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

Jorrit Zwijnenburg (OECD)

Paper prepared for the 34th IARIW General Conference

Dresden, Germany, August 21-27, 2016

Session 2A: Integrating Micro and Macro Approaches to National Income Analysis

Time: Monday, August 22, 2016 [Afternoon]

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

Jorrit Zwijnenburg (OECD) [12]

Paper prepared for the 34th General IARIW Conference
Dresden, Germany, August 21 – 27, 2016

Session 2A: Integrating Micro and Macro Approaches to National Income Analysis
Monday, August 22, Afternoon

Discussant: Sigita Grundiza (Eurostat)

[1] The views expressed in this paper are those of the author, and should not be considered as representing the official views of the OECD or of its member countries.

[2] The author would like to thank Mr. Roberto Astolfi and Mr. Matthew De Queljoe for their contributions to the work on the nowcasting.

Abstract

Recent years have seen an increased interest in distributional data. The Expert Group on Disparities in National Accounts (EG DNA) has developed methodology to breakdown national accounts aggregates for the household sector using distributional information from micro sources. This paper focuses on two issues that pose challenges in the compilation of distributional results. The first issue deals with gaps between micro and macro aggregates that need to be bridged in the compilation process. The paper discusses possible reasons for these gaps and presents a framework for allocating these gaps to the relevant households. A second issue concerns the timeliness of the results. As the timeliness of the underlying micro data used in the compilation process is often poor, current distributional results suffer from a substantial time lag. The paper discusses possible nowcasting techniques to arrive at more timely distributional estimates.

Keywords: Households, Distributional results, National Accounts, Nowcasting

1. Introduction

1. In the light of the recommendations made by Stiglitz, Sen and Fitoussi (2009), the OECD and Eurostat launched a joint Expert Group in 2011 to develop methodology to compile distributional measures of income, consumption and savings across household groups within the framework of the national accounts. The expert group worked out a preliminary methodology on the basis of which first experimental results were compiled in 2013. The work has been followed up by an OECD expert group on Disparities in National Accounts (EG DNA) to further improve the methodology and to resolve some of the most relevant issues. On the basis of enhanced methodology a second exercise was conducted in 2015.

2. The methodology used by the expert group combines national accounts data with micro data sources on income, consumption and savings. As a consequence, the quality of the distributional results is highly dependent on the quality and the alignment of these micro and macro results. Especially in case of large gaps between the micro and macro results, the way in which these gaps are bridged may significantly affect the distributional results. Therefore, more guidance is needed on how countries should deal with these gaps. The first part of this paper discusses possible reasons for gaps between micro and macro aggregates and explains how the allocation to the relevant households may vary depending on the underlying reasons. A framework is presented to analyse these gaps and to allocate them to the relevant households, and results are shown from an exercise in which EG DNA members applied this framework for the items that showed the largest gaps in their countries.

3. The second part of the paper focuses on techniques to compile more timely distributional estimates. As the methodology to compile distributional results within the national accounts framework relies heavily on micro data sources that often only become available with a certain delay, distributional results currently suffer from a substantial time lag as well, usually coming available only two to three years after the reference period. From the users' side, there is considerable demand for more timely data. Nowcasting techniques may be used to arrive at more timely results. These methods combine forecasting techniques with newly available data that strongly relate to the target variable. This part of the paper presents different techniques that have been identified within the expert group to arrive at more timely distributional results and presents some experimental results for some of these techniques.

4. The paper is structured as follows. First, section 2 describes the current methodology that is used within the EG DNA to compile distributional results within a national accounts framework. It shows the role of micro data in the compilation process and how the methodology deals with gaps between micro and macro results. Section 3 then provides information on how these gaps may affect the distributional results and presents a framework on how countries may deal with these gaps. Section 4 discusses the various nowcasting techniques that have been identified to arrive at more timely distributional results. On the basis of available distributional data, experimental results are presented for some of these techniques. The paper concludes with some main lessons learned and ideas for future work in section 5.

2. Methodology of the EG DNA

5. The methodology for the compilation of distributional results within the framework of the national accounts uses a step-by-step approach combining data from national accounts with micro data sources (see Figure 1).

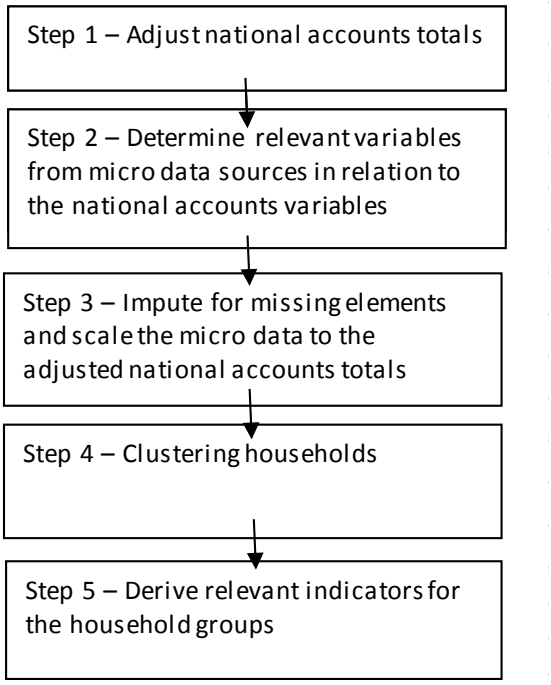
6. As the household sector in the national accounts may be broader than domestic private households as covered by micro data sources, in a first step, adjustments are made to the national accounts totals to exclude parts that do not relate to domestic private households. This may relate to non-profit institutions serving households (NPISHs) that may be published together with data for the household sector, expenditures of non-resident households on the national territory that may be included in the consumption column in the supply-and-use tables, and the income and consumption of people living in non-private dwellings (such as prisons, retirement homes and boarding schools).

7. After this step, micro variables need to be identified that can be linked to the national accounts items in step 2. Different data sources may be used, depending on which micro data sources provide the best link to the national accounts items. Furthermore, as micro statistics and national accounts may be based on different definitions, reclassification of sub-items may be needed to correctly match the micro and macro items. For example, wages and salaries paid while an employee is on sick leave are often included in compensation of employees in the micro statistics, whereas they are regarded as social benefits in the national accounts.

8. In the third step, the underlying micro data are scaled to the adjusted national accounts totals. Usually the two will not be equal as concepts may differ and as they are the result of different compilation processes. The micro aggregates are the results of a process of editing, imputing and weighting micro data, whereas the national accounts totals are the result of a process of confronting and balancing different data sets at an aggregated level within the system of national accounts. As a consequence, the micro and macro totals will often differ in levels, even when the underlying concepts are the same. In those cases, the micro data need to be aligned to equal the national accounts results. Furthermore, in this third step imputations need to be made for items for which micro data is lacking. This may relate to items that are specific to the system of national accounts (such as imputed items like investment income disbursements or FISIM), but also to certain elements that may be missing from the micro data, such as the underground economy or illegal activities.

9. After the imputation and alignment of distributional results per item, households can be clustered into quintiles on the basis of their household disposable income in the fourth step and results can be derived for the various main aggregates and distributional measures in the fifth. More detailed information on the methodology can be found in Fesseau and Mattonetti (2013) and in the forthcoming paper by Zwijnenburg, Bournot and Giovannelli (2016).

Figure 1. A step-by-step approach for the estimation of distributional information



10. The step-by-step procedure shows that micro data play an important role in the compilation process. They constitute the basis for determining the distributional patterns in the data, whereas the national accounts data provide the totals to which these micro data need to be aligned. Ideally, all national accounts items can be linked to micro data and all items show perfect alignment between the micro and the national accounts aggregates. However, as some items are specific to the system of national accounts, imputations will be needed. Furthermore, micro and macro data usually differ in size and alignments are needed to close any gaps between the two. This implies that the third step in the procedure is a very relevant one, which may have a substantial impact on the distributional results. This will of course mainly depend on the number of items for which imputations are needed and on the size of the gaps between the micro and the macro aggregates. Section 3 focuses on the second issue and presents a framework how countries may deal with these gaps.

11. Furthermore, ideally, the micro data become available within a relative short period of time and are available for each year. In that case, recent national accounts data can be combined with recent micro data to compile timely distributional results on an annual basis. However, most countries predominantly rely on household surveys that only become available with a time lag of two to three years (due to the substantial amount of time it takes to conduct and process the results) and that may not be available for each year. Administrative data usually have better timeliness and coverage, but will only be applicable for a restricted number of variables. That is why there is a need to explore techniques to compile more timely estimates and results for time periods for which no micro data is available. Section 4 presents various techniques that can be applied in this regard.

3. Dealing with gaps between micro and macro data

12. This section focuses on the specific issue of micro-macro gaps and presents a framework that countries may use to bridge these gaps. First, on the basis of results from an exercise by the EG DNA conducted in 2015, the relative importance of these gaps in the compilation of distributional results is explained, and the most relevant items in this regard are shown. Subsequently, possible reasons for the gaps and ways to allocate them to relevant households are discussed. These are at the basis of the framework. After the presentation of the framework, some results are presented from an exercise in which some members of the EG DNA used this framework to allocate parts of the gaps to underlying quintiles. The section ends with the main conclusions and ideas for further work.

3.1 The role of micro-macro gaps in compiling distributional results

13. As described in section 2, the distributional outcomes are the result of the underlying micro data, the alignment of these micro data to the macro aggregates and the imputations for items for which micro data is lacking. In almost all countries the distributional results for adjusted disposable income and actual consumption expenditure are to the largest extent based on micro data. Looking at adjusted disposable income, micro data sources cover more than 70 percent of the underlying flows on average, whereas this is more than 60 percent for actual consumption expenditure[3]. However, the impact of imputation and alignments on the distributional results may still be substantial.

14. The impact of the imputations and alignments on the distributional results can best be reviewed by presenting the size of these adjustments in absolute terms as percentage of the balancing items. This provides insight in the maximum amount that has to be allocated to the various household. Whereas positive and negative adjustments may (partly) cancel out at the level of the household sector as a whole, their overall impact on distributional results may still be significant, especially when they are allocated to quintiles at the opposite end of the spectrum.

15. Figures 2 and 3 present the impact of imputations and alignments on adjusted disposable income and actual consumption expenditure respectively, for the household sector as a whole. It shows that for most of the countries the impact of alignments is larger than the impact of imputations. This is related to the fact that most countries have micro data available for the majority of items in the exercise and only need to rely on imputations for few of them. In that regard, it should also be mentioned that the number of items for which countries report imputations and alignments varies across countries. This has to be borne in mind when analysing the results.

[3] The coverage rate has been calculated by looking at the relative shares of micro data, imputations and alignments in the absolute flows that constitute adjusted disposable income and actual consumption expenditure. As adjusted disposable income consists of positive and negative items, the absolute flows have been considered to get a correct view of the contributions of each of the three factors to this aggregate. The relative shares for the imputations and alignments have been calculated by dividing their absolute values by the absolute flows constituting adjusted disposable income and actual final consumption. In deriving the share of the micro data, it has to be borne in mind that that micro totals can exceed the macro aggregates, so simply looking at the sum of the micro totals as percentage of the absolute flows would not provide a correct picture. Therefore, the share of the micro total in the balancing items is derived as a residual, i.e. after deduction of the shares of the imputations and the alignments.

Figure 2. Size of the absolute alignments and imputations as percentage of adjusted disposable income for the household sector as a whole

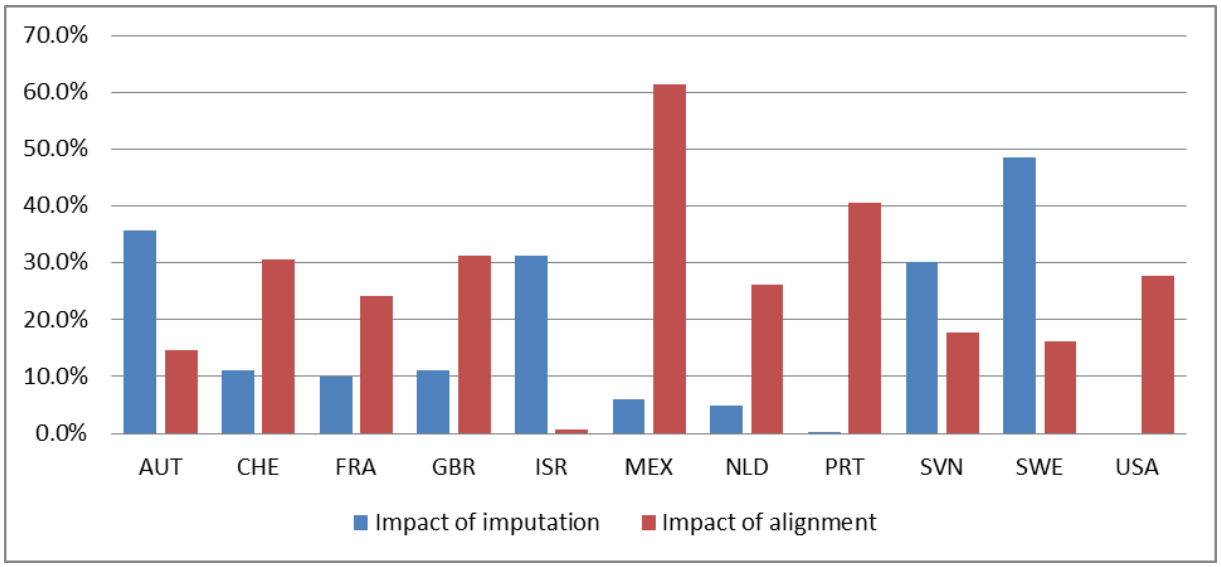
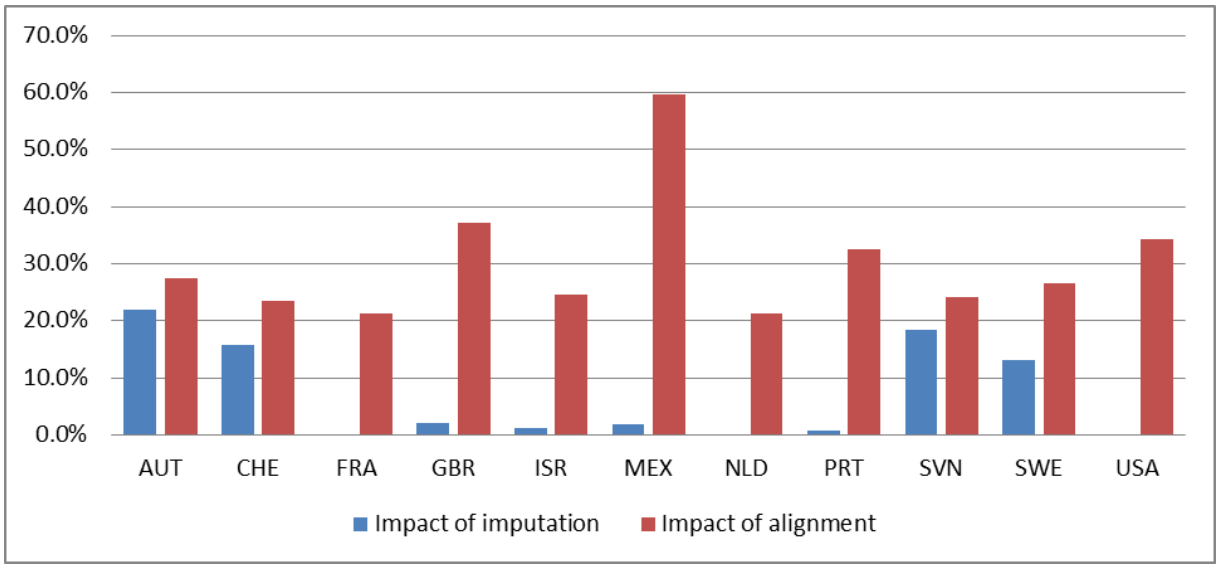


Figure 3. Size of the absolute alignments and imputations as percentage of actual consumption expenditure for the household sector as a whole



16. The impact of alignments turns out to be particularly large in Mexico. The sum of the positive and negative alignments amounts to more than 60 percent of adjusted disposable income for the household sector as a whole and almost 60 percent for actual consumption expenditure. Furthermore, the impact of alignments is relatively large in Portugal, the United Kingdom and the United States, as well as in Switzerland for actual consumption expenditure. When looking at the average for all countries, the impact of alignments is 26.5 percent for adjusted disposable income and 30.2 percent for actual consumption expenditure.

17. Looking at the impact of the alignments on the distributional results, one has to bear in mind that this amount needs to be allocated to the underlying households. This means that for Mexico more than 60.0 percent of adjusted disposable income has to be allocated to the relevant households. It is self-evident that the way in which this is done, may seriously affect the distributional results. If all positive alignments would be allocated to the households with the

highest incomes and the negative ones to the households with the lowest incomes, the inequality will be much larger than when the amounts would be allocated proportionally to all households. Of course, such an extreme allocation would only be possible from a theoretical point of view[4], but it shows how the allocation of these gaps may significantly alter the distributional results.

3.2 Overview of the gaps between the micro and macro aggregates

18. The previous subsection showed that the alignment of gaps between micro and macro data has a large impact on the distributional results in most of the countries. This implies that for some items large gaps exist between the micro aggregates and the national accounts totals. Table 1 shows the adjustment coefficients for the main income components on the basis of the most recent year available from the exercise conducted in 2015. The adjustment coefficient shows by how much the micro results need to be adjusted to align them with the adjusted national accounts total. It is calculated as the adjusted national accounts aggregate divided by the micro aggregate. The table shows the number of countries for which an adjustment coefficient could be calculated (i.e. micro data was available to compile the distributional results), the average value of the coefficient, the median value and the minimum and maximum values in the exercise[5].

Table 1. Adjustment coefficient for the main income components

Code	Instrument	Number of countries	Average		Median		Minimum		Maximum	
			most recent year	second most recent year	most recent year	second most recent year	most recent year	second most recent year	most recent year	second most recent year
B2	Operating surplus	6	1.47	*1.27*	1.44	*1.27*	0.47	*1.12*	2.43	*1.42*
B3	Mixed income	9	2.69	*1.79*	2.43	*1.79*	1.30	*1.67*	5.24	*1.91*
D1R	Compensation of employees	9	1.15	…	1.16	…	1.01	*0.00*	1.38	*0.00*
D41R	Interest (not adjusted for FISIM), received	8	2.08	*1.90*	1.56	*1.05*	0.66	*0.72*	6.40	*4.77*
D42R	Distributed income of corporations	7	5.06	*10.67*	1.88	*5.53*	0.70	*3.00*	17.76	*23.50*
D41P	Interest (not adjusted for FISIM), paid	9	3.58	*2.47*	2.94	*1.50*	1.02	*1.01*	11.31	*4.65*
D5P	Current taxes on income and wealth	10	1.18	*1.19*	1.18	*1.15*	0.78	*0.74*	1.54	*1.78*
D61P	Net social contributions	2	1.23	*2.01*	1.23	*2.01*	1.19	*1.28*	1.27	*2.73*
D62R	Social benefits other than STiK	10	1.22	*1.30*	1.15	*1.26*	0.97	*0.98*	1.55	*1.65*
D63R1	Social Transfers in Kind - Education	3	0.94	*0.88*	0.95	*0.88*	0.72	*0.78*	1.13	*0.98*
D63R2	Social Transfers in Kind - Health	3	1.36	*1.37*	1.18	*1.37*	1.16	*0.99*	1.73	*1.75*

19. Looking at the results, it can be observed that the gaps between the micro and macro data are often quite substantial. An adjustment coefficient that is close to 1 implies good alignment, but the table shows that the average values substantially differ from 1 for most of the income components. For the majority of the items the coefficient is above 1, meaning that the micro aggregates are lower than the macro aggregates. Only in a few cases it is the other way around. *Distributed income of corporations* turns out to have the highest adjustment coefficient on average (5.06), followed by *Interest paid* (3.58), *Mixed income* (2.69), and *Interest received* (2.08). *Distributed income of corporations* also records the highest maximum values. The maximum adjustment recorded for the most recent year even amounted to 17.76, implying that the micro results had to

[4] Due to several linkages between various items in the system, some alignments and allocations will have to be done in conjunction. Furthermore, it may be assumed that more general causes will underlie the gaps, leading to a more balanced allocation of the gaps.

[5] The table has been taken from a paper that has been presented at the meeting of the EG DNA in May 2015.

be multiplied by 17.76 to arrive at the adjusted national accounts total. It is obvious that such an adjustment can have a large impact on overall distributional results.

Table 2. Adjustment coefficient for the main expenditure components

Code	Instrument	Number of countries	Average		Median		Minimum		Maximum	
			most recent year	second most recent year	most recent year	second most recent year	most recent year	second most recent year	most recent year	second most recent year
CP010	Food and non-alcoholic beverages	10	1.48	*1.53*	1.34	*1.40*	1.06	*0.95*	2.87	*2.76*
CP020	Alcoholic beverages, tobacco and narcotics	9	3.60	*5.37*	2.51	*2.52*	1.68	*1.13*	12.00	*21.03*
CP030	Clothing and footwear	10	1.57	*1.70*	1.25	*1.40*	1.09	*1.03*	2.90	*2.80*
CP040	Housing, water, electricity, gas and other fuels	9	1.23	*1.16*	1.06	*0.94*	0.84	*0.87*	2.47	*2.30*
CP050	Furnishings, households equipment & house maint.	10	1.60	*1.71*	1.41	*1.59*	1.15	*0.96*	2.93	*2.93*
CP060	Health	9	2.47	*2.72*	2.16	*2.27*	1.22	*1.15*	4.78	*4.74*
CP070	Transport	8	1.56	*1.59*	1.34	*1.36*	0.98	*0.95*	3.18	*2.87*
CP080	Communications	10	1.25	*1.53*	1.26	*1.34*	0.71	*1.08*	2.28	*2.50*
CP090	Recreation and culture	10	1.90	*1.85*	1.45	*1.65*	1.14	*1.01*	4.05	*3.50*
CP100	Education	10	1.09	*0.92*	1.08	*1.05*	0.19	*0.09*	1.87	*1.51*
CP110	Restaurants and hotels	10	1.54	*1.32*	1.52	*1.29*	0.97	*1.06*	2.20	*1.64*
CP120	Miscellaneous goods and services	7	1.89	*1.88*	2.06	*1.78*	0.97	*1.13*	2.63	*2.85*

20. Looking at the consumption components (Table 2), smaller differences across components are observed compared to income. The average coefficients are between 1.09 and 2.47 in the most recent year, and between 0.92 and 2.72 in the second most recent year, when excluding the item *Alcoholic beverages, tobacco and narcotics* (CP020). The coefficient for the latter item has an average value of 3.60 and 5.37, respectively.

21. In conclusion, the gaps between micro and macro data are quite substantial for a significant number of items. As the way in which these gaps are allocated to the underlying households may have a large impact on the main aggregates and the distributional measures, as was shown in subsection 3.1, there is a need to gain more insight in the reasons for these gaps and the best way in which these gaps may be allocated to the underlying households.

3.3 Possible reasons for the gaps

22. Possible reasons for the differences between the micro results and the adjusted national accounts totals are related to the first three steps of the step-by-step approach presented in section 2, focusing on the quality of the data and of the assumptions used in the exercise:

Step 1: Adjustment of the national accounts totals:

- *i. The quality of the national accounts totals*
- *ii. The quality of the adjustments to the national account totals*

Step 2: Linking micro data source variables to the national accounts variables:

- *iii. Assumptions regarding the conceptual and classification differences*

Step 3: Imputation for missing elements and aligning data to national accounts totals

- *iv. The quality of the correction for the underground economy and illegal activities*
- *v. The quality of the micro data – Estimation errors*
- *vi. The quality of the micro data – Measurement errors*

23. The reasons for the gaps are discussed below, in accordance with the above categorisation.

3.3.1 The quality of the national accounts total

24. A first possible reason for the gap between the micro and macro results may be quality issues related to the national account totals. The national accounts totals are the product of a balancing framework in which data from various data sources are combined and confronted. Often source data need to be adjusted to arrive at consistency and comprehensiveness. In that process choices have to be made that may cause differences from the direct data sources. The quality of the data that are used in the system and the strength of the assumptions made in the balancing process will determine the quality of the final results. Gaps between micro and macro data may point to possible quality issues in this process.

3.3.2 The adjustments to the NA total

25. In the compilation process to arrive at distributional results, national accounts totals may have to be adjusted to exclude NPISHs, people living in non-private dwellings, and consumption expenditure by non-residents. In some cases, specific information will be available to make these adjustments, but in other cases these adjustments will have to be based on assumptions. Gaps between micro and macro results may be due to quality issues in making these adjustments.

3.3.3 Conceptual differences and classification issues

26. Gaps may also appear as a consequence of conceptual differences and classification issues between micro and macro data. Sometimes the definition of the national accounts may vary from the one used in the survey or administrative data source, and (part of the) transactions may be classified differently. Also, the time of recording may differ between the national accounts totals and the micro results. The latter often focus on a certain point in time (e.g. end of the quarter or end of the year), whereas the national accounts aim to capture all transactions within a certain time frame. This may also give rise to differences between the micro and macro results, for instance related to changes in the population or for specific economic events that may have occurred during the period.

3.3.4 The correction for the underground economy, illegal activities and other missing elements

27. The gaps may also be related to the imputation for the underground economy, illegal activities and other elements that may be missing from the micro data. The underground economy relates to activities that in principle are legal, but are deliberately concealed from public authorities. Illegal activities are forbidden by law or become illegal when carried out by unauthorised persons. Some other elements may be missing from the micro data as well. This may concern specific items (such as imputed items), but may also concern specific parts of items. Administrative data may for instance exclude some groups of households that fall below certain thresholds or that are exempted from taxation.

28. Estimations for the underground economy, illegal activities and other missing elements are usually made within the national accounts division. These estimates are based on their knowledge of the coverage of the micro data sources and on information that may be available on specific missing elements. Usually part of these estimates are based on assumptions. Therefore, in checking possible reasons for the gaps between micro and macro data, it is important to analyse the reliability of these specific estimates.

3.3.5 The quality of the micro data – Estimation errors (sampling and coverage)

29. Just as the macro aggregates may turn out to be incorrect, micro estimates may also be subject to quality issues. Estimation errors are a first type of errors that may affect the results. These relate to the extrapolation of the survey results to the target population and can be linked to the sample size, the representativeness of the sample and the magnitude of the non-response. The errors related to the sample size are referred to as the standard sampling error, implying that the smaller the survey sample, the larger the sampling error, as less data underlie the ultimate estimates. The other two issues are referred to as coverage errors. These occur in the case of the sampling frame being different from the target population and in the case of selective non-response. All these aspects may lead to higher variability of the weighted estimates and therefore to possible gaps with the macro results. Especially survey data may suffer from estimation errors. Administrative data sources tend to have broad coverage and are therefore less prone to these kinds of errors. However, it has to be realised that administrative data may also suffer from non-random under-coverage, especially in the case of reporting thresholds.[6]

3.3.6 The quality of the micro data – Measurement errors

30. Finally, errors may also occur due to mistakes in the data reported in the surveys. These are referred to as measurement errors, and may relate to item non-response or the reporting of incorrect data. Meyer, Mok and Sullivan (2008) show that a lot of statistics have to deal with measurement error and that these kinds of errors have increased over time, at least for some specific items. Especially questions on income are usually understood to be relatively sensitive and prone to higher non-response rates or larger measurement errors. The confrontation of survey results with administrative data, as well as the confrontation of income and consumption data on a micro-economic level may provide more insight in possible measurement errors.

3.4 Allocation of the gaps to the relevant households

31. After the gaps have been attributed to probable causes, the related estimates have to be allocated to the relevant households or household groups. As the allocation may differ per cause, this more detailed approach will lead to more accurate results than allocating the full gap in one turn. For all causes that concern micro data underlying the distributional results, specific solutions have to be found. Looking at the reasons that were presented in the previous section, this relates to categories iii to vi. Categories i and ii only concern the adjusted national accounts totals with which the micro data have to be aligned, so they only affect the distributional results in an indirect way.

32. It obviously depends on the cause what kind of information is available to properly allocate the relevant amounts. Sometimes the micro data itself may already show what kind of information is missing, for instance in the case that the match of income and consumption data shows implausible results for certain households[7]. In other cases compilers may rely on results from recent research into specific causes, for instance via the confrontation of survey data with administrative data (Törmälehto, 2014), analysis into which household types are more likely to be involved in the underground economy or illegal activities (Carson, 1984), or research to check consistency of micro

[6] More information on how the quality of survey results may be checked can be found in Vermeulen (2014), Coli and Tartamella (2014) and Törmälehto (2014).
[7] See Accardo et al. (2009) and Sabelhaus et al. (2012).

results within the same survey (D'Alessio and Faiella (2002) and D'Alessio and Neri (2015)). Of course, it depends on the items, the data sources and the country characteristics to what extent these results from literature will apply to specific country cases, but it may provide insights in how to approach the allocation question and may help in coming up with the most appropriate allocation for the specific reasons.

33. The allocation of the amounts to the underlying households should ideally be done on the level of the micro statistics, i.e. by making adjustments to the survey data, applying imputations on the micro level, or by adjusting the survey weights to arrive at the relevant aggregates. This will lead to improved micro data that underlie the new distributional measures and will make sure that the quintile classification is re-adjusted on the basis of these improved data. However, an alternative is to allocate the amounts on a meso-economic level. In that case, the quintile allocation on the basis of the 'unadjusted' micro data is taken as starting point and the amounts that have been attributed to the various causes are allocated to the quintiles. It is clear that the distributional results on the basis of this meso approach will not be as accurate as in the case of processing the corrections on the micro-economic level, but in the end may lead to better results than simply applying a proportional allocation.

3.5 A framework to allocate the micro-macro gaps

34. To assist compilers in discussing possible reasons for the gaps and to allocate them to the quintiles, a framework has been developed on the basis of the reasons expressed in the previous subsection. This framework consists of two parts. The first part focuses on assigning parts of the gap to possible underlying causes. This part is presented in Table 3. The first block (block I) in the table focuses on the derivation of the adjusted national accounts estimate, starting from the national accounts total and adjusting for NPISH, non-private households and expenditures of non-resident households on the territory respectively. The first column in this block shows the original estimates that were used to derive the adjusted national accounts figure. The second column provides the possibility to correct any of these original figures to close part of the gap between the micro and the macro results. The final result is presented in the third column.

35. The second block of the framework (block II) confronts the adjusted national accounts result with the micro aggregate, showing the gap between the two. The initial macro-micro gap is presented in the first column. The third column shows the gap that still remains after corrections have been made to the adjusted national accounts aggregate. This remaining gap still needs to be attributed to other reasons. This is done in block III. This block lists possible causes related to conceptual or classification differences, missing items or errors with regard to the micro data. In addition to the reasons presented in the previous section, it also contains an item for reasons that are not covered by the other categories. The block ends with the gap that still remains after attributing parts of the gap to the underlying reasons. Ideally, the amount of this remaining gap is zero, which would imply that the complete gap is explained by the various causes.

Table 3. Framework for attributing micro-macro gaps to underlying causes

	Item xx.	Original estimate	Correction	Ultimate Estimate
I	National account total (A)	…	…	…
	- Adjustment for NPISH (B1)	…	…	…
	- Adjustment for non-private households (B2)	…	…	…
	- Adjustment for expenditures of non-resident households on the territory (B3)	…	…	…
	= Adjusted NA total (C=A-B1-B2-B3)	…	…	…
II	Micro total (D)	…		…
	= Macro-Micro gap (E=C-D)	…	…	…
III	Conceptual or classification issues (F)	…	…	…
	Underground and illegal activities (G)	…	…	…
	Other elements missing in micro data (H)	…	…	…
	Estimation errors (under-/overcoverage) (I)			…
	Measurement errors (under-/overreporting) (J)			…
	Reasons n.e.c. (K)			…
	= Remaining gap (L=E-F-G-H-I-J-K)			…

36. After the attribution of the macro-micro gaps to the underlying causes, the related amounts should be allocated to the relevant household groups. Table 4 presents a framework for this step. Block IV focuses on the allocation on the basis of revised micro data, which, as was explained in the previous section, is the preferred option. In that case, corrections are processed on the micro level and new results are derived following the standard step-by-step approach. However, in some cases this may be deemed too time-consuming or too complex. In those cases, corrections may be allocated on an aggregated level. This can be done in block V which provides the opportunity to allocate the remaining gaps at the quintile level. Finally, block VI deals with allocating the remaining gap that could not be linked to any of the possible causes. The sum of the corrected micro data and the consecutive meso-corrections leads to the distributional results for the quintiles.

Table 4. Framework for allocating of the gaps to the quintiles

	Item xx.	Estimate	Q1	Q2	Q3	Q4	Q5
IV	*Allocation on the basis of micro data*						
	Original micro aggregate (P)	…	…	…	…	…	…
	Revised micro aggregate (Q)	…	…	…	…	…	…
V	*Allocation on the basis of meso corrections*						
	Conceptual or classification issues (R)	…	…	…	…	…	…
	Underground and illegal activities (S)	…	…	…	…	…	…
	Other elements missing in micro data (T)	…	…	…	…	…	…
	Estimation errors (under-/overcoverage) (U)	…	…	…	…	…	…
	Measurement errors (under-/overreporting) (V)	…	…	…	…	…	…
	Reasons n.e.c. (W)	…	…	…	…	…	…
VI	Alignment of remaining gap (X=C-Q-R-S-T-U-V-W)	…	…	…	…	…	…
	Final estimate (Y=Q+R+S+T+U+V+W+X)	…	…	…	…	…	…

3.6 Results from a recent request

37. In November 2015 a request was sent out to EG DNA members to apply the framework for the five items that appear to be most relevant for their country. The relevance was based on the relative size of the gaps and its relative weight with regard to the main balancing items. In the end,

eight countries responded to this request, of which five actually allocated parts of the gap to underlying reasons[8]. The other three countries expressed that it turned out to be too difficult to derive the exact reasons for gaps between the micro and macro aggregates.

38. Table 5 shows the number of countries that reported on the specific items and which underlying causes they selected per item. It shows that most countries were not able to allocate the full gap to underlying reasons. For 13 out of the 22 items a remaining gap still needed to be allocated. Looking at the allocation to underlying reasons, conceptual and classification issues are selected most frequently to explain part of the gap, followed by measurement errors and revisions of national accounts totals. Almost all of the items identified as possible reasons for the gaps between micro and macro aggregates have been selected by the countries, except for errors in the adjustment for non-private households and for estimation errors. The latter is particularly remarkable as it is usually regarded as one of the most likely reasons for gaps between micro and macro aggregates, especially in the case when micro aggregates are based on survey data.

Table 5. Reported reasons for micro-macro gaps as part of a questionnaire among EG DNA countries

Codes	Selected items in the exercise	Number of countries reporting on item	Revision of NA total	Revision of adj. for NPISH	Revision of adj. for non-private hh's	Revision of adj. for expenditure of non-residents	Revision of micro total	Conceptual and classification issues	Underground and illegal activities	Other elements missing in micro data	Estimation errors (under-/overcoverage)	Measurement errors (under-/overreporting)	Reasons n.e.c.	Alignment of the remaining gap
B2A	Operating surplus from owner occupied dwellings	1								1				1
B3	Mixed income	1						2	1					
B3C	Mixed income excl. undergr. production and own account production	1												1
D1R	Compensation of employees	1	1											
D11R	Wages and salaries	1						1						
D41R	Interest received (not adjusted for FISIM)	1												1
D41P	Interest paid (not adjusted for FISIM)	1								1				1
D42R	Distributed income of corporations received	1		1										1
D75P	Miscellaneous current transfers paid	1												1
CP010	Food and non-alcoholic beverages	2	1			1		1				2		
CP020	Alcoholic beverages, tobacco and narcotics	2	1			1		1				2		
CP060	Health	1												1
CP070	Transport	2	1			1	1			1		1	1	2
CP090	Recreation and culture	1					1	1						1
CP110	Restaurants and hotels	2					1	1				1		1
CP120	Miscellaneous goods and services	1	1			1	1			1		1	1	1
P31DC	Final domestic consumption expenditure	1						1	1					1
TOTAL		21	5	1	0	4	4	8	2	4	0	7	2	13

39. To have more insight in the contributions of the possible reasons to explain the micro-macro gaps, Table 6 shows an overview of how the gaps have been allocated to the underlying reasons. The table shows the shares for each of the possible reasons on the basis of a simple average of all 21 items. The table also shows the maximum share that has been allocated to each of the reasons.

[8] Four countries reported information on five items and one country reported data on one item.

Table 6. Allocation of the gap to the various reasons, as percentage of the total gap

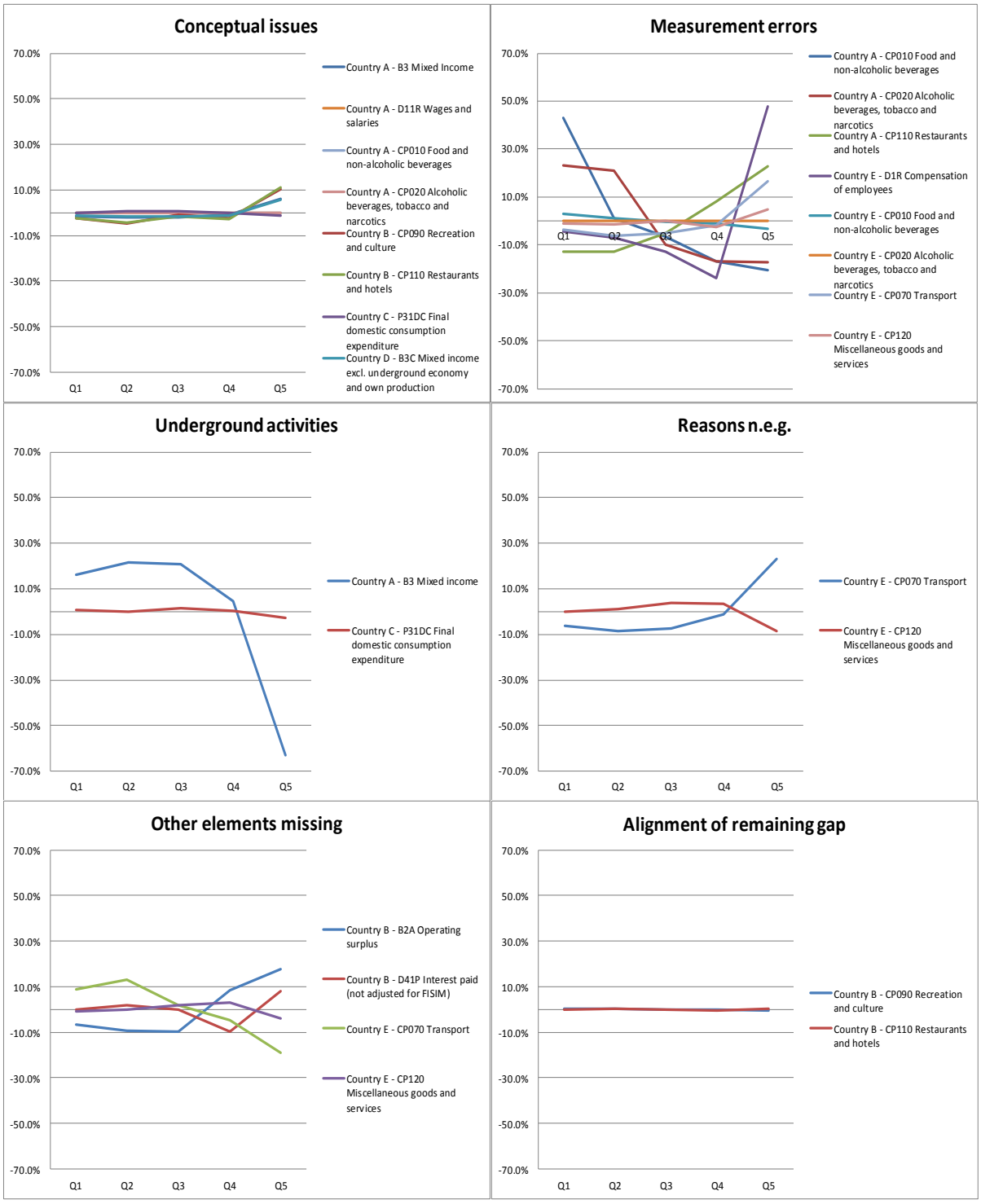
Reasons for gaps between micro and macro aggregates	Average	Maximum
As percentage of original gap		
Correction of NA total	-1.1%	0.0%
Correction of adjustment for NPISH	0.2%	3.9%
Correction of adjustment for non-private hh's	0.0%	0.0%
Correction of adjustment for expenditure of non-residents	1.0%	9.1%
Correction of micro total	4.6%	82.5%
As percentage of gap after revision of adjusted NA totals		
Conceptual and classification issues	12.5%	100.0%
Underground and illegal activities	3.4%	55.9%
Other elements missing in micro data	11.2%	69.7%
Estimation errors (under-/overcoverage)	0.0%	0.0%
Measurement errors (under-/overreporting)	25.0%	100.0%
Reasons n.e.c.	1.9%	23.7%
Alignment of the remaining gap	46.1%	153.6%

40. The table shows that only a small part of the gap is attributed to changes in the adjusted national accounts aggregates. Only one country corrected its national accounts data, which was probably not related to correct for part of the gap between the micro and macro aggregate (especially as it led to a further widening of the gap), but more likely concerns a regular revision of the national accounts data. Furthermore, one country corrected its adjustment for NPISH for one item and another country corrected its adjustment for expenditure of non-residents in the domestic economy for four items. However, the related amounts only explain small part of the gap for these items. Also, the revisions of the micro totals have been small on average. Two countries both corrected the micro aggregates for two items. For three of the items the revision was relatively small. For one item it was substantial; however it was reported that this was due to an error in the earlier transmission of the distributional results.

41. The lower part of Table 6 shows how the remaining gap between the adjusted national accounts totals and the micro aggregates has been attributed to the possible underlying reasons. It shows that on average 46.1 percent of the gaps was not attributed to one of the reasons. However, it can also be observed that in some cases 100 percent of the gaps was attributed to an underlying reason. This concerned *wages and salaries* (D11R) for which one country could link the full gap to conceptual and classification errors and to *compensation of employees* (D1R) for which one country ascribed the full gap to measurement errors.

42. To analyse whether different reasons for the gaps may also lead to different allocations to underlying households, countries were also asked to report the distributions of the gap over the quintiles per underlying reason. Figure 4 presents for each of the underlying reasons the difference between the allocation for that specific part of the gap and the distribution according to the micro data source. If the differences would be zero, the allocation for that specific reason is equal to the distribution according to the micro data source and would not lead to different distributional results. If the differences are positive, then the specific correction allocates a higher share to the relevant quintile than the micro data, if the differences are negative then the allocation leads to a lower share. In those cases, the overall distribution will change. The impact would depend on the size of the gap that can be explained by the specific reason and the differences from the original distribution.

Figure 4. Differences (in percentage-points) between relative distribution according to the various reasons for micro-macro gaps and according to the related micro data



43. The graphs show that the allocation across quintiles indeed differs across the various reasons and that in most cases they differ from the distributions according to the micro data. On the basis of the responses to the questionnaire the differences turn out to be particularly large for 'measurement errors' and 'underground activities'. The latter may be related to the non-inclusion of underground economy in earlier results and shows the importance of a separate estimation of these transactions. Furthermore, estimation errors may significantly alter the distribution across

households for specific items. The related figures shows that the distribution of the measurement error for *food and non-alcoholic beverages* and *alcoholic beverages, tobacco and narcotics* assigns more to the lower income groups and less to the higher income groups. For *transport*, *restaurants and hotels* and *miscellaneous goods and services* it seems to be the other way around. One country reported measurement errors for *compensation of employees*. The correction for this error would lead to a relatively higher income for the higher income groups in comparison to the lower income groups. In interpreting these results, it has to be borne in mind that the number of countries in the exercise is limited, so that no general patterns should be derived from them.

3.7 Conclusions and further work

44. This section showed that the alignment of the micro data to the national accounts aggregates may have a large impact on the results. That is why it is important to look at the most likely reasons for these gaps and to allocate the gaps to the underlying households accordingly. The framework that was presented in subsection 3.5 may be helpful in that regard. Results from an exercise in which some EG DNA countries applied this framework showed that the allocation to the relevant households usually differs across the various reasons and that these allocations would lead to different results than simply applying a proportional allocation of the gaps. As the number of countries involved in the exercise was small, no general guidance can be derived on how to allocate part of the gaps in case of specific underlying reasons for the gap. This will probably also largely depend on the type of data source (survey or administrative data) and its specific characteristics in terms of (population and item) coverage, imputation techniques and editing processes, as well as on country-specific particularities (e.g. with regard to the size and the concentration of the underground economy).

45. Currently, little information is available on the reasons for gaps between the micro and macro aggregates; therefore, it is recommended to conduct further research in this area. Analyses in which results from various data sources (such as administrative data sources with survey results) are confronted may be very helpful in this regard. Also, the linking of micro data on income, consumption and wealth (if available) at the micro level may be useful, as it may reveal inconsistencies at the most detailed level that may be at the basis of some of the micro-macro gaps. Furthermore, regular discussion of gaps between micro and macro aggregates between specialists from both sides is encouraged as it will add to the awareness of these gaps and as the exchange of expertise may provide useful insights in how to deal with them. This would not only be relevant for projects in which micro and macro results are combined, but would also be beneficial to properly explain to users why the results of micro and macro statistics on similar subjects may deviate.

4. Compiling early estimates

46. The methodology for the compilation of distributional results relies heavily on micro data sources that often only come available with a certain time lag. For more recent years, national accounts totals will usually be available, but recent information on the distribution among household groups may often be lacking. For some items micro data may be available from administrative data sources or specific timely surveys, but this will usually not be enough to compile

a complete set of distributional results for disposable income or final consumption expenditure. That is why extrapolation techniques are explored to arrive at more timely distributional results.

47. Nowcasting techniques use newly available data that strongly correlate with the target variables to arrive at early estimates for the present or for the recent past. When looking at the compilation of early distributional estimates, these can be based on the distributional results of previous years combined with newly available data from national accounts. Furthermore, auxiliary variables may be used if they strongly correlate with distributional results.

48. This section describes the results of the exploration of nowcasting techniques within the EG DNA project. The first subsection explains the three approaches that can be applied to arrive at early estimates. Subsection 4.2 then discusses one of these approaches in more detail, i.e. the top-down approach starting from aggregated results of previous years. Results on the basis of the various techniques within this top-down approach are presented in subsection 4.3. The section ends with some conclusions and ideas for further work in subsection 4.4.

4.1 Possible approaches

49. There are various ways to arrive at distributional results for a more recent year:

A) A top-down approach starting from 'macro' distributional results for previous years: If no micro data are available, one may start from the aggregated distributional results of previous years to extrapolate new distributional results. Doing so, distributional results are directly targeted at the aggregated level instead of building them up from the underlying micro data. This has the advantage that no re-clustering will be needed as dynamics are already incorporated in the macro-results. The downside, however, is that it is applied on a rather high level of aggregation and that it does not incorporate detailed information that may be available.

B) A bottom up approach starting from micro data: If micro data are available or can be extrapolated, that can provide the basis for new distributional results. New distributional results are then built up from the nowcasted or newly available micro data. This bottom up approach has the advantage that individuals can be assigned to the various quintiles or household groups on the basis of the updated micro data. The downside, however, is that it will be difficult to find good information to arrive at new micro results.

C) A meso approach starting from data on household groups: The starting point in this method is not the historical macro results by income quintile but by more detailed groups of households, i.e. at a less aggregated level then the macro results. Especially when homogeneous groups can be created within the population, it may be easier to derive more timely distributional results of good quality than on the basis of rather heterogeneous quintile groups. Sometimes demographic information may be available on the basis of which the number of persons in these groups can be extrapolated and by linking them back to the relevant quintiles it may lead to more accurate nowcasts. The downside is that more assumptions may be needed to link the extrapolated results to the quintiles and that a solution needs to be found to deal with dynamics between household groups and quintiles.

50. Within the EG DNA project, only aggregated distributional data are available. That is why the focus of the EG DNA has been on the top-down approach so far. The techniques and some preliminary results are the subject of the next two subsections. It is envisaged to also start exploring the other two approaches in the near future, when data becomes available to test the various techniques.

4.2 The top-down (macro) approach in more detail

51. This section presents the various methods available to arrive at early distributional results on the basis of a top down (macro) approach. The steps that need to be taken to arrive at nowcast results on the basis of the macro approach are similar to the ones presented in section 2. The next subsections describe these steps in more detail.

4.2.1 Step 1: Preliminary adjustment of national accounts totals

52. In order to arrive at distributional results for a more recent year that are consistent with past distributional results, the more recent national accounts aggregates have to be adjusted in order to exclude parts that do not relate to domestic private households. In the aggregated approach proposed here, the correction is based on the average adjustments (as percentage of original national accounts totals) recorded over the past estimates.

4.2.2 Step 2: Extrapolation of distributional results for the quintiles

53. In the next step, distributional results by quintile can be extrapolated to arrive at distributional results for a more recent year (table 7).

Table 7. Aim of the nowcasting: Distributional results per transaction

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL
Transaction x						

54. In doing so, two options are available with regard to the target variables: the first approach consists in extrapolating the shares of each quintile to the total value of the transaction. This approach allows estimating a set of distributional keys to be applied to the total value of the transaction at hand for the most recent year. The main benefit of this approach is that the keys sum up to one, as a consequence of which no further adjustments will be required. The second approach, instead, focuses on the levels of the quintiles, not necessarily leading to results that are consistent with the adjusted national accounts totals. If this is not the case, an adjustment will be needed to align the data with these totals per transaction (see step 3 in subsection 4.2.3).

55. Furthermore, it should be noted that nowcast estimates for the distribution of income and consumption can be derived using two different strategies:

1. The direct approach consists of directly estimating each transaction, regardless of their level of aggregation. This means that all detailed components would be estimated directly, but also the balancing items, like the balance of primary income and household disposable income. The direct approach may lead to accounting inconsistencies in that the sum of the estimated 'children' transactions may not add up to the corresponding 'parent' transactions

(vertical inconsistency). In such a case re-balancing techniques should be applied to reconcile the estimations, i.e. aligning the underlying results to the results of the aggregated items.

2. In the indirect approach, 'child' transactions at the highest possible level of disaggregation are estimated directly and 'parent' transactions are obtained by summing up these estimated child transactions. In this case the consistency constraint is always met by definition and the aggregated items are the result of the sum of the underlying items.

56. Although the main interest of users will be on the distributional results for main aggregates, it is probably best to use the indirect approach, focusing on the underlying components, and to subsequently derive the main aggregates. This is deemed to lead to the most accurate results, as the distributions over the quintiles will differ from one instrument to the other, and as the weights of the various underlying instruments will show different developments over time. It also has to be borne in mind that the dynamics in balancing items may be more volatile than that of its underlying components, dependent on the way the balancing item is constructed[9]. In subsection 4.3 results are shown on the basis of the indirect approach.

57. Within the top-down approach, five nowcast techniques are currently distinguished to arrive at more timely distributional results.

A. Naïve method (t-1 or latest available shares)

58. The naïve method, only using the most recent shares, is the simplest approach. Such a method represents the sole alternative when only data for a single year are available. The naïve method relies on the assumption that the distribution of income transactions across quintiles remains unchanged from one year to the other. Therefore, all changes in the distribution of total income will be the result of the evolution of the relative weight of the transaction to the total income.

Table 8. The naïve method

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL
…						
t-1						
T						

59. For instance, if the share of quintile 1 in *compensation of employees* in t-1 was 10 percent, one tenth of the new adjusted national accounts total for compensation of employees will be attributed to quintile 1 in the year t.

[9] If the balancing item is a simple summation of its underlying components, some of the changes in the underlying components may cancel out and the balancing item will show relatively low fluctuation over time in comparison to its components. However, if the balancing item is a construct of summations and subtractions (e.g. resources and uses), the level of the balancing item will be closer to zero and fluctuations may have a larger impact. In that regard, it may show relatively large fluctuation in comparison with its underlying components. Especially in the latter case, it may prove more difficult to nowcast the balancing item directly and the results may differ to a large extent from the results that would be obtained via the indirect approach.

60. Table 9 shows an example of the impact of differences in changes for underlying transactions. In the example, 'compensation of employees' for all households increases by five percent. Keeping the relative distribution unchanged, the entry for Q1 moves from 100 to 105. In the same example, 'net property income' decreases by the same percentage, so that the amount assigned to Q1 becomes 4.8, which corresponds to five percent of total net property income recorded in t-1. As a result, the overall share of Q1 will increase by 0.1 percent, due to the relative weight of the two transactions.

Table 9. Example of the t-1 Method

Income		t-1 METHOD							
		t-1				T			
		Q1	**Q2-Q5**	**Tot**	*Weight*	**Q1**	**Q2-Q5**	**Tot**	*Weight*
D1R	Compensation of employees	100	800	900	*90.0%*	105	840	945	*90.9%*
	Shares	*11.1%*	*88.9%*			*11.1%*	*88.9%*		
D4N	Net property income received	5	95	100	*10.0%*	4.8	90.3	95	*9.1%*
	Shares	*5.0%*	*95.0%*			*5.0%*	*95.0%*		
	Share to total	*10.5%*	*89.5%*			*10.6%*	*89.4%*		

61. This method has the merit to be relatively easy to apply, but at the cost of imposing flat dynamics in the distribution process, which could represent a too strong assumption. As mentioned above, this approach may be the sole alternative when no additional data is available, or it could be a valid alternative in the presence of a structural break in the time series by which older data may be of no use. However, if more information is available, it would be wise to analyse whether incorporating this information could improve the estimates.

B. Average shares

62. If distributional results are available for a few years, a valid alternative is to calculate the average shares recorded over time and to use those to arrive at distributional results for the year t. If the share of quintile 2 for *social benefits* was 16.5 percent in t-3, 18.3 in t-2 and 17.7 in t-1, an average share of 17.5 percent could be used for the share of quintile 2 in year t. The number of years that are taken into account will depend on the time series available and the years that are deemed to be relevant in the nowcasting. Alternatively, one could also opt for a weighted approach in which the most recent years get a higher weight than earlier years, the weights still adding up to 1.

Table 10. Average shares

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL
…						
t-3						
t-2						
t-1						
t						

63. The advantage of this method is that it is easy to apply and takes into account multiple years of information. Moreover, as in the case of the previous method, the key distribution meets the

consistency constraint as their sum across quintiles is equal to one. Hence, if the shares tend to be relatively stable over time and the volatility of the quintile shares could be attributed to random effects, this technique may gain good results. However, if the shares tend to fluctuate over time and there appears to be a trend, the 'average' method may lose important information.

C. *Linear trend Method*

64. If distributional information is available for multiple years, and if the shares of the various quintile results tend to systematically increase or decrease over time, a regression against a linear trend may be an effective alternative to predict the distribution at time t. For that purpose, one has to look at the evolution over time, preferably using longer time series. Graphical aids may help in detecting certain trends.

Table 11. Linear trend method

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL
…						
t-3						
t-2						
t-1						
t						

65. The method can be applied to the levels of the transactions per quintile as well as to their shares. A reasonable criterion to choose between the use of the share or the levels could be the highest coefficient of determination (R^2). The advantage of this method is that it captures the dynamic of the evolution of the distributional information over time, though in a very simple way. The longer the time series, the more robust the estimation of the parameter will be.

D. *Regression against the adjusted national accounts total*

66. Another option is to nowcast the quintile results on the basis of their correlation to the adjusted national account totals for the household sector as a whole for specific components. If the shares or the totals of the quintiles tend to correlate with this total value, this may be used to extrapolate the distributional information. In some cases, the coefficients found will be almost equal to the average share, but in some cases the coefficients may deviate. In that case, increases or decreases will alter the distribution over the quintiles, but in a steady pattern in the case that correlation can indeed be found.

Table 12. Regression against adjusted national accounts totals

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL
…						
t-3						
t-2						
t-1						
t						

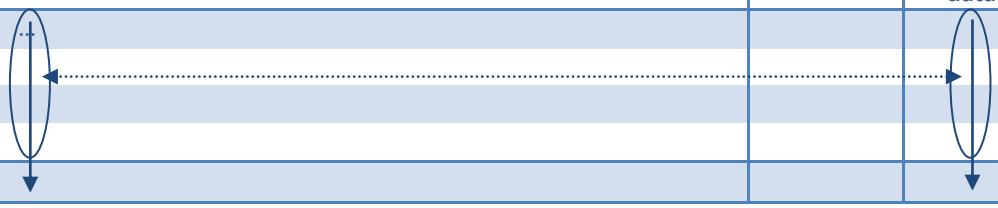
67. The regression against the adjusted national accounts total does not ensure meeting the consistency criteria. In that event, estimations need to be adjusted using a matrix balancing procedure. Subsection 4.2.3 addresses this issue.

E. Regression against auxiliary data

68. It may also be the case that values of the quintiles for certain transactions correlate with auxiliary data. The choice of the auxiliary data depends on the nature of transaction and in some cases may even depend on the quintiles. For instance, compensation of employees for certain quintiles may be related to employment rates and wage developments in specific industries. Values for some income or consumption components for certain quintiles may be related to socio-demographic information.

Table 13. Regression against auxiliary data

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	TOTAL	Auxiliary data
…							
t-3							
t-2							
t-1							
t							



69. Auxiliary information can only be used if long time series are available. In the present exercise, no such auxiliary information has been identified so far to nowcast disposable income or final consumption expenditure.

4.2.3 Step 3: Alignment of the results to the adjusted national accounts totals

70. As there are multiple ways to nowcast distributional results, even for the various quintiles within a specific transaction, results for the quintiles need not necessarily add up to 100 percent or to the adjusted national accounts totals for the specific items. In this case, the sum of the quintiles has to be aligned to the national accounts totals. Should the estimation for all quintiles be equally reliable, the double proportional (Stone et al., 1942) procedure or more advanced methods (Lenzen et al., 2009) can be applied to overcome the issues due to presence of negative or null entries. However, should the estimation of the share of certain quintiles be considered to be more reliable than the others (e.g. if its variability is significantly smaller) a different strategy could be applied in that the quintile that is deemed to be the most accurate can be kept as it is and the discrepancy can be proportionally attributed only across the remaining quintiles.

4.2.4 Step 4: Derivation of indicators

71. Once the data have been aligned to the national accounts totals, the relevant indicators can be derived. As a macro approach is used, there is no need to re-cluster households across the quintiles. This would, however, be relevant in the bottom-up and the meso approach.

4.3 Results for the top down (macro) approach

72. This section presents preliminary results on the basis of the macro approach as described in the previous section. Results are presented for Australia for which distributional time series information could be derived from their website. At the time of the calculations data were available for five years; 2003, 2005, 2007, 2009 and 2011. In assessing the various nowcasting methodologies, income and consumption results are compared against the most recent available year of the time

series, i.e. 2011. In that way, four years of the available time series have been used to nowcast the results for the fifth year.

4.3.1 Step 1: Preliminary adjustment of national accounts totals

73. Preliminary adjustment of the national accounts totals is required whenever the national accounts totals include information other than private households. The Australian data have been derived from their website for which the sum of the quintiles equalled the totals that are used in the exercise, so therefore, no adjustment was required.

4.3.2 Step 2: Extrapolation of distributional results for the quintiles

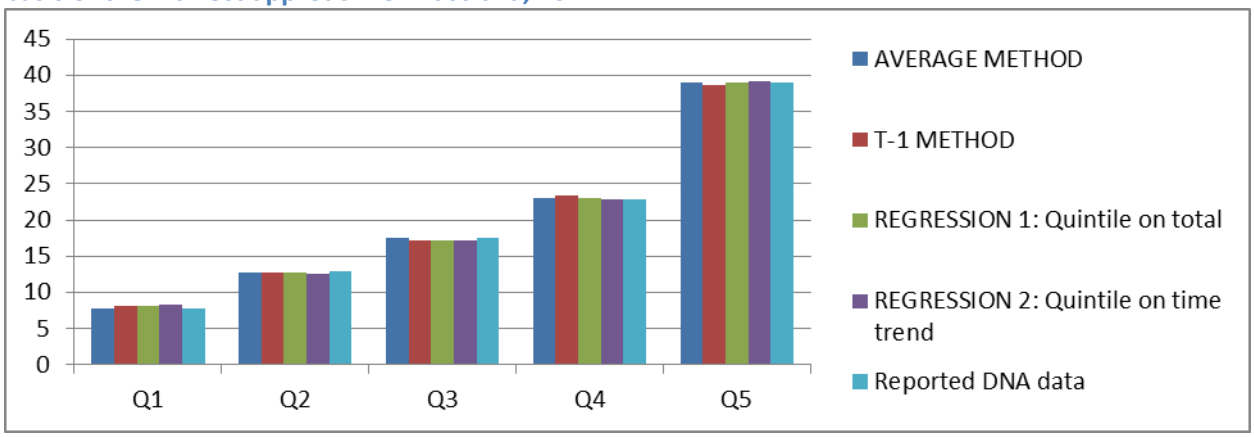
74. This section presents the preliminary results obtainable by applying the methods described in subsection 4.2.2 (except for regression against auxiliary data). Comparative results on the dispersion of estimates are first presented for the totals of disposable income and consumption. The analysis focuses on the quintile shares according to the various extrapolation methods in comparison with the actual results, and on the growth rates of the level of disposable income and consumption per quintile. Subsequently, the focus shifts to the results for the underlying items of disposable income and consumption expenditure, using the various techniques. As it may be assumed that no single method will prove to be best for all items, it is interesting to see which methods turn out to perform best for which items and to what results that would lead for the main aggregates.

A. Results for the main aggregates

Household disposable income

75. Figure 5 present the results for the quintile shares for *disposable income* (B6) for the various nowcast methods. The same method is applied to all underlying components and the results are added up to indirectly obtain the aggregate. Nowcast results from the four methods are compared against the actual data for the reference year (i.e. 'reported DNA data').

Figure 5. Relative shares per quintile from nowcast methods for disposable income (B6) on the basis of the indirect approach for Australia, 2011

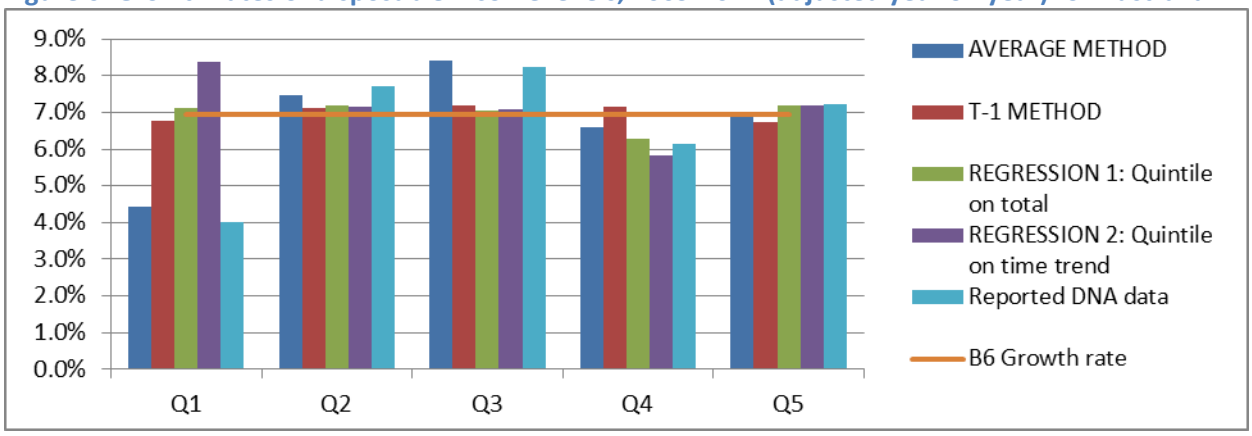


76. At a first graphical inspection the results for the four methods seem relatively close for the quintile shares and also aligned with the actual results. The maximum difference with the actual results is recorded for the first quintile, using regression against a time trend (0.63 percentage-point). The average method turns out to be the most accurate for the first and the third quintiles,

whereas for the fourth and fifth quintile regression against the adjusted national accounts totals come closest to the reported values.

77. Looking at the growth rates of the levels of disposable income provides more insight in how accurate the nowcasts are in capturing the changes. Figure 6 shows the growth rates of disposable income levels for all quintiles according to the various methods and the actual values reported on the basis of the indirect approach. All methods lead to a growth of disposable income for all quintiles. This is due to the fact that the growth rate for the household sector as a whole was quite substantial. The figure also shows that whereas the gaps with the actual growth rates are small for most quintiles, large gaps can be observed for the first quintile, where only the average method comes close to the actual results.

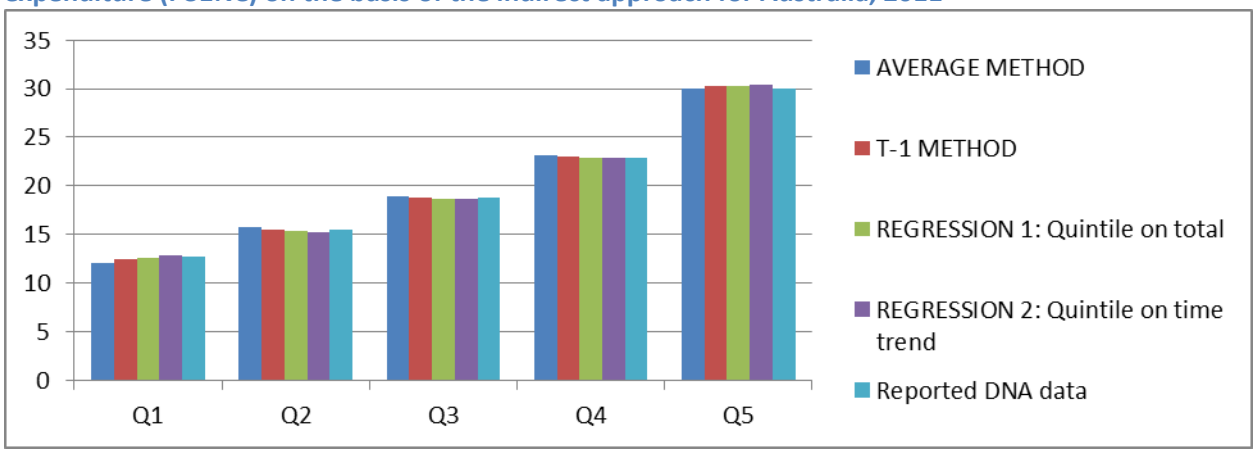
Figure 6. Growth rates of disposable income levels, 2009-2011 (adjusted year on year) for Australia



Final consumption

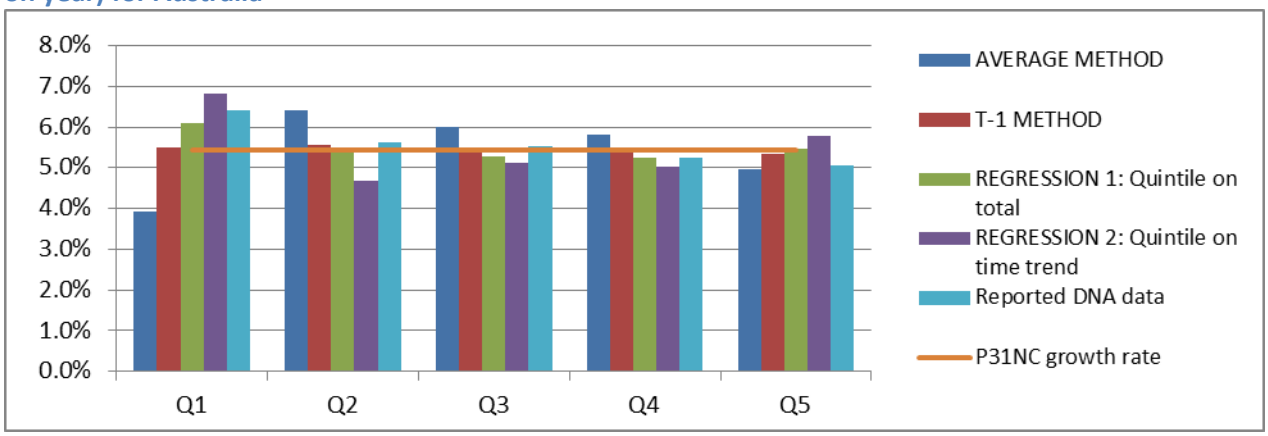
78. Figure 7 presents the results for the quintile shares for *final national consumption expenditure* (P31NC) for the various nowcast methods on the basis of the indirect approach. As was the case with the results for disposable income, the results for the shares for the various quintiles of expenditure on final consumption seem to be relatively close to one another and seem to be close to the actual values. The largest gaps can be observed for the first and fifth quintile. The maximum difference from the reported values is 0.59 percentage-point, for quintile 1 using the average method. This is comparable to the maximum gap for disposable income.

Figure 7. Relative shares per quintile from nowcast methods for final national consumption expenditure (P31NC) on the basis of the indirect approach for Australia, 2011



79. Figure 8 shows the level growth rates of final consumption expenditure for all quintiles according to the various methods and the actual results reported by Australia. The gaps with the actual results are of similar size as the ones for disposable income and depend on the method that is chosen per quintile. The average method and regression against time trend show the largest dispersions on average.

Figure 8. Growth rates of final national consumption expenditure levels, 2009-2011 (adjusted year on year) for Australia



B. Results for the underlying components

80. As the best nowcast technique may differ across the various components, it is worth investigating the relative performance of the four methods at transaction level to see how they contribute to the changes in disposable income and final consumption expenditure[10]. This provides more insight in the accuracy of the methods in nowcasting the specific transactions and how the selection of different methods across items may affect nowcast results for disposable income and consumption expenditure.

Household disposable income

81. Figure 9 presents the main contributors to the change in household disposable income for Australia. It presents the contribution of each of the components to the change of the aggregate per method. For example, 0.99 percent-points of the 4.30 percent growth rate in disposable income for the first quintile via the average method was caused by *Operating surplus* (B2). The figure shows that *Compensation of employees* (D1R) is a main driver in the change of disposable income for all quintiles. Furthermore, *Operating surplus* (B2), *Mixed income* (B3) and *Social benefits other than STiK* (D62) also contribute to a large extent to the change in disposable income for the first two quintiles. For the higher income quintiles *Current taxes on income and wealth (D5)* plays an increasing role.

82. When looking at the main contributors to the gaps with the actual results, it can be observed that especially *Operating surplus* (B2), *Mixed income* (B3), *Compensation of Employees (D1)*, and, to a lesser extent, *Social benefits other than STiK* (D62) are the most relevant ones in this respect. For

[10] In analysing the performance of the methods on the level of the underlying components, one has to consider that the various components will have different impacts on the aggregate, depending on their relative share. Therefore, one should not only look at the isolated results for the items themselves, but also how they contribute to the overall results depending on their relative share.

Figure 9. Growth rate (%) of the quintile levels for selected underlying transactions of gross disposable income (including their share in B6) for Australia, 2011

B2: Operating surplus; **B3:** Mixed income; **D1:** Compensation of Employees; **D41R:** Interest received (not corrected for FISIM); **D42R**: Distributed income of corporations; **D41P:** Interest paid (not corrected for FISIM); **FISIM:** Correction for financial intermediation services indirectly measured; **D5P:** Current taxes on income and wealth; **D61P:** Net social contributions; **D62R:** Social benefits other than in kind; **D71P:** Non-life insurance premiums; **D72R:** Non-life insurance claims.

all methods, they show significant deviations from the actual results. It will depend on the specific method how much the deviation will be, but in any case it will contribute to any observed gaps on the aggregated level.

Final Consumption

83. Figure 10 present results for the main contributors to the changes in final consumption expenditure. As was the case with the income items, for some transactions the nowcast results come very close to the actual values, whereas for others the gaps turn out to be quite substantial.

84. The items that show largest relative gaps between the nowcast and the actual results are *Health* (CP060), *Recreation and culture* (CP090) and *FISIM* (CP1261). These are also the items that show largest dispersion between the results of the various nowcast methods. However, the figure also shows that for each of these items, there are still methods that come relatively close to the actual results. That means that if the best method is chosen, the results would probably not deviate much from the actual ones.

C. Scoring the methods per component

85. As was explained before, the most appropriate nowcast method will differ per transaction. Furthermore, the optimal choice may differ from year to year. This will largely depend on the stability of the patterns in the data. At this point in time, the optimal method can only be checked for one single year by confronting the nowcast estimates with the actual results. By looking at the gaps between the results for the shares according to the various methods and the actual share results, it can be determined which method gains the best nowcast results in practice for this specific year. Under the assumption that the distance from the actual percentage share per quintile is a good measure of the accuracy of the methods, the sum of these gaps (in percentage-points) can be used to compare methods' performances.

Results for household disposable income

86. Table 14 presents results for household disposable income. For each item the sum of the absolute gaps for the quintiles of the nowcast result with the actual result is shown for each of the methods. It shows that in some cases specific methods lead to large differences, whereas they come really close for others. This is in line with the observations in the previous subsection. It underlines that it really matters which method is used for which instrument.

87. Looking at the results it turns out that the average method and the t-1 method are the most preferred options (in 13 out of 16 cases). They come closest to the actual values for most of the instruments. The two regression options only gain the best results in three of the cases (i.e. *Interest received*, *Rent paid* and *Other current transfers received*). The table also shows that for most of the transactions the results of the various methods are rather close to one another. Only for *Operating surplus* (B2) and *Current transfers received* (D75R) larger gaps can be observed. This probably relates to the fact that only short time series are available for the analysis. Furthermore, the results for the available years are relatively stable over time for most of the transactions, except for the latter two.

Figure 10. Growth rate (%) of the quintile levels for selected underlying transactions of Final national consumption expenditure (including their share in P31NC) for Australia, 2011

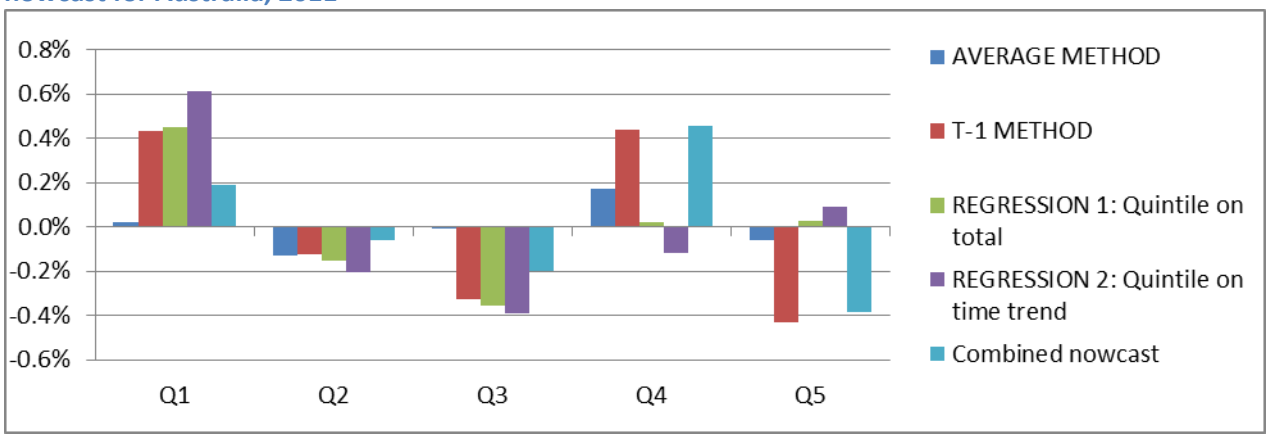
CP010: Food and non-alcoholic beverages; **CP041:** Actual rentals on housing; **CP042:** Imputed rentals on housing; **CP045:** Electricity, gas and other fuels; **CP060:** Health; **CP072:** Operation of personal transport equipment; **CP073**: Transport services; **CP090**: Recreation and culture; **CP110:** Restaurants and hotels; **CP12x:** Miscellaneous goods and services (excl. FISIM and insurance services); **CP125**: Insurance services.

Table 14. The gaps between the nowcast results and the actual values for the various underlying transactions of household disposable income for Australia, 2011

	Instrument	Average method	t-1 method	Regression 1	Regression 2	"Best" choice
B2	Operating surplus	7.34	49.04	50.85	50.19	Average method
B3	Mixed income	10.20	11.79	12.66	13.35	Average method
D1R	Compensation of employees	1.39	0.94	1.79	1.96	t-1 method
D4N	Net property income	18.35	13.32	16.07	17.82	t-1 method
D41'R	Interest received (excl. FISIM)	4.91	5.68	3.98	2.70	Regression 2
D42R	Distributed income of corporations	8.43	10.40	9.08	9.65	Average method
D45R	Rent received	10.33	9.14	11.69	9.26	t-1 method
D41'P	Interest paid (excl. FISIM)	4.05	5.25	4.21	6.89	Average method
D45P	Rent paid	10.33	9.15	8.97	9.27	Regression 1
D5P	Current taxes on income and wealth	3.57	1.82	3.74	2.88	t-1 method
D61P	Net social contributions	2.09	0.44	1.41	1.28	t-1 method
D62R	Social benefits other than STiK	4.30	3.07	3.41	4.79	t-1 method
D71P	Non-life insurance premiums	1.42	0.29	0.38	0.97	t-1 method
D72R	Non-life insurance claims	1.42	0.29	0.46	0.97	t-1 method
D75R	Misc. current transfers received	14.95	20.08	13.67	25.78	Regression 1
D75P	Misc. current transfers paid	6.29	1.95	2.30	3.69	t-1 method

88. Figure 11 shows the results for disposable income using the indirect approach for the various methods in relation to the actual values, including the method that would combine all 'best' choices as presented in Table 14. The latter is referred to as the hybrid-indirect approach, as it combines different methods for the various transactions. It turns out that the combined approach would lead to rather accurate results for the first, second and the third quintile in comparison with most of the other nowcast methods, but that it leads to higher differences for the fourth and the fifth quintile. In that regard, the combined nowcast approach on the basis of different techniques per item does not immediately qualify as best approach from the viewpoint of all quintiles. In that regard, further research is needed into the most appropriate selection criteria to select the best method, and also whether applying different techniques within an item for the various quintiles may lead to better results.

Figure 11. Difference between nowcast results and reference DNA data for the quintile shares of disposable income on the basis of the indirect approach, including the best choice ('combined') nowcast for Australia, 2011



Results for consumption expenditure

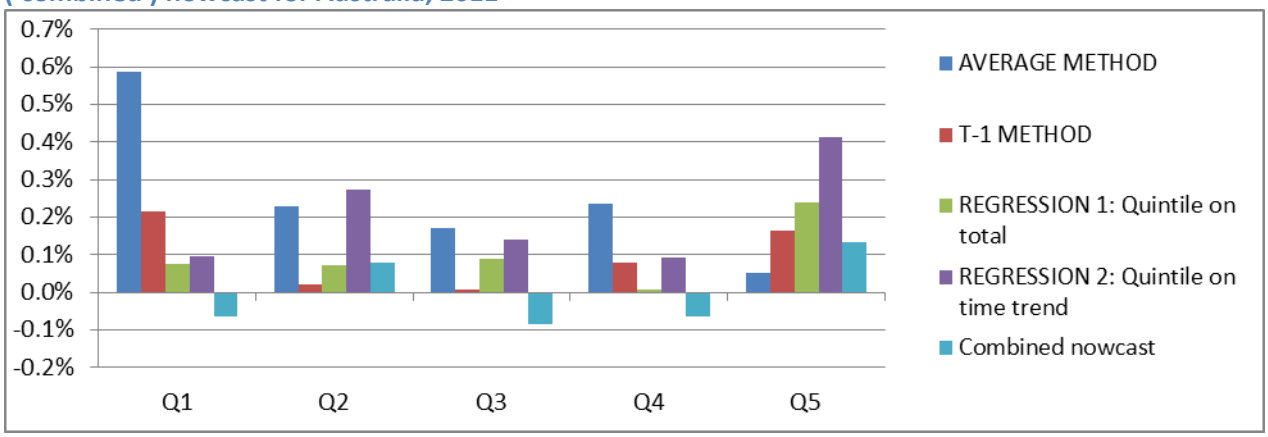
89. Table 15 presents the gaps on the basis of the various methods for the underlying items of consumption expenditure. When the results are compared with the gaps for the income components, it can be noted that the gaps for consumption are smaller on average. This implies that the distribution of the consumption items is more stable over time than that of the income items. Looking at the specific results, the two regression methods turn out to gain the best results for most of the items (in 9 out of 12 cases). There are only three cases where the t-1 method or the Average method leads to the best results (i.e. *Housing, water*, *Electricity, gas and other fuels*, *Restaurants and hotels*, and *Miscellaneous goods and services*).

Table 15, The gaps between the nowcast results and the actual values for the various underlying transactions of consumption expenditure for Australia, 2011

		Average method	t-1 method	Regression 1	Regression 2	"Best" choice
CP010	Food and non-alcoholic beverages	1.65	0.60	0.23	0.24	Regression 1
CP020	Alc. beverages, tob. and narcotics	5.59	2.43	0.64	0.69	Regression 1
CP030	Clothing and footwear	6.42	2.15	2.32	1.26	Regression 2
CP040	Housing, water, electricity, …	0.52	0.15	0.17	0.16	t-1 method
CP050	Furnishings, households equip., …	5.12	1.77	0.80	0.75	Regression 2
CP060	Health	11.60	3.52	0.36	3.13	Regression 1
CP070	Transport	2.59	0.87	1.30	0.47	Regression 2
CP080	Communications	2.75	0.91	0.89	0.51	Regression 2
CP090	Recreation and culture	4.79	1.37	0.71	1.62	Regression 1
CP100	Education	6.80	1.92	1.04	2.42	Regression 1
CP110	Restaurants and hotels	3.71	0.93	1.19	1.56	t-1 method
CP120	Misc. goods and services	2.57	2.63	3.23	4.52	Average method

90. Figure 12 shows the results for consumption expenditure using the indirect approach for the various methods in relation to the actual results, including the method that would combine all 'best' choices as presented in Table 15. In this case, the best choices would lead to an improvement of the nowcasts for consumption expenditure, although the t-1 method and the regression on total do not differ by much. For quintiles 1 and 4 the combination would lead to the best results, but for quintiles 2 and 3 this would be the t-1 method and for quintile 5 the average method.

Figure 12. Difference between nowcast results and reference DNA data for the quintile shares of consumption expenditure on the basis of the indirect approach, including the best choice ('combined') nowcast for Australia, 2011



4.4 Conclusions and further work

91. There is a strong demand for timely distributional data on income and consumption within the framework of national accounts. As timely micro data may be lacking, nowcasting techniques may be used to arrive at more timely distributional results. This section explained that there are generally three approaches to compile these estimates: starting from macro results, micro data or meso results from previous years and combining them with newly available national accounts results. It also presented five nowcasting techniques within the macro approach and showed results for four of them on the basis of time series information for one country. The results showed that the nowcasting techniques come rather close to the actual results, although more dispersion can be observed when looking at growth rates. Furthermore, the investigation of the underlying transactions showed that for certain items the gaps were particularly high depending on the selected method. The preferred approach will depend on the patterns that can be found in the data and the presence of additional information that can be incorporated in the nowcasts.

92. For further research it would be good if the techniques presented in this section could be applied to a wider variety of countries and could be based on longer time series. That would help in testing the validity of the nowcast techniques and to advise on which techniques should be applied under which circumstances. Furthermore, the availability of micro- and meso-data would help in extending the analyses towards the use of the micro and the meso approaches. If micro data time series would become available that underlie the distributional results or that may correlate with the distributional results and have good timeliness to use in the nowcasts, it would provide possibilities to explore nowcasting techniques within the bottom-up approach. More detailed information on specific and more homogeneous household groups would be helpful in further exploring nowcasting techniques within the meso-approach.

5. Conclusions and further work

93. This paper has shown that the methodology to compile distributional results within the framework of the national accounts is very dependent on the quality and the alignment of micro and macro data used in the process. Any gaps between these data need to be bridged and allocated to the relevant households, as a consequence of which the overall quality of the distributional results heavily depends on the size and information on these gaps. As currently not a lot of information is available on the reasons for these gaps, most countries apply a proportional allocation to the underlying households. However, the paper has shown that there may be various reasons underlying these gaps, that the allocation to households may vary across these reasons and that this may differ from the distributions derived from the micro data. That is why it is important to obtain more insight in the reasons for these micro-macro gaps and how they may affect the distributional results. A framework has been presented via which micro and macro experts can allocate the gaps to underlying reasons and via which the related amounts can be allocated to the relevant households. It is recommended that micro and macro experts use this framework to regularly discuss the gaps between the micro and macro aggregates and that compilers of distributional results use its results in the compilation process of these results. This should be preferred to a simple proportional allocation and would lead to improved distributional results. Moreover, to have more insight into

the exact reasons for the gaps, more research is needed, amongst others via comparisons of different data sources and via linkages of data across different data sources.

94. The second part of the paper presented possible approaches to compile more timely distributional results via nowcasting techniques. Distributional results can theoretically be nowcasted using a top-down, a bottom-up or a meso-approach. Within those approaches various techniques can be applied to arrive at distributional results for a more recent year. These techniques vary from taking the distribution of the last available year, to using the average over a set of recent years, to looking at regression over time trend, or regression against an auxiliary variable. Preliminary results on the basis of these techniques within the top-down approach showed that nowcasts come relatively close to the reported shares of income and consumption by quintile. However, larger differences can be are observed when analysing the growth rates. Furthermore, depending on the exact nowcasting technique applied, gaps for underlying items may also turn out high. An appropriate combination of methods across all the underlying components may in that case lead to more accurate results. More research will be needed in this domain to test which technique should best be used in which case. Finally, it has to be borne in mind that the results presented in this paper only relate to a single country on the basis of relatively short time series. The research should be broadened to test these techniques for a broader range of countries and on the basis of longer time series. Furthermore, on the basis of additional data sets, the applicability of the micro and meso approach should also be explored.

95. In addition to some other issues that still need attention, the Expert Group on Disparities in National Accounts will continue its work on both the micro-macro gaps and the nowcasting. The issue of the micro-macro gaps will be further discussed at the next meeting of the expert group in October 2016 in which best practices on bridging the gaps will be exchanged between member countries and further guidance will be sought on how to deal with these gaps. Furthermore, together with the European Central Bank, the Expert Group wants to start exploring possibilities to also include the wealth dimension in the work on distributional results. In addition to constituting very relevant information in itself, it would also provide more insight in the plausibility of the savings results currently derived in the EG DNA project by confronting results from income and consumption with results from the wealth side. This may point to implausibilities in the data and provide direction on how parts of the gaps may be allocated to the various household groups (also revealing possible reasons for these gaps). With regard to nowcasting, the Expert Group has started to explore the availability of additional data sources that may be used to test the various techniques. In that, it is also looking for cooperation with other projects involved in nowcasting to see whether the number of techniques for nowcasting timely distributional results may be further extended. A working paper on these techniques is foreseen to be published in the course of 2017.

References

Accardo, Jérôme; Bellamy, Vanessa; Consalès, Georges; Fesseau, Maryse; Le Laidier, Sylvie; and Raynaud, Émilie, "Inequalities between households in the national accounts – Breakdown of household accounts", 2009.

Carson, Carol S., "The Underground Economy: An introduction", *Survey of current business*, Volume 64, Number 5, p21-37, 1984.

Coli, Alessandra and Tartamella, Francesca, "Using administrative and survey data to analyse tax evasions from unregistered labour", *IARIW-paper*, 2014.

D'Alessio, Giovanni and Faiella, Ivan, "Non-response behaviour in the bank of Italy's survey of household income and wealth", *Working papers*, number 462, 2002.

D'Alessio, Giovanni and Neri, Andrea, "Income and wealth sample estimates consistent with macro aggregates: some experiments", *Occasional paper Banca D'Italia*, no. 272, 2015.

Fesseau, Maryse and Mattonetti, Maria L., "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*, 2013.

Lenzen, Manfred; Gallego, Blanca and Wood, Richard, "Matrix balancing under conflicting information", *Economic Systems Research*, Vol. 21, Issue 1, 2009.

Meyer, Bruce D.; Mok, Wallace K.C.; and Sullivan, James X., "The under-reporting of transfers in household surveys: its nature and consequences", 2008.

Sabelhaus, John; Johnson, David; Ash, Stephen; Garner, Thesia; Greenlees, John; Henderson, Steve; and Swanson, David, "Is the consumer expenditure survey representative by income?", 2012.

Törmälehto, Veli-Matti, "High incomes and affluence: Evidence from EU-SILC", *paper prepared for the 2014 EU-SILC International Conference*, held in Lisbon on 16-17 October 2014.

Stiglitz, Joseph E.; Sen, Amartya and Fitoussi, Jean-Paul, "Report by the Commission on the Measurement of Economic Performance and Social Progress", 2009.

Stone, R., Champernowne, D.G. and Meade, J.E., "The Precision of National Income Estimates", *Review of Economic Studies*, 9: p111–125, 1942.

Vermeulen, Philip, "How fat is the top tail of the wealth distribution?", *ECB Working paper series*, N.1692, 2014.

Zwijnenburg, Jorrit; Bournot, Sophie and Giovannelli, Federico (2016), "Expert Group on Disparities within a National Accounts Framework – Results from a recent exercise", *forthcoming OECD Working Paper*, 2016.