Unequal distributions:
EG DNA versus DINA approach

By JORRIT ZWIJNENBURG*

In the past years, projects have been initiated to develop distributional results in line with national accounts totals. Their results differ from those obtained from micro statistics, but also show differences across projects, due to different underlying concepts and methodology to derive the results. This paper explains the differences between the work by the OECD-Eurostat Expert Group on Disparities in National Accounts (EG DNA) and the work on Distributional National Accounts (DINA) as developed in the context of the World Wealth and Income Database. It shows that different concepts and assumptions in the compilation process may significantly affect distributional results.

* Organisation for Economic Co-operation and Development (OECD). Email: Jorrit.Zwijnenburg@oecd.org.

1. Introduction

The past years have seen an increased interest in distributional information on the household sector. In response, several projects have been initiated to further improve the availability and the quality of distributional measures. A significant number of these initiatives further explore the use of micro economic statistics, but two initiatives specifically focus on the development of distributional results in line with national accounts totals. Their aim is to provide a more comprehensive overview of the distribution of household income, consumption and wealth, consistent with economy wide aggregates such as GDP and household disposable income.

The first initiative concerns the development of methodology to compile distributional measures of household income, consumption and saving within the framework of the national accounts by an OECD-Eurostat Expert Group on Disparities in a National Accounts framework (EG DNA) that was set up in 2011. The expert group already engaged in two exercises (see Fesseau and Mattonetti (2013) and Zwijnenburg, Bournot and Giovannelli (2017)), and at the end of 2018 seven countries had already started to publish distributional results in line with this methodology. The work of the group is still ongoing, trying to address specific methodological issues and exploring possibilities to improve the timeliness of the distributional results.

The second initiative relates to work in the context of the World Wealth and Income Database (WID.world) to develop annual estimates of the distribution of income and wealth consistent with national accounts concepts, not only compiling results at an aggregated level, but also including the production of synthetic income and wealth micro-files, providing the opportunity to publish results at a very granular level of detail. This work is known as the Distributional National Accounts (DINA) project. The project already covers data for more than 60 countries

and work is still ongoing to further improve the results and to broaden the range of countries included in the project.

As both projects aim to compile distributional results in line with national accounts totals they share a lot of similarities. However, there are also several differences, which may give rise to differences in results. This paper provides an overview of the main differences, focusing on the distribution of household income. In addition to broadening the knowledge on both initiatives, the paper aims to open up discussion on pros and cons of using certain concepts and assumptions in the compilation of distributional results. This should give further impetus to a correct understanding and interpretation of the results, as well as to their quality and usefulness.

The paper is structured as follows. Section 2 discusses differences in scope between the two projects. Section 3 provides an overview of conceptual differences, focusing at target population, unit of analysis, and income concepts. Section 4 zooms in on differences in methodology. The paper ends with some conclusions in Section 5.

2. Scope

Both projects aim to compile distributional results in line with national accounts concepts. However, while the DINA project focuses on income and wealth, the EG DNA project currently focuses on income, consumption and saving, planning to include the wealth component in a second phase. In addition to providing users with comprehensive distributional information, this also has the advantage for compilers to crosscheck the consistency of the results (i.e. checking the consistency of the full set of accounts per household group). As the methodology to arrive at distributional results depends on several assumptions and on the quality of the underlying data, this possibility to crosscheck the results is expected to add to the quality of the results.

Both projects also differ in the level of detail with which they aim to compile distributional results. While the EG DNA project aims to arrive at breakdowns of the household sector at an aggregated level (the level of detail depending on the quality and the available detail from the underlying micro data and the robustness of the methodology), the DINA project also aims at arriving at synthetic micro-files. On the basis of these files, the DINA project also publishes distributional results at much more granular levels, e.g. at the percentile level and even at more detailed levels for the top percentiles.

3. Concepts

3.1   Target population

The target population slightly differs for both projects. The DINA project applies the adult individual as the benchmark, aiming to estimate the distribution of income and wealth for all individuals of 20 years old and above, which means that all individuals below 20 years old are excluded from their inequality measures. The EG DNA project, on the other hand, looks at private households, excluding transactions by institutional households such as people living in prison or retirement homes, as these households are assumed to have little or no autonomy of decision in economic matters, and are usually not covered by micro data sources. Although it is not possible to say upfront how these differences in target population will affect inequality results, it is clear that they may give rise to different outcomes between the two projects.

3.2   Unit of analysis

Both projects apply different units of analysis. The DINA project looks at inequality on the basis of the income and wealth of the *individual*, whereas the EG DNA project looks at inequality on the basis of the *equivalized* income and consumption of the *household*. The main reason for the DINA project to focus on the individual seems related to the objective of arriving at homogenous results on the basis of fiscal data. As fiscal units may concern single persons as well as multi-person households with different consumption needs, focusing on results at the fiscal level would not provide a fair comparison, particularly in times of changing household compositions. Two approaches are applied to arrive at individual results in case of fiscal units that consist of more than one adult. The "equal-split-adults series" split income and wealth equally between the adult household members, whereas the "individualistic-adults series" attribute income and wealth to each individual income earner and wealth owner.

The EG DNA project focuses on the *equivalized* income and consumption of the *household*, based on the assumption that income is shared within the household and that certain types of goods (mainly housing and food) are consumed collectively by the household. Equivalence scales are used to correct for differences in size and composition of households, taking into account economies of scale for additional household members. In this way, the EG DNA arrives at homogenous results for households of different size and composition.

It is clear that these different approaches may lead to different income results and consequently inequality measures. The latter will largely depend on the composition of households across the distribution. If these are distributed rather equally, the impact will be negligible. However, if the composition differs across the distribution, DINA and EG DNA are likely to come up with different results.

3.3   Income concepts

Both projects also focus on different income concepts when analyzing inequality. The EG DNA project focuses on household disposable and adjusted disposable income, whereas the DINA project looks at pre-tax factor income, pre-tax national income, post-tax disposable income, and post-tax national income. These latter concepts all align to national income, and are thus broader than the income of the household sector as defined in the system of national accounts. In addition to the income of the household sector, they also include the income of the other domestic sectors (i.e. the non-financial corporations sector, the financial corporations sector, the general government sector, and the non-profit institutions serving households sector), under the assumption that in the end all income within the economy is benefiting households. As a result, the DINA income measures generally exceed their equivalents in the system of national accounts as defined for the household sector.

 Pre-tax factor income is the income accruing to the individual owners of the production factors, labour and capital, before any taxes or transfers have taken place. It is equal to the primary income of the household sector plus the primary incomes of all other domestic sectors in the economy, corrected for any taxes (less subsidies) on production as paid by the various sectors. Because of its broader coverage, pre-tax factor income exceeds primary income of the household sector in all of the countries, ranging from 14% in the United States to 48% in Sweden, the main driver for this difference being undistributed profits of non-financial corporations and the adjustment to correct for paid taxes (less subsidies) on production (see Zwijnenburg (2017)).

Pre-tax national income is equal to pre-tax factor income, but taking into account the operation of the pension system. This means that pension contributions are deducted from pre-tax factor income and pension benefits are added. Furthermore, as contributions and benefits may not be equal, any resulting gap is also attributed to households to arrive at the same values as pre-tax factor income.

Post-tax disposable income is equal to pre-tax national income, but after taking into account the operation of the tax/transfer system, i.e. deducting taxes on production, income and wealth, and taking into account non-pension social contributions and benefits in cash. As the focus is on national income, this does not only concern amounts as paid and received by the household sector, but also those paid and received by the other domestic sectors. Whereas post-tax disposable income comes close to disposable income (for the domestic economy) as defined in the national accounts, it omits the impact of other current transfers. These cover non-life insurance premiums and claims, but also transfers such as remittances, which may be an important form of income for certain household groups. Consequently, their exclusion may significantly affect the distributional results as derived in DINA.

Finally, post-tax national income is equal to post-tax disposable income plus social transfers in kind, collective consumption expenditure and government primary surplus. It comes close to adjusted disposable income (for the domestic economy) as defined in the national accounts, except for the inclusion of collective consumption expenditure and government primary surplus. As this may concern large amounts, their inclusion may have a significant impact on the income measures. Post-tax national income exceeds adjusted disposable income of the household sector as used in EG DNA for all countries, the difference ranging from 9.3% in Portugal to 36.3% in Denmark.

As explained above, the main argument for taking into account national income for the economy as a whole is that ultimately any income in the economy will benefit domestic households (see Alvaredo (2016)). However, some reservations can be made when looking at some specific items. Firstly, although there may indeed be arguments to assign undistributed profits of domestic corporations to their owners, it has to be borne in mind that not all portfolio equity of these corporations will be held by domestic households and that domestic households may also own portfolio equity in foreign corporations. Incorporating information from the revaluation account of the household sector would lead to results that are more accurate in that regard. Moreover, it would also provide information on other holding gains and losses as earned by households on their assets and liabilities.

Secondly, when looking at the allocation of any government surplus or deficit, it is questionable whether it makes sense to attribute the full amount to the current population. Usually, any balancing of surpluses or deficits is spread over time, most likely involving future generations. Due to the variation of surpluses and deficits over time, the various generations involved, and the difficulty to predict who in the end is going to pay for a deficit or benefit from a surplus, their inclusion and allocation to underlying households seems arbitrary and complex.

Thirdly, it is questionable whether it makes sense to attribute any gaps between pension contributions and benefits to specific individuals, particularly when considering that the pension system to a large extent concerns re-distribution over time at the individual level, in which an individual makes contributions to save for his own retirement. The correct allocation of the gap would in that case imply offsetting the initial pension transactions, returning to pre-tax factor income levels at the individual level.

Fourthly, the inclusion of collective consumption is debatable as by definition it concerns consumption that benefits the community at large and consequently cannot be allocated to specific individuals. This is also acknowledged in the Guidelines when it is explained that "it is extremely difficult to do this type of imputation, and it is not even clear whether it really makes any sense to attribute public spending such as roads or police to individuals".

Finally, the exclusion of other current transfers is regarded as a major shortcoming in the post-tax income measures in DINA, as for some households they may form a very important income source. When looking at the results of the 2015 exercise conducted by the EG DNA (see Zwijnenburg, Bournot and Giovannelli (2017)), net other current transfers in cash constitute 20.8% of disposable income of the lowest income quintile in Mexico, 16.1% in Israel, and 8.8% in Portugal. Not including the impact of these transfers may thus lead to an overestimation of inequality.

4. Methodology

Differences in methodology may also lead to different distributional outcomes. This may concern the use of different data sources, but also relate to different techniques to correct for any gaps between the micro data and the national accounts aggregates, and to allocate the amounts across households for those items for which micro data are missing.

The data sources as used by both projects will often be the same. The DINA project relies heavily on tax data, accompanied by information from surveys and 'rich lists', whereas in the EG DNA project it largely depends on the data availability in the various countries. Many countries use administrative data in combination with survey data for items for which either no tax information is available or for parts of the population that may be excluded from administrative data (e.g., because they are exempted from taxation). On the other hand, some countries still rely predominantly on survey data in deriving their distributional results. The use of different data sources may lead to differences in distributional results, also depending on the coverage of the micro data, their underlying concepts, and the quality of the data. In that regard, adjustments for conceptual or classification differences and corrections for estimation and measurement errors may also give rise to differences in the distributional results, even when the same micro data sources are used.

As not all national accounts' items will have an equivalent in the micro data and as micro data will often not perfectly match the macro aggregates, adjustments will be needed to impute for missing items and to align the micro data to the national accounts' totals, which may also lead to differences in results between the two projects. This will largely depend on the size of the imputations and of the gaps between the micro data and the national accounts aggregates in the two projects, and the techniques used to allocate the amounts to the relevant households or individuals. Information from the EG DNA project showed that the size of imputations and alignments may be significant. In the 2015 exercise, the sum of the two adjustments ranged from 25% of adjusted disposable income in the United States to almost 70% in Mexico. When bearing in mind that these amounts need to be allocated to underlying individuals or households, it is clear that the way in which this is done may significantly affect the distributional results. Ideally, additional information is available for the correct allocation of the amounts to the relevant individuals or households, but often (part of) the allocation needs to be done on the basis of assumptions, possibly giving rise to significant margins of error surrounding the results. The latter is one of the main reasons why the EG DNA project is currently only targeting publication of distributional results at the quintile level.

Micro-macro gaps are often allocated in proportion to the available micro data, implicitly assuming a proportional misreporting by all respondents. However, as gaps may be caused by different reasons and may relate to specific households, it is questionable whether this leads to the most accurate results. In that regard, it would be better to explore the main underlying reasons and to try to allocate them to the relevant households accordingly. This may significantly affect the distributional results, particularly in case of large micro-macro gaps (see Zwijnenburg (2016)). For these reasons, it is useful when DINA and EG DNA provide more insight in the size of the gaps and on the techniques that are used to allocate them to the relevant households. In addition to explaining possible differences between the results of the two projects, it will also provide users with more insight in possible margins of error surrounding the results, providing them with better guidance on how to use and interpret the results.

With regard to imputations, it will depend on the number of items for which microdata is lacking and on the robustness of the assumptions to allocate the amounts to the relevant individuals or households, how this will affect the distributional results. In that regard, both projects include several items for which no direct micro data will be available. For DINA it may be assumed that this is the case for undistributed income of domestic corporations, government surplus or deficit, the gap between pension contributions and benefits, and for some specific national accounts items such as investment income disbursements. On the basis of this assumption, the amount of imputations in DINA adds up to 30.5% of post-tax national income for the United States and to more than 55% for Denmark and the Netherlands (see Zwijnenburg (2017)). It is clear that the way in which these amounts are allocated to underlying households, may significantly affect the distributional results and may lead to differences between the DINA and the EG DNA results.

When looking at the allocation techniques as used in DINA for some of these imputed items, some issues can be raised. For example, for a proper allocation of undistributed profits to underlying individuals, ideally information is available on the individual equity holdings of individuals so that undistributed profits can be attributed to the relevant individual. However, as this information is often lacking, the DINA Guidelines apply an alternative approach in which undistributed profits are allocated on the basis of the distribution of wealth, assuming equal rates of return on wealth. This would imply more or less equal amounts of undistributed profits across corporations, which is very unlikely, as corporations usually show a large dispersion in profits and will also differ in dividend policies. Secondly, in the DINA project the distribution of wealth itself is to a large degree derived on the basis of the underlying capital income flows (i.e. the capitalization method). Any errors in the assumptions to derive wealth results on the basis of capital income flows, as well as any errors in the underlying micro data, will automatically lead to errors in the wealth distribution, and consequently also affect the allocation of undistributed profits.

Another issue concerns the allocation of the amounts for social transfers in kind and for public spending on collective goods and services to underlying households. The DINA Guidelines explain that there are generally two ways to allocate the amounts to individuals, i.e. in proportion to post-tax disposable income or via a lump-sum method in which the same average monetary value is attributed to each adult individual. Whereas the latter method is used for social transfers in kind related to health, the proportional method is used for all other in-kind transfers and collective expenditures. It is highly debatable whether this is the best approach. Apart from the discussion whether some of these items should be included at all (as discussed before), it could also be argued that a flat distribution for some of these items may work better than an allocation in proportion to post-tax disposable income. In that regard, the

EG DNA recommends an actual use approach for other social transfers in kind (in which the values are allocated to those household groups that directly benefit from the relevant goods and services) which will lead to a distribution that is relatively flat across income quintiles. As in most countries these transfers constitute an important part of household income, a different allocation may significantly alter inequality measures.

5.   *Conclusions*

This paper has explained that, whereas both the EG DNA and the DINA project aim to compile distributional results in line with national accounts aggregates, they may end up with significantly different results, due to differences in scope, the use of different concepts in measuring inequality, and possible deviations in methodology to arrive at the distributional results. By focusing on the main differences, this paper has tried to broaden the knowledge on both projects, and to initiate a discussion on the pros and cons of certain choices in compiling distributional results. The latter is important with regard to further improving the methodologies used in both projects and to properly explain the main strengths and weaknesses of the specific indicators.

REFERENCES

ALVAREDO, FACUNDO, ANTHONY B. ATKINSON, LUCAS CHANCEL, THOMAS PIKETTY, EMMANUEL SAEZ AND GABRIEL ZUCMAN. 2017. "Distributional National Accounts (DINA) Guidelines". Version 9 June 2017. WID.world Working Paper series No. 2016/1.

BELLAMY, VANESSA, GEORGES CONSALÈS, MARYSE FESSEAU, SYLVIE LE LAIDIER AND ÉMILIE RAYNAUD. 2009. "Une decomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003". Working Paper G 2009/11, INSEE.

EUROPEAN COMMISSION, IMF, OECD, UN, WORLD BANK. 2009. "System of National Accounts 2008".

FESSEAU, MARYSE AND MARIA L. MATTONETTI. 2013. "Distributional measures across household groups in a national accounts framework – results from an experimental cross-country exercise on household income, consumption and saving". OECD Statistics Working Paper series.

OECD. 2017. "Distributional information on Household income, consumption and saving - guidelines". Version 22 June 2017.

PIKETTY, THOMAS. 2014. "Capital in the twenty-first century", Cambridge, Massachusetts London, England.

STIGLITZ, JOSEPH E., AMARTYA SEN AND JEAN-PAUL FITOUSSI. 2009. "Report by the Commission on the Measurement of Economic Performance and Social Progress".

ZWIJNENBURG, JORRIT. 2016. "Further enhancing the work on household distributional data – techniques for bridging gaps between micro and macro results and nowcasting methodologies for compiling more timely results". Paper prepared for the 34th IARIW General Conference, August 212-27 in Dresden.

ZWIJNENBURG, JORRIT, SOPHIE BOURNOT AND FEDERICO GIOVANNELLI. 2017. "Expert Group on Disparities within a National Accounts framework – Results from a 2015 exercise". OECD Working Paper No. 76.

ZWIJNENBURG, JORRIT. 2017. "Unequal distributions? – A study on differences between the compilation of household distributional results according to DINA and EG DNA methodology". Paper prepared for the WID-world conference, Paris, 14-15 December 2017.