Methodological Remarks – *Data Access Research Tool (DART)*

Aggregates

Aggregate *Equivalised Disposable Household Income* refers to cash and non-cash income from labour, income from capital, income from pensions (including private and public pensions) and non-pension public social benefits stemming from insurance, universal or assistance schemes (including in-kind social assistance transfers), as well as cash and non-cash private transfers, after deduction of the amount of income taxes and social contributions paid. *Disposable Household Income* is equivalised at individual level as the total amount divided by the square root of household members. Before equivalisation, top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Disposable Household Income*: at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Aggregate *Equivalised Market Household Income* refers to income received by the households before public redistribution takes place; it includes cash and non-cash income from labour, income from capital, income from private pensions, as well as cash and non-cash private transfers, before deduction of income taxes and social contributions paid. *Market Household Income* is equivalised at individual level as the total amount divided by the square root of household members. Before equivalisation, top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Market Household Income*: at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Aggregate *Equivalised Market Household Income (including contributory transfers)* refers to income received by the households before public redistribution takes place, whereby contributory transfers are treated as deferred labour income; it includes cash and non-cash income from labour, income from capital, income from private pensions, cash and non-cash private transfers, as well as income from contributory transfers, after deduction of social contributions paid, before deduction of income taxes paid. *Market Household Income (including contributory transfers)* is equivalised at individual level as the total amount divided by the square root of household members. Before equivalisation, top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Market Household Income (including contributory transfers)*: at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Aggregate *Gross Wages* refers to monetary payments received from regular and irregular dependent employment. This includes cash wage and salary income and monetary supplements to the basic wage, such as overtime pay, employer bonuses, 13th month bonus, profit-share, tips before deduction of income taxes and social contributions. Estimates refer to wage earners aged 16-64 who have non-*zero* wages. Top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Gross wages:* at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Indicators available in '*trends'*, '*scatter plots'*, and '*maps'*

Indicator *Gini Index* measures the extent to which the distribution of the specified aggregate among individuals or households within an economy deviates from a perfectly equal distribution. The Gini index measures the area between the Lorenz curve and the hypothetical line of absolute equality. A Gini index of *zero* represents perfect equality and 1, perfect inequality.

Indicator *Atkinson Index* is a welfare-based measure of inequality, representing the percentage of total income that a given society would have to sacrifice in order to have (more) equally distributed incomes (more equal shares of income between its citizens). This measure depends on the degree of society aversion to inequality, where a higher value entails greater social utility or willingness by individuals to accept smaller incomes in exchange for a more equal distribution. The aversion parameter is set to 1.5.

Indicator *90/10 Percentile Ratio* represents the income by individuals at the 90th percentile compared to the one of individuals at 10th percentile.

Indicator *90/50 Percentile Ratio* represents the income by individuals at the 90th percentile compared to the one of individuals at the median.

Indicator *50/10 Percentile Ratio* represents the income by individuals at the median compared to the one of individuals at 10th percentile.

Indicator *Average (PPP in $US)* is the arithmetic mean, i.e. the sum of all income divided by the number of valid observations. Conversion to International Dollars follows the PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

Indicator *Median (PPP in $US)* is the amount that divides the selected population into two equal groups, half with income above this value and half below. Conversion to International Dollars follows the PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

Indicator *Relative Poverty Rate at 50% of the Median* is the percentage of the population whose income falls below the poverty line as defined as half of the median of *equivalised disposable household income*. In combination with *decompositions*, the ratio refers to the percentage of each group, whose income falls below the above defined poverty line.

Indicator *Relative Poverty Rate at 60% of the Median* is the percentage of the population whose income falls below the poverty line as defined as 60 % of the median of *equivalised disposable household income*. In combination with *decompositions*, the ratio refers to the percentage of each group, whose income falls below the above defined poverty line.

Indicator *Share of Low Income Workers (< 50% of Median)* is the percentage of wage earners whose wage falls below half of the median of *gross wages*. In combination with *decompositions*, the ratio refers to the percentage of each group, whose income falls below the above defined ratio.

Indicator *Share of Category in Total Population* is the percentage of persons in a sub-population for which the aggregate variable is defined.

Indicators available in '*distributions*'

Indicator *Lorenz Curve* is a graph that shows the extent of inequality in income by plotting the cumulative share of population on the horizontal axis against the cumulative share of the total income on the vertical axis. The extent to which the curve falls below a straight diagonal line indicates the degree of inequality.

Indicator *Density Function* is a graph determining the shape of distributions; the area under the density curve between two points corresponds to the probability that the income falls between those two values. For estimation of probability density functions, kernel density estimation is applied. Data are censored at the bottom and top 1 % of the distribution. Conversion to International Dollars is applied, following PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

Decompositions in combination with aggregates *Equivalised Disposable Household Income* and *Equivalised Market Household Income*

Decomposition by *Income Deciles* splits the total population in ten population groups equal in size, while applying household weighting factors multiplied by the number of persons. The *1st* income decile is defined as the population group with the lowest *equivalised disposable household income*, respectively the *10th* income decile refers to the population group with the highest.

Decomposition by *Household Type* splits the total population in five groups: *One person households, Couples without children, Couples with children, Single parents,* and *Other family / non-family households*. Children are defined as own children, step-children, or foster children of the head aged 17 or younger. Households with only the nuclear family (head & spouse plus children / head plus children), where both *children aged 17 or younger* and *children 18 or older* are present, are still included in *Couples with children* / *Single parents*. When there are nuclear families with only adult children of the head, then they are included in category *Other family / non-family households*. Households which consist of nuclear families and other relatives or non-relatives are all included in category *Other family / non-family households*.

Decomposition by *Homeownership* splits the total population in two groups: *Homeowners* and *Not homeowners*.

Decomposition by *Urban vs. Rural* splits the total population according to their geographic area: those living in *Urban area* and those living in *Rural area*. Categorisation in urban vs. rural areas follows typically national definitions.

Decomposition by *Age Group* splits the total population in three groups: *Children (aged 17 or below), Working age adults (aged 18 - 64),* and *Elderly (aged 65 or above)*.

Decompositions in combination with aggregate *Gross Wages*

Decomposition by *Wage Deciles* splits the population of wage earners in ten groups equal in size, while applying weighting factors. The *1st* income decile is defined as the wage earner group with the lowest *gross wages*, respectively the *10th* income decile refers to the population group with the highest.

Decomposition by *Sex* splits the population of wage earners in two groups: *Men* and *Women*.

Decomposition by *Age Group* splits the population of wage earners in five groups: *16-24 years, 25-34 years, 35-44 years, 45-54 years,* and *55-64 years*.

Decomposition by **Education Level** splits the population of wage earners in three groups: those with *Low education,* those with *Medium education,* and those with *High education.* The categories are constructed following UNESCO's ISCED 1997 / 2011 education recoding standards, where *Low education* refers to ISCED levels 0 (early childhood educational development and pre-primary), 1 (primary), and 2 (lower secondary), *Medium education* to ISCED levels 3 (upper secondary) and 4 (post-secondary, non-tertiary), and *High education* to ISCED 1997 levels 5 (first stage of tertiary) and 6 (second stage of tertiary) or respectively ISCED 2011 levels 5 (short-cycle tertiary), 6 (bachelor or equivalent), 7 (master or equivalent), and 8 (doctorate or equivalent). *Low education* includes also persons who never attended or did not complete any education level.

Decomposition by **Immigration** splits the population of wage earners in two groups: *Not immigrants* and *Immigrants.* Immigrant status refer, either to self-defined immigration status, years of residence in the country, country of birth, citizenship, or a combination of country of birth and citizenship.

WEALTH DATABASE

Aggregates

Aggregate **Equivalised Disposable Household Income** refers to cash and non-cash income from labour, income from capital, income from pensions (including private and public pensions) and non-pension public social benefits stemming from insurance, universal or assistance schemes (including in-kind social assistance transfers), as well as cash and non-cash private transfers, after deduction of the amount of income taxes and social contributions paid. *Disposable Household Income* is equivalised at individual level as the total amount divided by the square root of household members. Before equivalisation, top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Disposable Household Income*: at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Aggregate **Equivalised Gross Household Income** refers to cash and non-cash income from labour, income from capital, income from pensions (including private and public pensions) and non-pension public social benefits stemming from insurance, universal or assistance schemes (including in-kind social assistance transfers), as well as cash and non-cash private transfers, before deduction of income taxes and social contributions paid. *Gross Household Income* is equivalised at individual level as the total amount divided by the square root of household members. Before equivalisation, top and bottom coding has been applied by setting boundaries for extreme values of log transformed *Gross Household Income*: at the top Q3 plus 3 times the interquartile range (Q3-Q1), and at the bottom Q1 minus 3 times the interquartile range.

Aggregate **Disposable Net Worth** refers to the sum of non-financial and financial assets, excluding pension assets and other long-term savings, minus total liabilities. The unit of analysis is the household. *Zero* and negative values are kept. No top and bottom coding has been applied.

Aggregate **Value of Principal Residence** refers to the current value of the main dwelling occupied by the household and owned or partly owned by one or more of its members. The land on which the residence is located, as well as outbuildings that belong to the principal residence, are included in the value. Houseboats and mobile homes that are considered the principal residence are included. The unit of analysis is the household. Estimates refer to households who possess principal residence. *Zero* values are excluded. No top and bottom coding has been applied.

Aggregate **Financial Assets** refers to the market value of financial investments (bonds and other debt securities, stocks and other equity, investment funds and alternative investments), deposit accounts, cash and other financial assets owned by household members. Aggregate *Financial Assets* excludes pension assets and other long-term savings. *Zero* values are excluded. No top and bottom coding has been applied.

Aggregate **Total Debt** refers to the outstanding balance of debt for current obligations taken by household members for the principal residence and other real estate, as well as the current value of all non-housing liabilities such as investment loans, consumer goods loans, education loans and other liabilities. *Zero* values are excluded. No top and bottom coding has been applied.

Indicators **available in *'trends', 'scatter plots',* and *'maps'***

Indicator **Gini Index** measures the extent to which the distribution of the specified aggregate among individuals or households within an economy deviates from a perfectly equal distribution. The Gini index measures the area between the Lorenz curve and the hypothetical line of absolute equality. A Gini index of *zero* represents perfect equality and 1, perfect inequality.

Indicator **Atkinson Index** is a welfare-based measure of inequality, representing the percentage of total income that a given society would have to sacrifice in order to have (more) equally distributed incomes (more equal shares of income between its citizens). This measure depends on the degree of society aversion to inequality, where a higher value entails greater social utility or

willingness by individuals to accept smaller incomes in exchange for a more equal distribution. The aversion parameter is set to 1.5.

Indicator *90/10 Percentile Ratio* represents the income / assets / liabilities by individuals at the 90th percentile compared to the one of individuals at 10th percentile.

Indicator *90/50 Percentile Ratio* represents the income / assets / liabilities by individuals at the 90th percentile compared to the one of individuals at the median.

Indicator *50/10 Percentile Ratio* represents the income / assets / liabilities by individuals at the median compared to the one of individuals at 10th percentile.

Indicator *Average (PPP in $US)* is the arithmetic mean, i.e. the sum of all income / assets / liabilities divided by the number of valid observations. Conversion to International Dollars follows the PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

Indicator *Median (PPP in $US)* is the amount that divides the selected population into two equal groups, half with income / assets / liabilities above this value and one below. Conversion to International Dollars follows the PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

Indicator *Relative Poverty Rate at 50% of the Median* is the percentage of the population whose income falls below the poverty line as defined as half of the median of *equivalised disposable or gross household income*. In combination with *decompositions*, the ratio refers to the percentage of each group, whose income falls below the above defined poverty line.

Indicator *Relative Poverty Rate at 60% of the Median* is the percentage of the population whose income falls below the poverty line as defined as 60 % of the median of *equivalised disposable or gross household income*. In combination with *decompositions*, the ratio refers to the percentage of each group, whose income falls below the above defined poverty line.

Indicator *Participation Rate* refers to the share of households who are represented in the estimates. This means the share of households which have non-*zero* values in the aggregate.

Indicator *Income and Asset Poor in %* refers to the population share living in households whose *equivalised gross household income* falls below the relative income poverty line at 50% of the median and who do not have sufficient *financial assets* to cover consumption for at least 3 months (i.e. equivalised financial assets below 25% of the relative income poverty line).

Indicator *Not Income Poor, Asset Poor in %* refers to the population share living in households whose *equivalised gross household income* is above the relative income poverty line at 50% of the median, but who do not have sufficient *financial assets* to cover consumption for at least 3 months (i.e. equivalised financial assets below 25% of the relative income poverty line). These households are referred to as *economically vulnerable.*

Indicator *Income Poor, not Asset Poor in %* refers to the population share living in households whose *equivalised gross household income* is falls below the relative income poverty line at 50% of the median, but who have sufficient *financial assets* to cover consumption for at least 3 months (i.e. equivalised financial assets above 25% of the relative income poverty line).

Indicator *Median Debt-to-Asset Ratio* (*leverage ratio*) refers to the ratio of the sum of all debts by the household, divided by the sum of non-financial and financial assets (excluding pensions). The estimate reports the median of the ratios for all households who had debt at the moment of collection of assets and liabilities.

Indicator *Median Debt-to-Income Ratio* refers to the ratio of the sum of all debts by the household, divided by total yearly gross income. The estimate reports the median of the ratios for all households who had debt at the moment of collection of assets and liabilities.

Indicator *Median Debt Payment-to-Income Ratio* refers to the ratio of to the sum of all debt payments (principal and interest on mortgage and other loans) by the household in the income reference year, divided by total yearly gross income. The estimate reports the median of the ratios for all households who had debt payments in the income reference year.

Indicator *Share of Category in Total Population* is the percentage of persons in a sub-population for which the aggregate variable is defined. For *Wealth Database* aggregates *Disposable Net Worth*, *Value of Principal Residence*, *Financial Assets*, and *Total Debt* the estimates are calculated as the percentage of households.

Indicators **available in '*distributions*'**

Indicator **Lorenz Curve** is a graph that shows the extent of inequality in income / assets / liabilities by plotting the cumulative share of population on the horizontal axis against the cumulative share of the total income / assets / liabilities on the vertical axis. The extent to which the curve falls below a straight diagonal line indicates the degree of inequality.

Indicator **Density Function** is a graph determining the shape of distributions; the area under the density curve between two points corresponds to the probability that the income / asset / liability falls between those two values. For estimation of probability density functions, kernel density estimation is applied. Data are censored at the bottom and top 1 % of the distribution for income variables, and at the bottom 0.1 % and top 5 % for wealth variables. Conversion to International Dollars is applied, following PPP conversion factors for 2017 and Consumer Price Indexes (CPIs) from the World Bank Development Indicators; values refer to 2017 PPPs (in 2017 $US).

In order to better visualise the density function, we recommend users to select one country at a time.

Decompositions in combination with aggregate *Equivalised Disposable Household Income*

Decomposition by **Income Deciles** splits the total population in ten population groups equal in size, while applying household weighting factors multiplied by the number of persons. The *1st* income decile is defined as the population group with the lowest *equivalised disposable household income*, respectively the *10th* income decile refers to the population group with the highest.

Decomposition by **Wealth Percentiles** splits the population of households in five population groups, while applying household weighting factors. The following five percentile groups are constructed *1st-50th* (equivalent to the bottom 50% of households with the lowest disposable net worth), *51st - 75th, 76th - 90th, 91st - 95th,* and *96th - 100th*.

Decomposition by **Household Type** splits the total population in five groups: *One person households, Couples without children, Couples with children, Single parents,* and *Other family / non-family households*. Children are defined as own children, step-children, or foster children of the head aged 17 or younger. Households with only the nuclear family (head & spouse plus children / head plus children), where both *children aged 17 or younger* and *children 18 or older* are present, are still included in *Couples with children / Single parents*. When there are nuclear families with only adult children of the head, then they are included in category *Other family / non-family households*. Households which consist of nuclear families and other relatives or non-relatives are all included in category *Other family / non-family households*.

Decomposition by **Homeownership** splits the total population in two groups: *Homeowners* and *Not homeowners*.

Decompositions in combination with aggregate *Equivalised Gross Household Income*

Decomposition by **Income Deciles** splits the total population in ten population groups equal in size, while applying household weighting factors multiplied by the number of persons. The *1st* income decile is defined as the population group with the lowest *equivalised gross household income*, respectively the *10th* income decile refers to the population group with the highest.

Decomposition by **Wealth Percentiles** splits the population of households in five population groups, while applying household weighting factors. The following five percentile groups are constructed *1st-50th* (equivalent to the bottom 50% of households with the lowest disposable net worth), *51st - 75th, 76th - 90th, 91st - 95th,* and *96th - 100th*.

Decomposition by **Household Type** splits the total population in five groups: *One person households, Couples without children, Couples with children, Single parents,* and *Other family / non-family households*. Children are defined as own children, step-children, or foster children of the head aged 17 or younger. Households with only the nuclear family (head & spouse plus children / head plus children), where both *children aged 17 or younger* and *children 18 or older* are present, are still included in *Couples with children / Single parents*. When there are nuclear families with only adult children of the head, then they are included in category *Other family / non-family households*. Households which consist of nuclear families and other relatives or non-relatives are all included in category *Other family / non-family households*.

Decomposition by **Homeownership** splits the total population in two groups: *Homeowners* and *Not homeowners*.

Decompositions in combination with aggregate *Disposable Net Worth, Value of Principal Residence, Financial Assets,* and *Total Debt*

Decomposition by **Income Deciles** splits the total population in ten population groups equal in size, while applying household weighting factors multiplied by the number of persons. The *1st* income decile is defined as the population group with the lowest *equivalised gross household income*, respectively the *10th* income decile refers to the population group with the highest.

Decomposition by **Wealth Percentiles** splits the population of households in five population groups, while applying household weighting factors. The following five percentile groups are constructed *1st-50th* (equivalent to the bottom 50% of households with the lowest disposable net worth), *51st - 75th, 76th - 90th, 91st - 95th,* and *96th - 100th*.

Decomposition by **Sex** splits the population of household heads in two groups: *Men* and *Women*.

Decomposition by *Household Type* splits the total population in five groups: *One person households, Couples without children, Couples with children, Single parents,* and *Other family / non-family households*. Children are defined as own children, step-children, or foster children of the head aged 17 or younger. Households with only the nuclear family (head & spouse plus children / head plus children), where both *children aged 17 or younger* and *children 18 or older* are present, are still included in *Couples with children / Single parents.* When there are nuclear families with only adult children of the head, then they are included in category *Other family / non-family households.* Households which consist of nuclear families and other relatives or non-relatives are all included in category *Other family / non-family households.*

Decomposition by *Education Level* splits the population of wage earners in three groups: those with *Low education,* those with *Medium education,* and those with *High education.* The categories are constructed following UNESCO's ISCED 1997 / 2011 education recoding standards, where *Low education* refers to ISCED levels 0 (early childhood educational development and pre-primary), 1 (primary), and 2 (lower secondary), *Medium education* to ISCED levels 3 (upper secondary) and 4 (post-secondary, non-tertiary), and *High education* to ISCED 1997 levels 5 (first stage of tertiary) and 6 (second stage of tertiary) or respectively ISCED 2011 levels 5 (short-cycle tertiary), 6 (bachelor or equivalent), 7 (master or equivalent), and 8 (doctorate or equivalent). *Low education* includes also persons who never attended or did not complete any education level.

Decomposition by *Homeownership* splits the population of household heads in two groups: *Homeowners* and *Not homeowners.*

Decomposition by *Age Group* splits the population of household heads in five groups: *Below 35 years, 35 - 44 years, 45 -54 years, 55 - 64 years,* and *65 years or older.*

GENERAL REMARKS

Income reference year: In the 'wealth' database, the income reference year may not be the same year as the point in time for wealth information. We keep the reference in our income items in line with the reference year for the wealth collection. Please consult METIS for further information.

Minimum number of observations: For estimates to be shown in 'trends' and 'scatter plots' a minimum number of 20 observations needs to be available. For estimates to be shown in 'distributions' a minimum number of 1000 observations needs to be present in order to calculate percentiles.

Country selection: Country selection is limited to countries where at least one estimate is available in DART. In the screens 'scatter plots', 'distributions', and 'maps' selection is restricted to one specific year. The selection tool follows an algorithm of +1 / -1 / +2 / -2 year(s) in order to keep various countries in the charts and tables, where the selected year is not available. Countries with data 3 years or more apart from the selected year are indicated with 'No data available' and dimmed grey in the selection menu.

Colouring: Colouring in the tool is pre-defined, following world regions. To change the colour, one can click on the ISO-3 legend on the left-hand selection panel.

Exporting and Importing: By clicking the 'export' button, graphs, tables, and documents can be exported as hard copy to your disk. Likewise, through exporting the .json criteria, a file can later on be recreated, when the .json criteria are imported.