirical part, however, we focus on netting down the more narrow concept of *earnings* from dependent employment; as we will explain, the challenges of comparing gross and net data are more apparent with earnings. Program syntax is available in an online appendix.

2 Comparing gross and net income when using the LIS data

Comparing results based on gross versus net income data can be of substantive interest; carrying out research using a mix of the two types of data can also present challenges with respect to comparability.

The difference between gross and net income is of substantive interest, and can be assessed directly, when a single dataset contains information on both gross and net income, when information on taxes is available, and when additional information on social transfers is available. This allows LIS users to compare pre-tax-pre-transfer income to post-tax-post-transfer income, and thereby to answer a set of research questions about how taxes, and also social transfers, affect income distributions. This approach has been applied to study the effects of social welfare policies on poverty rates (see, e.g., Kenworthy, 1999). Other studies have evaluated how taxes and transfers affect poverty rates among specific subgroups, such as among children (see, e.g., Gornick and Jäntti, 2012), working-age populations (Gornick and Milanovic, 2015), single parents (Maldonado and Nieuwenhuis, 2015a,b), and migrant households (Sainsbury and Morissens, 2012). For such 'redistribution studies', the actual differences between gross and net income are of substantive interest, and both are compared *within* a single dataset.

The difference between gross and net income becomes a challenge in comparability, when comparing datasets of which some are gross and others are net. This is the case in country-comparative analyses and / or in trend analyses. It has been shown that country-comparative studies based on different earnings concepts across countries can be "*seriously misleading*" (Atkinson and Brandolini, 2001, p. 777).

The issue of comparability between gross and net datasets in LIS also applies to the above-mentioned redistribution studies. This is clarified using an example of a typical redistribution study on the comparison between *pre-tax-pre-transfer* income (referred to as 'market household income') to *post-tax-post-transfer* income (referred to as 'disposable household income') (Gornick and Jäntti, 2012). Market income is reported gross of income taxes and contributions in some LIS datasets, and net of income taxes and contributions in other LIS datasets. Without correction, this would have understated the poverty/inequality reduction in the net datasets, as the comparison between market income and disposable household income in these datasets only captures the effects of transfers, whereas in gross datasets this comparison would capture the combined effect of taxes and transfers. Thus, here too, comparisons of redistributive efforts drawing on a mix of gross and net datasets can be improved by netting down the

gross datasets.

2.1 Netting down, or grossing up?

An alternative to netting down gross income data would be to 'gross up' net income data. With LIS, however, grossing up is not possible, drawing on the microdata alone, as most net datasets do not contain information on taxes. To then estimate the gross income would require country-specific details on the tax system, which is beyond the scope of this technical note. Detailed simulations to this end are available for European countries through the Euromod project (Sutherland and Figari, 2013).

3 Netting down person-level earnings

So far we have discussed netting down gross *income*. In this section, we introduce practical complexities involved with netting down one specific type of income: that is, earnings. Earnings, of course, are typically studied at the level of individuals rather than households. Hence, we shift our focus to developing and evaluating two procedures designed specifically for netting down *person-level earnings*.

As noted earlier, when researchers studying redistribution across households are faced with a mix of gross and net datasets, they sometimes 'net down' the gross income data to enable meaningful comparisons across the two types of datasets. In short, by shifting to only net income they restrict their comparisons to the effects of transfers only (not of transfers combined with taxes).

Likewise, researchers concerned with earnings are often faced with a mix of gross and net datasets. They could simply mix the two (as some researchers have done), but that risks arriving at results that are incomparable across datasets. For example,if one is studying gender gaps in earnings and comparing results in a country with gross data (e.g. the United States) with results in a country with net data (e.g. Hungary), the results will be problematic.In the United States, with gross data, the researcher is capturing gaps in pay levels as set by employers; in countries with net data, such as Hungary, these same gender pay gaps have been reduced (in most cases) due to progressive taxation. The researcher thus is working with 'apples and oranges', and cannot know how much of the difference between the two countries is 'real' versus an artefact of the data.We argue that researchers ought never mix gross and net earnings data whether they are studying earnings disparities between groups, or other outcomes related to the distribution of earnings. The techniques laid out in this paper offer researchers one option for avoiding that kind of 'mixing'. In short, 'netting down' restricts comparisons across all datasets to after-tax earnings; thus, researchers are comparing 'apples to apples'.

In addition, in some cases, researchers want to assess net earnings.For example, some argue that net earnings are more appropriate than gross earnings in studies of intra-household bargaining.If partners are negotiating (for example, who does what domestic work) on the grounds of how much money each

'brings in', it is arguably net earnings that matter.Our netting down procedure offers researchers a way to 'net down' gross earnings, including in those cases when researchers prefer to work with net earnings for substantive reasons, and not just to maximise cross-country comparability.

3.1 Earnings as a specific income source

There are three basic sources of income: labor, capital, and transfers. Netting down the income from a specific source, such as earnings, is challenging both conceptually and practically. The conceptual problem with net earnings lies in the fact that countries can apply different tax rates to income from, for instance, labour and capital. As these tax rates are progressive based on total income, and only the total amount of paid income taxes is recorded in the data, it must be assumed that the average tax rate applied equally to all separate sources of income.

The practical problem with net income from separate sources is that as a result of the above, only information on total taxes is available. Therefore, to calculate the net income from a separate source, given the information available in the microdata, the assumption that income from each source was taxed at the same rate is required. This assumption is likely violated as most countries have different tax rates for income from labour and capital. However, as few households pay taxes on capital income, it remains to be seen how much this violated assumption leads to a biased approximation of net earnings. [7]

3.2 Earnings as a person-level concept

In order to calculate net earnings at the person level, person-level income taxes and social contributions must be subtracted from person-level gross earnings. In countries with joint taxation, however, this is conceptually challenging because the amount of taxes to be paid is determined at the level of the household. This also means that individual 'personal' earnings depend on the earnings of other household members. Joint taxation often pertains to the head of the household and her/his spouse, with separate taxation of the income from additional earners such as older children or relatives living in the household.[8] If no person-level tax variables are available, netting down person-level earnings requires the assumption that the taxes paid at the household level were paid by each household member proportionally to the share of the total household income received by that member. This assumption is likely violated in joint taxation regimes, but it remains an empirical question to what extent this leads to a biased approximation of net earnings.

[7] In the country-samples used in our analyses only 3% of individuals lived in a household in which capital income represented more than 10% of gross household income, with 1% in Estonia 2004 and 7% in Belgium 1997.

[8] In the country-samples used in our analyses approximately 16% of individuals, other than the household-head or spouse, contributed more than 10% of total household earnings. This percentage ranged from 12% in the United Kingdom 2004 to 25% in Ireland 2004.

3.3 Two netting down procedures

We developed two programs that perform netting down procedures, available for STATA, SPSS, SAS, and R. One procedure uses information on taxes at the person-level. If these are not available, the other procedure can be used based on household-level tax information. The LIS website has a table providing information on whether datasets are gross or net. Datasets classified as mixed should be treated with more caution, as the earnings reported in these datasets can be gross of income taxes but net of contributions, or vice versa. This is reported in detail in the LIS data documentation per country. All LIS datasets also contain a variable named `grossnet`, providing information on how earnings (and other income variables) were reported.

The person-level netting down procedure can only be used when person-level variables on taxes (LIS variable `pmxiti`) and (self-paid) social security contributions (`pmxitss`) are available. It first calculates the proportion of earnings in the total taxable income: gross earnings (`pmile`), self employment (`pmils`), unemployment compensation benefits (`pmitsisun`), short-term sickness and work injury benefits (`pmitsissi`), family leave benefits (`pmitsisma`), and pensions (`ppension`). Then it calculates net earnings by subtracting from the gross earnings the value of taxes paid, proportional to the amount of total income obtained from earnings (`propearnings`). This assumes that the total amount of taxes was distributed proportionally over all sources of income. As taxable income is made up of different components across countries, this procedure is based merely on an approximation of taxable income. The calculation is shown in equation 1.

$$\begin{aligned} propearnings = & \quad \frac{pmile}{pmile + pmils + pmitsisun + pmitsissi + pmitsisma + ppension} \\ net = & \quad pmile - ((pmxiti + pmxitss) * propearnings) \end{aligned} \tag{1}$$

The household-level netting down procedure can be used when tax information is available only at the household level. It calculates the percentage of the total monetary household income (`hmi`) that remains after taxes (`hmi - hmxit`) and multiplies gross person-level earnings (`pmile`) by this percentage. This assumes that this percentage is equal across all members of the household, and applies equally to all sources of income. The calculation is shown in equation 2.

$$net = pmile * \frac{hmi - hmxit}{hmi} \tag{2}$$

It should be noted that these netting down procedures are deliberately simple, in the sense that no country-specific rules were applied. The benefit is that these procedures can be applied to all gross LIS datasets. At the same time, if users wish to modify these procedures to account for specific countries' tax systems, they can do so.

4 Method and data

4.1 Method

A select number of LIS datasets has both gross and net earnings variables, as well as information on taxes and social security contributions on both the person-level and the household-level. This provides a unique opportunity for evaluating netting down procedures.

To evaluate a netting down procedure, we applied it to a gross earnings variable, and compared the resulting 'netted-down' variable to the original net earnings variable in the LIS dataset. We calculated bias for each percentile in the earnings distribution:

$$Bias(\%) = \frac{X_{nd} - X_n}{X_n} \times 100\% \tag{3}$$

in which X_{nd} represents the earnings in the 'netted-down' earnings variable, and X_n represents the net earnings reported in the LIS dataset. The resulting bias is expressed as a percentage of the reported net earnings. So, a bias of 0% means that the results based on the 'netted-down' earnings variable are identical to those reported in the original net earnings variable. If the bias % is larger than 0, this means that the netted-down earnings are higher than those based on the reported net earnings; a percentage below 0 indicates that the netted-down results are lower.

In addition to calculating bias for the earnings levels of different percentiles, we also calculated bias (again based on equation 3) for commonly used measures of inequality: the ratio of the 75th to the 25th percentile of earnings, the Theil index, the Coefficient of Variation, the Gini, the low earnings rate (defined as the percentage of earners with earnings below 2/3 of median earnings), and the gender gap in earnings (defined as: (male earnings - female earnings) / male earnings).

4.2 Data

The netting down procedures described here can be applied to LIS datasets harmonised using the 'new' (post-2011) template that are classified as gross (LIS, 2016). The evaluation of these netting down procedures, however, required the availability of both gross and net earnings variables in the data, which could only be the case when using the 'old' template (pre-2011). The required earnings variables, as well as person- and household-level variables on taxes and social contribution were available in 7 datasets: Austria 2004, Belgium 1992, Belgium 1997, Estonia 2004, Ireland 2004, United Kingdom 1999, and United Kingdom 2004. We restricted our analyses to those observations with valid information on both the gross and net earnings variables. This ensured that our measurement of bias was not affected by the possibility that gross and net earnings variables were based on different observations.

5 Results

Figure 1 shows bias incidence curves, representing the amount of bias associated with a netting down procedure for each percentile in the earnings distribution. The solid lines represent the scenario in which no netting down was applied, i.e. a direct comparison between gross and net earnings. Of course, the difference between gross and net does not necessarily indicate bias, as they represent different earnings concepts. However, the lines represent a reference point to evaluate the performance of netting down procedures compared to no netting down at all. In all countries, the results show that the differences between gross and net earnings increase at higher percentiles, which of course results from progressive taxation.

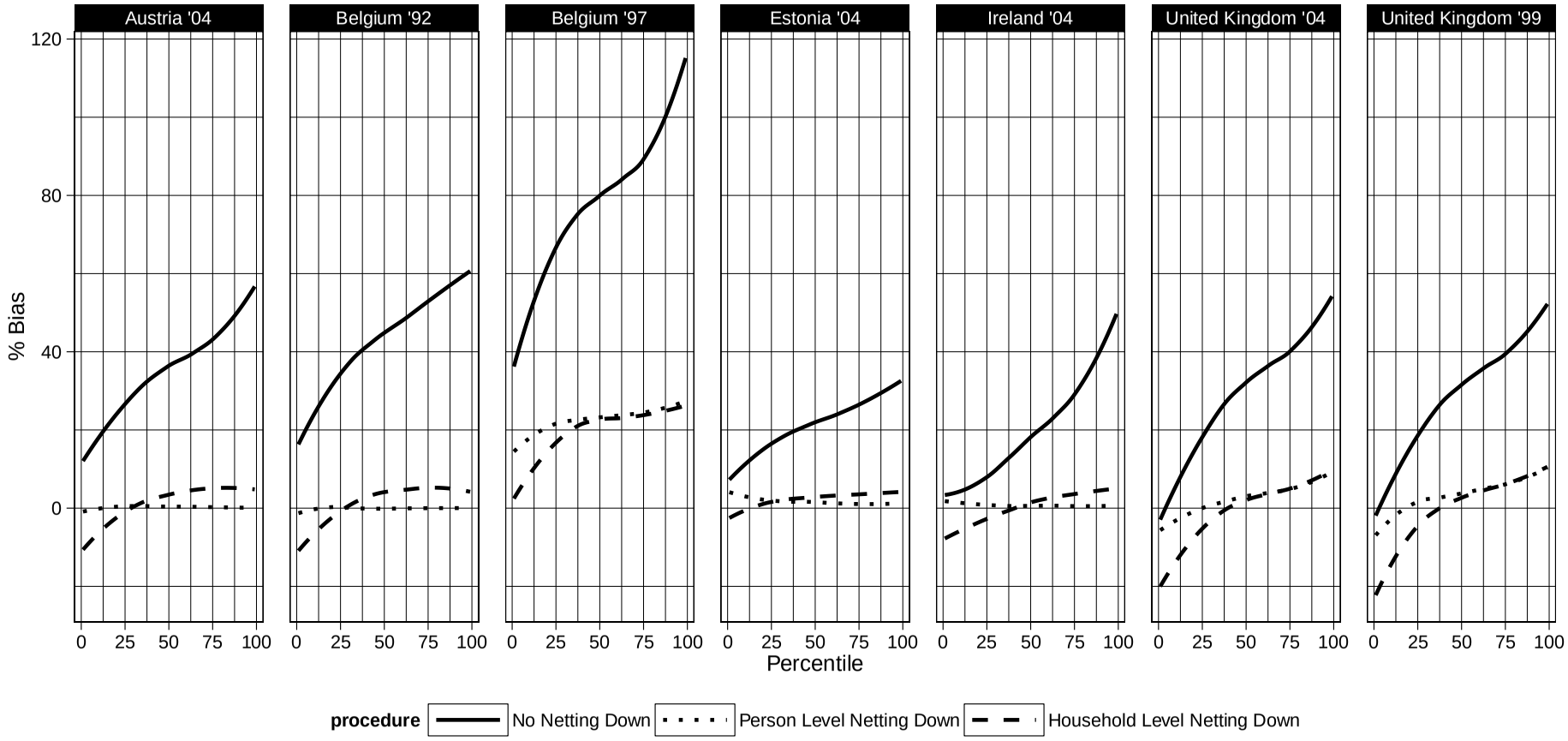
The dotted lines represent the bias resulting from the netting down procedure that used person-level tax information. The results suggest that this procedure typically underestimates earnings levels at lower percentiles (bias < 0) and overestimates at higher percentiles (bias > 0). Bias levels are substantially smaller than not correcting at all (the solid line), but reaches levels above 25% at higher percentiles in the Belgium 1997 and both UK datasets. In the other datasets, levels of bias are close to 0 at all percentiles.

The dashed lines represent the bias resulting from the netting down procedures that used household-level tax information. While the patterns are similar as described above, the household-level netting down procedure typically performs less well than the person-level procedure.

These results demonstrate that applying netting down procedures is preferable over not correcting for the difference between gross and net earnings in comparative research, and that preferably the person-level procedure is applied. Nevertheless, in some datasets substantial amounts of bias remained, particularly at the higher percentiles. As the net earnings tend to be under-estimated at lower percentiles and over-estimated at higher percentiles, the bias incidence curves further suggest that various estimates of inequality based on netted down earnings variables would be biased upwards. The extent of this bias, however, is difficult to assess from these curves. Therefore, Tables 1 and 2 present estimates of bias for six commonly applied measures of inequality based on the person-level and household-level netting down procedures, respectively.

The bias of the person-level netting down procedure (in Table 1) typically was below 10%, with the clear exception of the Theil index and the Coefficient of Variation in the United Kingdom. In that country, the bias incidence curve of the person-level netting down procedure continued sloping upwards at higher income percentiles. The bias of the household-level netting down procedure (in Table 2) was typically higher, with many estimates upwards of 10%. Again, the Theil index and Coefficient of Variation in the United Kingdom show exceptionally high levels of bias, up to 51%. In some cases the netting down procedures were associated with a negative bias, indicating that using the netted down earnings variable resulted in a under-estimate of inequality. An example is the share of low earners in Estonia 2004: the

Figure 1: Bias incidence curves for two netting down procedures compared to no netting down



person-level netting down procedure is associated with a bias of −2%. Overall, comparing the results presented in Tables 1 and 2 suggests that the person-level netting down procedure outperformed the household-level netting-down procedure.

Table 1: Quantifying Bias (%) in Person-Level Netting Down Procedure

Dataset	75p/25p	Theil Index	Coefficient of Variation	GINI	Low Earnings	Gender Gap
Austria '04	-0.8	-0.1	0.2	-0.1	-0.2	0.0
Belgium '92	0.0	-0.8	-0.3	-0.5	-1.3	-0.5
Belgium '97	0.4	8.4	4.8	4.3	1.4	2.0
Estonia '04	-1.2	-1.2	-0.8	-0.5	-2.3	-1.3
Ireland '04	-0.8	-0.5	-0.5	-0.2	-0.4	-0.2
United Kingdom '99	4.2	13.9	17.3	5.0	3.5	2.9
United Kingdom '04	4.8	30.7	48.7	7.9	4.9	5.0

Table 2: Quantifying Bias (%) in Household-Level Netting Down Procedure

Dataset	75p/25p	Theil Index	Coefficient of Variation	GINI	Low Earnings	Gender Gap
Austria '04	5.7	6.2	2.8	3.2	6.0	6.4
Belgium '92	5.9	8.8	3.0	5.5	29.0	8.1
Belgium '97	6.4	12.2	3.6	7.7	20.5	13.2
Estonia '04	1.4	1.9	0.5	1.2	-2.5	2.3
Ireland '04	6.6	5.9	1.4	3.4	3.3	2.8
United Kingdom '99	11.0	21.7	20.8	8.8	9.2	8.2
United Kingdom '04	10.0	38.0	51.3	11.5	9.6	10.0

6 Conclusion

This technical note presented guidelines for comparing gross and net income datasets, which were tailored to use with the LIS data but which apply to a wider range of income datasets. Two netting down procedures were introduced that approximate net earnings from information regarding gross earnings, in conjunction with data on taxes and social contributions paid by the household. Using these netting down procedures reduced bias in comparisons of earnings between gross and net LIS datasets. Generally, this suggests that applying a netting down procedure is preferable over not netting down. Using the person-level procedure was desirable over using the procedure based on household-level tax variables.

Data availability will often dictate which of the two netting down procedures users can apply. It should be noted, however, that it is to be expected that the household-level netting down procedure performs better in a country with joint-taxation, relative to countries with separate taxation. Furthermore, in both netting down procedures it is assumed that all sources of income are taxed at the same rate. From this, the expectation follows that the procedures will perform better in countries with a single, rather than a dual tax system in which separate tax rates exist for capital income and other income.

The netting down procedures performed more poorly in the datasets for Belgium and the United Kingdom, compared to other datasets. The LIS dataset on Belgium in 1997 was based on the Socio-Economic Panel. In this original dataset, the information on holiday- and end-of-year bonuses was

available net of taxes, and unavailable gross of taxes. Hence, in calculating the yearly gross earnings when preparing the LIS dataset, the monthly earnings were multiplied by 13,85 (approximating the average bonuses). For the net yearly earnings the information on bonuses was available in the original data. Hence, whereas in the harmonised LIS dataset the net yearly earnings account for person-level variation in bonuses independent of other earnings, in the gross yearly earnings the bonuses that affect such person-level variation were not accounted for. Second, the LIS datasets from the United Kingdom (both in 1999 and 2004) were based on the Family Resources Survey. During the recoding of these datasets to the LIS templates, the gross earnings were specified to include income from odd jobs, while net earnings could not be specified to include this source of income. Hence, the difference between gross and net yearly earnings is an overestimation of the 'real difference. Therefore, the netted down results may actually be a better representation of persons true net earnings than the net earnings reported in the data. It should be noted, that within the scope of this paper it was not possible to empirically test this statement.

When comparing a large number of gross and net datasets, users may want to statistically control for the different netting down procedures used. In regression-based analyses, for instance, this could be done by adding dummy variables indicating the observations derived from datasets netted down with the person-level procedure, and another dummy for the observations from datasets that were netted down using the household-level procedure (with observations from net datasets as the reference category).

To conclude, country-comparative and trend analyses of earnings based on both gross and net datasets should be done with caution. The netting down procedures presented here typically improve comparability in studies based on the LIS data. However, depending on the outcome measure of interest, and especially when no person-level tax variables are available, netting down procedures based solely on household income can sometimes result in substantially biased approximations of net earnings.

7 Appendices

A detailed description of the differences between net and gross datasets in LIS

Income surveys widely differ in whether income is measured gross or net of taxes and contributions.[9] Many surveys collect income gross of income taxes and social security contributions (the part paid by the employee only). Among these 'gross' income surveys, we find some further variation as to the way that data on taxes and contributions are collected. Arguably, the highest quality datasets are those based on register data. In these datasets information on gross income is perfectly mirrored by the information on the actual amount of taxes and contributions paid, as registered in the tax administration records. Examples are most Nordic Income Surveys, which are fully based on income information from tax registers, as well as surveys such as the Canadian Survey of Labour and Income Dynamics - SLID, in which taxes and social contributions are collected directly from tax records. In some datasets, the amounts of taxes and contributions are provided by the respondent. For instance, this is the case in the Irish and the Slovak Survey of Income and Living Conditions - SILC. The most typical situation is that in which taxes and contributions are imputed by the data provider based on the information collected in the survey. This is the case in the US Current Population Survey - CPS, the Australian Survey of Income And Housing - SIH, and the German Socio Economic Panel - SOEP. Yet, in other cases the data provider does not provide any measure of taxes and contributions, so that it is left to the user to impute such an amount in order to construct a measure of disposable income. This is notably the case of surveys in some Latin American countries, such as the Brazilian National Household Sample Survey - PNAD, the Colombian Great Integrated Household Survey - GEIH, and the Panamanian Continuous Household Survey - ECH. For those surveys LIS has had to carry out a simulation of taxes and contributions in order to get a measure of disposable household income. In the LIS Database, these surveys are obviously labeled as 'gross'.

On the other hand, there are a number of surveys that collect income data net of taxes and contributions. This is common in countries where respondents would know their net incomes better than the gross ones. In those cases, income taxes and social security contributions may either be collected separately, so that it is possible to create a measure of gross incomes by adding the two components, or not. Among the first group are cases such as the UK Family Resources Survey - FRS, in which incomes are asked net of taxes and contributions (this survey uses the notion of 'take-home pay'), and then several questions are asked in order to come up with an exact measure of the taxes and contributions withheld or successively paid on each of the incomes received. A similar procedure is applied in some household budget surveys where net incomes are asked and the amounts of taxes and contributions are

[9] Although the authors are associated with LIS, the presentation of this method does not represent official LIS work nor an official LIS recommendation; LIS data users should feel free to utilise the method presented here or any other.

then collected among the expenditures, split by the type of income on which they are being paid (such as the Polish Household Budget Survey). In these cases, it was possible to retrieve gross income amounts at the level of the subcomponents, and hence the datasets have been labeled as 'gross' in LIS.

In other net income surveys, information on taxes and contributions is available based on successive imputation carried out by the data provider. This is the case in the Greek Survey of Income and Living Conditions - SILC. In most of those cases, whereas it is possible to come up with a measure of total gross income by adding the several net incomes and the amount of taxes and social contributions, it is impossible to get measures of gross income at the subcomponent level (as taxes and contributions are most likely available for the overall income, and not by income source). For this reason, LIS codes these datasets as 'mixed': the total gross income is available, but all the subcomponents of income are net and not gross. However, the most common situation with respect to net surveys is that there is no information at all on taxes and contributions, so that it is not possible to have gross income amounts (not even for total household income). This is the case of many household budget surveys that only collect consumption expenditures (e.g., Georgia, Mexico, Slovenia, Serbia), as well as many other income surveys (such as the Italian Survey of Income and Wealth - SHIW, the Hungarian Household Monitor Survey, and the Russian Longitudinal Monitoring Survey - RLMS). In all those cases, labeled as 'net' in LIS, as soon as the researcher wishes to depart from the concept of disposable household income, the comparability between gross and net datasets becomes an issue.

Finally, there are datasets that are neither purely gross, nor purely net. First of all, this happens in all the cases where both gross and net incomes are collected (this is the case in the Spanish and Luxembourg Survey of Income and Living Conditions - SILC), or when the respondent is given the choice whether to report gross or net incomes and the data provider then imputes whatever the respondent has not given (this is the case in the Estonian Social Survey). In those cases, all information needed is available (gross and net incomes at the subcomponent level as well as at the total household income level), so that there is no comparability issue. Those cases are coded as 'gross' in LIS, as all the income subcomponents report gross measures, and the total household income is also measured in gross terms. However, there are other mixed situations where it is not possible to come up with a comparable measure of gross income, not even at the total household level, and then comparability again becomes an issue. This can happen in cases when information is available only for taxes but not for contributions (this is the case in the French Household Budget Survey, where incomes are collected net of social security contributions but, because taxes are not withheld at source, gross of taxes), or in cases when the information on taxes and contributions was only available for certain subcomponents of total income (e.g., the cases in which taxes and contributions are only available for wage income, such as in the Peruvian National Household Survey - ENAHO and the South African National Income Dynamics Study - NIDS), or for some subsets of the respondents (in the Chinese Household Income Survey Project CHIP, amounts are gross for rural

and urban sub-samples, and net for migrants sub-sample). Those datasets are flagged as being 'mixed' income datasets in LIS.

STATA Program Person-Level netting down

```
* Check whether a program with name 'persontax' exists
* Removes the program from memory if needed:
capture program drop persontax

* Start definition of the person-level netting down procedure:
program define persontax

* Removes variables from memory if needed (e.g. from previous runs)
capture drop propwage
capture drop earnings

* Replace the missing values into zeros for variables that are completely empty:
quietly sum pmils
replace pmils=0 if  r(N) == 0
quietly sum pmitsisun
replace pmitsisun=0 if  r(N) == 0
quietly sum pmitsissi
replace pmitsissi=0 if  r(N) == 0
quietly sum pmitsisma
replace pmitsisma=0 if  r(N) == 0
quietly sum ppension
replace ppension=0 if  r(N) == 0
quietly sum pmxiti
replace pmxiti=0 if  r(N) == 0
quietly sum pmxitss
replace pmxitss=0 if  r(N) == 0

* Calculate the proportion of earnings into total taxable earnings:
gen propwage = pmile / (pmile+pmils+pmitsisun+pmitsissi+pmitsisma+ppension)

* Calculate net earnings by subtracting from the gross earnings,
* the income taxes and contributions in the same proportion
* as earnings into total taxable earnings:
gen earnings = pmile - ((pmxiti + pmxitss) * propwage)

* End definition of the person-level netting down procedure:
end
```

STATA Program Household-Level netting down

```
* This program assumes that household and person-level datasets are merged
* Check whether a program with name 'household' exists
* Removes the program from memory if needed:
capture program drop householdtax

* Start definition of the household-level netting down procedure:
program define householdtax

* Removes variables from memory if needed (e.g. from previous runs)
capture drop netpercentage
capture drop earnings

* Calculate the proportion of net household earnings
* as percentage of gross household earnings:
gen netpercentage = (hmi - hmxit) / hmi if hi!=0

* Calculate net person-level earnings by multiplying
* gross earnings with the percentage net earnings:
gen earnings = pmile * netpercentage

* End definition of the household-level netting down procedure:
end
```

R Program Person-Level netting down

```
# Start definition of the person-level netting down procedure:
persontax <- function(x)
{

# Replace the missing values into zeros for variables that are completely empty:
if(sum(is.na(x$pmils)) ==length(x$pmils))  x$pmils <- 0
if(sum(is.na(x$pmitsisun)) ==length(x$pmitsisun))  x$pmitsisun <- 0
if(sum(is.na(x$pmitsissi)) ==length(x$pmitsissi))  x$pmitsissi <- 0
if(sum(is.na(x$pmitsisma)) ==length(x$pmitsisma))  x$pmitsisma <- 0
if(sum(is.na(x$ppension)) ==length(x$ppension))  x$ppension <- 0
if(sum(is.na(x$pmxiti)) ==length(x$pmxiti))  x$pmxiti <- 0
if(sum(is.na(x$pmxitss)) ==length(x$pmxitss))  x$pmxitss <- 0

# Calculate the proportion of earnings into total taxable earnings:
x <- within(x, propwage <-
pmile / (pmile+pmils+pmitsisun+pmitsissi+pmitsisma+ppension))

# Calculate net earnings by subtracting from the gross earnings
# the taxes and contributions in the same proportion
# as earnings into total taxable earnings:
x <- within(x, earnings <- pmile - ((pmxiti + pmxitss) * propwage))

# End definition of the person-level netting down procedure:
return(x)
}
```

R Program Household-Level netting down

```
# This program assumes that household and person-level datasets are merged
# Start definition of the household-level netting down procedure:
householdtax <- function(x)
{

# Calculate the proportion of net household earnings
# as percentage of gross household earnings:
x <- within(x, netpercentage <- (hmi - hmxit) / hmi)
x <- within(x, netpercentage[hi==0] <- NA)

# Calculate net person-level earnings by multiplying
# gross earnings with the percentage net earnings:
x <- within(x, earnings <- pmile * netpercentage)

# End definition of the household-level netting down procedure:
return(x)
}
```

SPSS Program Person-Level netting down

```
* Start definition of the person-level netting down procedure:.
define persontax() .

* Replace the missing values into zeros for variables that are completely empty: .

COMPUTE ismissing = nmiss(pmils) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmils = 0 .

COMPUTE ismissing = nmiss(pmitsisun) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmitsisun = 0 .

COMPUTE ismissing = nmiss(pmitsissi) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmitsissi = 0 .

COMPUTE ismissing = nmiss(pmitsisma) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmitsisma = 0 .

COMPUTE ismissing = nmiss(ppension) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) ppension = 0 .

COMPUTE ismissing = nmiss(pmxiti) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmxiti = 0 .

COMPUTE ismissing = nmiss(pmxitss) .
AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES overwrite=YES
 /allmiss=MEAN(ismissing).
IF (allmiss=1) pmxitss = 0 .

* Calculate the proportion of earnings into total taxable earnings: .
COMPUTE totalinc = pmile + pmils + pmitsisun + pmitsissi + pmitsisma + ppension .
IF (totalinc ~= 0) propwage = pmile / totalinc .
```

```
* Calculate net earnings by subtracting from the gross earnings,
* the income taxes and contributions in the same proportion
* as earnings into total taxable earnings: .

COMPUTE earnings = pmile - ((pmxiti + pmxitss) * propwage) .
IF (missing(propwage)) earnings = $sysmis .

* End definition of the person-level netting down procedure: .

!enddefine .
```

SPSS Program Household-Level netting down

```
* Start definition of the household-level netting down procedure:.
* This program assumes that household and person-level datasets are merged .

define householdtax () .

* Calculate the proportion of net household earnings
* as percentage of gross household earnings: .
if hi~=0 netpercentage = (hmi - hmxit) / hmi .

* Calculate net person-level earnings by multiplying
* gross earnings with the percentage net earnings: .
compute earnings = pmile * netpercentage .

* End definition of the household-level netting down procedure: .
!enddefine .
```

SAS Program Person-Level netting down

```
%MACRO indNetDown ;
/*-------------------------------------------------------------------*/
/*              INDIVIDUAL-LEVEL NETTING DOWN PROCEDURE              */
/*-------------------------------------------------------------------*/
/* NOTE -- This program assumes that you replace the macro variable */
/* &CCYYp by the valid LIS individual (SAS) dataset name such       */
/* as &ie04p.                                                       */
/*-------------------------------------------------------------------*/

DATA myPFile (DROP=i j k epmils epmitsisun epmitsissi epmitsisma eppension epmxiti epmxitss totalinc) ;
* Keep the value of the former record ;
RETAIN epmils epmitsisun epmitsissi epmitsisma eppension epmxiti epmxitss ;
* Loop over the entire dataset ;
IF _N_ = 1 THEN DO i = 1 TO all;
* Open the dataset ;
SET &CCYYp (KEEP=pmils  pmitsisun  pmitsissi  pmitsisma  ppension  pmxiti  pmxitss pmile) NOBS=all ;
* Create arrays with a new lsit of control variable (VarExist1-varExist7) to count the number of ;
* times a variable is missing ;
ARRAY varList  {7} pmils   pmitsisun pmitsissi  pmitsisma   ppension  pmxiti pmxitss ;
ARRAY varExist {7} epmils epmitsisun epmitsissi epmitsisma eppension epmxiti epmxitss (7*0) ;
* Count the number of times each selected variable (Array: varList) get a missing value;
DO j = 1 TO DIM(varList) ;
IF varList{j} = . THEN varExist{j} + 1 ;
END ;
END ;

* Re-open the dataset including the counters ;
SET &CCYYp  NOBS = total;
* Replace missing by 0 when the entire variable is not filled;
DO k = 1 TO DIM(varList) ;
IF varExist(k) = total  THEN varList(k) = 0 ;
END ;
* Calculate the proportion of earnings into total taxable earnings ;
totalinc = pmile + pmils + pmitsisun + pmitsissi + pmitsisma + ppension;
propwage = . ;
IF (totalinc ne 0) THEN propwage = pmile / totalinc  ;
/* Calculate net earnings by subtracting from the gross earnings, */
/* the income taxes and contributions in the same proportion as   */
/* earnings into total taxable earnings:                          */
earnings = . ;
IF (propwage ne .) THEN earnings = pmile - ((pmxiti + pmxitss) * propwage) ;
RUN ;
%MEND indNetDown ;
```

SAS Program Household-Level netting down

```
%MACRO hldNetDown ;
/*-------------------------------------------------------------------------------*/
/*              HOUSEHOLD-LEVEL NETTING DOWN PROCEDURE                           */
/*-------------------------------------------------------------------------------*/
/*  NOTE -- This program assumes that your targeted LIS household dataset and   */
/*  the person-level dataset created in the MACRO indNetDown (myPFile) are      */
/*  merged into a dataset called myMerge.                                        */
/*-------------------------------------------------------------------------------*/

DATA myMerge ;
SET myMerge ;
* Calculate the proportion of net household earnings as ;
* percentage of gross household earnings  ;
netpercentage = . ;
IF (hi ne 0) THEN netpercentage = (hmi - hmxit) / hmi ;
* Calculate net person-level earnings by multiplying gross earnings ;
* with the percentage net earnings ;
earnings = pmile * netpercentage ;
RUN ;

%MEND hldNetDown ;
```

References

Atkinson, A. and A. Brandolini, "Promise and pitfalls in the use of secondary data-sets: Income inequality in OECD countries as a case study", *Journal of Economic Literature*, 39, 771–799, 2001.

Gornick, J. C. and M. Jäntti, "Child poverty in cross-national perspective: Lessons from the Luxembourg Income Study," *Children and Youth Services Review*, 34, 558–568, 2012.

Gornick, J. C. and B. Milanovic, "Income Inequality in the United States in Cross-National Perspective: Redistribution Revisited," *LIS Center Research Brief*, 1, 2015.

Kenworthy, L. "Do social-welfare policies reduce poverty? A cross-national assessment," *Social Forces*, 77, 1119–1139, 1999.

Luxembourg Income Study (LIS) Database (2016). http://www.lisdatacenter.org *(multiple countries; Database Accessed June 2015 - December 2015)* Luxembourg: LIS.

Maldonado, L. C. and R. Nieuwenhuis, "Family Policies and Single Parent Poverty in 18 OECD Countries, 1978-2008," *Community, Work & Family*, 18:395–415, 2015a.

_____ "Single-Parent Family Poverty in 24 OECD Countries: A Focus on Market and Redistribution Strategies," *LIS Center Research Brief*, 2, 2015b.

Sainsbury, D. and A. Morissens, "Immigrants' Social Rights across Welfare States," in Sainsbury, D., *Welfare States and Immigrant Rights. The Politics of Inclusion and Exclusion*, Chapter 6, Oxford University Press, Oxford, 2012.

Sutherland, H. and F. Figari, "EUROMOD: the European Union tax-benefit microsimulation model". *International Journal of Microsimulation*, 6, 4–26, 2013.