Extension of the EUROMOD-ITT Tool
JRC Project no. JRC/SVQ/2020/OP/1373
FINAL REPORT

Bart Capéau[⋆,†], *André Decoster*[⋆], *& Duygu Güner*[⋆]
[*]Department of Economics KU Leuven
[†]ECARES, ULB

March 2022

Contents

1 Introduction **4**

2 Data **5**
2.1 Data availability . . . 5
2.2 Data preparation . . . 5
2.2.1 Harmonisation of the definition of units . . . 6
2.2.2 Harmonisation of reference period . . . 6
2.2.3 Harmonisation of existing variables and derivation of new variables . . . 7
2.3 Processing the expenditure data . . . 10
2.4 Data evaluation . . . 11

3 Imputation **16**
3.1 Methodology . . . 16
3.2 Implementation . . . 19
3.2.1 Definition of the 20 broad categories . . . 19
3.2.2 Explanatory variables . . . 19
3.2.3 Treatment of outliers and functional forms . . . 19
3.2.4 Saving and expenditures as derived variables . . . 20

4 Imputation results **22**
4.1 Data issues . . . 22
4.2 Regression results . . . 23
4.3 Ventile graphs . . . 25
4.4 Difference in correlation structure . . . 30
4.5 Macro validation . . . 32

5 Tax incidence parameters and changes in the ITT structure 38
5.1 Tax incidence . . . 38
5.1.1 Tax incidence parameters . . . 39
5.1.2 Implications for indirect tax revenues . . . 41
5.1.3 Welfare implications . . . 42
5.2 Implementation of the tax incidence parameter in ITT . . . 43
5.2.1 The expanded set of indirect tax parameters . . . 44
5.2.2 Calculation of the post reform prices . . . 48
5.3 Changes in the ITT output . . . 54
5.4 Addition of the baseline extension . . . 55
5.5 Other changes in the ITT module . . . 58
5.6 Updates to the Statistics Presenter . . . 61
5.7 Using ITT in STATA . . . 64

6 Application: greening food taxation 66
6.1 Introduction . . . 66
6.2 Baseline simulation . . . 67
6.3 Tax reform scenarios . . . 70
6.3.1 Top-down: a uniform carbon price in the agricultural sector . . . 70
6.3.2 Bottom-up: VAT rate differentiation according to CO2eq emission intensity . 71
6.3.2.1 Increasing VAT on the most polluting goods . . . 72
6.3.2.2 Combination of VAT increases with VAT reductions . . . 73
6.3.2.3 Revenue recycling . . . 73
6.3.2.4 Producer price adaptations modelled through the tax incidence parameter . . . 75
6.4 Reform simulation results . . . 78
6.4.1 Aggregate effects . . . 78
6.4.2 Distributional and welfare effects . . . 78
6.4.3 Sensitivity with respect to the tax incidence parameter . . . 82

7 Conclusion 86

8 Bibliography 87

Addendum: pending issues in ITT 88

Appendices 89

A Country names and codes 89

B Summary files of the imputations 90

C Definition of the 20 broad expenditure categories 94

D Covariates used in different countries 97

E Detailed list of excise goods 98

F Notes on ITT parameters 100

F.1 ITT consumer prices . . . 102

F.2 ITT excise parameters . . . 103

G Consumer price changes simulated by the CAPRI Model 106

H CAPRI-COICOP mapping of commodity aggregates 107

I Income and expenditure shares of CAPRI aggregates 108

J Summary statistics 109

J.1 Summary statistics for Austria . . . 109

J.2 Summary statistics for Bulgaria . . . 114

J.3 Summary statistics for Estonia . . . 118

J.4 Summary statistics for Croatia . . . 122

J.5 Summary statistics for Luxembourg . . . 126

J.6 Summary statistics for Latvia . . . 130

J.7 Summary statistics for Malta . . . 134

J.8 Summary statistics for the Netherlands . . . 137

J.9 Summary statistics for Sweden . . . 141

1 Introduction

This is the final report of the project Extension of the EUROMOD-ITT tool, JRC/SVQ/2020/OP/1373. It is the follow-up project of A new indirect tax tool for EUROMOD (JRC/SVQ/2018/B.2/0021/OC), and the first objective of the project is to extend the coverage of the EUROMOD Indirect Tax Tool (ITT). The previous version of EUROMOD-ITT, henceforth ITTv3 , enriched the EUROMOD input EU-SILC data by imputing information on expenditures, using EUROSTAT Household Budget Surveys (HBS) and enabled EUROMOD users to simulate the effects of changes in indirect taxes (VAT and excises). Currently, ITT runs for 18 countries and this project realises full EU coverage by the inclusion of Austria, Bulgaria, Estonia, Croatia, Luxembourg, Latvia, Malta, the Netherlands, and Sweden.

The second major contribution of the current project is to implement tax incidence parameters into the EUROMOD ITT. This new set of parameters allows for changes in indirect tax instruments not to be any more fully reflected in the new consumer prices, but being partially mitigated by letting producer prices to react to the indirect tax changes. Welfare and tax revenue implications for consumers of this producer price flexibility can now be assessed with the new ITT.

The report is structured as follows. Section 2 gives a summary of the data that were used in this project, along with a brief discussion of the data preparation process for the imputation of expenditure information from the HBS's into the SILC-data. The section also provides a detailed data evaluation for each country. Section 3 describes the imputation methodology and Section 4 evaluates the imputation results using the tools developed in ITTv3. The implementation of the tax incidence parameters and other changes in the ITT module are detailed in Section 5. Section 6 illustrates the use of ITT by means of a number of policy reform simulations aiming at greening indirect taxes on food. Section 7 concludes.

2 Data

This section provides an overview of the data collection, the preparation of the datasets for the final imputations. In the following subsections, we first provide information on the availability of datasets. Afterwards, we describe the data harmonisation process that SILC and HBS micro-datasets are subjected to. There are several inconsistencies present in expenditure information in HBS. Section 2.3 summaries these inconsistencies and outlines how these issues were addressed. Finally, we provide a detailed evaluation of the HBS and SILC datasets for each country.

2.1 Data availability

Since, the aim of the project is to extend ITTv3 for the remaining countries, the priority was to use the data coming from the year 2010 to ensure consistency with the current structure of ITTv3. Hence, we used 2010 HBS data when available and opted for the closest year when not. Similarly as was the case for ITTv3, we use HBS surveys harmonized by EUROSTAT whenever possible. However, EUROSTAT harmonized 2010 HBS datasets for Austria and the Netherlands were not available and the data of Luxembourg did not contain information on income. For Austria and Luxembourg, we used national HBS data, to which we gained access via the national statistic offices. For Luxembourg, we used 2013 national HBS as this was the first year where the income information was adequate for our imputation method. For the Netherlands, the use of national version of HBS was not a feasible option.[1] Since EUROSTAT harmonized HBS data were released in May 2021, and the Netherlands was included this time, contrary to what was the case for 2010, we opted for using EUROSTAT 2015 HBS dataset.

The project description foresaw in the imputation of expenditure information in the EUROMOD input EU-SILC data of 2010, and when available, also on the 2019 versions. For 8 countries we obtained the EU-SILC of the corresponding year 2010, and for Croatia, we obtained the EU–SILC of 2012. In September 2021, JRC published the EUROMOD version of 2019 EU-SILC datasets for Austria, Bulgaria, Estonia, Croatia, Luxembourg, Latvia, Malta and the Netherlands. As of 17.12.2021, Sweden data was still not released, hence, we performed the imputations on 2019 SILC dataset only for the other 8 countries. This report does however not discuss the imputation results for 2019 datasets. The country summary files for 2019 imputations will be submitted along with the other supplementary documents.

2.2 Data preparation

Prior to imputation, the two datasets, EU–SILC and HBS were harmonised for each country. Variable names and units as well as the scope and detail of information available, differ. Therefore, the two

[1] National HBS of the Netherlands is distributed by Statistics Netherlands (CBS). The application procedure of CBS is rather complicated and time consuming. Moreover, because of the services provided by CBS by default, the use of HBS is rather expensive. Finally, CBS imposes a set of restrictions on data use which further impedes the use of this dataset to impute expenditures into EU-SILC.

datasets are initially subjected to a harmonisation process. The harmonisation process is identical to ITTv3 and it follows the work procedures proposed by D'Orazio et al. (2006) and advocated by EUROSTAT (Leulescu and Agafitei, 2013; Lamarche, 2017; Serafino and Tonkin, 2017).

During this process, the common variables on household characteristics are identified and their names and units are standardised. Where necessary, information is converted from the individual to the household level for both datasets. Following the harmonisation process, the common variables of the SILC and HBS micro–data are listed and their distributions are compared for each country.

2.2.1 Harmonisation of the definition of units

Both datasets, HBS and SILC, contain information at the individual level (such as age, education level, gender etc.) and at the household level (such as household type, disposable income, the region the household resides in). The definitions of 'household' in both datasets draw on EUROSTAT's household concept.[2] Thus, there is no need for harmonisation of units. However, since the expenditure information in HBS is recorded, and thus will be imputed, at the household level, the information at the individual level need to be transformed into household level too.

We convert individual characteristics into household characteristics. When a person has a certain individual characteristic (age, say), the household she belongs to is characterised by the presence of a person with that characteristic. For instance, one can record the number of individuals in the household that fall under a certain group (such as the number of male household members, the number of household members with higher education and so on).

For the countries not belonging to the euro area in 2010, monetary variables in EU–SILC were converted into Euro, as the corresponding values in the HBS were reported in Euro.[3] The exchange rates are stored in the tab `Parameters` of the file `country parameters.xlsx` that is delivered together with the imputation code.

2.2.2 Harmonisation of reference period

While the HBS data officially refer to 2010, not all countries did collect a HBS in that year. EUROSTAT allows the national statistical offices of each country to deliver household budget survey data collected at most two years earlier than the reference year 2010. In addition, it was allowed to all countries to increase sample sizes of the delivered data by merging budget surveys of two preceding years. Information on how data were processed to be representative for the 2010 is scarce.[4]

The reference period for the income variables in the 2010 SILC datasets is 2009. Therefore, the

[2] According to the glossary of *EUROSTAT Statistics Explained*, this concept corresponds to 'either one person living alone or a group of people, not necessarily related, living at the same address with common housekeeping, i.e. sharing at least one meal per day or sharing a living or sitting room (`ec.europa.eu/Glossary:Household_social_statistics`).

[3] These countries are Bulgaria, Estonia, Croatia and Latvia

[4] For more information, see also the "Quality report of the 'Household Budget Surveys' 2010" downloadable from `ec.europa.eu/publications`.

income levels in the SILC 2010 datasets were adjusted using 2010 inflation rates. You can find these inflation rates in the tab `Parameters` of the file `country parameters.xlsx` that is delivered together with the imputation code. These inflation rates are downloaded from the HICP–inflation rates database of EUROSTAT (`ec.europa.eu/hicp`).

For Croatia, the closest year to 2010 for which a SILC dataset is available, is 2012. For that country, the incomes of SILC were rescaled such that the mean income of the SILC data equals that of the HBS of 2010.

2.2.3 Harmonisation of existing variables and derivation of new variables

Harmonisation of existing variables and derivation of new variables are the most vital steps in preparing the datasets for imputation. In order to achieve that, we first identified the common variables in both datasets.

1) Common variables at the household level.

- The region the households resides in: determined with respect to NUTS1 code.
- Disposable income.

  In SILC, total disposable household income is computed as follows.
  Gross personal incomes of all household members are added together. Components of gross personal income are:

  - gross employee cash or near cash income,
  - company car,
  - gross cash benefits or losses from self-employment (including royalties),
  - pensions received from individual private plans (other than those covered under ESSPROS),
  - unemployment benefits,
  - old–age benefits,
  - survivor' benefits,
  - sickness benefits,
  - disability benefits, and
  - education–related allowances.

  To this sum, gross income components at household level are added. These consist of:

  - income from rental of a property or land,
  - family/children related allowances,
  - social exclusion not elsewhere classified,
  - housing allowances,
  - regular inter-household cash transfers received,
  - interests, dividends, profit from capital investments in unincorporated business, and

- income received by people aged under 16.

From that the following amounts are subtracted:

- regular taxes on wealth,
- regular inter–household cash transfer paid,
- tax on income,
- the tax adjustments–repayment/receipt received or paid during the income reference period, and
- the social insurance contributions paid during the income reference period.

For the HBS, no information on the construction and components of the disposable income variable is available at Eurostat (Eurostat). We only know that it concerns monetary income. Therefore, we subtract from the SILC disposable income variable (`yds`) the fringe benefits for company cars (`kfbcc`). We have no variables on other non–monetary income possibly included in the SILC disposable income concept. We proceeded by assuming that the thus corrected concept of disposable income in SILC (`yds-kfbcc`) coincides with that of the HBS.

2) Common Variables at the individual level.

- Citizenship: national, other EU, non–EU.
- Gender: male, female.
- Education level (completed).
- Current education level.

  The two last variables are determined with respect to the ISCED1 code:

  a. less than primary education,
  b. primary education,
  c. lower secondary education,
  d. upper secondary education,
  e. post–secondary non–tertiary education,
  f. tertiary education.

- Employment status:

  a. manual worker except agriculture,
  b. worker except agriculture,
  c. self–employed person or agricultural worker ,
  d. unemployed,
  e. retired,
  f. other inactive,
  g. not applicable (legal age to work not attained).

- Age.

  The SILC dataset contains the exact age of each individual whereas the HBS dataset contains only the intervals in which the age of the individual falls. These intervals are defined as follows:

  a. between 0 and 14,
  b. between 15 and 29,
  c. between 30 and 44,
  d. between 45 and 59,
  e. 60 or more.

  Another set of age–related variables that are available in the HBS dataset are at the household level. They are as follows:

  a. number of household members between the ages of 0 and 4,
  b. number of household members between the ages of 5 and 13,
  c. number of household members between the ages of 14 and 15,
  d. number of household members between the ages of 16 and 24,
  e. number of household members between the ages of 16 and 24 who are students,
  f. number of household members between the ages of 25 and 64,
  g. number of household members between who are 65 years old or older.

  However, values of this second set of variables are not always available for all households in all countries. Therefore, in order to maintain as much as possible the same regression model across countries, we will use the first set of age–related variables.

The variables above are harmonised and used for deriving new common variables at household level. The final set of common variables that will be used for the imputation is as follows:

- disposable household income (EUR/year),
- the region the households resides in (NUTS1 level),
- whether the reference person is a farmer or not,[5]
- number of male household(HH) members over the age of 14,
- number of HH members under the age of 14,
- number of HH members between the ages of 15 and 29,
- number of HH members between the ages of 30 and 44,
- number of HH members between the ages of 45 and 59,
- number of HH members over the age of 60,
- number of disabled HH members,

[5] This is considered to be an indicator of an agricultural household.

- number of employed HH members,
- number of unemployed HH members,
- number of non–EU citizen HH members,
- number of pensioner HH members,
- number of student HH members over the age of 14,
- number of HH members with higher education.

A set of descriptive statistics of these variables for both the HBS and SILC datasets of each country, is contained in the sheet `descriptive statistics XX` of the file `summary XX.xlsx` (see Appendix B for the complete content of these `Excel` files) where `XX` stands for the country code (see Appendix A for the country codes).

2.3 Processing the expenditure data

There are several problems and inconsistencies with HBS expenditure data. First of all, not every broad expenditure category was disaggregated at the same level. Expenditures in the HBS data are recorded at four different levels of aggregation. For some categories, however, expenditures are not disaggregated until the fourth level. In most of such cases, the dataset contains a single category at the fourth level which carries the same name and value as the expenditure category at the third level of aggregation.

Furthermore, we observe three types of inconsistencies concerning different aggregation levels. In some cases, the expenditure on a category at the third level with only one sub–category is recorded at either the third or the fourth level, but not at both. For some households total expenditure on a broader category is larger than the sum of expenditures on its sub–categories and the reverse holds true for some other observations.[6]

To overcome these issues, we started by creating additional variables where needed and filling them with the values observed at the lower level of disaggregation. This provided the data a consistent structure. Since our objective is to impute HBS expenditure data at the most detailed, i.e. the fourth, level of disaggregation into SILC, we took the expenditures at this level as our base. The only exception to this are the cases where *all* recorded values for a category at a certain level of disaggregation are zero. In such cases, we chose the highest level of disaggregation that contains non–zero records as our base level.

For the categories at the third aggregation level for which have only one subcategory, there are cases where only one of these variables contains non–zero values while the other only contains zeros. We replace values at both levels with the maximum of both records. This means that if we observe a

[6]Readers may refer to Akoğuz et al. (2020) for a detailed discussion of these issues.

positive expenditure for that category at either of the disaggregation levels, we take this value for granted, and overrule the zero record at the other level with this value.

After these operations, we took the values recorded at the highest level of disaggregation as base. We construct expenditures at lower disaggregation levels as sums of the subcategories at a higher level of disaggregation. This ensures that we get a dataset where values are consistently aggregated, contrary to the current state of the EUROSTAT HBS data.

Households with negative expenditures on goods or good aggregates are excluded from the imputation process. We report the number of observations excluded in cells `3Y:4Z` of the sheet `descriptive statistics XX` of the file `summary XX.xlsx` which we provide for each country `XX` (see Appendix A for the country codes).

2.4 Data evaluation

This section evaluates the remaining nine countries on the basis of the availability of mutual variables in HBS and SILC datasets and the overlap between their distributions. Both HBS and SILC claim to be representative for the same population. Therefore, we expect them to perform similarly as it comes to making inference on the population distribution of overlapping variables. In order to evaluate whether this is indeed the case, we use weighted statistics and compare the distribution of common variables by constructing

- frequency tables for a set of categorical and discrete numerical variables,
- a Q–Q plot where the disposable income percentiles of both datasets are plotted against each other.

In the coming subsections, we first provide a summary of the information missing either in SILC or HBS dataset for each country. Thereafter, we discuss our inferences regarding the data quality of each country on the basis of the tables and plots described above. These tables and plots can be found in country specific summary statistics provided in Appendix J.

Austria

EUROSTAT does not publish the harmonised HBS surveys for Austria. Hence, we perform the imputations for Austria using the 2010 national HBS distributed by Statistik Austria.

The marginal distributions of common variables match extremely well for the HBS and SILC 2010 datasets of Austria. The only sizeable difference between the HBS and SILC samples is observed among the households where the household head is a farmer. SILC dataset indicates a relatively small share of households where the household head is a farmer (0.9%), while their share observed in HBS dataset is as high as 3.6%. This is not only higher than the SILC figure but it is also above the numbers we observe in other countries we investigate with Croatia being the exception (7.2%).[7]

[7] The inconsistency between HBS and SILC figures might be purely due to variable definitions. National HBS does not allow us to differentiate between household heads working in agriculture and forestry. The inclusion of

The Q–Q plot follows the 45 degree line quite well while we observe that HBS dataset contains many records with exactly same income values. Starting from the 8^{th} decile, the zigzag shape becomes more visible. Income information gathered from SILC is relatively higher for the highest 3 deciles of the income distribution. The share of households where at least one household member has a higher education degree is 2.2 percentage point higher in SILC (25.6%) than in HBS (23.4%). Moreover, there is also a 2.8 percentage point gap between SILC and HBS with respect to the share of households where at least one household member is employed. The last two observations might be the underlying reasons of the slightly higher income levels observed in SILC dataset.

Bulgaria

SILC dataset of Bulgaria does not provide information on the reference person being a farmer. Hence, this covariate is ignored when discussing the quality of Bulgaria datasets. The marginal distributions of other common variables match poorly for several important variables for the HBS and SILC 2010 datasets of Bulgaria. These variables include *household size*, *household type*, *number of employed household members* and *number of pensioner household members*. Moreover, the difference in distribution of disposable incomes according to SILC and HBS is substantially larger than in most other countries. The Q–Q plot has a much higher slope than 45 degree starting from the first decile.

Estonia

Estonia is not divided at NUTS–1 level either. Moreover, neither SILC nor HBS dataset of Estonia do have information on reference person being a farmer, hence, this variable is excluded from the list of covariates.

The marginal distributions of common variables match extremely well for the HBS and SILC 2010 datasets of Estonia. However, the distributions of net income are highly different from each other. The zigzag shape of the Q–Q plot and the parts where it advances vertically are particularly noticeable. This occurs since the HBS dataset contains many records with exactly same income values indicating that the households did not report their exact incomes but rather rounded values or were offered income brackets to respond. Moreover, the slope of the Q–Q plot is substantially higher than 45 degrees.

Croatia

For Croatia, we could not obtain the 2010 SILC micro–data and we used 2012 SILC instead.[8] The 2012 SILC dataset does not provide information on the reference person being a farmer. Therefore, this variable is excluded from the list of covariates. Moreover, Croatia is not divided at NUTS–1 level, hence, region will not be used as a covariate.

self–employed household heads working in forestry might be boosting the numbers observed in HBS.

[8] When imputing expenditure information into SILC–data from an HBS dataset of another year we rescaled incomes of the SILC–data.

Most likely due to this two year period between HBS and SILC datasets, the marginal distributions of some variables do not match very well for Croatia. The gap hovers around 2 to 4 percentage points for *household size*, *number of household members between 0 and 4 years old* and *household type* variables. Moreover, we observe a significant HBS–SILC difference in economic activity measures:

– households where none of the household members are employed account for 40.5 percent of the households in 2010 HBS sample while this number is as high as 45 percent in 2012 SILC,
– the share of households where none of the household members are unemployed is 79.8 percent in 2010 HBS and this number is only 72.2 percent in 2012 SILC.

This is not surprising given the economic stagnation in Croatia during 2010-2012 period. Respondents of the 2012 SILC survey report higher income levels than the respondents of the 2010 HBS survey despite the constant decrease in real GDP in Croatia from 2009 to 2012. That being said, the gap observed in Q–Q plot is rather narrow and it only widens above the 7^{th} decile.

Luxembourg

Statistical office of Luxembourg, STATEC, started collecting income information as a continuous variable only in 2011. Previously, income information was recorded in brackets. In 2011, STATEC was still allowing the respondents to declare their income in brackets and as a result, the income information in the 2011 dataset contained both discrete and continuous values. Moreover, the income variable also had a considerable amount of missing values. Hence, EUROSTAT did not report income information in the harmonised version of 2010 HBS dataset of Luxembourg. Since income is a crucial information, we contacted STATEC to get access to the national HBS version where income data is available. The earliest HBS dataset where income information is available with a consistent variable definition and negligible amount of missing values is the year 2013. The national HBS of 2013 is a dataset covering the pool of the annual HBS surveys conducted in 2012, 2013 and 2014, where the income and consumption information is published in 2013 prices.

Luxembourg is not divided at NUTS–1 level and marginal distributions of the rest of the common variables match rather well in the 2013 HBS and 2010 SILC samples. That being said, we observe a slight variation in the share of single–person households; in the 2013 HBS dataset the share of single–person households is 33.3 %, 3.9 percentage point higher than their share in 2010 SILC (29.4%). This is reflected in the household type figures as well. Nevertheless, this trivial increase is most likely driven by the sampling rather than a possible demographic shift. The Q–Q plot follows the 45 degree line almost perfectly.

Latvia

The 2010 HBS survey of Latvia does not provide information on non-EU household members, hence, we could not generate the variable *number of non-EU household members*. Moreover, the HBS survey does not report household type either. Normally, one can use the HBS household member dataset to construct the household type variable. However, in the absence of educational attainment information, it is impossible to identify dependent children, and hence, the household type. On the other hand, Latvia SILC does not have information on reference person being a farmer. Finally, Latvia is not divided at NUTS–1 level.

The marginal distributions of common variables match well in general for the HBS and SILC 2010 datasets of Latvia. However, the distributions of disposable income are quite different from each other. The Q-Q plot deviates upwards from the 45 degree line substantially after the 1st quantile. Additionally, the match between the frequency tables of the variable 'household type' is not very strong.

Malta

The 2010 HBS survey of Malta does not provide information on the education level of the household members, hence, we could not control for households educational endowment. Moreover, the HBS dataset only provides limited information on the economic activity of the household members which further hinders the construction of a pensioner variable. Malta is not divided at NUTS–1 level. Neither SILC nor HBS dataset of Malta do have information on reference person being a farmer, hence, this variable is excluded from the list of covariates.

Apart from this missing information problem, the marginal distributions of common variables match quite well for the HBS and SILC 2010 datasets in general. One exception is the age composition of the households where SILC sample suggests a shift towards 65 years and older compared to HBS sample. On the other hand, the disposable income distributions match extremely well. Despite a minor upward bias starting approximately at the fourth decile, the Q–Q plot mostly follows the 45 degree line until the last decile. It deviates slightly upwards in the last decile indicating that SILC contains higher disposable income values, however the difference is not substantial.

The Netherlands

Due to aforementioned problems (Section 2.1), we did not request access to the 2010 national HBS dataset for the Netherlands. Instead we used the EUROSTAT harmonized HBS survey of 2015. The 2015 HBS survey of the Netherlands does not provide information on non-EU household members, hence, we could not generate the variable *number of non-EU household members*. Moreover, the HBS survey does not report education level either. Finally, the EU-SILC data does not contain region information and information on the reference person being a farmer or not.

Household age composition matches reasonably well, despite the five year gap between the SILC and HBS datasets. However, there is a mismatch of the economic activity figures on *number of employed household members* and *number of unemployed household members*, and, even more so, in *number of pensioner household members*. Households where none of the household members are employed account for 40 percent of the households in the 2015 HBS sample while this number is only 35.9 percent for 2010 SILC. The share of households where no one is unemployed is 90% in 2015 HBS and 95.9% in 2010 SILC. The gap is 10 percentage points for the share of households where no one is pensioner (69.6% in 2015 HBS and 79.8% in 2010 SILC). The Q–Q plot of disposable income is parallel to, but lies slightly above the 45 degree line.

Sweden

The 2010 HBS survey of Sweden does not provide information on the education level and economic activity of the household members, hence, we could not generate the associated variables.[9] Moreover, HBS survey does not report household type either. Normally, one can use the HBS household member dataset to construct the household type variable. However, in the absence of educational attainment information, it is impossible to identify dependent children, and hence, the household type. Neither the SILC nor the HBS dataset of Sweden do have information on reference person being a farmer, hence, this variable is excluded from the list of covariates. SILC Sweden further does not provide information on region.

Similar to the Malta data, Sweden SILC sample suggests a shift towards 65 years and older compared to HBS sample. The Q–Q plot of disposable income follows the 45 degree line quite well, with a slight upward bias which grows with income level but still remains rather negligible.

[9] These variables are *number of employed household members*, *number of unemployed household members*, *number of pensioner household members*, *number of disabled household members* and *number of household members with higher education*.

3 Imputation

This section outlines the imputation methodology developed in Akoğuz et al. (2020). The main idea behind this method is to use regressions of broad expenditure categories on the common variables (household characteristics) to obtain a match between the households in HBS and SILC datasets. The method uses the fitted values of the regressions as inputs in the distance function, and then matches the households with the smallest distance.

First, we employ a two-step regression model; a binary regression which determines the probability of positive expenditures for a given expenditure category and then, for each category, a linear regression of positive expenditures, expressed as shares of disposable income, on the same variables as the binary counterpart. Once the two–step regression is estimated, income shares of the twenty broad categories are fitted for each household, both in the source and recipient data.. Afterwards, these fitted values are fed into a distance function in order to calculate the distance between all possible household pairs. The model uses the Mahalanobis distance function to compute the distances between all possible household pairs and identifies matches. A household h in the source dataset is matched to a household g in the recipient dataset with the smallest distance to h. The model then assigns the observed values of household h, income shares of expenditures at the lowest level of commodity aggregation, as imputed values for household g's expenditures.

3.1 Methodology

We now give a step–by–step description of our imputation method.

1) A household h's expenditure on a good i in the source dataset (the HBS in our case), indexed by s), denoted by e_{shi}, is converted into a share, w_{shi}, of disposable income (y_{sh}):

$$w_{shi} = \frac{e_{shi}}{y_{sh}}, \qquad i \in \mathcal{N}, \tag{1}$$

with $\mathcal{N}$ the set of indices of goods at the most detailed level in the HBS.

2) These income shares of expenditures on detailed goods are aggregated under broader categories. These categories should be big enough to reduce the occurrence of zeros due to infrequent expenditures. We index these categories by $A, B, \ldots$. Consequently, the indices $A, B, \ldots$ denote non–overlapping and non–empty subsets of $\mathcal{N}$, denoted by $\mathcal{N}_A, \mathcal{N}_B, \ldots$, whose union equals $\mathcal{N}$. Thus, the income share of expenditure category X for $X = A, B, \ldots$, say W_{shX} , equals:

$$W_{shX} \equiv \sum_{i \in \mathcal{N}_X} w_{shi}. \tag{2}$$

3) The purpose is to develop a multidimensional Predictive Mean Matching (PMM) method by constructing a distance function that takes values of income shares of aggregate expenditure categories fitted from a regression model as inputs.

   Some of the broad categories may still contain a significant number of zero observations. For a sufficiently high level of aggregation, we can consider these as true zeros, i.e. they are not a consequence of the infrequent expenditures problem. Therefore, we propose a two–step approach for modelling these aggregates. The probability that a household exhibits positive expenditures on commodity aggregate X ($X = A, B, \ldots$) is modelled by a binary model, more specifically, a probit model, using the common variables in the source and destination data as explanatory variables. Formally,

$$\Pr\left(W_{shX} > 0\right) = 1 - \Phi\left(-\boldsymbol{\gamma}_X' \mathbf{x}_{sh}\right) = \Phi\left(\boldsymbol{\gamma}_X' \mathbf{x}_{sh}\right), \tag{3}$$

where Φ denotes the standard normal distribution function, $\mathbf{x}_{sh}$ is the vector of values of explanatory variables for household h in the source dataset s, and the vector $\boldsymbol{\gamma}_X$ contains parameters to be estimated.

Next, an ordinary continuous regression model is formulated for assessing the relation of *positive income shares of broad expenditure categories with the common variables*:

$$W_{shX} = \boldsymbol{\beta}_X' \mathbf{x}_{sh} + \varepsilon_{hX}, \qquad \text{for } W_{shX} > 0. \tag{4}$$

As we are only after correlations, no sample selection correction terms are added to this equation.

The model is estimated on the source dataset, and estimated parameters are denoted with a hat: $\widehat{\boldsymbol{\gamma}}_X$ for the probit models, and $\widehat{\boldsymbol{\beta}}_X$ for the linear regression models.

4) Using the estimated models, values are fitted for the income shares of expenditures on the broad categories $A, B, \ldots$, for all households in *both* the source *and* the recipient datasets, indexed by s, respectively r. These fitted values are denoted by $\widehat{W}_{dhX}$ and defined as follows:

$$\widehat{W}_{dhX} = \Phi\left(\widehat{\boldsymbol{\gamma}}_X' \mathbf{x}_{dh}\right) \widehat{\boldsymbol{\beta}}_X' \mathbf{x}_{dh}, \qquad \text{for } d = s, r, \tag{5}$$

where the first factor on the RHS corresponds to the estimated probability that household h has positive expenditure on the aggregate category X, while the second factor corresponds to the estimated income share that household h spends on aggregate expenditure category X, given that this share is positive.

Before we construct a distance function on the basis of these fitted values, we want to assess the extent to which the estimated two–step model is able to explain households' expenditure behaviour. Thereto we construct a pseudo–R^2 value:

$$\text{pseudo}-R^2\left(X\right) = 1 - \frac{\sum_h \left(W_{shX} - \widehat{W}_{shX}\right)^2}{\sum_h \left(W_{shX} - \overline{W}_{sX}\right)^2}, \tag{6}$$

where $\overline{W}_{sX} = \sum_h W_{shX}/H_s$, and H_s is the number of households in the source dataset, and sums run over all observations in the source dataset.

Only categories exhibiting a 'reasonable' fit according to this pseudo–R^2 are retained as inputs for calculating the distance between two households. Including variables with a low fit would imply comparing households on the basis of only a small fraction of their true expenditure behaviour. So, we could consider two households to be close to each other on that basis, while in reality they might fall far apart, or *vice versa*. We discuss the choice of the threshold level for the pseudo–R^2's in Section 3.2.3.

5) Denoting a vector of fitted shares retained as input for the distance by $\widehat{\mathbf{W}}_{dh} \equiv \left(\widehat{W}_{dhA}, \widehat{W}_{dhB}, \ldots\right)$ $(d = s, r)$, and using the Mahalanobis distance metric, the distance between a household h in the source data, and a household g in the recipient data is defined as:

$$dist\,(h, g) = d\left(\widehat{\mathbf{W}}_{rg}, \widehat{\mathbf{W}}_{sh}\right) = \sqrt{\left(\widehat{\mathbf{W}}_{rg} - \widehat{\mathbf{W}}_{sh}\right)' \Sigma^{-1} \left(\widehat{\mathbf{W}}_{rg} - \widehat{\mathbf{W}}_{sh}\right)}, \tag{7}$$

where Σ here stands for the variance covariance matrix of the vector $\widehat{\mathbf{W}}$, using data from both source and recipient.

6) A match for household g in the recipient dataset is defined as the household h in the source dataset that has the smallest distance to household g, where this distance is measured in terms of Equation (7).

7) For each match (h, g), income shares of expenditures *at the most detailed level* of good disaggregation $i \in \mathcal{N}$ for the recipient household g, are obtained from the corresponding values of the source household h:

$$w_{rgi} = w_{shi}. \tag{8}$$

This method successfully imputes values at a detailed level of aggregation by using the observed values rather than the fitted ones. Moreover, the method also exploits the relation between household characteristics (that is, the explanatory variables in the regression) and expenditures (dependent variables) in the dataset. Nevertheless, there are certain limitations to this method.

First of all, the method is not able to impute expenditures for households with non–positive income. We use logarithm of income as a covariate in the two step regression model.[10] Since the logarithm of a non–positive number is unidentified, the two step model omits these households and the fitted values are never calculated for them. Without fitted values, the method lacks the necessary information to calculate the distance function and hence, the method fails to match these households.[11]

Furthermore, in order to make reasonable predictions, the model heavily relies on expenditures being sufficiently aggregated at reasonably broad categories. Since the match function only uses information coming from the aggregated categories, there is no guarantee that households in the source dataset will be matched to household with similar characteristics in the recipient dataset. The model actually might match two households with very distinct characteristics, and when this

[10] Section 3.2.2 details the set of explanatory variables.

[11] In fact our approach makes only sense for households with a sufficiently high and positive income. Expenditure behaviour of agents with negative or extremely small positive income, do not fit into our model. Indeed, the concept of an income share in terms of which our model is specified, makes not much sense in case of negative incomes, is not defined in case of zero incomes, and may yield extreme values in case of incomes close to zero.

happens, there is no reason to expect these two households to behave similarly when allocating their budget for broad categories to the commodities belonging to those categories.

3.2 Implementation

3.2.1 Definition of the 20 broad categories

The HBS expenditure information, which is available up to the COICOP level 4 detail (5 digits, 193 subclasses), is aggregated into 20 broad expenditure categories. To each of these 20 broad expenditure categories, the two-step regression model is applied separately for each country. Appendix C provides a full list of these 20 categories and their corresponding COICOP subcategories.

For each country we provide sample summary statistics on expenditures and income shares of expenditures on these broad categories, total expenditure and saving (that is, disposable income minus total expenditures) in cells `34R:60AN` of the sheet `descriptive statistics XX` of the file `summary XX.xlsx` with summary tables of the imputation which we provide for each country (`XX` stands for the country code, and the legend of country codes can be found in Appendix A).

3.2.2 Explanatory variables

Our basic regression model correlates income shares of expenditures on the 20 broad categories defined in the previous section, with household characteristics. Even though we do not give any structural interpretation to the regression model, the selection of covariates is very much inspired by the specification of Engel curves. More specifically, a third degree polynomial in the log of incomes, and a rich set of household composition characteristics were included, containing detailed information on the number of household members in different socio–demographic groups, such as gender, labour market status, and age.

A list of all potential covariates can be found in Appendix D. A covariate is excluded from the regression models for a specific country when the information on the variable is absent in either SILC or HBS datasets, or not relevant for a particular country. Appendix D contains more detailed information on which variables are excluded for which country. This information can also be found in cells 1D:21F of the sheet `regression results XX` of the file `summary XX.xlsx`). Note that, for each country, the covariates of the probit and linear regression are the same.

3.2.3 Treatment of outliers and functional forms

The linear part of our regression model (Equation 4) is known to be sensitive to outliers. Such outliers in the dependent variable (income shares of expenditures) might occur especially in case of incomes close to zero, yielding extremely high income shares. In a first step, we removed observations from the source dataset with extremely low incomes (*i.e.* lower than or equal to 100 € per year). Still then, outliers in the income shares occur. In what follows, we treat an observation on an income share as an outlier when it surpasses 500%. When that is the case, a linear model and estimation method (such as OLS) will try to fit the outlier values well, and will basically use no information of the bulk of the observations for which the model was constructed. To deal with the outlier problem we decided to do the imputation for all countries on the basis of a log share transformation of the shares as a dependent variable in the linear regression equations (4). Our final regression equations thus become:

$$\ln W_{shX} = \boldsymbol{\beta}'_X \mathbf{x}_{sh} + \varepsilon_{hX}, \qquad \text{for } W_{shX} > 0. \tag{9}$$

Notice that the outlier problem not only affects the estimates, but also the R^2 of those regressions.[12] As explained in the fourth step of the description of the imputation methodology in Section 3.1, the pseudo–R^2 of fitted expenditures on broad categories serves as a measure of how well the regression model (the linear part and the probit together) can explain the true expenditure behaviour for a given category. This measure is equally sensitive to the outlier fitting problem just explained for regular R^2's in a linear regression.[13] We therefore calculated and report the pseudo–R^2's excluding outliers.

Since we used a log share transformation, the definition of the pseudo–R^2 has to be slightly adapted. Indeed, the concept of fitted log share is ill defined as it would amount to the estimated probability of positive expenditures times the fitted value of the log share conditional on positive expenditures, plus the estimated probability of a zero share times the log of zero, which is minus infinity.

We therefore converted the fitted values of the linear regression (Equation 9), which are logs of shares, back into shares by taking an exponent. That is, $\widehat{W}_{dhX}$, defined in Equation (5), becomes:

$$\widehat{W}_{dhX} = \Phi\left(\widehat{\boldsymbol{\gamma}}'_X \mathbf{x}_{dh}\right) \exp\left(\widehat{\boldsymbol{\beta}}'_X \mathbf{x}_{dh}\right), \qquad \text{for } d = s, r. \tag{10}$$

We do realise that this is a biased estimate of the shares. Yet, we did not attempt to correct for this.

Finally, in the distance function we only retained fitted values of aggregates with pseudo–R^2 above or equal to 0.1. In cells 1B:21B of the sheet `regression results` of the files `summary XX.xlsx` with summary information of the imputation for each country, the pseudo–R^2 values for each of the 20 broad categories' regression models are reported. Categories that surpass the 0.1 threshold, and for which the fitted value thus enters the distance function, are highlighted in green. The same sheet also includes more detailed information of the regression results for the 20 probits and linear regressions.

3.2.4 Saving and expenditures as derived variables

In our data we observe disposable income in both source and recipient data, respectively denoted by y_{sh} for a household h in the source data and y_{rg} for a household g in the recipient data. The

[12] If a variable accounting for much of the variance is fitted well, a high R^2 will result, even though the regressions might result in a bad fit for the bulk of the observations. *Vice versa*, a bad fit for the outlier, might cause a low R^2 for a regression that fits the bulk of the observations quite well.

[13] In case there are no zero observations and there would have been no log transformation of the dependent variable in the continuous linear regression (Equation 4), the R^2 of the continuous regression and the pseudo–R^2 defined in Equation (6) would coincide.

source dataset also contains observations on expenditures on specific detailed goods, indexed by $i, j, \ldots \in \mathcal{N}$. These expenditures for household h, denoted by e_{shi}, are non–negative and sum to total expenditures, denoted by E_{sh}. Saving, S_{sh}, is defined as the difference between disposable income and total expenditures. So, the source dataset contains information on:

$$E_{sh} \equiv \sum_{i \in \mathcal{N}} e_{shi}, \tag{11}$$

$$S_{sh} \equiv y_{sh} - E_{sh}. \tag{12}$$

We can then define shares of detailed and total expenditures in disposable income as

$$w_{shi} = \frac{e_{shi}}{y_{sh}}, \tag{13}$$

$$W_{sh} = \frac{\sum_{i \in \mathcal{N}} e_{shi}}{y_{sh}} = \frac{E_{sh}}{y_{sh}}. \tag{14}$$

Consequently, the share of saving in disposable income equals:

$$\theta_{sh} = \frac{S_{sh}}{y_{sh}} = 1 - W_{sh}. \tag{15}$$

For a matched couple (h,g), we imputed the income shares of detailed expenditures and saving of the household h in the source dataset, into the record of household g in the recipient. Detailed expenditures, saving and total expenditures for the recipient household g can then be calculated as:

$$e_{rgi} = y_{rg} w_{shi} = y_{rg} \frac{e_{shi}}{y_{sh}}, \tag{16}$$

$$S_{rg} = y_{rg} \theta_{sh} = y_{rg} \frac{S_{sh}}{y_{sh}}, \tag{17}$$

$$E_{rg} = \sum_{i \in \mathcal{N}} e_{rgi} = y_{rg} \frac{E_{sh}}{y_{sh}}. \tag{18}$$

This method distributes the difference in disposable income between a pair of matched household across all expenditures and saving by rescaling the observed values with the ratio of the disposable income of household g to that of household h as shown in equations (16), (17) and (18). A drawback is that, when a household with low income and negative saving is matched to a household with higher income, the imputed dis–saving level of the latter will be even higher.

4 Imputation results

This section presents the results of the statistical matching procedure for the nine remaining countries. First, we discuss the data issues that emerged during imputations. We briefly describe the issues and how we tackled them. Then, we provide the list of expenditure categories whose fitted values enter as inputs in the distance function. We evaluate the imputation results by means of the tools developed within ITTv3 (Akoğuz et al., 2020). The first tool is the ventile graphs, where we discuss the mean income shares of expenditures per income ventile, then we briefly comment upon the performance of the imputation in terms of safeguarding the correlation structures between variables.

For each country, we provide an Excel file `summary XX.xlsx` (where `XX` stands for the country code defined in Appendix A) which contains all outputs discussed in this report. A detailed overview of the content of these files is given in Appendix B.

Finally, this section also discusses the comparison of estimated total expenditures and tax revenues on the basis of the imputation in the SILC with similar estimates from the HBS (for expenditures) and national accounts (for expenditures and tax revenues).

4.1 Data issues

Before discuss the imputation output, we outline the data issues needed to be tackled before finalizing the imputations.

- **Adjusting monetary variables for Austria HBS:** The field work for the Austria 2010 HBS is done in 2009 and 2010. Nevertheless, the expenditures and income information collected in 2009 are not adjusted in the raw data. We inflate the values for these observations on expenditures and income to 2010 nominal values, using the CPI.[14]
- **Missing information on important variables:** The HBS's of Estonia and Sweden did not provide detail on energy consumption (EUR_HE045). However, this information is rather important for simulation of excise revenues.[15]
  In the Estonia HBS, all observations were zero for EUR_HH04521 and EUR_HH04522 while EUR_HH0452 has positive values for 20% of the households. Since different excise rates apply to the sub–categories of EUR_HH0452, we split the expenditures reported under EUR_HH0452 over EUR_HH04521 and EUR_HH04522, using the mean shares observed in two neighbouring countries, namely Latvia and Lithuania.

[14] Another important note on Austria HBS is that the expenditure information in HBS data corresponds to monthly consumption. While this does not pose a great challenge, since we are working with income shares rather than levels, it might still have implications for imputations.

[15] The HBS COICOP category EUR_HH045: Electricity, gas and other fuels contains the following sub–categories:
EUR_HH04511: Electricity
EUR_HH0452: Gas, EUR_HH04521: Town gas and natural gas, EUR_HH04522: Liquefied hydrocarbons
EUR_HH04531: Liquid fuels
EUR_HH04541: Solid fuels

Similarly, details on EUR_HH045 were missing in the Sweden HBS as well. We divided the aggregate expenditures EUR_HH045 over its sub–categories proportional to the mean consumption levels we received from the National Statistical Office of Sweden.

– **Negative expenditures:** Similarly as was done for the countries that were the subject of ITTv3, HBS households recording negative expenditure on at least one commodity are excluded from the imputation process. For most of the countries, none of the households report negative expenditures (Austria, Bulgaria, Croatia, Luxembourg and Malta). For some other countries exclusion of households with negative expenditures resulted in losing less than 5% of the observations (71 households in Estonia and 11 households in Latvia). However, for Sweden and the Netherlands the numbers were higher. In Sweden, in total 173 households report negative expenditures corresponding roughly to 10% of the whole sample. In the 2015 HBS of the Netherlands the number of such observation were suspiciously high: 5335 households out of 14408 (or *37%*) reported negative expenditures in one or more consumption items and negative expenditures were scattered across different commodities.[16]

– **Low household disposable income:** For observations with very low household disposable income, the functional form of our two stage estimation could lead to very high fitted expenditure shares in the SILC data (order of magnitude 10^{20}). This causes the variance-covariance matrix in the Mahalanobis distance function to be badly scaled. This was indeed the case for Sweden; 2 observations with incomes lower than 10 euros (per year) were distorting the computation of the distance function. Hence, for Sweden, we imposed the sample selection rule in the HBS of minimally 100 euros of disposable also on the SILC dataset. In total, 6 households whose annual household disposable income was less than 100 euros were thus excluded from the imputation procedure for Sweden.

4.2 Regression results

For each country, summaries of both, the probit and linear regression models, for each broad expenditure category, can be found in the sheet `regression results XX` of the file `summary XX.xlsx`.

In order to evaluate the goodness of fit of our two–step regression model as a whole, we look at the pseudo–R^2, as defined in Equation (6) of Section 3.1 and Equation (10) in Section 3.2.3. An expenditure category is selected as input in the distance function if the pseudo–R^2 value of the corresponding estimates is higher than 0.1. Table 1 shows for each country the pseudo-R^2 values for the categories whose fitted values are used as input for the distance function. The last four columns of Table 1 give the number of times each category entered the distance function and the mean pseudo-R^2 value conditional on being selected, for both ITTv4 (columns 10 and 11) and ITTv3 (columns 12 and 13).

The regression models for five expenditure categories, namely *food and non–alcoholic beverages*,

[16] For the countries covered by ITTv3, the number of households with negative expenditures was 275 in Denmark, 165 in France, 25 in Czechia, 10 in Finland and 5 in Slovenia. For all other countries we did not encounter this problem.

Table 1: Pseudo–R^2 values of good categories surpassing the 0.1 threshold (per country)

												ITTv3	
Category	**AT**	**BG**	**EE**	**HR**	**LU**	**LV**	**MT**	**NL**	**SE**	**#**	**mean Pseudo–R^2**	**#**	**mean Pseudo–R^2**
Food and non-alc beverages	0.34	0.51	0.38	0.43	0.23	0.48	0.53	0.43	0.44	9	0.42	17	0.45
Utilities	0.29	0.36	0.28	0.50	0.27	0.51	0.40	0.64		8	0.41	15	0.41
Communications			0.21	0.22	0.16	0.24		0.56	0.32	6	0.28	14	0.36
Tobacco	0.16		0.11			0.39	0.47		0.12	5	0.25	6	0.16
Housing and rental	0.28				0.31		0.11	0.54	0.51	5	0.35	11	0.24
Insurance								0.80	0.32	2	0.56	4	0.43
Culture and recreation				0.11					0.48	2	0.29	8	0.21
Clothing and personal items		0.23								1	0.23	1	0.15
Public transportation						0.16				1	0.16	6	0.24
Alcoholic beverages						0.15				1	0.15	5	0.29
Personal care				0.13						1	0.13	8	0.22
Restaurants												5	0.19
Health and care												4	0.13
Private transportation												4	0.29
House goods and services												3	0.16
Education												2	0.17
House durables												0	-
Other												0	-
Travelling and holiday												0	-
Vehicles												0	-

utilities, *communications*, *tobacco* and *housing and rental* seem to perform better in comparison to the rest of the categories.

A peculiar result is the high pseudo–R^2 value for the regression of the *insurance* category in the Netherlands. Basic health insurance, *basisverzekering* is universally mandatory in the Netherlands. Hence, unlike in other countries, all households in the 2015 Dutch HBS report positive expenditures on the item *EUR_HE1253: Insurance connected with health*. Although, the basic insurance policy is rather standard, there are different types of policies allowing the individuals to chose the optimal coverage depending on their preferences.[17] Nevertheless, the health insurance market is still a very homogeneous market relative to many other commodity markets within the scope of this project. Although, the payments are issued on a monthly basis, the contract is annual and consumers rarely switch between providers and policies. The high number of positive observations and the characteristics of this particular market explain the high explanatory power of our two step estimation procedure for insurance expenditures in the Netherlands.

Similar to the results of ITTv3, the pseudo–R^2 values for the categories that are expected to suffer the strongest from the infrequent expenditures problem, namely *travel and holidays*, *house durables*, and *vehicles*, do not exceed the 0.1 threshold, and therefore do not enter the distance functions. The same holds true for expenditure categories which are more prone to the zero expenditures problem

[17] The compulsory basic insurance costs approximately 100 euros per month and supplemental insurance adds on between 2 to 150 euros per month depending on the policy coverage.

with the exception of emphTobacco. The pseudo–R^2 value of *Tobacco* exceeds the threshold for more than half of the countries. As was the case for ITTv3 countries, The category of *other* which gathers all expenditure variables that fall under none of the other categories, is never used in the distance function. This result is once again in line with our expectations as it is not plausible to expect a strong relation between the expenditures made on such a diverse collection of goods and services, and household characteristics.

4.3 Ventile graphs

In this section, we evaluate the ventile graphs of mean income shares of expenditures against income for each country in the frame of the three performance aspects developed in Section 3.1 of Akoğuz et al. (2020), *i.e.* similarity of patterns, similarity of expectations conditional on income, and unbiasedness of the imputations. To recall, these graphs plot mean imputed (SILC) and observed (HBS) income shares of expenditures on the 20 broad categories against mean disposable income per ventile. We examine the imputation results for all broad categories, mentioning the categories that perform the best and the worst in terms of all three aforementioned aspects. We also discuss the derived imputations of total expenditure, defined as the sum of income shares of all expenditures, and saving (one minus the income share of total expenditure).[18]

Similar to the findings of Akoğuz et al. (2020), imputations for the categories of *food and non–alcoholic beverages* and *utilities* perform by far the best across all nine countries. This can be explained by the fact that these expenditures do not contain many records with zero expenditures.

Austria

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

Imputations seem to be unbiased for most categories, however there are some exceptions. The imputations of *housing and rental* is upward biased for low income ventiles while the opposite holds true for high income ventiles. The imputation for *food and non–alcoholic beverages*, *utilities*, *communication*, *culture and recreation*, *personal care* and *tobacco* are downward biased for higher income ventiles.

The similarity between the average observed and imputed income shares of expenditure conditional on income is quite high for almost all categories, with only few exceptions, namely *housing goods*

[18] As ventile graphs demonstrate, the first income ventile might contain substantial outliers where households combine extremely low income levels with regular expenditure levels. This remains the case even when we exclude households with disposable incomes lower than 100 euro per year. These outliers may have a serious impact on ventile means, and so might easily lead to large difference in the imputed and observed means for a given ventile, depending on whether or not, or how frequently, SILC households are matched with such outliers. Therefore, while comparing imputation performances for different categories, we will not put emphasis on the first ventile. The category *other* will not be taken into consideration as this variable stands for the total expenditure on a very diverse collection of goods and services.

and services, *travelling and holiday*.

The categories with best performing imputations are *food and non–alcoholic beverages* and *utilities*. The categories with worst performing imputations are *health and care* and *house durables*.[19]

The imputation of income shares of total expenditure and saving perform fairly good in terms of preserving the observed relation between net household income and income share of expenditures. However, the imputations of saving (total expenditures) seem to be biased upwards (downwards) for the first half of the income distribution and downwards (upwards) in the second half. The similarity between the average of observed and imputed income shares of total expenditure conditional on income is very good.

Bulgaria

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

There are several categories for which imputations seem to be biased either generally or for a certain part of the distribution. There is an upward bias over the whole income distribution in *house durables* expenditure imputations. We observe a systematic bias in *travelling and holiday* as well, imputations for this category tend to overshoot the observed values considerably for most of the income ventiles. On the other hand, for *utilities* a slight downward bias is observed only in high income ventiles.

The similarity between the average of observed and imputed income shares of expenditure conditional on income is fairly good for approximately one quarter of the categories. There are only two categories where the average observed and imputed income shares of expenditure are rather similar, namely *housing and rental* and *housing goods and services*. For the rest of the categories the overlap between average observed and imputed values is mediocre.

The categories with best performing imputations are *food and non–alcoholic beverages* and *utilities*. The categories with worst performing imputations are *education* and *house durables*.

The imputation of income shares of total expenditure and saving seem to be unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. The similarity between the average observed and imputed income shares of total expenditure conditional on income is fairly good.

Estonia

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

[19] *Vehicles* is another category where imputations perform very poorly, but this poor performance is mainly driven by the very small number of observations with positive expenditures on this category. This holds true for most countries.

For the majority of the categories, the overlap between the average of observed and imputed income shares of expenditure conditional on income is fairly good. For *housing and rental* a downward bias is present across the majority of income ventiles. An upward bias is observed in *vehicles* for low to mid income ventiles.

For *culture and recreation*, *tobacco*, *travelling and holiday*, *education* and *restaurants*, the overlap between the average of observed and imputed income shares of expenditure conditional on income is extremely well. The overlap is fairly well for the rest of the categories. The only exception is *housing and rental.*

The categories with best performing imputations are *food and non–alcoholic beverages* and *communication.* The categories with worst performing imputations are *housing and rental*, *education* and *travelling and holiday.*

The imputation of income shares of total expenditure and saving perform well in terms of preserving the observed relation between net household income and income share of expenditures. However the imputations seem to be slightly downward biased (upward biased for saving) for the middle sections of the income distribution.

Croatia

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

Imputations seem to be unbiased for the majority of the categories. However, the imputations for the *alcoholic beverages* and *insurance* are upward biased for the low income ventiles while the imputations for the *personal care*, *health and care* and *restaurants* categories seem to be downward biased for the high income ventiles.

The similarity between the average observed and imputed income shares of expenditure conditional on income is quite high for more than half of the categories.

The similarity between the average of observed and imputed income shares of expenditure conditional on income is fairly good for approximately one quarter of the categories. There are only three categories where the similarity of the average observed and imputed income shares of expenditure are rather low, namely *housing and rental*, *tobacco* and *public transportation.*

The categories with best performing imputations are *food and non–alcoholic beverages* and *utilities.* The categories with worst performing imputations are *housing rental*, *public transportation* and *education.*

The imputation of income shares of total expenditure and saving seem to be unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. The similarity between the average observed and imputed income shares of total expenditure conditional on income is fairly good.

Luxembourg

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures with the exception of the category *education* for which imputed expenditure shares often fails to follow the trend in the observed expenditure shares and *restaurants* for which imputed expenditure shares for the lower end of the income distribution can not replicate the negative correlation between income and observed expenditure shares.

Overall, imputations seem to be unbiased for the most of the categories, though the imputations for the *food and non–alcoholic beverages* and *insurance* categories are downward biased for the high income ventiles.

The similarity between the average observed and imputed income shares of expenditure conditional on income is quite high for approximately half of the categories. For the rest of the categories the overlap between average observed and imputed values is mediocre. Overlap is very low for *restaurants*, *housing goods and services* and *education*.

The categories with best performing imputations is *housing and rental*. The categories with worst performing imputations are *culture and recreation*, *education* and *clothing and personal items*.

The imputation of income shares of total expenditure and saving does not exhibit a bias and overall performance in terms of preserving the observed relation between net household income and income share of expenditures is fair. The similarity between the average observed and imputed income shares of total expenditure conditional on income is good.

Latvia

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

For the majority of the categories, the overlap between the average of observed and imputed income shares of expenditure conditional on income is fairly good.

The overlaps between the average of observed and imputed income shares of expenditure per ventile are extremely well for majority of the categories and still quite very good for the rest.

The categories with best performing imputations are *food and non–alcoholic beverages* and *communication*. The categories with worst performing imputations are *house durables*, *education* and *travelling and holiday*.

The imputation of income shares of total expenditure and saving seem to be unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. The similarity between the average observed and imputed income shares of total expenditure conditional on income is fairly good.

Malta

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

Most categories seem to be imputed in an unbiased manner except for the *private transportation* category which is slightly over–imputed for the first half of the income distribution.

The similarity between the average observed and imputed income shares of expenditure conditional on income is quite high for almost all categories except *education*.

The categories with best performing imputations are *food and non–alcoholic beverages*, *utilities* and *communication*. The categories with worst performing imputations are *education* and *house durables*.

The imputation of income shares of total expenditure and saving seem to be unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. The similarity between the average observed and imputed income shares of total expenditure conditional on income is very good.

The Netherlands

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures.

Imputations seem to be unbiased for the majority of the categories. However, the imputations for the *housing and rental* is upward biased while the imputations for the *culture and recreation*, *alcoholic beverages* and *private transportation* categories seem to be downward biased.

The similarity between the average observed and imputed income shares of expenditure conditional on income is quite high for approximately half of the categories. For *housing and rental* the gap between average observed and imputed income shares of expenditure conditional on income is quite high. For the rest of the categories the overlap between average observed and imputed values is mediocre.

The categories with best performing imputations are *utilities* and *insurance*. The categories with worst performing imputations are *house durables* and *health and care*.

The imputation of income shares of total expenditure and saving is unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. However, the similarity between the average observed and imputed income shares of total expenditure conditional on income is relatively weak most likely driven by the disparities in the lowest ventile.

Sweden

Imputations for almost all categories perform fairly well in terms of preserving the observed relation between net household income and income share of expenditures with the exception of the category *personal care* for which imputed expenditure shares for the higher end of the income distribution can not replicate the negative correlation between income and observed expenditure shares.

Imputations seem to be unbiased for the majority of the categories. However, the imputations for the *food and non–alcoholic beverages* seems to be slightly upward biased. On the other hand, the imputations for the *house durables* category seem to be upward biased for the second half of the income distribution.

The similarity between the average observed and imputed income shares of expenditure conditional on income is fairly well for almost all categories with the exception of *tobacco*.

The category with best performing imputations is *housing and rental*. The categories with worst performing imputations are *house goods and services* and *house durables*.

The imputation of income shares of total expenditure and saving seem to be unbiased and perform well in terms of preserving the observed relation between net household income and income share of expenditures. The similarity between the average observed and imputed income shares of total expenditure conditional on income is very good.

4.4 Difference in correlation structure

In the sheet `correlation differences XX` of the `summary XX.xlsx` files, we report the differences in the correlation matrices. As indicated in Section 3.2 of Akoğuz et al. (2020), the difference in the correlation matrices can be decomposed into three components: the mean absolute difference in correlations among socio–demographic characteristics (`within covariates`), the mean absolute difference in correlations among income shares of expenditures (`within expenditures`), and the mean absolute difference in correlations between socio–demographic characteristics and income shares of expenditures (`between covariates and expenditures`).

Figure 1 and Figure 2, show the mean absolute difference of the last two, `within expenditures` and `between covariates and expenditures`, against the first one (`within covariates`). Red dots with country label represent the results for the new countries, while blue dots refer to the values for the countries involved in ITTv3.

Larger differences in `within covariates` might lead to larger values for the other two figures which are not to be ascribed to the imputation method *per se*, but might stem from the datasets not being representative for the same population. Therefore, we say that an imputation has relatively lower quality if it exhibits relatively large values for `within expenditures` or `between covariates and expenditures` for given values of `within covariates`.

We see that this is the case for `within expenditures` against `within covariates` for the Netherlands (see Figure 1). The Netherlands exhibits similar correlation statistics in HBS and SILC for the socio–demographic characteristics as for example Czechia, Lithuania and Portugal, but is doing much worse than those countries in safeguarding the correlation structure among income shares of expenditures in the imputation. The same applies to the rest of the countries to varying extend except to Austria. Austria, the country with the second lowest variation in `within covariates`, has the lowest variation in `within expenditures`.

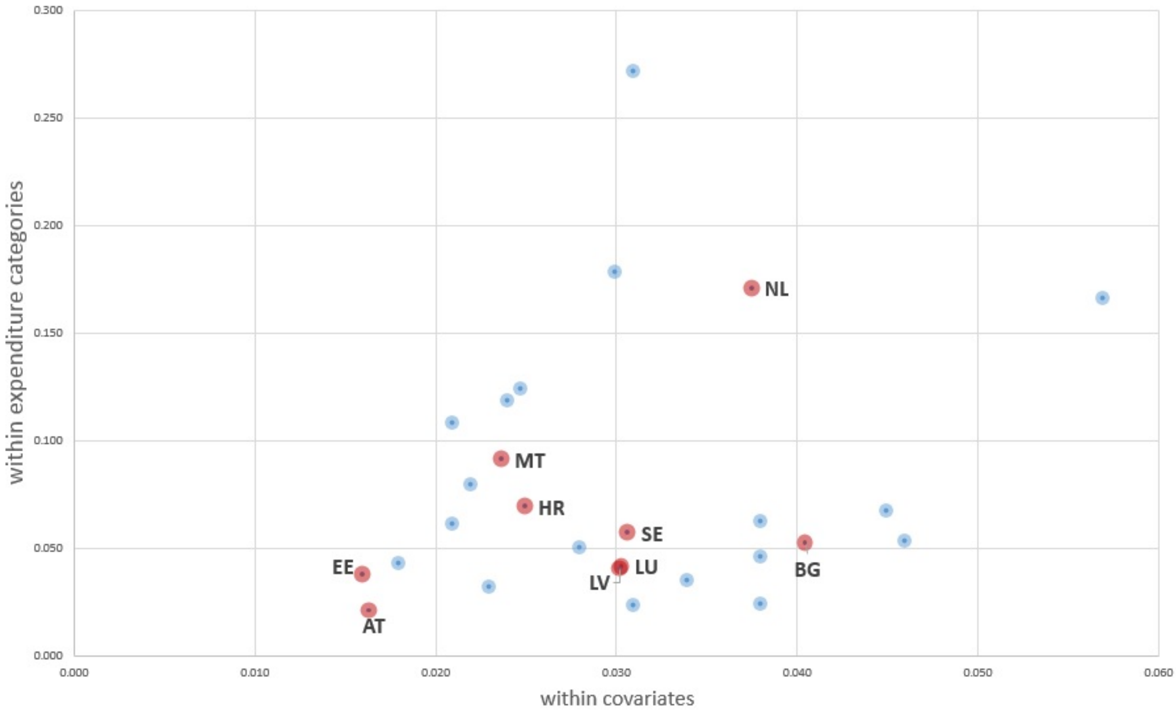


Figure 1: Scatterplot of mean diff. in correlation 'within cov' *versus* 'within exp'

Figure 2, on the other hand, shows that the correlation structure between socio–demographic characteristics and income shares of expenditures seems to be much less affected by `within covariates`. Nevertheless, we still observe that the Netherlands exhibits a larger difference in the correlation patterns between covariates and income shares of expenditures of HBS and SILC compared to Czechia, Lithuania and Portugal. Figure 2 further indicates that the relationship between Austria and Estonia slightly reversed when `between covariates and expenditures` variation is examined instead of `within expenditures`. Austria performs (slightly) worse than Estonia in terms of the mean absolute difference in correlations between socio–demographic characteristics and income shares of expenditures (`between covariates and expenditures`), despite their similarities in the mean absolute difference in correlations among socio–demographic characteristics (`within covariates`).

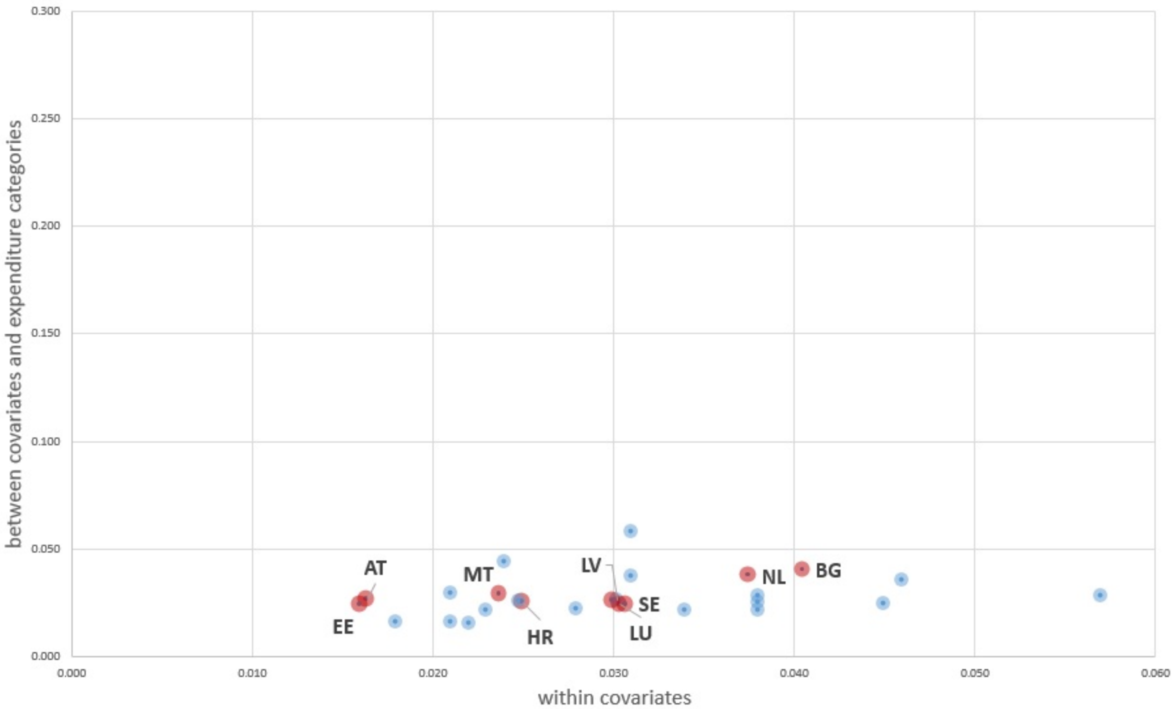


Figure 2: Scatterplot of mean diff. in correlation 'within cov' *versus* 'between cov and exp'

4.5 Macro validation

In this section we analyse the results of the ITT baseline simulations. These baselines refer to the year 2010 for all countries except Croatia, for which simulations are performed for 2013.[20] We thereto provide a comparison between the simulated values and national account figures. This gives an assessment of the ability of our model to regenerate indicators such as expenditures and indirect tax revenues at the macro level.[21] First, we show results for expenditures. Table 3 compares

[20] Indirect tax policy parameters and some of the consumer prices are not available for Croatia in 2012. Hence, we implemented ITT in the 2013 EM policy system for Croatia, using the 2012 EM-SILC as the input dataset.

[21] Detailed findings are summarised in the Excel file `Macro validation.xlsx`. The sheet `Expenditures` contains the information on the expenditures, and the sheet `Indirect taxes` covers indirect tax revenues. The sheet `Expenditures` is organised as follows:

- cells `2A:16K`: simulated total household expenditures in SILC on 12 COICOP categories;
- cells `18A:32K`: estimated total household expenditures on 12 COICOP categories based on the HBS;
- cells `34A:49K`: estimated total household expenditures on 12 COICOP categories based on the NA;
- cells `51A:67L`: coverage of HBS expenditure levels by simulated SILC expenditures;
- cells `69A:85L`: coverage of NA expenditure levels by HBS expenditure levels;
- cells `87A:103L`: coverage of NA expenditure levels by simulated SILC expenditures.

The entries of the coverage tables are shaded: darker colors refer to higher deviations from observed HBS and NA figures, blue for underestimation, red for overestimation. These tables are augmented with a row and column, indicating the number of COICOP aggregates (per country) and the number of countries (per COICOP aggregate) respectively for the ITT simulations of aggregate household expenditures for the twelve COICOP 1 (two digits) commodity categories with the corresponding 2010 national accounts figures (2013 for Croatia), while in Table 2, ITT simulations are compared with estimates on the basis of the *reported* expenditures in the corresponding HBS used for imputation.[22]

Table 2: Household expenditures coverage (Simulation/ HBS)

	AT	BG	EE	HR	LU	LV	MT	NL	SE
Food & non-alcoholic beverages	102.9%	110.1%	102.5%	92.9%	97.3%	107.2%	112.4%	102.7%	108.4%
Alcoholic beverages & tobacco	104.0%	147.9%	104.1%	88.6%	87.9%	113.6%	118.6%	100.8%	107.8%
Clothing & Footwear	105.4%	187.1%	99.5%	97.0%	88.7%	114.8%	101.5%	105.4%	93.3%
Housing, water, energy	86.9%	107.0%	92.7%	92.6%	106.8%	107.4%	113.8%	82.7%	113.7%
Furnishings, equipment & maintenance	108.6%	166.6%	115.5%	99.6%	97.1%	113.1%	102.4%	104.7%	100.6%
Health	99.0%	99.5%	88.9%	94.1%	93.8%	106.8%	118.1%	106.6%	83.7%
Transport	95.1%	165.4%	126.1%	94.7%	98.6%	112.2%	119.6%	96.6%	91.9%
Communications	108.4%	129.2%	93.9%	93.2%	105.5%	113.5%	123.7%	98.9%	93.2%
Recreation and culture	102.6%	197.9%	108.7%	93.9%	98.7%	113.6%	103.0%	104.6%	87.7%
Education	107.5%	242.6%	100.0%	90.3%	128.2%	110.6%	94.0%	94.3%	149.6%
Restaurants and hotels	106.5%	175.2%	1158.3%	100.8%	92.7%	114.4%	118.9%	106.0%	114.7%
Miscellaneous goods and services	103.4%	188.9%	116.4%	95.0%	84.8%	109.5%	99.7%	101.3%	114.2%
All	99.9%	130.7%	107.5%	93.9%	96.6%	110.0%	110.9%	98.2%	101.9%

For the category *Housing, water, energy*, the reported statistics and simulated values are excluding virtual rents for owner–occupiers.

We highlight the following observations from comparison between the ITT simulation and HBS estimations in Table 2.

- HBS–coverage is mainly driven by the variation between EUROMOD simulated disposable incomes and observed HBS incomes, as the SILC values are based on imputed income shares from the HBS. Recall that, for Bulgaria, Estonia and Latvia, there is a significant discrepancy between the HBS and SILC disposable income levels (Figure 21, Figure 22 and Figure 25 in Appendix J). Among these countries, the deviation in disposable income levels between both surveys is the largest for Bulgaria, and this is reflected in the HBS coverage (Table 2). While the HBS coverage is reasonably well for Estonia and Latvia, almost all consumption categories

which the deviation of simulated and observed values are within the 10% range.

Macro validation tables report the total HBS expenditures in policy year prices for which the simulations are performed. For all countries, except Croatia and the Netherlands, simulations are performed for the same year as the HBS survey used for imputation. Hence, no price adjustment is needed. However, for Croatia, the simulations are performed for policy year 2013 while the expenditure share information is imputed using the HBS survey of 2010. Hence, 2010 HBS expenditures of Croatia are adjusted using the CPI, so that the HBS figures represent total expenditures in 2013 prices. Similarly, for the Netherlands, the HBS survey of 2015 is used to impute the expenditure shares, while the simulations are performed for policy year 2010. Therefore, for the Netherlands, the HBS expenditure levels of 2015 are deflated to 2010 prices.

[22] National accounts' household expenditure levels are collected from the `Final consumption expenditure of households by consumption purpose (COICOP 3 digit)` available at `eurostat/final-consumption`.
For Croatia, the simulations are compared to 2013 national account figures.

are overestimated for Bulgaria. Simulated aggregates almost double the HBS figures in one fourth of the consumption categories, and are two and half times higher for *education*.

- The results for Malta and the Netherlands cannot be explained by the difference in disposable income. The HBS coverage of both countries is approximately 10% off, despite HBS and SILC disposable income levels being closely related.
- The total of simulated expenditures is almost identical to the HBS estimates for Austria, the Netherlands, and Sweden, despite some variation across different consumption categories.

Table 3: Household expenditures coverage (Simulation/ national accounts)

	AT	**BG**	**EE**	**HR**	**LU**	**LV**	**MT**	**NL**	**SE**
Food & non-alcoholic beverages	101.5%	86.7%	71.7%	79.0%	71.4%	83.2%	104.6%	87.8%	75.8%
Alcoholic beverages & tobacco	59.3%	43.9%	22.2%	24.5%	10.7%	24.2%	48.7%	82.6%	41.2%
Clothing & Footwear	79.6%	85.1%	38.6%	71.2%	79.0%	69.9%	82.9%	77.4%	58.7%
Housing, water, energy	94.1%	99.5%	117.3%	78.5%	112.2%	111.2%	117.3%	105.3%	87.5%
Furnishings, equipment & maintenance	87.0%	37.9%	101.4%	52.2%	58.1%	71.3%	84.0%	82.1%	72.0%
Health	74.3%	61.6%	58.8%	38.9%	63.9%	77.8%	124.3%	40.9%	40.0%
Transport	86.2%	28.6%	60.8%	57.1%	67.7%	66.6%	81.8%	81.7%	59.2%
Communications	70.0%	51.0%	75.8%	62.9%	100.0%	92.8%	97.0%	82.6%	61.0%
Recreation and culture	103.8%	35.0%	74.0%	31.3%	70.0%	54.4%	60.6%	65.5%	73.0%
Education	112.3%	51.6%	81.5%	35.7%	36.6%	58.3%	58.6%	156.6%	6.9%
Restaurants and hotels	41.4%	63.0%	27.4%	9.5%	67.9%	47.0%	32.1%	76.5%	50.1%
Miscellaneous goods and services	75.0%	60.3%	42.4%	51.5%	46.1%	57.5%	81.6%	88.3%	42.5%
All	81.5%	58.9%	62.5%	51.4%	64.0%	69.4%	77.2%	83.0%	64.8%

For the category *Housing, water, energy*, the reported statistics and simulated values are excluding virtual rents for owners–occupiers.

National accounts coverage of the HBS is quite low. The HBS covers less than two thirds of the national account numbers for six out of the nine countries (BG, EE; HR, LU, LV, and SE). For Austria, Malta and the Netherlands, the national account coverage rate of HBS is 82%, 70%, and 92% respectively. Bulgaria has the lowest HBS-NA coverage (45%).[23] This should be born in mind in the following discussion of Table 3 which compares ITT simulations of expenditures with national account figures.

- Under–coverage of the national account figures by the HBS is reflected in the ITT simulations. When comparing simulated SILC expenditures with national accounts, we observe an overall underestimation of expenditures. Few exceptions are the categories where HBS-NA coverage was exceptionally high, such as *housing, water, electricity, gas & other fuels* in Estonia, *health* in Malta and *education* in the Netherlands.
- For Bulgaria, the excess coverage of HBS expenditures caused by higher SILC incomes (131%)

[23] For the countries covered by ITTv3, HBS total expenditures covered less than two thirds of the corresponding National Account figures for 8 out of 18 countries. The lowest coverage rate was for Romania (46%), while the highest one was obtained for Finland (89%).

attenuates the low HBS-national accounts coverage (45%). ITT simulation for Bulgaria is able to account for 59% of the national account numbers.

- The overall average coverage varies between 51.4% (HR) and 83% (NL).
- As expected, the best coverage is observed among categories with few zero expenditure observations (*housing, water, electricity, gas & other fuels*, and *food & non–alcoholic beverages*).
- Under–coverage is the largest for *alcoholic beverages & tobacco* and *hotel & restaurants*.

For the indirect tax validation, we collected figures from EUROSTAT for the year 2010 on the VAT component of indirect tax revenues and from the European Commission on the excise components.[24]

Table 4 displays the coverage of indirect tax revenues. The cells correspond to the ratio of simulated to national figures for the respective indirect tax revenues.[25] The first row of Table 4 shows the coverage rate for the VAT revenues. The following rows are the coverage rates of excise revenues. Finally, the very last row shows the estimated size of the household sector's share in total VAT revenues. National tax revenues are not given at the same level of detail for all countries. The indirect tax revenues are therefore aggregated for some excise goods when calculating the coverage rates. For example, in Austria and Croatia all mineral oil revenues are aggregated into one category, and so we report coverage rates only up to that level of detail for all countries. Only Estonia reports excise revenues from cigars and cigarillos separately. So, we aggregate the excise revenues from these sources. For some countries, the level of aggregation is in even broader categories. The empty entries in Table 4 mean that this is the case and excise revenues from the group of commodities to which the empty cell refer are included into another category of goods.

- For the VAT–component, the coverage ranges between 23.7% (Luxembourg) and 86.8% (Latvia). The coverage rate of excise revenues varies in an even wider range. VAT coverage is especially low for countries where the simulated/national account ratio of expenditures are low, Bulgaria being the exception. Although under–coverage of the total expenditures (with respect to national accounts) is partially responsible for the low VAT and excise coverage rates, the main reason behind the low coverage remains to be the fact that ITT only covers the household sector. This becomes clear if we compare the VAT coverage rate of EUROMOD simulations with the estimated share of the household sector in the VAT revenues (last row of Table 4).[26] For Austria, Bulgaria, Malta, and the Netherlands the simulated/national account coverage

[24] VAT revenues are collected from `Main national accounts tax aggregates`, accessible at `eurostat/TAXAG`. Excise tax revenues can be found at `ec.europa.eu/excise_duties`. Excise duty receipts of Member States for the period 2008 to 2020 are listed at `excise_duties_alcohol.pdf`, `excise_duties_tobacco.pdf`, and `excise_duties_energy_products.pdf`.

[25] For details, the user may refer to the tab `Indirect taxes` of the Excel file `Macro validation.xlsx`. The collected tax revenues from VAT are reported on row 25 of the tab `Indirect taxes`, and for excises on rows 26 to 45 of the same sheet. Simulated indirect taxes for policy year 2010 (2013 for Croatia) can be found on cells `B1:21K` of the same sheet. Cells `B100:114K` of that sheet report the average coverage rate of simulated indirect taxes.

[26] Household sector's share in the VAT revenues are estimated by JRC.B2.

Table 4: VAT and excise revenues coverage (Simulation/ National accounts)

	AT	BG	EE	HR	LU	LV	MT	NL	SE
VAT	68.8%	66.9%	53.1%	44.5%	23.7%	86.8%	73.4%	57.5%	35.8%
EXCISES									
Ethyl Alcohol	55.1%	35.8%	17.1%	41.8%	18.5%	26.2%	22.9%	71.1%	4.1%
Wine (still and sparkling)			22.9%			57.7%		61.2%	46.9%
Beer	42.0%	36.0%	20.4%	20.3%	27.0%	16.8%	0.0%	49.2%	33.1%
All alcoholic beverages	49.4%	35.9%	18.5%	26.3%	19.5%	27.3%	20.2%	60.0%	27.4%
Cigarettes	68.9%	56.8%	31.2%	38.2%	5.7%	62.3%	49.9%	81.9%	23.9%
Cigar and cigarillos		0.0%	0.0%		0.0%	0.0%		116.7%	122.0%
Other smoking tobacco		18.6%	8.2%		1.2%	0.0%	55.3%	73.9%	317.2%
All tobacco goods	68.9%	56.0%	30.4%	38.2%	4.9%	60.9%	50.0%	80.2%	40.4%
All mineral oil	42.4%	25.6%	27.7%	31.3%	10.1%	25.0%	67.6%	61.6%	48.1%
Natural gas	50.0%		4.8%	0.0%	102.3%	317.5%		54.6%	0.0%
Coal & coke		509.1%	520000%	13076.9%		493.8%		87.0%	300.6%
Electricity		0.0%	30.3%	248.5%	35.3%	0.0%	65.2%	165.7%	35.1%
All energy goods	43.6%	26.6%	27.9%	32.3%	10.6%	27.1%	67.5%	73.5%	44.2%
Estimated share of household sector in VAT revenues									
	69.2%	66.6%	63.7%	81.0%	61.6%	70.9%	73.8%	54.8%	54.4%

rates matches the estimated share of the household sector quite well. The variation is within a 3 percentage point band for these four countries.

- The effect of excluding the corporate sector is even more visible in simulated excise revenues. The share of companies in excises on *energy products* is rather sizable. The national revenue statistics include these revenues, while the ITT-model only covers the households sector. Since excises from *electricity* and *all mineral oil* make up the brunt of energy excise revenues, they inevitably drive the coverage statistics. In Bulgaria, household consumption of electricity is not subject to specific excises. Hence, simulated specific excises are zero for households' expenditures on electricity. Nevertheless, the country has 32.14 million BGN excise revenues from electricity. Another example is Sweden. Households do not use natural gas, but national account statistics report 792 million SEK excise revenue from natural gas.
- Under-coverage of excises on alcohol and tobacco is most probably due to under-reporting of expenditures on these categories. HBS expenditures cover less than 50% of the NA-counterparts for seven of the nine countries. For some countries, cross-border purchases of these goods might be a cause too: if alcohol and tobacco are cheaper in one particular country (mostly due to differences in indirect tax policies), people abroad living relatively close to the border might chop across the border.
- The high overcoverage of excise revenues from coal and coke has to do with the weighting rule used for excises on solid fuels (COICOP category 04541). This commodity group comprises,

besides coal and coke other solid lubricants such as products like briquettes, firewood, wood pellets, and nutshells. However, we implemented the excise rate for coal and coke to total expenditures on solid fuels. Appendix F gives further information on the principles of weighting prices and excise rates for goods subject to excises.

- ITT cannot always address all aspects of a country's excise policy. For example, in Sweden, regional excise tariffs are in place, and excise levels are considerably lower in the north. However, ITT is not constructed in a way that it can control for household characteristics and hence, ITT is not able to account for the variations in excise tariffs that depend on the household characteristics.

5 Tax incidence parameters and changes in the ITT structure

5.1 Tax incidence

The EUROMOD ITT tool's main objective is to enable distributional and welfare analysis of joint indirect and direct tax reforms. The main effect of a change in indirect taxation is a change in consumer prices. It is thus necessary to model how consumers react to changes in consumer prices, when designing such a tool. Analogously to what is assumed by the EUROMOD direct tax calculator, the most elementary way to do this, is to assume that demand remains constant under price changes.[27] ITTv3 allows for such simulations under *constant quantities* (CQ) assumption. This assumption implies that own and cross price elasticities, and income elasticities are all equal to zero. When used stand alone, such an assumption fails to satisfy the budget constraint. In combination with a direct tax model, it implies that the monetary impact of indirect tax reforms is completely absorbed by changes in saving.

ITTv3 implemented also two alternative assumptions on consumers' behavioral reactions to changes in commodity prices: *constant income shares* (CS_Y) and *constant* expenditure shares (CS_E). Both assumptions imply that there are no cross price effects, that own price elasticities are all equal to minus one, and that income, respectively, expenditures elasticities are all equal to one. Under CS_E, saving remains constant at the baseline level, while under CS_Y, saving is considered as being just another commodity (future consumption, say) next to the other commodities, also exhibiting unit income elasticities, zero cross price effects, and own price elasticities equal to minus one.[28]

In the ITTv3 setting, producer prices were assumed to be fixed at their baseline levels. As a result, indirect tax policy changes were fully reflected on the consumer prices. Formally this is equivalent to assuming that the percentage change in consumer prices due to a change in indirect taxes equals the percentage point change in the implicit indirect tax rate (that is the wedge between consumer and producer price relative to the producer price) divided by one plus the baseline indirect tax rate (*cf.* Equation 24 below).

The latest release of ITT, ITTv4, relaxes this fixed producer price assumption, also known as the full pass-through assumption. By introducing a tax incidence parameter, θ, ITTv4 thus allows for producer prices to adapt to indirect tax changes. It enables to analyse the welfare cost implications of a change in indirect taxes for consumers under variable assumptions on the pass-through.[29] This section first portrays the theoretical framework of the tax incidence parameter and then describes how this new parameter is implemented in EUROMOD.

[27] In the direct tax calculator of EUROMOD, labor supply is assumed to be fixed. Of course EUROMOD's direct tax module can be combined with an external labor supply model. It is even technically possible to call EUROMOD within the estimation phase of a labor supply model, when the estimation method would require so.

[28] Akoğuz et al. (2020) provides a detailed review of the implications of the different behavioral assumptions.

[29] The effect of variable producer prices on producer surpluses is beyond the scope of EUROMOD ITT, which is a micro-model of consumers, and does not include a production module.

5.1.1 Tax incidence parameters

Table 5 summarizes the notation for prices and taxes on individual commodities that will be used in this section. A specific excise, a_k, is a monetary amount of taxes to be paid per quantity unit of commodity k. An *ad valorem* excise rate, v_k, is expressed as the percentage of consumer price q_k to be paid. Similarly, a VAT rate t_k, is expressed as the percentage of the producer price p_k, augmented with excises $a_k + v_k q_k$ for goods on which excises are payable.

Table 5: Notation for prices and taxes of a specific commodity k

Consumer (unit) price	q_k	per (quantity) unit
Producer (unit) price	p_k	per (quantity) unit
VAT rate	t_k	% of producer price plus excises
Specific excise	a_k	per (quantity) unit
Ad valorem excise	v_k	% of consumer price
Implicit tax rate	τ_k	% of producer price
Quantities	${x_k}^h$	of household h
Income share	${w_k}^h$	of household h
Expenditure share	${\omega_k}^h$	of household h

The consumer prices q_k is an implicit function of producer prices p_k and indirect tax parameters:

$$q_k = (1 + t_k)(p_k + a_k + v_k q_k) = (1 + \tau_k)p_k, \tag{19}$$

where τ_k is the implicit indirect tax rate, defined as the wedge between consumer and producer price caused by indirect taxes, relative to the producer price.

Rearranging Equation (19) gives a formula for the implicit tax rate τ_k:

$$\tau_k = \frac{1 + t_k}{1 - (1 + t_k)\left(v_k + \frac{a_k}{q_k}\right)} - 1. \tag{20}$$

Once the fixed producer price assumption is relaxed, p_k becomes a function of the implicit tax rate τ_k and Equation (19) becomes:

$$q_k = (1 + \tau_k)\, p_k(\tau_k). \tag{21}$$

Equation (21) is a generic formula that applies to all commodities. Hence, we will drop the subscript k wherever possible, for notational simplicity. When variables are subscripted by $t = 0, 1$, the index does not refer to commodities, but to the values of the variable in the pre- and post-reform situation.

Using Equation (21), we can write the proportional change in consumer price following a change in the implicit tax rate from τ_0 to τ_1 as follows:

$$\frac{q_1 - q_0}{q_0} = \frac{(1 + \tau_1)p(\tau_1) - (1 + \tau_0)p(\tau_0)}{(1 + \tau_0)p(\tau_0)}. \tag{22}$$

In the sequel, $p(\tau_t)$ $(t = 0, 1)$ will be abbreviated as p_t.

Rearranging Equation (22) enables to formulate the change in consumer price as a linear function of the (normalized) change in the implicit tax rate. The proportionality factor of this linaer function, indicates how the change in τ is reflected in the consumer price. It is denoted by θ and will be called the tax incidence parameter:

$$\frac{\mathrm{d}\,q}{q_0} = \underbrace{\left(1 + \frac{\mathrm{d}\,p}{p_0}\frac{1+\tau_1}{\mathrm{d}\,\tau}\right)}_{\theta}\frac{\mathrm{d}\,\tau}{1+\tau_0}, \tag{23}$$

where a variable $d\,x$ stands for $x_1 - x_0$.

Under the constant producer prices assumption, $\mathrm{d}\,p$ is equal to zero and Equation (23) reduces to:

$$\frac{\mathrm{d}\,q}{q_0} = \frac{\mathrm{d}\,\tau}{1+\tau_0}. \tag{24}$$

In that case $\theta = 1$, and the change in consumer price is equal to the (normalized) change in the implicit tax rate ($1 + \tau_0$ being the normalization factor).

When the constant producer prices assumption is relaxed, the change in consumer price q as a response to a change in the implicit tax rate τ will no longer solely depend on the change in the implicit tax rate (and thus on the tax parameters), but also on the degree to which producer prices react to a change in the implicit tax rate τ. The value of the tax incidence parameter, θ, quantifies this producer price sensitivity to the tax parameters. If $\theta < 1$, the producer price p decreases with an increase in τ. The post reform consumer price q_1 can be written as:

$$q_1 = q_0\left(1 + \theta\frac{\mathrm{d}\,\tau}{1+\tau_0}\right), \tag{25}$$

where the effect of the variable producer price is embodied in the value of the tax incidence parameter θ.

The new implicit indirect tax rate τ_1 equals:

$$\tau_1 = \frac{1+t_1}{1-(1+t_1)\left(v_1 + \frac{a_1}{q_1}\right)} - 1. \tag{26}$$

For commodities which are not subject to specific excises, Equation (26) boils down to:

$$\tau_1 = \frac{1+t_1}{1-(1+t_1)v_1} - 1. \tag{27}$$

Hence, τ_1 no longer depends on the consumer prices q_1, and Equation (25) and (27) can be solved immediately. For commodities which are subject to specific excises, Equation (25) and (26) form a system of two equations in two unknowns (q_1 and τ_1), which can be solved by the following iterative procedure:[30]

$$\tau_1^{(n)} = \frac{1+t_1}{1-(1+t_1)\left(v_1 + \frac{a_1}{q_1^{(n-1)}}\right)} - 1. \tag{28}$$

[30] In the case of fixed producer prices, an explicit solution is possible and was implemented in previous versions of ITT: From Equation (19) it follows that $q_1 = \frac{(1+t_1)\left(\frac{q_0}{1+\tau_0}+a_1\right)}{1-(1+t_1)v_1}$. Then, $\tau_1 = \frac{q_1(1+\tau_0)}{q_0} - 1$.

$$q_1^{(n)} = q_0\left(1 + \theta\frac{\tau_1^{(n)} - \tau_0}{1 + \tau_0}\right). \quad (29)$$

A solution is found when $q_1^{(n-1)} = q_1^{(n)}$.

We followed here a completely model free approach: no structural interpretation is given to the tax incidence parameter. This way, the implementation of the tax incidence parameter remains valid irrespective of the modelling of producer price dependency on indirect taxes.

5.1.2 Implications for indirect tax revenues

The effect of the tax incidence parameter θ on indirect tax revenues runs through its impact on the consumer price and varies with the assumptions on demand behavior with respect to prices.

Under the behavioral assumptions of constant income shares (CS_Y) or constant expenditure shares (CS_E), VAT liabilities and *ad valorem* excises remain unaffected by θ, as expenditures on a commodity k, e_k^h, are equal to the constant share (respectively w_k and ω_k) times disposable income, respectively total expenditures. They are thus not dependent on the consumer prices. Revenues on specific excises, on the other hand, do depend on θ. The level of θ affects the consumer price q_1. The quantity demanded, defined as expenditures on k divided by the consumer price, $x_k^h = e_k^h/q_k$, is therefore a function of θ.

The opposite holds true for the constant quantities assumption (CQ). Under CQ, expenditures on commodity k are not fixed, but equal the constant quantity times the consumer price: $e_k^h = q_k x_{k,0}^h$. Consequently, VAT and *ad valorem* excise revenues depend on the consumer price, and, hence, are affected by θ. On the other hand, the specific excise revenues are independent of θ under CQ, as they depend only on quantities, which are by definition fixed to the baseline level under CQ.

Table 6 summarizes the different components of indirect tax revenues for the three behavioral assumptions of ITT. Here, y represents disposable household income, which is exogenous in the ITT demand models. Saving is denoted by S, and is exogenous and fixed under CS_E. Under CQ, quantities, x_0, are exogenous and fixed.

Table 6: Indirect tax revenues and the tax incidence parameter

	behavioral assumption		
Indirect tax component	CS_Y	CS_E	CQ
VAT	$t_1\frac{wy}{1+t_1}$	$t_1\frac{\omega(y-S)}{1+t_1}$	$t_1\frac{x_0 q_1}{1+t_1}$
Ad valorem excises	$v_1 wy$	$v_1\omega(y-S)$	$v_1 x_0 q_1$
Specific excises	$a_1\frac{wy}{q_1}$	$a_1\frac{\omega(y-S)}{q_1}$	$a_1 x_0$

Household and commodity indices are dropped for notational simplicity.

5.1.3 Welfare implications

In order to analyse the impact of price changes due to tax reforms on consumer welfare (neglecting labour supply effects), ITT uses household disposable income or expenditures divided by a price index, as a welfare measure of consumption.

Under constant income or expenditure shares assumptions, the Stone price index can be used. The resulting *real* disposable income (disposable income divided by the Stone price index) equals the minimal amount of disposable income a household needs at reference prices equal to one for all commodities, in order to be equally well off as in the situation under consideration.[31]

A Stone price index, $P_s(\mathbf{q})$, is a weighted geometric mean of prices:

$$P_s(\mathbf{q}) = \prod_k (q_k)^{s_k}, \tag{30}$$

where $\mathbf{q}$ is the vector of consumer prices, q_k $(k = 1, 2, \ldots)$, and s_k is a set of non–negative weights, summing to one.

Under CS_Y it is appropriate to use income shares of expenditures, w_k, as weights:

$$P_w(\mathbf{q}) = \prod_k (q_k)^{w_k}. \tag{31}$$

Recall that these income shares are household specific, and thus, so is the Stone price index.

The corresponding welfare metric is:

$$Q_y(\mathbf{q}) = \frac{y}{P_w(\mathbf{q})}. \tag{32}$$

Under CS_E, total expenditure, E, can be deflated by a Stone price index using expenditure shares as weights:

$$Q_E = \frac{E}{P_\omega(\mathbf{q})}, \tag{33}$$

where E denotes total expenditure, which is exogenous in the CS_E case. Again, the price index is household specific.

Under CQ, the additional amount of income needed to buy the constant quantity bundle at the new consumer prices, can serve as an appropriate measure of the welfare cost of the indirect tax reform:

$$C_q = (\mathbf{q}_1 - \mathbf{q}_0)' \mathbf{x}_0. \tag{34}$$

A corresponding welfare index is equal to:

$$Q_q = \frac{\mathbf{q}_0' \mathbf{x}_0}{\mathbf{q}_1' \mathbf{x}_0} E.^{32} \tag{35}$$

[31] Given our normalisation of producer prices, this is equivalent to use baseline producer prices as reference prices.
[32] An alternative is to subtract C_q from E.

In this index, expenditures are divided by a household specific Laspeyres price index: $\frac{\mathbf{q}_1'\mathbf{x}_0}{\mathbf{q}_0'\mathbf{x}_0}$. Prices (identical across households) are weighted with fixed household specific baseline quantities.

Table 7 shows the quantity and welfare levels under the alternative behavioral assumptions of ITT.

Table 7: Welfare effects of the tax incidence parameter

	CS_Y	CS_E	CQ
Quantity	$x_1 = w\frac{y}{q_1}$	$x_1 = \omega\frac{E}{q_1}$	$x_1 = x_0$
Welfare	$Q_y = \frac{y}{P_w(\mathbf{q}_1)}$	$Q_E = \frac{E}{P_\omega(\mathbf{q}_1)}$	$Q_q = \frac{\mathbf{q}_0'\mathbf{x}_0}{\mathbf{q}_1'\mathbf{x}_0}E$

Household and commodity indices are dropped for notational simplicity.

As all of these measures depend on the (post-reform) consumer prices, except for the quantities under CQ of course, these measures also depend on the tax incidence parameter. Comparing the welfare of a simulation of a reform with θ equal to one for all commodities, with the results for the same reform, specifying θ different from one for all (or some commodities) affected by the indirect tax reform, thus allows to quantify the welfare cost (or gain) of different assumptions on the tax incidence parameter. In case of a welfare gain for the consumer, it is said that part of the cost of an indirect tax increase is passed on to the producer. The welfare implications for the producer cannot be calculated by EUROMOD ITT which is a micro-model of consumers and therefore does not contain a production sector model.

If one wants to perform welfare analysis at the individual, instead of at the household level, it deserves recommendation to adjust the suggested welfare indicators by means of an equivalence scale.

It is furthermore important to make a sharp distinction between the welfare implications of variable pass-through and the implications on payable taxes. Under CS_Y and CS_E, e.g., variable pass-through does not affect VAT and *ad valorem* excise liabilities, but it can have considerable welfare effects, both for consumers and producers. These welfare effects do not have to add up (the welfare gain of the consumers is not necessarily equal to the welfare loss of the producers, or vice versa). EUROMOD ITT is able to calculate both, the impact, if any, on tax liabilities and the welfare effects for the consumers.

5.2 Implementation of the tax incidence parameter in ITT

The TCO policy module `tco_cc` (`cc` denotes the country) contains all the indirect tax parameters and functions calculating the prices, expenditures, indirect taxes, and other variables of interest. The implementation of the tax incidence parameters mostly does not alter the TCO policy module. However, there are still considerable changes in ITTv4 compared to ITTv3.

First of all, the tax incidence parameter, θ, is introduced as a new extension, $XTip$. When it is turned off, simulations will bypass the iteration routine. It is recommended to switch it on only when doing simulations where the tax incidence parameter is specified different from one for at least one commodity. When $XTip$ is switched off the simulation will run as if $\theta = 1$ for all commodities, irrespective of their values specified in the policy sheets.

The rest of the changes can be classified into two groups: i) the expansion of the set of indirect tax parameters and ii) changes in the calculations of producer and consumer prices. Once the consumer prices are calculated, the calculation of expenditures, indirect tax liabilities and other TCO output variables remain unchanged. Hence, in the following section we only discuss the changes in the parameter space and the price calculation procedure.

5.2.1 The expanded set of indirect tax parameters

In ITTv4, the set of indirect tax parameters is considerably larger. This expansion does not require new input information. ITT solves the system of equations (Equation (25) and Equation (26)) using a `Loop` function that operates over pre- and post-reform parameters. Hence, the implementation of the tax incidence parameter necessitates the baseline and the reform parameters to be defined explicitly within the same policy system. This way, the `Loop` function can iterate between the baseline and reform implicit tax rates and consumer prices, and find a solution for the new consumer prices (Equation (25)). The indirect tax module `tco_cc` also requires the tax incidence parameter to be defined explicitly for each commodity. Figure 3 gives a complete overview of the parameters in the new `tco_cc` of ITTv4.

	Policy	Grp/No	BE_2010	Comment
38.1	DefConst		switch	Parameters
38.2	DefConst		switch	Parameters: vat rates (all goods)
38.3	DefConst		switch	Parameters: specific excises (all goods)
38.4	DefConst		switch	Parameters: ad valorem excises (all goods)
38.5	DefConst		switch	Tip- Parameters: tax incidence parameter (all goods)
38.6	DefConst		switch	Parameters: vat rates (excise goods)
38.7	DefConst		switch	Parameters: specific excises (excise goods)
38.8	DefConst		switch	Parameters: ad valorem excises (excise goods)
38.9	DefConst		switch	Parameters: excise goods (calculations)
38.10	DefConst		switch	Baseline parameters: vat rates (all goods)
38.11	DefConst		switch	Baseline parameters: specific excises (all goods)
38.12	DefConst		switch	Baseline parameters: ad valorem excises (all goods)
38.13	DefConst		switch	Baseline parameters: vat rates (excise goods)
38.14	DefConst		switch	Baseline parameters: specific excises (excise goods)
38.15	DefConst		switch	Baseline parameters: ad valorem excises (excise goods)
38.16	DefConst		switch	Baseline parameters: consumer prices (excise goods)
38.17	DefConst		switch	Baseline parameters: excise goods (calculations)
38.18	DefConst		switch	Total expenditures COICOP aggregates (million euros)
38.19	DefIl		switch	Parameter list: vat rates
38.20	DefIl		switch	Parameter list: specific excises
38.21	DefIl		switch	Parameter list: ad valorem excises
38.22	DefIl		switch	Parameter list: tax incidence parameter
38.23	DefIl		switch	Baseline parameter list: vat rates
38.24	DefIl		switch	Baseline parameter list: specific excises
38.25	DefIl		switch	Baseline parameter list: ad valorem excises
38.26	DefIl		switch	Commodity list: il_xs = income shares (all goods)
38.27	DefIl		switch	Commodity list: il_xs_exc = income shares of excise goods
38.28	DefIl		switch	Commodity list: il_xs_rest = income shares of non excise goods

Figure 3: The expanded set of indirect tax parameters

As in ITTv3, the first set of functions in `tco_cc` define the indirect tax and price parameters. The `Loop` function operates over the complete set of commodities. Hence, each commodity in the input dataset must be assigned a VAT rate, a specific excise rate, an *ad valorem* excise rate, and a tax incidence parameter. For goods for which a non-zero specific excise rate is specified, consumer prices for the baseline policy year are required as well.

	Policy	Grp/No	BE_2010	Comment
38	tco_be		on	**TAX: Commodities**
38.1	DefConst		switch	Parameters
38.1.1	$tco_t_std	1	21%	vat - standard rate
38.1.2	$tco_t_red1	2	12%	vat - reduced rate 1
38.1.3	$tco_t_red2	3	6%	vat - reduced rate 2
38.1.4	$tco_t_zero	4	0%	vat - 0% rate and exempted
38.1.5	$tco_theta1	5	20%	Theta- Tax incidence parameter level 1
38.1.6	$tco_theta2	6	50%	Theta- Tax incidence parameter level 2
38.1.7	$tco_theta3	7	80%	Theta- Tax incidence parameter level 3
38.1.8	$tco_theta0	8	100%	Theta=1
38.1.9	$tco_base_t_std	9	21%	Baseline vat rate categories - standard rate
38.1.10	$tco_base_t_red1	10	12%	Baseline vat rate categories - reduced rate 1
38.1.11	$tco_base_t_red2	11	6%	Baseline vat rate categories - reduced rate 2
38.1.12	$tco_base_t_zero	12	0%	Baseline vat rate categories - 0% rate and exempted

Figure 4: Baseline and reform VAT rate categories

1. The function `1.Parameters` defines a set of baseline and possible reform values for the VAT-rates and for the tax incidence parameter (Figure 4):

   – reform VAT rate categories: standard and reduced VAT rates `$tco_t_xxx`,

   – baseline VAT rate categories: standard and reduced VAT rates `$tco_base_t_xxx`,

   – a set of predetermined levels for the tax incidence parameter `$tco_theta_xxx` from which the user can select when running a reform scenario. We set four possible values for the tax incidence parameter from which one can be selected when running the reform: $\theta = 0.20, 0.50, 0.80, 1.00$. The user can change these levels if wanted. The user can also define additional levels for the tax incidence parameter using the `add parameter` option of the `DefConst` function.

2. For all commodities, the function `2.Parameters: vat rates (all goods)` assigns the reform VAT rates, `$tco_t_[xxxxx]`.

3. The functions `3.Parameters: specific excises (all goods)` and `4.Parameters: ad valorem excises (all goods)` generates specific excise tariffs `$tco_a_[xxxxx]` and *ad valorem* excise rates `$tco_v_[xxxxx]` for all commodities, irrespective of whether these commodities are subject to excises or not. These parameters are necessary for the iteration procedure called by the `Loop` function, since the same algorithm operates over the entire commodity set. Nevertheless, specific excise and *ad valorem* excise parameters are only relevant for excise goods. Hence, these parameters are set to zero for all commodities. The parameter values for excise goods will be assigned further down in the `tco_cc`.

4. The function `5.Parameters: tax incidence parameter (all goods)` assigns the commodity specific tax incidence parameter values, `$theta_[xxxxx]`. Full pass-through, that is $\theta = 1$ for all commodities, is the default setting. The user can alter this default setting by changing the assigned θ values. However, this function operates only when the tax incidence parameter

extension, *XTip* is set *on*. `$theta_[xxxxx]` variables will not be created in the baseline scenario.

5. The functions `6.Parameters: vat rates (excise goods)`, `7.Parameters: specific excises (excise goods)`, and `8.Parameters: ad valorem excises (excise goods)` overwrite the parameter values for excise goods. These replace the excise parameters which were initially set to zero. At the first glance, these functions may appear to be redundant since one could set the true parameter values from the start. However, the commodity list of ITT is rather long and the commodities are ordered alphabetically according to their COICOP label. There are in total 193 commodities and excise goods are scattered around this list which makes scrolling through it rather cumbersome. Hence, allocating a separate section where the user can adjust the excise good parameters improves user friendliness considerably.

6. The function `9.Parameters: excise goods (calculations)` contains the necessary mutations, such as taking (weighted) averages of the statutory parameters. It is common that the level of detail on the various commodities in the input data is more limited than the information on statutory excises (e.g. in most of the HBS surveys there is only one variable for *Car fuel*, while most countries tax gas oil and petrol differently).[33]

7. The following six functions, `10.Baseline parameters: vat rates (all goods)`, `11.Baseline parameters: specific excises (all goods)`, `12.Baseline parameters: ad valorem excises (all goods)`, `13.Baseline parameters: vat rates (excise goods)`, `14.Baseline parameters: specific excises (excise goods)`, and `15.Baseline parameters: ad valorem excises (excise goods)`, repeat the same procedure for the baseline parameter values `$tco_base_t_[xxxxx]`, `$tco_base_a_[xxxxx]` and `$tco_base_v_[xxxxx]`. In the absence of a reform, `Parameters` and `Baseline parameters` values are identical. When simulating an indirect tax change, the user must change only the content of the `Parameters` functions, while the `Baseline parameters` functions must remain unchanged.

8. The function `16.Baseline parameters: consumer prices (excise goods)` sets the baseline consumer prices, `$tco_base_q_[xxxxx]`, for the goods where excises apply.

9. The function `17.Baseline parameters: excise goods (calculations)` makes a set of calculations similar to `9.Parameters: excise goods (calculations)`. For some excise goods the baseline consumer prices are collected in a unit other than the unit in which specific excises are specified. For example, the baseline consumer price of wine is the price of wine per litre, while the specific excises on wine is defined for 100 litres. Unit conversion of baseline consumer prices are also done in this section.

10. The function `18.Total expenditures COICOP aggregates` contains the simulated and the national account aggregates for the policy year. The simulated baseline and the national

[33] Appendix F provides additional information on the weighting procedure.

account aggregates are used to adjust the simulation results to national account statistics.

11. For ease of modelling, a set of functions group the parameters into several lists.

    The set of `Parameter list:...` and `Baseline parameter list:...` functions define in total seven different parameter lists containing the values of VAT rates, specific and *ad valorem* excises both at the reform and baseline levels. Unlike the policy parameters, the tax incidence parameter is a constant and as such does not belong to the reform specification. Having the parameters of each group in a list allows the ITT to use the lists as vectors in operations.

    - `il_tco_t`: list containing the VAT rates of all goods,
    - `il_tco_a`: list containing the specific excises of all goods,
    - `il_tco_v`: list containing the *ad valorem* excises of all goods,
    - `il_tco_theta`: list containing the tax incidence parameters of all goods,
    - `il_tco_base_t`: list containing the baseline VAT rates of all goods,
    - `il_tco_base_a`: list containing the baseline specific excises of all goods,
    - `il_tco_base_v`: list containing the baseline *ad valorem* excises of all goods.

    The three functions `Commodity list:...` define three variable lists:

    - `il_xs`: list containing the income shares of all goods,
    - `il_xs_exc`: list containing the income shares of only the excise goods,
    - `il_xs_rest`: list containing the income shares of the non-excise goods.

5.2.2 Calculation of the post reform prices

We explain now how the ITT calculates the post reform producer and consumer prices for given baseline prices, baseline and reform policy parameters.

	Policy	Grp/No	BE_2010	Comment
38.28	DefIl		switch	Commodity list: il_xs_rest = income shares of non excise goods
38.29	IlArithOp		switch	Baseline producer prices (1st step, non-excise goods)
38.30	IlArithOp		switch	Baseline consumer prices (1st step, non-excise goods)
38.31	DefIl		switch	Join the baseline consumer prices in a one income list
38.32	IlArithOp		switch	Baseline producer prices (2nd step, excise goods)
38.33	DefIl		switch	Join the baseline producer prices in one income list
38.34	IlArithOp		switch	Baseline consumer prices (2nd step, excise goods, overwrites prices of the 1st step)
38.35	IlArithOp		switch	Baseline implicit tax rate (1st step, all goods)
38.36	IlArithOp		switch	Baseline implicit tax rate (2nd step, excise goods, overwrites implicit tax rates of the 1st step)
38.37	IlArithOp		switch	Consumer prices (1st step, all goods)
38.38	IlArithOp		switch	Implicit tax rate (1st step, all goods)
38.39	IlArithOp		switch	Consumer prices (2nd step, excise goods, overwrites prices of the 1st step)
38.40	IlArithOp		switch	Implicit tax rate (2nd step, excise goods, overwrites implicit tax rate of the 1st step)
38.41	DefConst		switch	Tip- Total variables (used in iteration)
38.42	Loop		switch	Tip- Iteration loop
38.43	IlArithOp		switch	Tip- Consumer prices (1st step, all goods)
38.44	IlArithOp		switch	Tip- Consumer prices (before iteration)
38.45	IlArithOp		switch	Tip- Implicit tax rate (2nd step, excise goods, overwrites implicit tax rate of the 1st step)
38.46	IlArithOp		switch	Tip- Consumer prices (2nd step, excise goods, overwrites prices of the 1st step)
38.47	Totals		switch	Tip- Creating totals (needed for iteration)
38.48	Totals		switch	Tip- Creating totals (needed for iteration)
38.49	DefIl		switch	Join the consumer prices in a new income list
38.50	IlArithOp		switch	Tip- Change in consumer prices (in percentage)
38.51	IlArithOp		switch	Tip- Change in implicit tax rate (in percentage points)
38.52	IlArithOp		switch	Producer prices
38.53	IlArithOp		switch	Producer prices, adjusted (p_base=1)
38.54	IlArithOp		switch	Baseline consumer prices, adjusted (p_base=1)
38.55	IlArithOp		switch	Consumer prices, adjusted (p_base=1)
38.56	IlArithOp		switch	Baseline specific excises, adjusted (p_base=1)
38.57	IlArithOp		switch	Specific excises, adjusted (p_base=1)
38.58	IlArithOp		switch	Baseline producer prices, adjusted (p_base=1)
38.59	IlArithOp		switch	Baseline implicit tax rate

Figure 5: Calculating the post reform producer and consumer prices

1. Once the parameters and associated income lists are created, ITT generates the baseline producer prices for the non-excise goods via the `29.Baseline producer prices (1`st `step, non-excise goods)` function. This function sets the baseline producer prices of non-excise goods at 1 euro.[34]

$$\texttt{\$tco\_base\_p} = 1.$$

Then, the function `30.Baseline consumer prices (1`st `step, non-excise goods)` creates baseline prices for non–excise goods using Equation (21) (see Section 5.1) for the case where excises are zero. This is the first step of a two step procedure and baseline consumer prices are defined as:

$$\texttt{\$tco\_base\_q} = 1 + \texttt{\$tco\_base\_t}.$$

The function `31.Join the baseline consumer prices in an income list` creates the

[34] The COICOP commodity suffix `_[xxxxx]` is dropped here for notational simplicity.

baseline consumer price income list `il_tco_base_q`. Recall that baseline consumer prices for the excise goods are already stored in tco_cc as a baseline parameter value. Hence, `il_tco_base_q` not only includes the baseline consumer prices of the non–excise goods, but also the baseline consumer prices of the excise goods.

2. In the second step, `32.Baseline producer prices (2nd step)` function calculates the baseline producer prices for the excise goods using:

$$\texttt{\$tco\_base\_p} = \texttt{\$tco\_base\_q}\left(1 - \texttt{\$tco\_base\_v} - \left(\frac{\texttt{\$tco\_base\_t}}{1 + \texttt{\$tco\_base\_t}}\right)\right) - \texttt{\$tco\_base\_a}.$$

   Then `33.Join the baseline producer prices in an income list` creates the baseline producer price income list `il_tco_base_p` which covers all commodities.

   Similarly, `34.Baseline consumer prices (2nd step, excise goods)` function calculates the baseline consumer prices for all goods using:

$$\texttt{\$tco\_base\_q} = \frac{\texttt{\$tco\_base\_p} + \texttt{\$tco\_base\_a}}{1 - \texttt{\$tco\_base\_v} - \left(\frac{\texttt{\$tco\_base\_t}}{1+\texttt{\$tco\_base\_t}}\right)}.$$

   Technically, this step overwrites the consumer prices of non–excise goods as well. However, in practice this does not make any difference since the formula above is identical to $\texttt{\$tco\_base\_q} = 1 + \texttt{\$tco\_base\_t}$ in the absence of excises ($\texttt{\$tco\_base\_a} = 0$, $\texttt{\$tco\_base\_v} = 0$) and producer prices equal to 1 ($\texttt{\$tco\_base\_p} = 1$).

3. The final set of baseline parameters are the baseline implicit tax rates, `$tco_base_tau`. `$tco_base_tau` is generated in a similar fashion as the prices, using a two step procedure. First, `35.Baseline implicit tax rate (1st step, all goods)` sets the implicit tax rate for all goods:

$$\texttt{\$tco\_base\_tau} = \texttt{\$tco\_base\_t}.$$

   `36.Baseline implicit tax rate (2nd step, excise goods, overwrites implicit tax rate of the 1st step))` overwrites the implicit tax rate values for the excise goods.

$$\texttt{\$tco\_base\_tau} = \frac{(1 + \texttt{\$tco\_base\_t})}{1 - (1 + \texttt{\$tco\_base\_t})\left(\texttt{\$tco\_base\_v} + \frac{\texttt{\$tco\_base\_a}}{\texttt{\$tco\_base\_q}}\right)}.$$

4. After creating the baseline parameters, `tco_cc` starts creating the post reform parameters.

   The function `37.Consumer prices (1st step, all goods)` creates consumer prices and set their value equal to baseline consumer prices:

$$\texttt{xq} = \texttt{\$tco\_base\_q}.$$

Similarly, `38.Implicit tax rate (1`st `step, all goods)` assigns `tau` its baseline value:

$$\texttt{tau} = \texttt{\$tco\_base\_tau}.$$

Then `39.Consumer prices (2`nd `step, excise goods, overwrites prices of the 1`st `step)` calculates the consumer price `xq` with the new indirect tax parameters:

$$\texttt{xq} = \frac{\texttt{\$tco\_base\_p} + \texttt{\$tco\_a}}{1 - \texttt{\$tco\_v} - \left(\frac{\texttt{\$tco\_t}}{1+\texttt{\$tco\_t}}\right)}.$$

Notice that at this stage, post reform consumer prices are still calculated under fixed producer price assumption (`$tco_base_p`). This assumption will be relaxed when when $XTip$ is switched *on* and θ is set at value different from one.

Next, `40.Implicit tax rate (2`nd `step, excise goods, overwrites implicit tax rate of the 1`st `step)` calculates `tau` using the reform indirect tax parameters:

$$\texttt{tau} = \frac{(1 + \texttt{\$tco\_t})}{1 - (1 + \texttt{\$tco\_t})\left(\texttt{\$tco\_v} + \frac{\texttt{\$tco\_a}}{\texttt{xq}}\right)}.$$

5. When $XTip$ is switched off, ITT bypasses the iteration procedure and directly calculates the producer prices (`52.Producer prices`).

   When $XTip$ is switched *on*, ITT runs over a set of functions which form the iteration procedure. The iteration procedure starts with generating the two variables over which the `Loop` function will iterate (`41.Tip-Total variables (used in iteration)`). The initial values of these two variables are set to zero:

$$\texttt{ittxq0\_total\_il\_ittq0} = 0,$$

   and

$$\texttt{ittxq1\_total\_il\_q1} = 0.$$

   Then, `42.Tip-Iteration loop` defines the iteration procedure (Figure 6) and `43.Tip-Consumer prices (1st step, all goods)` sets the consumer prices back at their baseline level.

$$\texttt{xq} = \texttt{\$tco\_base\_q}.$$

   The baseline consumer prices will be the starting values for the iteration procedure.

	Policy		Grp/No	BE_2010	BE_2019	Comment
38.40	fx Loop			switch	switch	Iteration loop
38.40.1		Loop_Id		Itr	Itr	
38.40.2		BreakC...		0.0001> abs((ittxq1_total_il_q1-ittxq0_total_il_ittq0)/ittxq0_total_il_ittq0)	0.0001> abs((ittxq1_total_il_q1-ittxq0_total_il_ittq0)/ ittxq0_total_il_ittq0)	
38.40.3		First_F...		5043119c-5a65-4f89-baf0-b7143d1b069f	9ee2d500-aaf8-4799-83f7-3b7d76a89dec	
38.40.4		Last_Func		ac12ace7-07f8-4a6e-83db-cf78e5832cd1	02d9a2cc-c7e2-47a9-b3d0-216dbe5689e4	

Figure 6: The `Loop` function

The `42.Tip-Iteration loop` function iterates between two functions whose `Identifiers` are given in the `First_Func` and `Last_Func` cells. The iteration procedure runs over functions `44.Tip-Consumer prices (before iteration)` till `48.Tip-Creating totals (needed for iteration)`, until the stopping rule is met. Notice that the `Identifier` values are policy system specific.

`BreakCond` defines the stopping rule of the loop: once the absolute percentage change in the totals, corresponding to $(|q_1^{(n)} - q_1^{(n-1)}|)/\ q_1^{(n-1)}$, is smaller than 0.0001, the `Loop` function stops iterating.

The `44.Tip-Consumer prices (before iteration)` creates a new consumer price variable `ittxq0`, `ittxq0` is the `tco_cc` equivalent of $q_1^{(n-1)}$ and its value is set to the consumer price.

$$\texttt{ittxq0} = \texttt{xq}.$$

Notice that just before starting the iteration, `xq` is equal to the baseline consumer price `$tco_base_q`, which means in the notation of Section 5.1.1 that $q_1^{(0)}$ is set equal to q_0.

Next, the `45.Tip-Implicit tax rate (`2^{nd}` step, excise goods, overwrites implicit tax rate of the `1^{st}` step)` function calculates `tau` according to Equation (28):

$$\texttt{tau} = \frac{(1 + \texttt{\$tco\_t})}{1 - (1 + \texttt{\$tco\_t})\left(\texttt{\$tco\_v} + \frac{\texttt{\$tco\_a}}{\texttt{xq}}\right)} - 1.$$

Hence, in the very first iteration the value of `tau` will be equal to:

$$\tau_1^{(1)} = \frac{1 + t_1}{1 - (1 + t_1)\left(v_1 + \frac{a_1}{q_1^{(0)}}\right)} - 1.$$

This `tau` value will be plugged in in `46.Tip-Consumer prices (`2^{nd}` step, excise goods, overwrites prices of the `1^{st}` step)`, which recalculates the consumer price `xq` with the updated `tau` value, in accordance with Equation (29):

$$\texttt{xq} = \texttt{\$tco\_base\_q}\left(1 + \texttt{\$tco\_theta}\left(\frac{\texttt{tau} - \texttt{\$tco\_base\_tau}}{1 + \texttt{\$tco\_base\_tau}}\right)\right).$$

For the very first iteration, this corresponds to:

$$q_1^{(1)} = q_0\left(1 + \theta\frac{\tau_1^{(1)} - \tau_0}{1 + \tau_0}\right).$$

The next two functions `47.Tip-Creating totals (needed for iteration)` and `48.Tip-Creating totals (needed for iteration)` creates the total values that are used in the stopping rule `BreakCond` in `42.Tip-Iteration loop` (Figure 7).

	Policy		Grp/No	BE_2010	BE_2019	Comment
38.44	fx Totals			switch	switch	Creating totals (needed for iteration)
38.44.1		TAX_UNIT		tu_individual_be	tu_individual_be	
38.44.2		WarnIfDuplicateD...		no	no	
38.44.3		Varname_Max		ittxq0_total	ittxq0_total	
38.44.4		Agg		il_ittq0	il_ittq0	
38.45	fx Totals			switch	switch	Creating totals (needed for iteration)
38.45.1		TAX_UNIT		tu_individual_be	tu_individual_be	
38.45.2		WarnIfDuplicateD...		no	no	
38.45.3		Varname_Max		ittxq1_total	ittxq1_total	
38.45.4		Agg		il_q1	il_q1	

Figure 7: `Creating totals (needed for iteration)` functions

Finally, `49.Join the consumer prices in a new income list` creates an income list which contains the final consumer prices, `il_q`.

6. Then, using `50.Tip-Change in consumer prices (in percentage)` and `51.Tip-Change in implicit tax rate (in percentage points)`, `tco_cc` creates two sets of variables defining the percentage change in consumer prices (`d_xq`) and percentage point change in implicit tax rates (`d_tau`):

$$\texttt{d\_xq} = \frac{\texttt{xq} - \texttt{\$tco\_base\_q}}{\texttt{\$tco\_base\_q}},$$

and

$$\texttt{d\_tau} = \texttt{tau} - \texttt{\$tco\_base\_tau}.$$

7. The last prices `tco_cc` calculates, are the post–reform producer prices, `52.Producer prices`:

$$\texttt{xp} = \frac{\texttt{xq}}{(1 + \texttt{tau})}.$$

8. Once all prices are calculated, a set of functions normalize the prices so that the baseline producer price (`$tco_base_p`) will be equal to 1 euro for all commodities, including the excise

goods. This is done by dividing the prices by the initial level of the baseline producer price (`$tco_base_p`). The normalization of prices is vital since ITTv4 reports baseline consumer and post–reform producer prices. In ITTv3 the producer prices were fixed at 1 and therefore, reporting producer prices was redundant. The introduction of tax incidence parameters in ITTv4 makes producer price information relevant, as it gives information on the implied changes in the producer prices by the specified value of the tax incidence parameters.

– `53.Producer prices, adjusted (p_base=1)`:

$$\texttt{xp} = \frac{\texttt{xp}}{\texttt{\$tco\_base\_p}}.$$

– `54.Baseline consumer prices, adjusted (p_base=1)`:

$$\texttt{xqbase} = \frac{\texttt{\$tco\_base\_q}}{\texttt{\$tco\_base\_p}}.$$

– `55.Consumer prices, adjusted (p_base=1)`:

$$\texttt{xq} = \frac{\texttt{xq}}{\texttt{\$tco\_base\_p}}.$$

– `56.Baseline specific excises, adjusted (p_base=1)`:

$$\texttt{xabase} = \frac{\texttt{\$tco\_base\_a}}{\texttt{\$tco\_base\_p}}.$$

– `57.Specific excises, adjusted (p_base=1)`:

$$\texttt{xa} = \frac{\texttt{xa}}{\texttt{\$tco\_base\_p}}.$$

– `58.Baseline producer prices, adjusted (p_base=1)`:

$$\texttt{xpbase} = \frac{\texttt{\$tco\_base\_p}}{\texttt{\$tco\_base\_p}} = 1.$$

9. Finally, the last function `59.Baseline implicit tax rate` creates `taubase`. This variable already exists in ITT but under a different name `$tco_base_tau`).

5.3 Changes in the ITT output

In ITTv4, the number of commodity specific variables created in the simulations increased as well. Nevertheless, the set of variables exported by default is kept identical to ITTv3 in order to optimize execution time. If needed, the user can choose to export the new variables as well.

To do so, the user can simply add the variable names in the `output_std_cc` policy either by adding the variables for a specific coicop commodity or for a variable group as a whole. Figure 8 illustrates.

To export a single variable, the user should add the exact variable name, $\$xp\_01111$. Then Euromod will export the producer price for rice. When the user specifies a variable group using the prefix of the variable group name followed by an asterisk, $\$xq\_*$, then EUROMOD will export the consumer prices for all commodities.

39.1	DefOutput		on	on	
39.1.1	File		BE_2010_std	BE_2019_std	Output - default
39.1.2	file		BE_2010_CIS_std	BE_2019_CIS_std	Output - plus ITT, constant income shares reform
39.1.3	File		BE_2010_CQ_std	BE_2019_CQ_std	Output - plus ITT, constant quantities reform
39.1.4	File		BE_2010_CES_std	BE_2019_CES_std	Output - plus ITT, constant expenditure shares reform
39.1.5	Var		saving	saving	
39.1.6	Var		hh_head	hh_head	
39.1.7	nDecimals		4	4	
39.1.8	Var		$xp_01111	n/a	
39.1.9	VarGroup		$xq_*	n/a	

Figure 8: Exporting additional variables

Table 8 provides a comparison of the commodity specific variables in ITTv3 and ITTv4. The variables which are created by ITT but not exported by default are marked with an asterisk. Notice that the variable name of consumer prices which was `xq[xxxxx]` in ITTv3, is replace by `xq_[xxxxx]` in ITTv4.

When the user runs a simulation for a θ-value other than 1, ITT will automatically create specific excise and *ad valorem* excise variables also for goods not subject to excises, because these variables are required for the iteration algorithm. However, this is simply a technical requirement, and they will play no role. The specific excise and *ad valorem* excise values will be set to zero for those goods. Some other variables, such as the tax incidence parameter θ, will be created only when the $XTip$ extension is set *on*. In Table 8 these variables are indicated with a diamond symbol.

5.4 Addition of the baseline extension

ITT simulates the baseline scenario using the information on the income shares of expenditures. Formally, the baseline simulation is a simulation under the CS_Y assumption, using the baseline parameters. Hence, in ITTv3, the execution of a baseline scenario and a reform simulation under CS_Y were practically identical. However, this setting was confusing and created some drawbacks for the user. When both, the baseline and reform simulations, were implemented on the same system, a reform under CS_Y would overwrite the output of the baseline scenario and vice versa. Hence, the user had to re-simulate the baseline scenario once a reform scenario was simulated underthe CS_Y assumption.

To make the conceptual difference between a baseline and a reform simulation clear, and to improve the user experience, ITTv4 includes a new extension, $XBase$, which specifies the baseline simulation. By the inclusion of $XBase$, a baseline scenario and a reform scenario under CS_Y are explicitly

Table 8: ITT output variable list

Variable name		**ITTv3** non–excise	**ITTv3** excise	**ITTv4** non–excise	**ITTv4** excise
Household variables					
xs[xxxxx]	Income shares	✓	✓	✓	✓
x[xxxxx]	Expenditure levels	✓	✓	✓	✓
xx[xxxxx]	Quantities	✓	✓	✓	✓
tva[xxxxx]	VAT payments	✓	✓	✓	✓
txv[xxxxx]	*Ad valorem* excise payments		✓	✓◇	✓
txa[xxxxx]	Specific excise payments		✓	✓◇	✓
x[xxxxx]_na	NA adjusted expenditures	✓	✓	✓	✓
tva[xxxxx]_na	NA adjusted VAT payments	✓	✓	✓	✓
txv[xxxxx]_na	NA adjusted *Ad valorem* excise payments		✓	✓◇	✓
txa[xxxxx]_na	NA adjusted specific excise payments		✓	✓◇	✓
Commodity variables					
xp_[xxxxx]	Producer price			*	*
xq_[xxxxx]	Consumer price	✓	✓	*	*
xa_[xxxxx]	Specific excise			*	*
theta_[xxxxx]	Tax incidence parameter			*◇	*◇
tau_[xxxxx]	Implicit tax rate			*	*
xpbase_[xxxxx]	Baseline producer price			*	*
xqbase_[xxxxx]	Baseline consumer price			*	*
xabase_[xxxxx]	Baseline specific excise			*	*
taubase_[xxxxx]	Baseline implicit tax rate			*	*
d_xq_[xxxxx]	Percentage change in consumer price			*◇	*◇
d_tau_[xxxxx]	Percentage point change in implicit tax rate			*◇	*◇

✓ Default, * Optional, ◇ *XTip on*

differentiated and the user is no longer required to repeat the baseline simulations after implementing a CS_Y reform.

The baseline extension, $XBase$, is set *on* by default for all countries and all systems. The remaining four extensions are set off by default. The user can easily switch between assumptions by adjusting the extensions accordingly (see Figure 9).

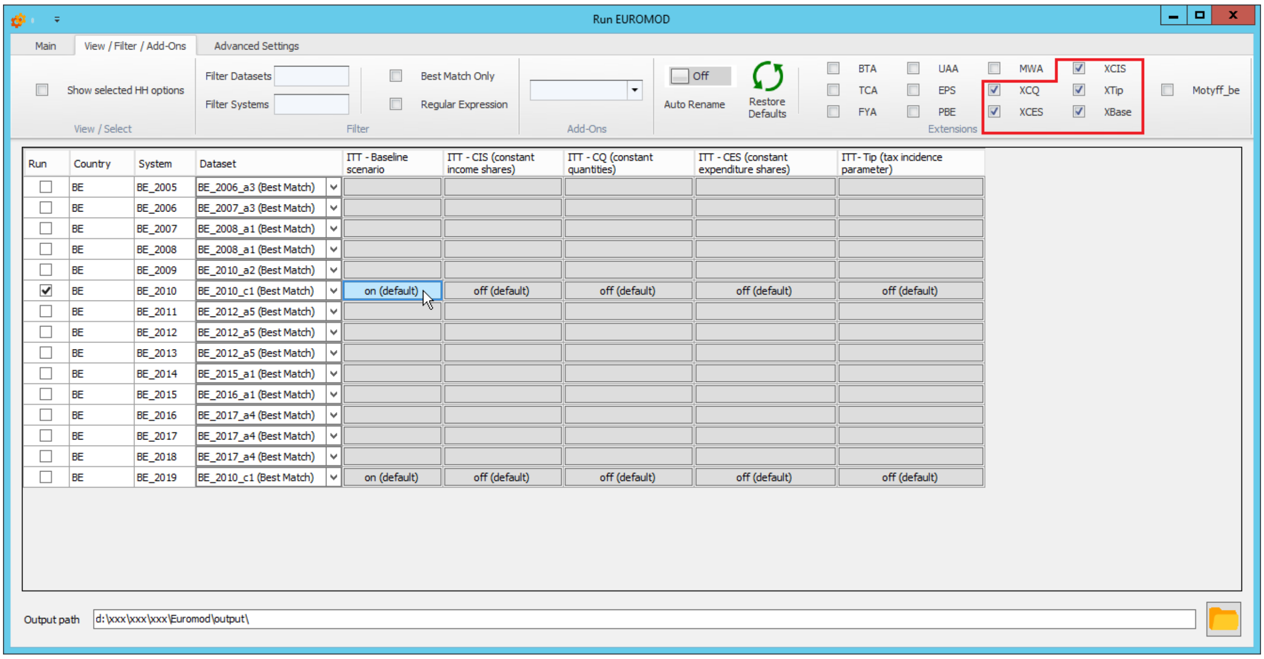


Figure 9: Addition of baseline extension

A similar issue was emerging with respect to the support files `tco_cq_txt` and `tco_ces_txt` which are required for the reform scenarios under CQ and CS_E respectively. In ITTv3, all three behavioral assumptions were producing these support files by default, and inevitably each simulation was overwriting the support files created in the previous simulation. Hence, the baseline simulation had to be repeated after each reform simulation to ensure that the support files were created by the baseline scenario. A way to overcome this, would be to use different systems for each reform scenario. However, implementing multiple systems would not always be practical, especially for cross-country analysis.

Hence, in ITTv4, the default output of the baseline and reform scenarios are rearranged. In ITTv4, the reform scenarios no longer produce the support files. Hence, the user is no longer required to re-simulate the baseline scenario after each reform simulation. Table 9 outlines the output structure of ITTv3 and the modifications in ITTv4.

Table 9: Baseline and reform simulations and corresponding output

	Baseline	**CS$_Y$**	**CQ**	**CS$_E$**
ITTv3				
Output file	cc_yyyy_**cis_std**.txt	cc_yyyy_**cis_std**.txt	cc_yyyy_**cq_std**.txt	cc_yyyy_**ces_std**.txt
Support file for CQ	cc_yyyy_**tco_cq**.txt	cc_yyyy_**tco_cq**.txt	cc_yyyy_**tco_cq**.txt	cc_yyyy_**tco_cq**.txt
Support file for CS_E	cc_yyyy_**tco_ces**.txt	cc_yyyy_**tco_ces**.txt	cc_yyyy_**tco_ces**.txt	cc_yyyy_**tco_ces**.txt
ITTv4				
Output file	cc_yyyy_**std**.txt	cc_yyyy_**cis_std**.txt	cc_yyyy_**cq_std**.txt	cc_yyyy_**ces_std**.txt
Support file for CQ	cc_yyyy_**tco_cq**.txt			
Support file for CS_E	cc_yyyy_**tco_ces**.txt			

5.5 Other changes in the ITT module

The monetary values in the EM output are always given in euros for countries which are currently in the euro area. This is irrespective of the currency of the EM SILC used as the input dataset for the simulations which could be either euro or national currencies depending on the collection year (Figure 10).

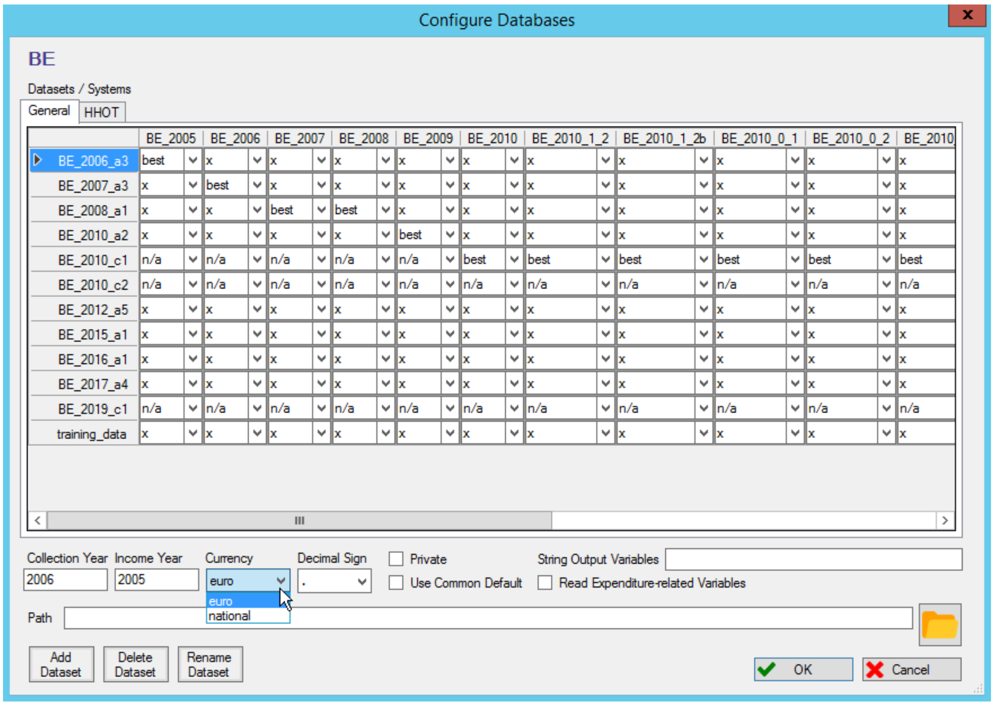
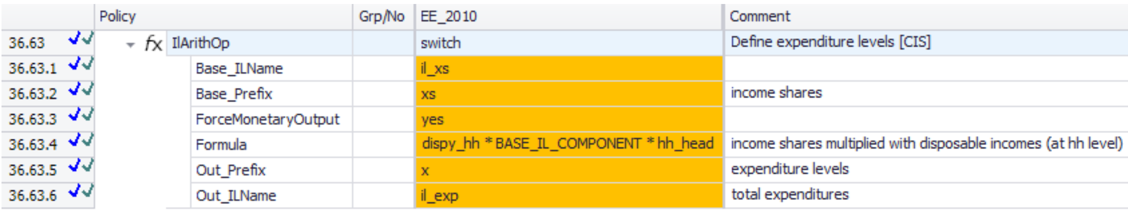


Figure 10: Database configuration

Moreover, although the default EM output for non-euro area member countries is in their national currencies, EUROMOD also has an option to export all monetary variables in euros.

`tco_cc` policy heavily uses `IlArithOp` function to create monetary variables. Nevertheless, `IlArithOp` function does not define its output variables as monetary by default.[35] This is why the `ForceMonetaryValues` parameter is introduced to the `IlArithOp` function and Figure 11 illustrates how this parameter should be utilized.

[35] This is an intentional choice since not all arithmetic operations with a monetary input would produce a monetary output.



	Policy	Grp/No	EE_2010	Comment
36.63	fx IlArithOp		switch	Define expenditure levels [CIS]
36.63.1	Base_ILName		il_xs	
36.63.2	Base_Prefix		xs	income shares
36.63.3	ForceMonetaryOutput		yes	
36.63.4	Formula		dispy_hh * BASE_IL_COMPONENT * hh_head	income shares multiplied with disposable incomes (at hh level)
36.63.5	Out_Prefix		x	expenditure levels
36.63.6	Out_ILName		il_exp	total expenditures

Figure 11: The `IlArithOp` function

In `63.Define expenditure levels [CIS]`, `tco_cc` multiplies a monetary variable (`dispy_hh`) with a number (`xs[xxxxx]`) but the product (`x[xxxxx]`) is not identified as a monetary variable by default. To do that, the user should set the `ForceMonetaryValues` parameter to "`yes`". This way, EUROMOD will acknowledge expenditures, `x[xxxxx]`, as a variable expressed in monetary terms and, when necessary, convert the values in euros.

In total, `tco_cc` creates four monetary variables using `IlArithOp` function, i) expenditure levels, ii) VAT, iii) *ad valorem* excise, and iv) specific excise liabilities. Table 10 provides the specific functions creating these four monetary variables under the three different behavioral assumptions.

Table 10: Specific `tco_cc` functions creating monetary variables

Variable		Behavioral assumption	`tco_cc` function	
`x[xxxxx]`	Expenditure levels	CS_Y	63.	`Define expenditure levels [CIS]`
		CQ	67.	`Define exp levels [CQ]`
		CS_E	76.	`Define expenditure levels [CES] - step 3 (combine step 1 & 2)`
`tva[xxxxx]`	VAT	All	77.	`Compute VAT`
`txv[xxxxx]`	*Ad valorem* excises	All	78.	`Calculate` *`ad valorem`* `excises`
`txa[xxxxx]`	Specific excises	All	79.	`Calculate specific excises`

While `tco_cc` routine creating indirect tax liabilities does not change under different behavioral assumptions, `tco_cc` uses different functions to compute the expenditure levels for each behavioral scenario. Hence, in total there are six functions creating these four monetary variables, and `ForceMonetaryValues` parameter should be added to all these six functions.

The `ForceMonetaryValues` parameter has an important role in the ITT simulation of non-euro area member countries Bulgaria, Czechia, Denmark, Hungary, Croatia, Poland, Romania and Sweden. Without this parameter, EUROMOD wouldn't be able to convert the above mentioned monetary variables in euros when "`All output in €`" option is set "`on`".

However, the use of `ForceMonetaryValues` is crucial for the simulations of Estonia, Lithuania and Latvia. These countries are currently in the euro area, but they entered the eurozone after 2010. Since they were still using national currencies in 2010, the monetary parameters in their 2010 systems, such as commodity prices and specific excises, are still given in their national currencies. Moreover, the monetary variables in their 2010 EM SILC dataset are also is in national currencies. Therefore, verification of their 2010 simulations requires the proper identification of the monetary variables since the default EM output of these countries is in euros like any other country which are currently in the euro area.

As of now, the income shares of expenditures are imputed on 2010 EM SILC and `tco_cc` is operational only for 2010 and 2019 systems. Hence, the currency conversion concerning `tco_cc` is relevant either for countries which entered the eurozone after 2010 or for countries which are still not in the euro area. The default setting of `tco_cc` covers all these cases.

However, if `tco_cc` would be expanded to cover pre-2010 systems, conversions might be needed for other countries depending on when these countries joined the eurozone. Table 11 shows the year the country joined the euro area and the earliest EUROMOD system year. When using `tco_cc` for a system prior to the country's euro area entry, the user should switch on `ForceMonetaryValues` as illustrated above.

Table 11: The year EU countries joined the eurozone

	Earliest EUROMOD system year	**Euro area entry year**
Austria	2007	1999
Belgium	2005	1999
Bulgaria	2007	
Cyprus	2006	2008
Czechia	2005	
Germany	2007	1999
Denmark	2007	
Estonia	2005	2011
Greece	2005	2001
Spain	2005	1999
Finland	2007	1999
France	2006	1999
Hungary	2005	
Croatia	2011	
Ireland	2006	1999
Italy	2005	1999
Lithuania	2005	2015
Luxembourg	2007	1999
Latvia	2006	2014
Malta	2007	2008
The Netherlands	2006	1999
Poland	2005	
Portugal	2006	1999
Romania	2007	
Sweden	2006	
Slovenia	2006	2007
Slovakia	2006	2009

5.6 Updates to the Statistics Presenter

The current Statistics Presenter (SP) already includes a module, `Indirect Taxes Analysis`, that provides a standard analysis of the ITT output. In the most recent EUROMOD release, the SP also provides an overview of the changes between the baseline and the reform scenarios, `Indirect Taxes Analysis-Reform`. `Indirect Taxes Analysis-Reform` not only provides budgetary and distributional statistics, and statistics by major COICOP categories, it also calculates the changes in output statistics.

To start, the user needs to select `Indirect Taxes Analysis-Reform` in the SP menu (Figure 12).

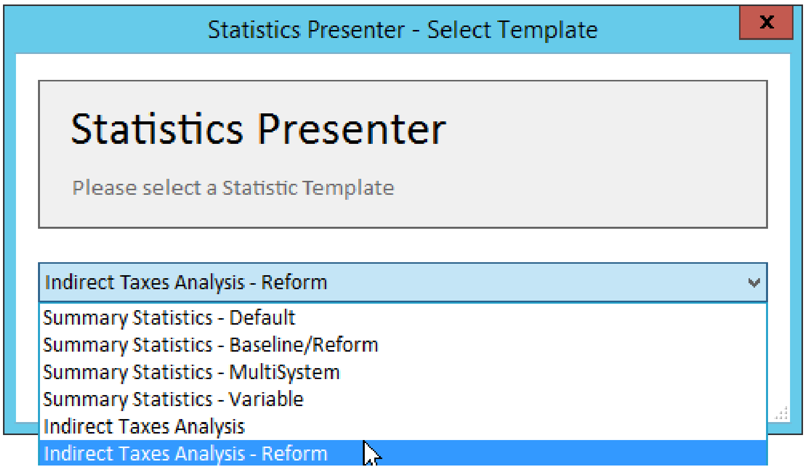


Figure 12: New Statistics Presenter module

Then the simulation output for the baseline and reform scenarios need to be selected (Figure 13).

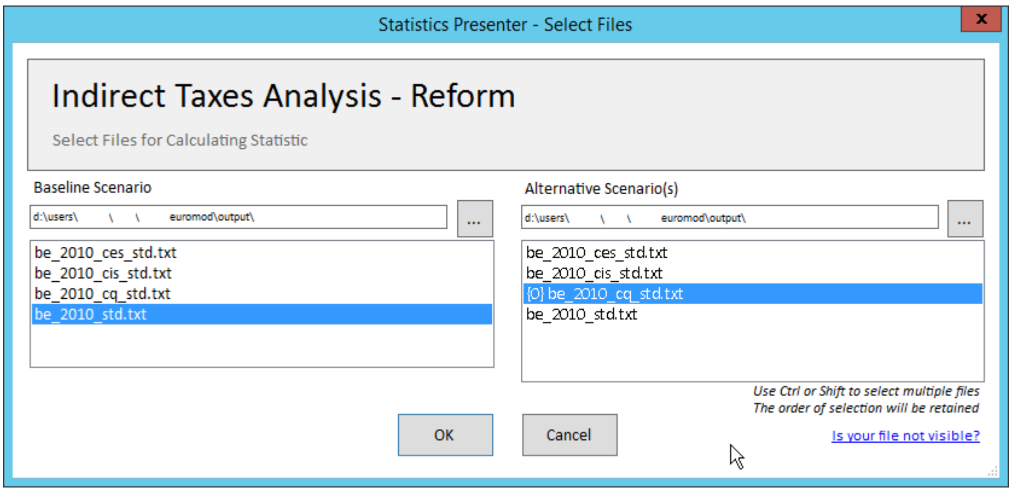


Figure 13: Selecting baseline and reform output

`Indirect Taxes Analysis-Reform` displays eight output tables identical to those of `Indirect`

`Taxes Analysis` (Figure 14). The new feature of `Indirect Taxes Analysis-Reform` is the addition of a set of tabs that makes it easier to navigate between different reform scenarios. These tabs are: `Baseline` (Figure 14), `Reform` (Figure 15), and `Difference` (Figure 16). The tables in the `Difference` tab display level changes in the indirect tax and budget components, from the baseline scenario to a reform scenario. SP allows the user to choose multiple reform scenarios as well (Figure 17, Figure 18). All tables can also be saved in Excel-files.

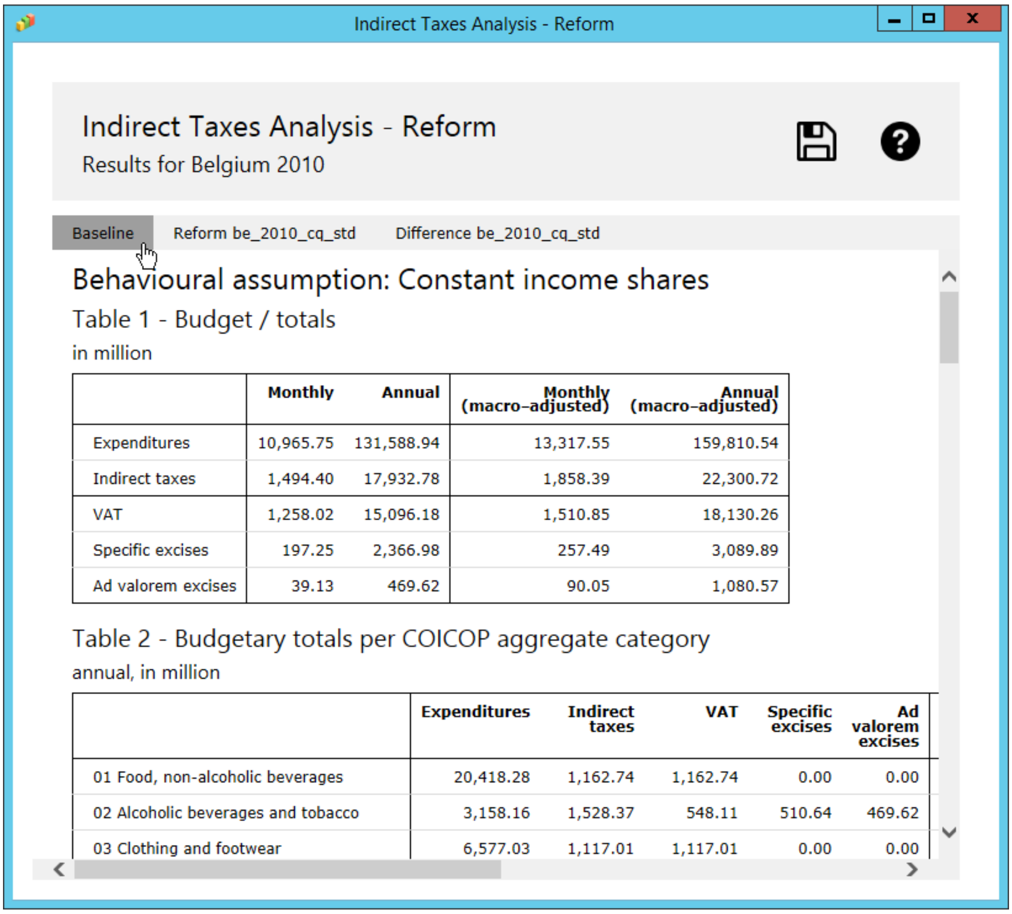


Figure 14: Baseline tab

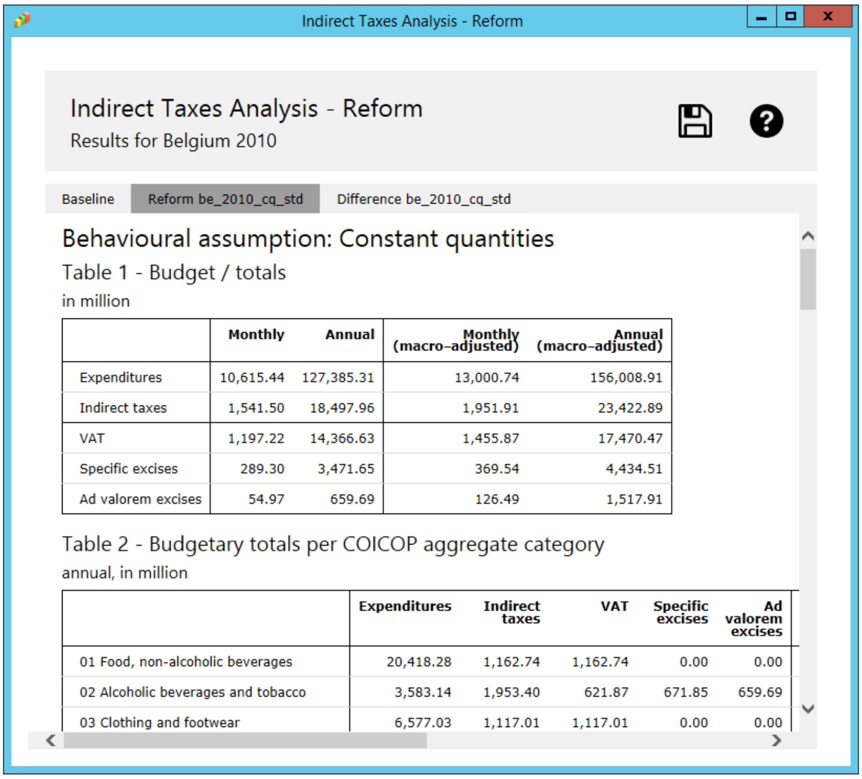


Figure 15: Reform tab

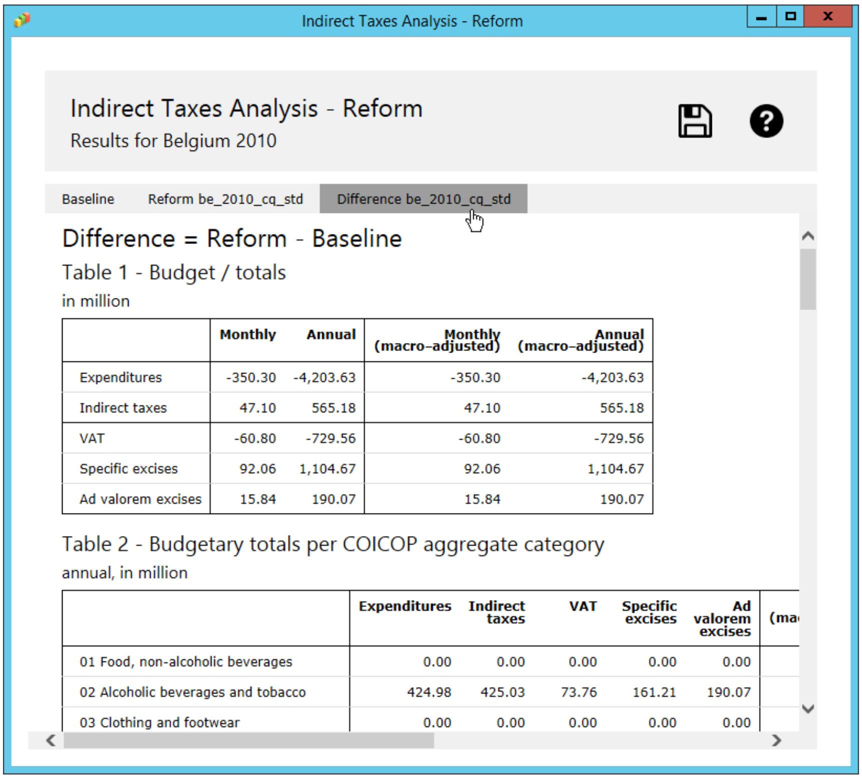


Figure 16: Difference tab

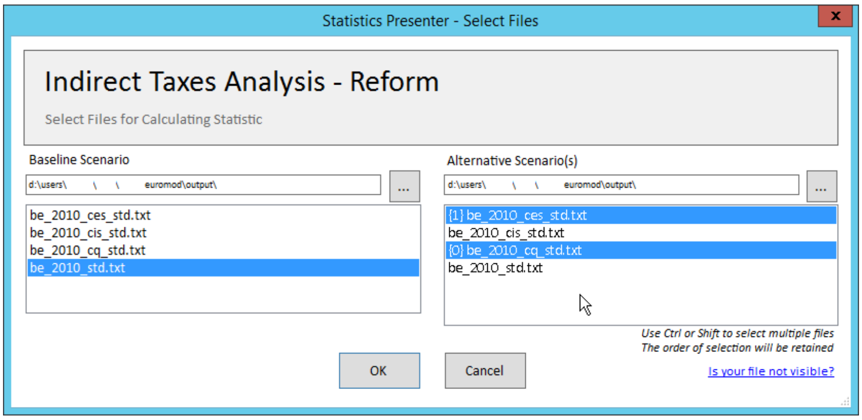


Figure 17: Selecting baseline and multiple reform output

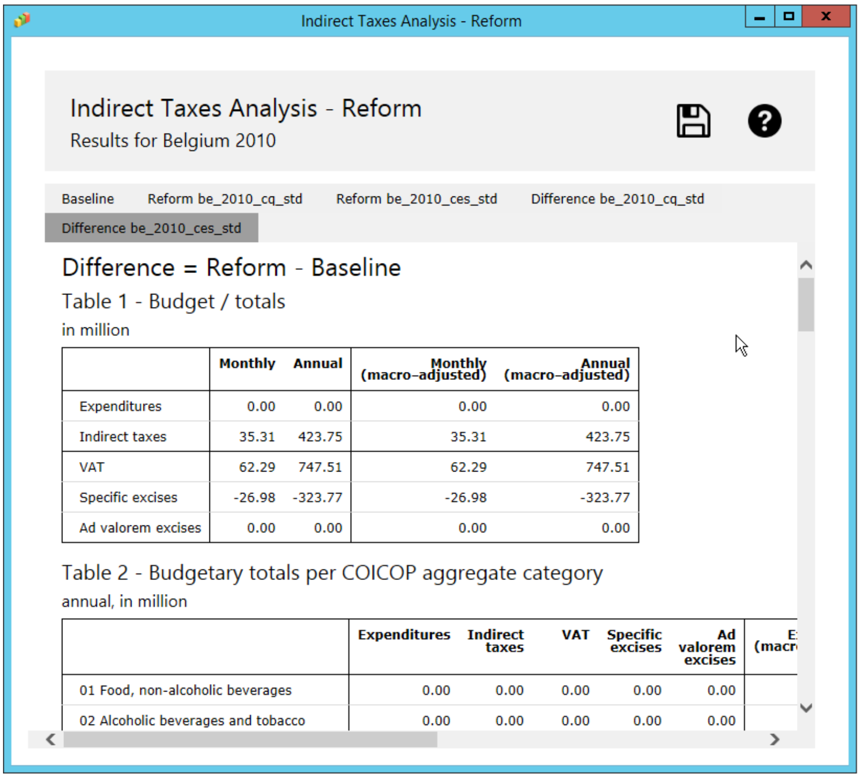


Figure 18: Difference tab with multiple reforms

5.7 Using ITT in STATA

ITT is compatible with STATA. The user can run ITT with command line parameters. To do that, the user should specify the path to the EUROMOD project and the name of the system to run a simulation in STATA. There are also several other optional parameters.

– $-emPath < project\_path >$: path to the EUROMOD project
– $-sys < system\_name >$: name of the system

optional:

– $-data < dataset\_name >$: name of the dataset;
– $-extSwitch < extension\_name >=< value >$: setting switches;
– $-outPath < output\_path >$: output folder;
– $-dataPath < data\_path >$: input data folder;
– ...

In order to use ITT in STATA, the user should set ITT switches, $XBase$, $XCIS$, XCQ, $XCES$ and $XTip$, accordingly. All switches can be added simultaneously. Figure 19 illustrates how a baseline simulation could be implemented using $STATA$.

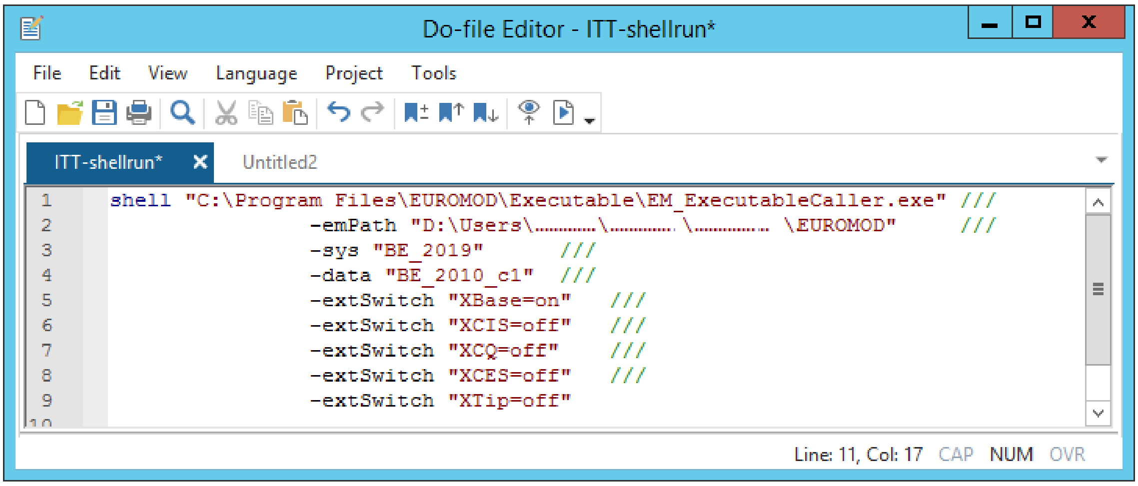


Figure 19: STATA shell run sample do file

6 Application: greening food taxation[36]

6.1 Introduction

In this last section we show the type of analyses the new ITT will allow for, by means of a number of policy reform simulations aiming at greening indirect taxes on food. At this stage, this exercise is mainly illustrative, and we thus focus on the method of analysis and broader insights the tool is able to provide, rather than on the concrete results.

There is a growing body of reports which stress that the ecological footprint of the food sector is non-negligible (see for example Crippa et al. (2021) and Ivanova et al. (2017)). Policies to mitigate this externality are currently largely absent, contrary to what is the case in many other sectors (see e.g. the EU Emissions Trading Scheme). Just as it is the case with existing policies to abate CO2eq emissions, the impact of the introduction of such measures in the food sector will be unequally distributed across households.[37] This might even be more of a concern in the food sector, as large(r) part of income goes to this expenditure category among the poorer persons in society: almost 30% of the household disposable income of the poorest decile in Belgium goes to food and non-alcoholic drinks, as compared to only slightly more than 9% for the richest decile (see infra and Table 13 for more details).

In the sequel, we perform two types of exercises. The first is a top-down approach, in which we investigate the distributional consequences of the introduction of a uniform carbon tax of 100 Euro per tonne of CO2eq emissions in the agricultural sector worldwide. Weitzman (2014) advocates a uniform carbon price as an efficient means to internalize the global warming externality, utilizing Pigouvian type of arguments. As is often the case, such Pigouvian efficiency arguments do not take into account distributional welfare costs. The possibility to assess these distributional costs is one of the trumps of EUROMOD's ITT.

Our second exercise is bottom-up. Instead of pricing carbon emissions directly, we evaluate the distributional consequences of increasing taxes on food commodities the production of which is deemed to be relatively more polluting. Theoretical reflections on the use(fulness) of indirect taxes as an additional carbon emission abatement instrument can be found in Jacobs and van der Ploeg (2019) and Stiglitz (2019). Here we follow a simulation based approach in three steps. First, we look at the distributional effects of the tax reform per se. Second, (partial) income recycling under the form of lowering tax rates on less polluting goods or through a lump sump redistribution is studied. Finally, we assess the impact of producer's reactions on the new tax rates through exploiting the newly introduced tax incidence parameter into ITTv4 (see Section 5.1).

[36] This section benefited from a collaboration with Jordan Hristov, Sofia Maier, Ignacio Perez Dominguez, Mattia Ricci, and Toon Vandyck from JRC-Sevilla, who provided us with the results of a simulation of the introduction of a carbon tax in the agricultural sector with the CAPRI-model.

[37] CO2eq stands for CO2-equivalent. Some other greenhouse gases affect global warming, and their impact is measured in CO2eq tonnes of emissions.

As the purpose of this section is mainly illustrative, we limit our scope of analysis throughout on one country: Belgium. Furthermore we limit ourselves here to simulations under one of the three behavioral assumptions on consumer demand implemented in ITT: constant income shares.[38]

In the sequel, we first discuss the baseline with respect to which we will compare the simulation results of different reform scenarios. We then give more detail on the different reform scenarios we have implemented. Finally, we turn to the analysis of the simulation results.

6.2 Baseline simulation

The baseline simulation uses the Belgian EUROMOD policy system for 2019. The input dataset is the Belgian 2010 EU–SILC, containing 2009 incomes. Hence, the EUROMOD up-rating tool will adapt all incomes in a way that reflects the average increases in market incomes and social benefits between 2009 and 2019.[39] Household demographics are not adjusted. The EUROMOD direct tax calculator thus produces 2019 household disposable incomes.

The ITT module requires an input dataset which contains for each EU-SILC household a vector of income shares of expenditures on a set of commodities.[40] These income shares are denoted by w_k^h, where k indexes the commodity. In this case we use thereto the imputation of the Belgian EUROSTAT 2010 HBS, which results in an imputed vector of income shares at the most detailed level available in this HBS (usually COICOP level 4). These income shares are parameters of the ITT module. They allow to calculate baseline expenditures on commodities at the same level of detail, VAT liabilities, *ad valorem* and specific excise liabilities, and to calculate individual welfare metrics.

Baseline expenditure levels for each commodity are constructed by multiplying the simulated household disposable incomes with the imputed income shares.

Akoğuz et al. (2020) describes how ITT calculates per commodity the tax liabilities. In our baseline the 2019 commodity specific statutory tax parameters, and the 2019 consumer prices for the goods subject to specific excises are used to perform the calculations.

The ITT baseline also computes the implicit indirect tax rate for each commodity k, τ_k , and uses these rates to derive quantities consumed of each commodity k for every household h, by dividing the household expenditures on good k, denoted by e_k^h, by $1 + \tau_k$.[41] Total household expenditures are calculated by summing expenditures on all individual commodities, and saving is the residual of disposable income and total expenditures.

[38] ITT operates under three different behavioral assumptions: i) constant income shares, ii) constant quantities, and iii) constant expenditure shares. Akoğuz et al. (2020) provides a detailed review of the implications of the different behavioral assumptions.

[39] The uprating of gross incomes and the simulation of household disposable incomes are explained in more detail in the EUROMOD Country Report for Belgium 2016–2019 available at `euromod.ac.uk/country-reports/BE.pdf`.

[40] Income shares of expenditures on a commodity k are defined as the household's expenditures on a commodity k, e_k^h, divided by householdis disposable income, y^h.

[41] Quantities are measured in monetary terms at 2019 producer prices.

The ITT baseline run produces a new individual level EUROMOD output file, that is identical to that of a standard EUROMOD run, but expanded with a large number of ITT-related variables. More specifically, the output file contains for each household and each commodity, the expenditures, VAT and excise liabilities, consumer prices, quantities, and if one chooses to, the tax parameters, the value of the tax incidence parameter, and implicit indirect tax rates and producer prices in baseline and reform.

Aggregating the outcomes to the national level yields a number of macro–statistics, which we present in Table 12.[42] Annual gross earned incomes amount to 184.7 • billion euros. Direct taxes sum up to 48.9 billion euros, whereas employee's social security contributions totalled 24.2 billion euros. All social benefits, including pensions, amounted to 66.4 billion euros. Disposable income and total household expenditures totalled 177.9 billion and 159.8 billion euros respectively. Saving amounts to 18.1 billion euros. Indirect tax revenues are 22.4 billion euros (14% of total expenditures) of which 81.7% levied through VAT, 16.3% through specific excises, and 2% through *ad valorem* excises.

Table 12: Macro indicators at the baseline (Annual, in million Euros)

Gross earned income	184 656.1
Direct taxes	48 929.1
Social security contributions	24 192.3
Benefits	66 405.8
Disposable household income	177 940.4
Total expenditures	159 840.6
Saving	18 099.8
VAT revenues	18 337.6
Ad valorem excises	455.0
Specific excises	3 650.9

When interpreting the imputed income shares of expenditures, w_k^h, as parameters of Cobb-Douglas preferences, one can construct a preference based welfare metric using the ITT calculated consumer prices, q_k (consumer prices are recollected in the vector $\mathbf{q}$). This welfare metric indicates the amount of disposable income a household would need when prices of all commodities were equal to one, in order to be equally well off as when facing the actual consumer prices, $\mathbf{q}$, with actual disposable income y^h.

This welfare index is defined as:

$$Q_y^h = \frac{y^h}{P_{w^h}(\mathbf{q})}, \tag{36}$$

[42] The figures in the table are grossed up at national level using the SILC survey weights, but not calibrated to national account statistics. There are negligible differences with an identical baseline run in Section 8 of (Akoğuz et al., 2020). This difference is driven by two households (idhh=2083 and idhh=9446). These households were excluded from the analysis in Section 8 of (Akoğuz et al., 2020) but they are kept in the current analyses. Unless otherwise indicated, all results below exclude households for which the EUROMOD direct tax calculator returns negative household disposable incomes.

where $P_{w^h}(\mathbf{q})$ is a Stone price index, defined as:

$$P_{w^h}(\mathbf{q}) = \prod_k (q_k)^{w_k^h}. \tag{37}$$

The Stone price index is thus a weighted geometric mean of consumer prices of each commodity. Weights are the household's income shares of expenditures of these commodities. The price of saving is included in this price index as well, but it is set equal to one. Note that these price indices are household specific since they are calculated based on income shares at the household level.

This welfare concept can be interpreted also as a quantity index, as it divides the disposable income by a price index. Hence, we will refer to this welfare metric in the sequel as real disposable income. In order to convert it to an individual welfare measure, we divide the measure defined in Equation (36) by an equivalence scale.[43] Our welfare analyses below are based on this equivalized real disposable income concept.

Table 13: Income shares of expenditures on broad commodity aggregates (COICOP level 1)

Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Food	28.26%	19.11%	17.81%	16.63%	15.19%	14.50%	13.08%	12.40%	11.17%	9.32%	15.75%
Alc. & Tob.	5.57%	3.33%	2.35%	2.25%	2.20%	1.98%	1.91%	1.51%	1.37%	1.19%	2.37%
Cloth	5.65%	4.24%	3.72%	5.26%	5.17%	5.36%	4.80%	4.91%	4.04%	3.90%	4.70%
Hous	41.41%	24.68%	20.90%	16.35%	14.31%	12.99%	11.36%	9.99%	8.96%	7.05%	16.80%
Equipment	11.24%	6.82%	6.25%	8.50%	5.55%	6.22%	5.99%	5.63%	5.35%	5.11%	6.67%
Health	7.52%	5.47%	6.01%	6.43%	5.74%	5.05%	4.45%	4.11%	3.82%	3.35%	5.20%
Transport	29.48%	17.05%	18.90%	15.59%	12.85%	13.26%	10.88%	11.12%	8.11%	8.50%	14.58%
Comunication	5.39%	3.90%	3.41%	3.20%	2.77%	2.86%	2.51%	2.31%	2.10%	1.65%	3.01%
Recreation	13.31%	9.25%	8.29%	8.60%	8.24%	9.20%	9.00%	8.75%	8.05%	6.56%	8.92%
Education	0.20%	0.42%	0.36%	0.42%	0.60%	0.70%	0.54%	0.28%	0.35%	0.41%	0.43%
Horeca	8.12%	5.48%	6.14%	5.79%	6.37%	6.79%	6.37%	6.30%	5.89%	5.43%	6.27%
Other	17.84%	12.09%	13.04%	12.12%	12.59%	11.26%	10.54%	10.58%	8.70%	8.48%	11.73%

Deciles are constructed with respect to the population of individuals, on the basis of equivalized real disposable incomes. Income shares are household expenditures on a commodity aggregate divided by household disposable income. Entries are means of the shares taken with respect to the population of individuals belonging to a certain cell. Households with negative disposable incomes are included.

As the income shares play such an important role in assessing the welfare effects of the simulation results, we produce a summary of them across the baseline welfare distribution in Table 13. Commodities are further aggregated into the twelve COICOP level 1 categories. Expenditures on food and non-alcoholic beverages turn out to be the second most important category in terms of income shares (on average 16%) only dominated by spending on housing (rents plus utilities). The three most important categories (Housing, Food, and Transport) also exhibit the strongest welfare gradient. This motivates an a priori concern for potentially adverse distributional effects of policy

[43] The modified OECD equivalence scale is used: The scale for a single person household is 1. Each additional adult counts for 0.5 persons, each additional child for 0.3 persons. An adult is any household member aged at least 14 years.

interventions in the food sector to mitigate CO2eq emissions. Table 33 in Appendix I gives more detail on the income shares within the food category.

Notice that average income shares of all COICOP 1 aggregates quite strongly decline from the first to the second welfare decile. This indicates in the first place that negative saving occurs predominantly in the lowest welfare decile. When we turn to expenditure shares, that is expenditures on commodity k divided by total expenditures on all commodities, this sharp drop indeed disappears for most good categories, except for housing, and to a lesser extent for transport, as shown in Table 14. Expenditure shares of food and non-alcoholic beverages are hardly declining through the welfare distribution. But notice that the expenditure shares of some other categories, such as clothing, recreation, restaurant and other, are increasing through the welfare distribution.

Table 14: Expenditure shares of expenditures on broad commodity aggregates (COICOP level 1)

Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Food	18.27%	19.36%	18.91%	18.28%	18.31%	17.73%	17.61%	17.28%	17.78%	16.67%	18.02%
Alc. & Tob.	3.38%	3.18%	2.45%	2.39%	2.47%	2.30%	2.36%	1.95%	2.02%	2.09%	2.46%
Cloth	3.11%	3.95%	3.73%	5.26%	5.52%	5.84%	5.78%	6.43%	5.89%	6.25%	5.18%
Hous	26.51%	23.56%	21.67%	17.90%	15.69%	15.22%	14.48%	13.35%	13.59%	12.24%	17.42%
Equipment	5.36%	5.79%	5.55%	6.74%	5.88%	6.32%	6.86%	6.79%	7.22%	7.90%	6.44%
Health	4.37%	5.06%	5.55%	6.76%	6.59%	5.69%	5.51%	5.32%	5.45%	5.46%	5.58%
Transport	12.33%	10.50%	11.69%	11.79%	11.75%	12.07%	11.54%	11.93%	10.89%	12.21%	11.67%
Comunication	3.46%	3.81%	3.59%	3.50%	3.30%	3.46%	3.33%	3.21%	3.32%	2.95%	3.39%
Recreation	6.86%	8.29%	8.22%	8.53%	8.96%	9.98%	10.85%	10.78%	11.43%	10.49%	9.44%
Education	0.16%	0.47%	0.38%	0.40%	0.69%	0.68%	0.65%	0.37%	0.50%	0.67%	0.50%
Horeca	4.26%	4.89%	5.83%	5.65%	6.91%	7.36%	7.59%	8.30%	8.65%	8.78%	6.82%
Other	10.36%	11.15%	12.44%	12.80%	13.95%	13.36%	13.43%	14.29%	13.27%	14.29%	12.93%

Deciles are constructed with respect to the population of individuals, on the basis of equivalized real disposable incomes. Expenditure shares are household expenditures on a commodity aggregate divided by total expenditures. Entries are means of the shares taken with respect to the population of individuals belonging to a certain cell. Households with negative disposable incomes are included.

6.3 Tax reform scenarios

6.3.1 Top-down: a uniform carbon price in the agricultural sector

The first simulation investigates the distributional consequences of the introduction of a uniform price of 100 euros per tonne of CO2eq emissions in the agricultural sector at a worldwide scale. The channel through which we will asses the distributional consequences of this tax, is through its effect on consumer prices. More specifically, the carbon tax will affect producer prices directly, and this will be reflected in consumer prices. These will in turn create a change in demand. A new equilibrium emerges when prices are such that supply equals demand. The EM-ITT model is not however an equilibrium model, as it does not cover the supply side of the market. To assess the effect of a global carbon tax in the food sector, a collaboration was set up with JRC Sevilla. Using their equilibrium model *Common Agricultural Policy Regionalised Impact* (CAPRI), JRC Sevilla simulated the effect of the introduction of a uniform carbon price of 100 euros per tonne of CO2eq emissions in the agricultural sector worldwide, on consumer prices in this sector.

The CAPRI model is a world scale static equilibrium model for the agricultural and primary processing sectors. Its global market module encompasses food markets. These commodities are aggregated in broad groups such as rice, cereals, different kinds of meat, milk, eggs, potatoes, tomatoes, other vegetables, and so on. The by the CAPRI model simulated percentage changes of consumer prices of those commodity aggregates in each of 27 countries (EU minus Luxembourg, UK still included), following the introduction of the carbon tax, are shown in Table 31 of Appendix G. A summary of these price changes for Belgium at a more aggregated level is shown in the first column of Table 15. The effects of the introduction of the carbon tax in the CAPRI model run mainly through import and export substitution: imported goods from countries with a more polluting production technology for that good, are substituted by locally produced variants of that commodity.

The CAPRI aggregates are linked to the COICOP classification of goods used in ITT. Information on this mapping can be found in Table 32 of Appendix H. We neglected goods with excises (wine and tobacco) because this would complicate things, without having much impacts on the final results. Furthermore, we do not have COICOP equivalents for some of the CAPRI goods. These products were therefore not considered in the analysis. We also neglected cocoa. Though, interestingly, it is the only commodity for which the introducing of a carbon tax as simulated by the CAPRI model, resulted in a price decline, income shares of expenditures on cocoa are hardly more than 0.01% (see Table 33 in Appendix I), so that its impact on the results is negligible.

Each CAPRI aggregate thus consists of one or more COICOP aggregates. We then adapted the consumer prices of each of those COICOP goods with the same percentage as simulated for the corresponding CAPRI aggregate. The thus obtained new set of consumer prices form the point of departure for assessing the distributional consequences of a uniform C02eq emissions' price with the EUROMOD ITT tool. We will call this in the sequel the CAPRI simulation. Notice that in this first exercise, VAT and excise rates remain unchanged with respect to the baseline (see column 3 of Table 15). We need, however, EUROMOD ITT to simulate the distributional consequences of the policy reform.

6.3.2 Bottom-up: VAT rate differentiation according to CO2eq emission intensity

In the bottom-up approach, we start from a simulation that intends to increase VAT on food commodities with a heavier CO2eq emissions footprint. As food is almost exclusively consumed by households, we can consider this as a differentiated household expenditure tax. The introduction of such a tax may however provoke behavioral reactions at the supply side. It is precisely to capture some of these reactions that a tax incidence parameter was introduced into EM ITTv4, and we will illustrate its use here.

Contrary to the previous simulation, we thus go from the micro-simulation model of individual household behavior to integrating the repercussions on the supply side. Hence the label 'bottom-up'.

Reversely, in the previous exercise we started from a simulation with the macro-equilibrium model CAPRI, and then investigate its consequences at the household level with the micro-model EM ITT.

6.3.2.1 Increasing VAT on the most polluting goods

Pollution is usually not measured at the final product level. Indeed, the production of final consumption goods implies the input of different intermediate goods, the production of which all imply their own emissions. Emissions are measured at a plant level basis, and in general it may be difficult to ascribe parts of the total emissions at the plant level to each of the products processed within that plant. But even for single product firms, it is not straightforward to compare emission intensity across firms, as it is not clear in which units the production level of such firms need to be compared, nor whether emissions are linear in the production levels.

To circumvent all these problems, we define goods to be 'more polluting' when the increase in the consumer price as a consequence of the introduction of a uniform carbon price (as simulated by the CAPRI model) is highest. In the first column of Table 15, these simulated percentage changes of consumer prices are shown. The second column shows the VAT rates which would result if we would drop the uniform carbon tax, and replace it by an increase in VAT rates which would mimick the consumer price changes expected from the introduction of the CO2eq emissions tax. As none of the goods involved in our simulations bears excises, this new VAT rate can simply be calculated as:

$$t_{new} = (1 + t_{baseline})(1 + \dot{q}) - 1, \tag{38}$$

where $\dot{q}$ is the percentage change in consumer price as simulated by CAPRI (first column of Table 15).

Consumer prices of sheep and goat meat, for example, are expected to increase with almost 39% when introducing the 100 euros per tonne CO2eq emissions tax. Correspondingly, one would need to increase the VAT rate from the baseline level of 6% to 47.1%. However, from a welfare point of view, such a scenario would not generate any other results as the CAPRI simulation, as for consumer's welfare only the final consumer prices matter, not how they are composed. In addition, from a policy point of view, the introduction of about 20 tax rates would not be very viable.

Therefore we introduced, next to the existing rate of 6% for food commodities, two additional VAT rates, 36% and 9%. The former is applied to sheep, goat, and beef, the commodities whose prices increased most due to the introduction of the CO2eq emissions tax. The rate of 9% is applied to commodities exhibiting moderate increases (at least 1.4%, and less than 10%): pork, poultry, butter, milk, and cheese products. The other food items are taxed at the same level as in the baseline (6% with the exception of margarines, for which it is 12%).

This scenario is labeled as the DIRECT simulation in the sequel, and the tax rates applied in this scenario are summarized in the fourth column of Table 15.

6.3.2.2 Combination of VAT increases with VAT reductions

The next simulation, labeled as BALANCED in the sequel, introduces next to the VAT increase of the DIRECT simulation, a VAT waiver for fruits and vegetables. This VAT waiver aims to increase the purchasing power of the consumers, thus releasing some of the pressure induced by the VAT increase. The welfare gradient of the income shares of these commodities is comparable to the one of meat products, which is the main category of commodities for which VAT was increased in the DIRECT simulation: expenditures on meat products represent on average almost 7% of disposable income for the lowest decile, while this value is only about 2.2% for the richest decile; for vegetables and fruits these figures are respectively 5% and 1.7%.[44]

The VAT rates used in the BALANCED scenario are listed in the fifth column of Table 15.

6.3.2.3 Revenue recycling

In the last bottom-up scenario, labeled as COMPENSATED in the sequel, the VAT rates are the same as in the DIRECT simulation. So, The VAT rates in column 6 of Table 15 are equal to those of column 4 (DIRECT). Extra VAT payments of each household are however reimbursed through a household specific lump sum transfer. In this way we keep the distribution of disposable income minus indirect tax payments constant compared to the baseline.

Table 16 summarizes the changes each of the alternative scenarios will cause according to the constant income shares version of the EUROMOD ITT model, as far as they can be determined a priori. For the scenarios where the disposable household income does not change (CAPRI, DIRECT, and BALANCED), total expenditures (E^h) and saving (S^h) levels will remain at their baseline levels. The same holds true for expenditure on a particular commodity k (e^h_k). On the contrary, the COMPENSATED scenario will result in an increase of total expenditures and of expenditure on each commodity k. Saving will increase if baseline saving is positive, and decrease otherwise.

As *ad valorem* excise liabilities on commodity k are an increasing function of the rates v_k, which remain unaltered throughout, and of expenditures on that commodity, e^h_k. Consequently, these *ad valorem* excise liabilities will change in the same direction as expenditures on that commodity, e^h_k. That is *ad valorem* excise liabilities will stay constant under CAPRI, DIRECT, and BALANCED, and they will increase under the COMPENSATED scenario. Of course, for goods with zero *ad valorem* rates, nothing will change.

VAT liabilities on commodity k depend positively on the VAT rate t_k and on expenditures on that commodity, e^h_k. For the scenarios for which expenditures on a commodity, e^h_k, do not change (CAPRI, DIRECT, and BALANCED), the change is thus in the same direction as the change in the VAT rate. For the CAPRI scenario VAT rates are unaltered, so the same holds true for VAT liabilities on all commodities. For the DIRECT scenario, VAT liabilities will increase for commodities whose

[44] These figures can be obtained from adding figures of the meat columns, on the one hand, and the vegetable and fruit columns on the other hand, for decile 1 and 10 in Table 33.

Table 15: Percentage change in consumer prices and VAT levels across different scenarios

	CAPRI model	Corresponding	VAT rates used in the simulations			
	% change cons. price	VAT rate	CAPRI	DIRECT	BALANCED	COMPENSATED
Sheep and goat meat	38.8	47.1	6.0	36.0	36.0	36.0
Beef	29.2	36.9	6.0	36.0	36.0	36.0
Pork meat	3.6	9.8	6.0	9.0	9.0	9.0
Poultry meat	3.5	9.7	6.0	9.0	9.0	9.0
Rice	1.3	7.3	6.0	6.0	6.0	6.0
Eggs	1.3	7.4	6.0	6.0	6.0	6.0
Milk and cheese	1.5	7.6	6.0	9.0	9.0	9.0
Fish and seafood	1.3	7.4	6.0	6.0	6.0	6.0
Vegetables	0.4	6.4	6.0	6.0	0.0	6.0
Bread and cereals	0.3	6.3	6.0	6.0	6.0	6.0
Butter	1.4	7.5	6.0	9.0	9.0	9.0
Oils and fats*	0.2	6.2	6.0	6.0	6.0	6.0
Fruit	0.2	6.3	6.0	6.0	0.0	6.0
Sugar	0.1	6.1	6.0	6.0	6.0	6.0
Potatoes	0.1	6.1	6.0	6.0	6.0	6.0
Coffee	0.3	6.3	6.0	6.0	6.0	6.0
Tea	0.6	6.6	6.0	6.0	6.0	6.0

The first column is the percentage increase of consumer prices due to the introduction of a uniform carbon tax. The second column shows the VAT levels which would correspond to mimicking this price changes by an increase in VAT-rates. The last four columns show the VAT rates actually implemented in each of the four simulations. Prices and VAT-rates of cocoa, alcoholic drinks and tobacco have not been changed throughout.

* The VAT rate for margarines is 12%. This rate is also applied in all reforms.

VAT rates are increased, and remain unaltered for those goods whose VAT rates remain constant. Total VAT liabilities will increase. For the BALANCED scenario, VAT liabilities will decrease for the goods whose VAT rate has been lowered (vegetables and fruits). So, the effect on total tax liabilities will be undetermined a priori. In the COMPENSATED case, which is a combination of certain VAT rates being increased, some remaining constant, and expenditures on all commodities being increased, we obtain an increase in VAT liabilities on all goods, and thus also an increase in total VAT liabilities.

Consumption (quantities) change in the opposite direction as the change in consumer prices and in the same direction as expenditures on the commodity. Recall that in the CAPRI simulation all food prices increased except for cocoa. As we did not implement the price change for cocoa, all quantities of food items will go down. For the DIRECT scenario, consumption of food items whose VAT rates have been increased, will go down. For the BALANCED scenario, consumption of vegetables and fruits will rise, as their VAT rates have been lowered. In the COMPENSATED scenario, the effect on consumption of goods whose VAT rates have been increased cannot be determined a priori, as the increased price lowers consumption, while the increase in expenditures raises it. For goods the VAT rates of which remain unaltered, consumption will increase.

As specific excise rates remain fixed at the baseline level throughout all the scenarios, specific excise liabilities change in the same direction as quantities. Since the VAT rate of none of the goods who bear in the baseline a positive specific excise tax, have been changed, specific excise liabilities will remain unchanged in the CAPRI, DIRECT, and BALANCED scenarios. In the COMPENSATED scenario specific excise liabilities for commodities with positive specific excise rates will increase as their consumption increases due to higher expenditures on those commodities, at unchanged consumer prices.

Table 16: Summary of theoretically determined effects of the four basic scenarios

Change in Variable	**Scenario** CAPRI	DIRECT	BALANCED	COMPENSATED
Expenditures (e_k^h, E^h)	0	0	0	$\nearrow$
Saving (S^h)	0	0	0	$\nearrow$ if $S^h > 0$ $\searrow$ if $S^h < 0$
Ad valorem excise liability ($T_{v_k}^h$)	0	0	0	$\nearrow$ or 0
VAT liability ($T_{t_k}^h$)	0	$\nearrow$ or 0	$\nearrow$ or 0 or $\searrow$	$\nearrow$
Quantity ($\tilde{x}_k^h$)	$\searrow$	$\searrow$ or 0	$\searrow$ or 0 or $\nearrow$	? or $\nearrow$
Specific excise liability ($T_{a_k}^h$)	0	0	0	$\nearrow$

6.3.2.4 Producer price adaptations modelled through the tax incidence parameter

Up to now, the bottom-up scenarios only involved the micro-component. Ideally, one would then run the tax scenarios also with the macro-model (CAPRI) in order to assess its implication on producer prices. We did not stretch the integration of the micro- and macro-component all the way however. At this stage we contented ourselves to a sensitivity analysis with the newly introduced tax incidence parameters in EUROMOD ITT. These parameters, θ_k , reflect the degree to which producer prices will adapt to an indirect tax reform.[45] At the benchmark, θ_k is one for all commodities, meaning that producer prices are unaltered. The lower the value of θ_k, the more producer prices will decrease following an indirect tax increase on commodity k. As producer prices decrease due to an indirect tax increase, the consumer's welfare cost of the tax lift will be lower than it were with fixed producer prices. Producers may then face a welfare cost under the form of decreasing profits.[46] The latter cannot be assessed quantitatively by the EUROMOD ITT. However, the impact on consumer welfare can be assessed through its impact on the welfare measure defined in Equation (36). Indeed, adapting the value of θ_k modifies the new consumer price level turning up in the Stone price index. Notice that adapting the value θ_k for a commodity k not subject to an indirect tax change, though technically possible, does not make much sense. Indeed, the tax incidence parameter captures the change in producer prices following a tax change.

[45] The θ_k parameters are implemented in the EUROMOD spine at the commodity level and can be set at one of four pre-defined values (0.20, 0.50, 0.80, 1). The user has the option to allocate different values for each commodity; the pre-defined levels can be adapted, and the number of pre-defined levels can be increased.

[46] Theoretically, a producer price decrease following an indirect tax increase may result in both profit increases and consumer price decreases. This would require negative values of θ_k. On the contrary, a value of θ_k larger than one implies overshooting: producers increase prices following a tax increase. The 'normal' case is where the tax incidence causes welfare costs of an indirect tax increase to decline for the consumer and to provoke profit losses for the producer.

We perform the following sensitivity analyses. First, we redo the DIRECT, BALANCED, and COMPENSATED scenarios with a value of θ_k equal to one half for all commodities k subject to a VAT change. For the DIRECT and COMPENSATED scenarios, this will imply a welfare gain compared to the corresponding scenario with fixed producer prices ($\theta_k = 1$ for all k). As the BALANCED scenario combines tax increases with tax cuts, the effect on consumer welfare is a priori undetermined. Indeed, a tax incidence parameter value lower than one (and higher than zero) for a commodity whose tax rate is decreased, means that producer prices are raised in response, implying a welfare loss for the consumer, compared to the fixed producer price case. Next, we redo the COMPENSATED scenario with an higher sensitivity of producer prices ($\theta_k = 0.2$ instead of 0.5).

In Table 17 we give an exhaustive list of the values of θ_k in the different scenarios. For all goods not mentioned in the table, $\theta_k = 1$ (as none of these goods is subject to an indirect tax reform in any of the scenarios).

Table 17: θ levels

COICOP	DIRECT and COMPENSATED	BALANCED	COMPENSATED
	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.2$
01111 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 1 Rice	1	1	1
01112 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 2 Bread	1	1	1
01113 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 3 Pasta products	1	1	1
01114 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 4 Pastry-cook products	1	1	1
01115 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 5 Sandwiches	1	1	1
01116 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 6 Other products	1	1	1
01121 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 1 Fresh, chilled or frozen meat of bovine animals	0.5	0.5	0.2
01122 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 2 Fresh, chilled or frozen meat of swine	0.5	0.5	0.2
01123 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 3 Fresh, chilled or frozen meat of sheep and goat	0.5	0.5	0.2
01124 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 4 Fresh, chilled or frozen meat of poultry	0.5	0.5	0.2
01125 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 5 Dried, salted or smoked meat and edible meat offal	0.5	0.5	0.2
01126 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 6 Other preserved or processed meat and meat preparations	0.5	0.5	0.2
01127 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 7 Other fresh, chilled or frozen edible meat	0.5	0.5	0.2
01131 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 1 Fresh, chilled or frozen fish	1	1	1
01132 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 2 Fresh, chilled or frozen seafood	1	1	1
01133 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 3 Dried, smoked or salted fish and seafood	1	1	1
01134 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 4 Other preserved or processed fish and seafood and fish and seafood preparations	1	1	1
01141 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 1 Whole milk	0.5	0.5	0.2
01142 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 2 Low fat milk	0.5	0.5	0.2
01143 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 3 Preserved milk	0.5	0.5	0.2
01144 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 4 Yoghurt	0.5	0.5	0.2
01145 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 5 Cheese and curd	0.5	0.5	0.2
01146 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 6 Other milk products	0.5	0.5	0.2
01147 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 7 Eggs	1	1	1
01151 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 1 Butter	0.5	0.5	0.2
01152 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 2 Margarine and other vegetable fats	1	1	1
01153 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 3 Olive oil	1	1	1
01154 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 4 Edible oils	1	1	1
01155 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 5 Other edible animal fats	1	1	1
01161 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 1 Citrus fruits (fresh, chilled or frozen)	1	0.5	1
01162 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 2 Bananas (fresh, chilled or frozen)	1	0.5	1
01163 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 3 Apples (fresh, chilled or frozen)	1	0.5	1
01164 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 4 Pears (fresh, chilled or frozen)	1	0.5	1
01165 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 5 Stone fruits (fresh, chilled or frozen)	1	0.5	1
01166 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 6 Berries (fresh, chilled or frozen)	1	0.5	1
01167 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 7 Other fresh, chilled or frozen fruits	1	0.5	1
01168 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 8 Dried fruit	1	0.5	1
01169 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 9 Preserved fruit and fruit based products	1	0.5	1
01171 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 1 Leaf and stem vegetables (fresh, chilled or frozen)	1	0.5	1
01172 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 2 Cabbages (fresh, chilled or frozen)	1	0.5	1
01173 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 3 Vegetables cultivated for their fruit (fresh, chilled or frozen)	1	0.5	1
01174 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 4 Root crops, non-starchy bulbs and mushrooms (fresh, chilled or frozen)	1	0.5	1
01175 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 5 Dried vegetables	1	0.5	1
01176 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 6 Other preserved or processed vegetables	1	0.5	1
01177 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 7 Potatoes	1	1	1
01178 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 8 Other tubers and products of tuber vegetables	1	0.5	1
01181 : 01 Food and nonalcoholic beverages - 1 Food - 8 Sugar, jam, honey, chocolate and confectionery - 1 Sugar	1	1	1

6.4 Reform simulation results

6.4.1 Aggregate effects

Recall that all reforms are simulated under the constant income shares behavioral assumption. The simulated aggregates of the household budget components for the four scenarios are given in Table 18. The upper panel of Table 18 displays the levels, while the lower panel shows the change in aggregates compared to the baseline. What we observe is in line with the theoretical predictions of Table 16.

When the disposable household income does not change (CAPRI, DIRECT, and BALANCED), we observe increases in VAT as a response to increased VAT rates, but the levels of all other budget components remains the same. The fruit and vegetable tax waiver introduced in the BALANCED scenario decreases the annual VAT revenues by 241 million euros compared to the DIRECT scenario. But VAT liabilities remain almost one billion euros higher than in the baseline. This reform is thus far from budget neutral.

By construction, the COMPENSATED scenario is (almost) budget neutral. There is a small increase in net indirect taxes, due to the income effect of the lump sum transfer, which increase expenditures on and consumption of all goods compared to the DIRECT scenario, implying an increase of the three indirect tax components as compared to the DIRECT scenario.

6.4.2 Distributional and welfare effects

Table 19 shows the household budget composition for the four scenarios throughout the welfare distribution. Table 20 outlines the level (mid panel) and percentage changes (lower panel) of the components of the household budget with respect to the baseline throughout the welfare distribution. The changes are displayed only for the budget components where baseline values do change.

For policies where the disposable income remains unchanged (CAPRI, DIRECT, and BALANCED), the expenditure, saving, *ad valorem* and specific excise levels remain at their baseline levels. Following the change in the VAT scheme, we observe increased VAT payments for all income deciles under DIRECT. It turns out the same holds true for the BALANCED scenario, though this is an empirical result in this scenario. It turns out that the tax waiver is not enough to compensate fully for the tax increase in the DIRECT scenario. Therefore, comparisons with the COMPENSATED scenario are difficult. In the COMPENSATED scenario, the levels of all budget components are increased (in absolute value).

The percentage change values in the lower panel of Table 20 shows that the lump sum transfer granted in the COMPENSATED scenario is higher as a percentage of baseline disposable income for the households to which the poorer individuals in society belong. Percentage changes in tax revenues exhibit a less clearly outspoken distributional pattern.

Tabel 21 contains an analysis of the welfare impact of the different scenarios. We switch to an analysis at the individual level now. That is, we consider equivalized real disposable income as an

Table 18: Simulated aggregates in response to different carbon tax reforms (Annual, in million euros)

	Baseline	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**
Gross earned income	184 656.1	184 656.1	184 656.1	184 656.1	184 656.1
Direct taxes	48 929.1	48 929.1	48 929.1	48 929.1	48 929.1
Social security contributions	24 192.3	24 192.3	24 192.3	24 192.3	24 192.3
Benefits	66 405.8	66 405.8	66 405.8	66 405.8	66 405.8
Household disposable income	177 940.4	177 940.4	177 940.4	177 940.4	179 118.3
Total expenditures	159 840.6	159 840.6	159 840.6	159 840.6	161 076.9
Saving	18 099.8	18 099.8	18 099.8	18 099.8	18 041.3
VAT revenues	18 337.6	18 337.6	19 515.4	19 274.5	19 667.7
Ad valorem excises	455.0	455.0	455.0	455.0	458.8
Specific excises	3 650.9	3 650.9	3 650.9	3 650.9	3 680.4
Level change	**Baseline**	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**
Gross earned income	184 656.1	0.0	0.0	0.0	0.0
Direct taxes	48 929.1	0.0	0.0	0.0	0.0
Social security contributions	24 192.3	0.0	0.0	0.0	0.0
Benefits	66 405.8	0.0	0.0	0.0	0.0
Household disposable income	177 940.4	0.0	0.0	0.0	1 177.8
Total expenditures	159 840.6	0.0	0.0	0.0	1 236.3
Saving	18 099.8	0.0	0.0	0.0	-58.5
VAT revenues	18 337.6	0.0	1 177.8	936.9	1 330.1
Ad valorem excises	455.0	0.0	0.0	0.0	3.7
Specific excises	3 650.9	0.0	0.0	0.0	29.5

Table 19: Household budget composition (Monthly, in euros)

Decile	1	2	3	4	5	6	7	8	9	10	All
Baseline											
Household disposable income	1 312.02	1 793.78	2 131.03	2 537.53	2 972.11	3 346.53	3 802.45	4 081.38	4 578.28	6 326.35	3 288.14
Total expenditure	2 366.01	2 147.33	2 423.03	2 650.50	2 845.91	3 085.13	3 178.28	3 262.73	3 176.17	3 844.25	2 897.93
Saving	-1 053.99	-353.56	-292.00	-112.97	126.20	261.40	624.16	818.65	1 402.11	2 482.10	390.21
VAT	275.33	242.75	279.45	302.55	320.56	353.68	368.64	380.10	362.09	440.10	332.52
Ad valorem excises	11.06	10.11	7.23	9.51	10.05	8.00	7.76	6.40	5.83	4.12	8.01
Specific excises	59.73	53.51	50.29	55.45	70.38	70.12	80.61	71.34	69.43	80.73	66.16
CAPRI											
Household disposable income	1 312.0	1 793.8	2 131.0	2 537.5	2 972.1	3 346.5	3 802.4	4 081.4	4 578.3	6 326.4	3 288.1
Total expenditure	2 366.0	2 147.3	2 423.0	2 650.5	2 845.9	3 085.1	3 178.3	3 262.7	3 176.2	3 844.3	2 897.9
Saving	-1 054.0	-353.6	-292.0	-113.0	126.2	261.4	624.2	818.6	1 402.1	2 482.1	390.2
VAT	275.3	242.7	279.4	302.5	320.6	353.7	368.6	380.1	362.1	440.1	332.5
Ad valorem excises	11.1	10.1	7.2	9.5	10.1	8.0	7.8	6.4	5.8	4.1	8.0
Specific excises	59.7	53.5	50.3	55.5	70.4	70.1	80.6	71.3	69.4	80.7	66.2
DIRECT											
Household disposable income	1 312.0	1 793.8	2 131.0	2 537.5	2 972.1	3 346.5	3 802.4	4 081.4	4 578.3	6 326.4	3 288.1
Total expenditure	2 366.0	2 147.3	2 423.0	2 650.5	2 845.9	3 085.1	3 178.3	3 262.7	3 176.2	3 844.3	2 897.9
Saving	-1 054.0	-353.6	-292.0	-113.0	126.2	261.4	624.2	818.6	1 402.1	2 482.1	390.2
VAT	291.5	259.5	297.9	323.0	342.5	376.7	392.1	403.6	385.9	465.7	353.9
Ad valorem excises	11.1	10.1	7.2	9.5	10.1	8.0	7.8	6.4	5.8	4.1	8.0
Specific excises	59.7	53.5	50.3	55.5	70.4	70.1	80.6	71.3	69.4	80.7	66.2
BALANCED											
Household disposable income	1 312.0	1 793.8	2 131.0	2 537.5	2 972.1	3 346.5	3 802.4	4 081.4	4 578.3	6 326.4	3 288.1
Total expenditure	2 366.0	2 147.3	2 423.0	2 650.5	2 845.9	3 085.1	3 178.3	3 262.7	3 176.2	3 844.3	2 897.9
Saving	-1 054.0	-353.6	-292.0	-113.0	126.2	261.4	624.2	818.6	1 402.1	2 482.1	390.2
VAT	287.8	256.0	294.3	318.9	338.2	372.2	387.5	399.0	381.2	459.9	349.5
Ad valorem excises	11.1	10.1	7.2	9.5	10.1	8.0	7.8	6.4	5.8	4.1	8.0
Specific excises	59.7	53.5	50.3	55.5	70.4	70.1	80.6	71.3	69.4	80.7	66.2
COMPENSATED											
Household disposable income	1 328.2	1 810.5	2 149.5	2 558.0	2 994.1	3 369.6	3 825.9	4 104.9	4 602.1	6 352.0	3 309.5
Total expenditure	2 404.7	2 167.8	2 443.9	2 672.6	2 867.4	3 106.8	3 198.2	3 281.8	3 193.2	3 861.0	2 919.7
Saving	-1 076.5	-357.3	-294.4	-114.6	126.7	262.8	627.7	823.1	1 408.9	2 491.0	389.7
VAT	296.2	261.9	300.5	325.8	345.1	379.4	394.5	406.0	388.0	467.8	356.5
Ad valorem excises	11.2	10.2	7.3	9.6	10.1	8.1	7.8	6.4	5.9	4.1	8.1
Specific excises	60.8	54.0	50.7	55.9	70.9	70.6	81.1	71.8	69.8	81.1	66.7

Deciles are constructed with respect to the population of individuals, on the basis of baseline equivalized real disposable incomes. Entries are means taken with respect to the population of households to which individuals in a particular decile belong.

Table 20: Changes in household budget composition (Monthly)

Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**All**
Baseline											
Household disposable income	1312.02	1793.78	2131.03	2537.53	2972.11	3346.53	3802.45	4081.38	4578.28	6326.35	3288.14
Total expenditure	2366.01	2147.33	2423.03	2650.50	2845.91	3085.13	3178.28	3262.73	3176.17	3844.25	2897.93
Saving	-1053.99	-353.56	-292.00	-112.97	126.20	261.40	624.16	818.65	1402.11	2482.10	390.21
VAT	275.33	242.75	279.45	302.55	320.56	353.68	368.64	380.10	362.09	440.10	332.52
Ad valorem excises	11.06	10.11	7.23	9.51	10.05	8.00	7.76	6.40	5.83	4.12	8.01
Specific excises	59.73	53.51	50.29	55.45	70.38	70.12	80.61	71.34	69.43	80.73	66.16
Level change											
DIRECT, VAT	16.21	16.70	18.49	20.46	21.96	23.04	23.42	23.53	23.80	25.65	21.33
BALANCED, VAT	12.51	13.23	14.82	16.39	17.67	18.47	18.84	18.94	19.08	19.77	16.97
COMPENSATED											
Household disposable income	16.21	16.70	18.49	20.46	21.96	23.04	23.42	23.53	23.80	25.65	21.33
Total expenditure	38.69	20.43	20.92	22.08	21.47	21.65	19.90	19.10	17.05	16.72	21.80
Saving	-22.48	-3.73	-2.42	-1.62	0.49	1.39	3.51	4.43	6.75	8.93	-0.47
VAT	20.89	19.17	21.10	23.22	24.58	25.73	25.89	25.92	25.93	27.72	24.01
Ad valorem excises	0.14	0.09	0.06	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.06
Specific excises	1.06	0.52	0.45	0.47	0.52	0.49	0.50	0.42	0.38	0.37	0.52
Percentage change											
DIRECT, VAT	5.89%	6.88%	6.62%	6.76%	6.85%	6.51%	6.35%	6.19%	6.57%	5.83%	6.41%
BALANCED, VAT	4.54%	5.45%	5.30%	5.42%	5.51%	5.22%	5.11%	4.98%	5.27%	4.49%	5.10%
COMPENSATED											
Household disposable income	1.24%	0.93%	0.87%	0.81%	0.74%	0.69%	0.62%	0.58%	0.52%	0.41%	0.65%
Total expenditure	1.64%	0.95%	0.86%	0.83%	0.75%	0.70%	0.63%	0.59%	0.54%	0.43%	0.75%
Saving	2.13%	1.05%	0.83%	1.43%	0.39%	0.53%	0.56%	0.54%	0.48%	0.36%	-0.12%
VAT	7.59%	7.90%	7.55%	7.68%	7.67%	7.28%	7.02%	6.82%	7.16%	6.30%	7.22%
Ad valorem excises	1.30%	0.91%	0.84%	0.81%	0.68%	0.76%	0.62%	0.60%	0.57%	0.44%	0.80%
Specific excises	1.77%	0.97%	0.89%	0.85%	0.74%	0.70%	0.62%	0.59%	0.55%	0.46%	0.78%

Deciles are constructed with respect to the population of individuals, on the basis of baseline equivalized real disposable incomes. Entries are means taken with respect to the population of households to which individuals in a particular decile belong.

individual variable, and all figures reported are averages with respect to the population of individuals.

We start our discussion with inspecting the last column of the lower panel: a fully compensated tax increase on polluting goods leads to a welfare loss. This stands to reason. A differentiated indirect tax increase is a distortionary intervention. So, the welfare cost of increasing tax revenues through an increase of indirect taxes on particular commodities is expected to be higher than the monetary amount of tax revenues. While the BALANCED approach in theory has the potential to generate welfare gains for some individuals, it turns out that the current implementation does not do so on average. Further detailed inspection of the results is needed to see whether this is also the case at the individual level. We already saw while discussing Table 18, that one of the reasons for this loss is that the tax waiver as conceived here, is not able to reach budget neutrality.

By construction, both the BALANCED and COMPENSATED scenarios cause lower welfare losses than the DIRECT scenario. While the VAT increases in the DIRECT scenario are not fully mimicking the consumer price changes obtained from the simulation of a carbon tax with CAPRI, the welfare costs of both scenarios are very similar.

All scenarios cause on average larger absolute welfare losses for the better off, the COMPENSATED scenario the least outspoken so. There is a small bend in that pattern between decile six and seven. In relative terms though (see the left hand side of the bottom panel of Table 22) the welfare cost is decreasing through the welfare distribution. Targeting the lump sum transfer towards the lowest deciles might sound an easy cure. But this is assuming that this targeting would not cause a(n additional) distortion which we cannot straightforwardly model through EUROMOD ITT, especially in the labor market.

Our tentative conclusion is that while the BALANCED scenario has the most promising perspectives from a theoretical point of view, as currently implemented, it fails to fulfill these expectations. Making the scenario budget neutral might bring relief. An additional compensating tax cut on e.g. cereals is the most obvious candidate for trying to improve upon the current scenario. Its income share is on average equal to that of beef (2.9% versus 3.2%) and the income shares exhibit a similar pattern through the income distribution (see Table 33 in Appendix I).

6.4.3 Sensitivity with respect to the tax incidence parameter

Table 22 displays the welfare effects of our scenarios under different values of the tax incidence parameter θ. The upper panel shows the levels, while the mid and lower panels present respectively the level and percentage changes. Changes are always calculated relative to the baseline simulation.

A lowering of the tax incidence parameter in the DIRECT and COMPENSATED scenario will cause a welfare improvement as compared to the corresponding simulation with fixed producer prices ($\theta = 1$).[47] A tax incidence parameter equal to 0.5 for all goods whose taxes changed in the

[47] Changing the tax incidence parameter in the CAPRI scenario makes no sense as tax rates are not altered in this case.

Table 21: Equivalized real disposable income under different scenarios (Monthly, in euros)

Decile	Baseline	CAPRI	DIRECT	BALANCED	COMPENSATED
1	680.71	671.20	671.62	673.38	679.19
2	1040.59	1028.88	1029.31	1031.33	1038.74
3	1223.32	1210.29	1210.81	1212.97	1221.27
4	1391.69	1378.05	1378.60	1380.85	1389.55
5	1562.50	1548.35	1548.86	1551.14	1560.28
6	1727.13	1712.47	1713.00	1715.42	1724.84
7	1910.35	1895.94	1896.44	1898.82	1908.11
8	2141.53	2126.36	2126.91	2129.37	2139.18
9	2466.26	2450.45	2451.05	2453.66	2463.82
10	3426.33	3409.13	3409.85	3413.12	3423.69
Benchmark: Baseline	**Baseline**	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**
1	680.71	-9.51	-9.08	-7.32	-1.52
2	1040.59	-11.71	-11.28	-9.26	-1.85
3	1223.32	-13.02	-12.50	-10.35	-2.05
4	1391.69	-13.64	-13.09	-10.85	-2.14
5	1562.50	-14.15	-13.64	-11.35	-2.22
6	1727.13	-14.66	-14.13	-11.71	-2.29
7	1910.35	-14.40	-13.91	-11.53	-2.24
8	2141.53	-15.17	-14.62	-12.16	-2.35
9	2466.26	-15.82	-15.22	-12.60	-2.45
10	3426.33	-17.20	-16.48	-13.21	-2.64

Deciles are constructed with respect to the population of individuals, on the basis of baseline equivalized real disposable incomes. Entries are means taken with respect to the population of individuals.

DIRECT and COMPENSATED scenario is enough to turn the welfare cost for consumers in the COMPENSATED scenario on average into a welfare gain. In relative terms these gains are larger for those being worse off in the baseline.

Notice that for the BALANCED scenario, lowering the tax incidence parameter on all goods affected by a VAT rate change has two counterbalancing effects. While for goods bearing a tax increase, a lower value of the tax incidence parameter causes in the 'normal' case a welfare cost decline for the consumer and a profit loss for the producer, the reverse holds true for goods bearing a tax cut. Part of the consumer's welfare gains due to the tax cut is annihilated by a producer price increase, possibly resulting in higher profits. Comparing the results for the BALANCED scenario with fixed producer prices with those of the BALANCED scenario with $\theta = 0.5$, shows that the decrease in welfare cost of the tax increases through lower producer prices, is higher than the loss in welfare gain due to higher producer prices of goods with lowered taxes. Indeed, the welfare cost decreases on average when the tax incidence parameter is lower.

The loss in welfare gain of lower taxes in the BALANCED scenario can be quantified by looking at the difference in welfare cost levels between the DIRECT and BALANCED scenarios under fixed producer prices ($\theta = 1$) with the same differences when $\theta = 0.5$. Indeed, this difference is the welfare gain from the tax waiver, and these gains are generally smaller when $\theta = 0.5$ as compared to when $\theta = 1$.

We do not reproduce the aggregate effects of the sensitivity analysis with respect to the tax incidence parameter, as these variables are unaffected by the tax incidence parameter. For example, since we only introduce VAT changes on goods not bearing specific excises, all indirect tax liabilities will be unaffected by the tax incidence parameter under the constant income shares behavioral assumption.

Table 22: Real disposable income under different carbon tax reforms with θ (Monthly)

						Tax incidence parameter lower than one			
Decile	**Baseline**	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**	**DIRECT**	**BALANCED**	**COMPENSATED**	
		$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.2$
1	680.71	671.20	671.62	673.38	679.19	675.76	676.63	683.41	686.30
2	1040.59	1028.88	1029.31	1031.33	1038.74	1034.46	1035.46	1043.95	1047.52
3	1223.32	1210.29	1210.81	1212.97	1221.27	1216.52	1217.59	1227.05	1231.00
4	1391.69	1378.05	1378.60	1380.85	1389.55	1384.58	1385.69	1395.60	1399.74
5	1562.50	1548.35	1548.86	1551.14	1560.28	1555.09	1556.22	1566.57	1570.88
6	1727.13	1712.47	1713.00	1715.42	1724.84	1719.45	1720.65	1731.35	1735.81
7	1910.35	1895.94	1896.44	1898.82	1908.11	1902.79	1903.96	1914.50	1918.88
8	2141.53	2126.36	2126.91	2129.37	2139.18	2133.58	2134.80	2145.90	2150.50
9	2466.26	2450.45	2451.05	2453.66	2463.82	2457.99	2459.29	2470.81	2475.60
10	3426.33	3409.13	3409.85	3413.12	3423.69	3417.37	3418.98	3431.26	3436.44
						Tax incidence parameter lower than one			
Level change	**Baseline**	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**	**DIRECT**	**BALANCED**	**COMPENSATED**	
		$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.2$
1	680.71	-9.51	-9.08	-7.32	-1.52	-4.94	-4.07	2.70	5.59
2	1040.59	-11.71	-11.28	-9.26	-1.85	-6.13	-5.13	3.36	6.93
3	1223.32	-13.02	-12.50	-10.35	-2.05	-6.79	-5.73	3.73	7.69
4	1391.69	-13.64	-13.09	-10.85	-2.14	-7.12	-6.00	3.91	8.05
5	1562.50	-14.15	-13.64	-11.35	-2.22	-7.41	-6.28	4.07	8.38
6	1727.13	-14.66	-14.13	-11.71	-2.29	-7.68	-6.48	4.22	8.68
7	1910.35	-14.40	-13.91	-11.53	-2.24	-7.56	-6.39	4.16	8.54
8	2141.53	-15.17	-14.62	-12.16	-2.35	-7.95	-6.73	4.37	8.97
9	2466.26	-15.82	-15.22	-12.60	-2.45	-8.27	-6.98	4.55	9.34
10	3426.33	-17.20	-16.48	-13.21	-2.64	-8.96	-7.35	4.93	10.11
						Tax incidence parameter lower than one			
Percentage change	**Baseline**	**CAPRI**	**DIRECT**	**BALANCED**	**COMPENSATED**	**DIRECT**	**BALANCED**	**COMPENSATED**	
		$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 1$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.5$	$\theta = 0.2$
1	680.71	-1.40%	-1.33%	-1.08%	-0.22%	-0.73%	-0.60%	0.40%	0.82%
2	1040.59	-1.12%	-1.08%	-0.89%	-0.18%	-0.59%	-0.49%	0.32%	0.67%
3	1223.32	-1.06%	-1.02%	-0.85%	-0.17%	-0.56%	-0.47%	0.30%	0.63%
4	1391.69	-0.98%	-0.94%	-0.78%	-0.15%	-0.51%	-0.43%	0.28%	0.58%
5	1562.50	-0.91%	-0.87%	-0.73%	-0.14%	-0.47%	-0.40%	0.26%	0.54%
6	1727.13	-0.85%	-0.82%	-0.68%	-0.13%	-0.44%	-0.38%	0.24%	0.50%
7	1910.35	-0.75%	-0.73%	-0.60%	-0.12%	-0.40%	-0.33%	0.22%	0.45%
8	2141.53	-0.71%	-0.68%	-0.57%	-0.11%	-0.37%	-0.31%	0.20%	0.42%
9	2466.26	-0.64%	-0.62%	-0.51%	-0.10%	-0.34%	-0.28%	0.18%	0.38%
10	3426.33	-0.50%	-0.48%	-0.39%	-0.08%	-0.26%	-0.21%	0.14%	0.30%

Deciles are constructed with respect to the population of individuals, on the basis of baseline equivalized real disposable incomes. Entries are means taken with respect to the population of individuals.

7 Conclusion

In this report, we summarized in detail the work completed for the project JRC/SVQ/2020/OP/1373. The first aim of the project was to extend the coverage of ITTv3 to all EU countries. This implied the addition of nine countries: Austria, Bulgaria, Estonia, Croatia, Luxembourg, Latvia, Malta, the Netherlands, and Sweden.

In the first half of the project, we completed the data collection. We obtained all datasets to be used within this project. We started by cleaning and preparing these data-sets for imputations. We then restructured the HBS data-sets of Austria and Luxembourg so that the data structure, variable definitions, and variable names are consistent with the harmonized EUROSTAT HBS data-sets. Adjustments were needed in the EUROSTAT 2015 HBS dataset for the Netherlands as well. The details of the restructuring process can be found in the corresponding `R` source. We also did collect the indirect tax policy rules for the remaining nine countries for the policy years 2010 and 2019. The imputation for three of the nine countries, Bulgaria, Croatia, and Estonia, was completed.

In the second half of the project, we finalised the imputations for Austria, Croatia, Luxembourg, Malta, the Netherlands, and Sweden as well. We integrated the tax policy information into EUROMOD for all remaining countries. With the inclusion of the nine remaining countries, ITT reached full EU coverage. We further provided a validation of the baseline EUROMOD simulations in terms of aggregate expenditures on the twelve COICOP level 1 aggregates and the indirect tax revenues broken down by categories of indirect taxes.

Furthermore, for the eight (Sweden being the exception) countries for which the 2019 EUROMOD version of EU-SILC was available, we also imputed expenditure shares from the same HBS as used for the 2010 imputations, in these more recent SILC data.

During the second half, we also implemented the tax incidence parameters in the `tco_cc` policy module of EUROMOD. This extension allows for assessing the welfare and tax revenue effects of potential producer price changes with respect to indirect tax reforms. We illustrate the power of this new instrument of analysis with an application on greening food taxation.

The ITT architecture has been adapted such that there is a full separation between a baseline simulation and a reform. Finally, the Statistical Presenter is enhanced by the inclusion of a reform module.

8 Bibliography

Akoğuz, C. E., B. Capéau, A. Decoster, L. De Sadeleer, D. Güner, and T. Vanheukelom (2020). A new indirect tax tool for euromod, final report. *JRC Project no. JRC/SVQ/2018/B.2/0021/OC*.

Crippa, M., E. Solazzo, D. Guizzardi, F. Monforti-Ferrario, F. Tubiello, and A. Leip (2021). Food systems are responsible for a third of global anthropogenic ghg emissions. *Nature Food 2*(3), 198–209.

D'Orazio, M., M. Di Zio, and S. M. (2006). *Statistical matching: Theory and practice.* Chichester: John Wiley & Sons Ltd.

Eurostat. Description of the data transmission for hbs (reference year) 2010.

Ivanova, D., G. Vita, K. Steen-Olsen, K. Stadler, P. Melo, R. Wood, and E. Hertwich (2017). Mapping the carbon footprint of eu regions. *Environmental Research Letters 12*(5).

Jacobs, B. and F. van der Ploeg (2019). Redistribution and pollution taxes with non-linear Engel curves. *Journal of Environmental Economics and Management 95*(C), 198–226.

Lamarche, P. (2017). Measuring income, consumption and wealth jointly at the micro–level. *Eurostat, mimeo*.

Leulescu, A. and M. Agafitei (2013). Statistical matching: a model based approach for data integration. *Eurostat Methodologies and Working papers*.

Serafino, P. and R. Tonkin (2017). Statistical matching of european union statistics on income and living conditions (eu–silc) and the household budget survey. *Eurostat Statistical Working Papers*.

Stiglitz, J. E. (2019). Addressing climate change through price and non-price interventions. *European Economic Review 119*(C), 594–612.

Weitzman, M. (2014). Can negotiating a uniform carbon price help to internalize the global warming externality? *Journal of the Association of Environmental and Resource Economists 1*(1), 29 – 49.

Addendum: pending issues in ITT

In the present section we summarize some issues we discovered after finishing the ITT-model building, and could not be patched any more within the scope of the present project.

- For the 18 countries for which an indirect tax module was constructed during the ITTv3 project, specific excise tariffs for the COICOP group **02122** *Other wines*, are not drawn from statutory tariffs for commodities belonging to that group, but from statutory tariffs for the group of *sparkling wines*, which are commodities that actually belong to COICOP category **02121** *Wines*. For the nine countries for which the module was constucted during the ITTv4 phase, and for Belgium, this has been repaired.
- Instead of specifying the convergence criterion of the iterative procedure in terms of the absolute value of the deviation of consumer prices between two rounds of iteration, it is implemented as the absolute value of the sum of deviations between consumer prices of two rounds of iterations. Negative and positive deviations might therefore cancel out. In practice, we did not discover any such problems up to now. Before trying to patch this, it might be worthwhile first to evaluate whether the iterative procedure cannot be circumvented (see next point).
- We made an extensive study of the iterative procedure. It turns out that there are generically two solutions to the set of Equations (28)-(29). We think, but are not sure, that in case there are two solutions, only one is economically meaningful. There is no guarantee that the iteration procedure always converges to the economic meaningful one. We provisionally advise to check carefully the producer and consumer price, and the implicit indirect tax rate output of a simulation where tax parameters of goods subject to specific excises are changed and the tax incidence parameter is set different from one, in order to see if they are strictly positive for prices, and non-negative for the indirect tax rates (as long as there are no subsidies involved). Only then results can validly be used. Up to now, we did not have encountered concrete problems. We are currently evaluating whether we cannot do without the iterative procedure at all. Once we have a solution, we will report it and we can discuss modalities of implementing the solution outside of the framework of this present project.

Appendices

A Country names and codes

Table 23: List of country abbreviations

Abbreviation	Country
ITTv4 countries	
AT	Austria
BG	Bulgaria
EE	Estonia
HR	Croatia
LU	Luxembourg
LV	Latvia
MT	Malta
NL	the Netherlands
SE	Sweden
ITTv3 countries	
BE	Belgium
CY	Cyprus
CZ	Czech Republic
DE	Germany
DK	Denmark
EL	Greece
ES	Spain
FI	Finland
FR	France
HU	Hungary
IE	Ireland
IT	Italy
LT	Lithuania
PL	Poland
PT	Portugal
RO	Romania
SI	Slovenia
SK	Slovakia
Other	
EU	European Union
UK	United Kingdom

B Summary files of the imputations

For each country we provide an Excel summary file `summary XX.xlsx` (where `XX` stands for the country code defined in Appendix A) with information on the imputation results. The content of this file is structured as follows.

- Sheet 1: `descriptive statistics XX`

  The sheet contains sample statistics for the HBS and SILC variables used in the imputation.

  For HBS, we give the following statistics:

  - HBS – number of households with non-positive incomes: cells `2R:4X`;
  - HBS – number of households with negative expenditures on a good or good aggregate: cells `3X:4Y`;
  - HBS – is the reference person a farmer? : cells `29X:32X`;
  - HBS – region : cells `2AA:13AB`;
  - HBS – numerical covariates : cells `14R:26AF`;
  - HBS – expenditure levels: cells `34R:46AN`;
  - HBS – income shares of expenditures : cells `48R:60AN`.

  For SILC, we give the following statistics:

  - SILC – number of households with non-positive incomes : cells `2A:4G`;
  - SILC – is the reference person a farmer? : cells `29G:32G`;
  - SILC – region : cells `2J:13K`;
  - SILC – numerical covariates : cells `14A:26O`.

- Sheet 2: `regression results XX`

  - Pseudo–R^2 values of regressions per broad category: cells `1A:21B`.
  - Covariates used in the regression: cells `1D:19D`.
  - Covariates not used in the regression: cells `1F:19F`.
  - Detailed regression results for 20 broad categories: probits start at cells `25A:25H`; linear regressions start at cells `25J:25T`.

  Categories retained as input for the distance function (pseudo–R^2>.1) are highlighted in green.

- Sheet 3: `ventile tables and graphs XX`

  Contains summary statistics of income shares of expenditures on broad categories including:

  - weighted mean, minimum and maximum values of income shares of expenditures, overall and per income ventile,
  - overall and per income ventile 5th, 25th, 50th, 75th and 90th population percentiles of income shares,
  - weighted mean household disposable income overall and per income ventile, and
  - an alternative mean income share, calculated as population total expenditure (overall or per ventile) divided by total income (overall or per ventile), denoted by `mean2` in the sheet.

Table 24: Position of the statistics for SILC (imputed expenditure values)

Food and non–alcoholic beverages:	cells `2A: 27K`,
Housing (rental):	cells `30A: 55K`,
Housing (goods and services):	cells `58A: 83K`,
Utilities:	cells `86A:111K`,
Communications:	cells `114A:139K`,
Culture and recreation:	cells `142A:167K`,
Personal care:	cells `170A:195K`,
Insurance:	cells `198A:223K`,
Alcoholic beverages:	cells `226A:251K`,
Tobacco:	cells `254A:279K`,
Private transportation:	cells `282A:307K`,
Public transportation:	cells `310A:335K`,
Travelling and holiday:	cells `338A:363K`,
Education:	cells `366A:391K`,
Vehicles:	cells `394A:419K`,
Housing (durables):	cells `422A:447K`,
Clothing and personal items:	cells `450A:475K`,
Health and care:	cells `478A:503K`,
Restaurants:	cells `506A:531K`,
Other:	cells `534A:559K`,
Total expenditure:	cells `562A:587K`,
Saving:	cells `590A:615K`.

Table 25: Position of the statistics for HBS (observed expenditure values)

Food and non–alcoholic beverages:	cells 20: 27Y,
Housing (rental):	cells 300: 55Y,
Housing (goods and services):	cells 580: 83Y,
Utilities:	cells 860:111Y,
Communications:	cells 1140:139Y,
Culture and recreation:	cells 1420:167Y,
Personal care:	cells 1700:195Y,
Insurance:	cells 1980:223Y,
Alcoholic beverages:	cells 2260:251Y,
Tobacco:	cells 2540:279Y,
Private transportation:	cells 2820:307Y,
Public transportation:	cells 3100:335Y,
Travelling and holiday:	cells 3380:363Y,
Education:	cells 3660:391Y,
Vehicles:	cells 3940:419Y,
Housing (durables):	cells 4220:447Y,
Clothing and personal items:	cells 4500:475Y,
Health and care:	cells 4780:503Y,
Restaurants:	cells 5060:531Y,
Other:	cells 5340:559Y,
Total expenditure:	cells 5620:587Y,
Saving:	cells 5900:615Y.

Plots of income shares of expenditures against disposable income are included to the right of the table for HBS (columns AB:AP).

- Sheet 4: `correlation differences XX`
  - Mean difference in weighted correlation in HBS and SILC:

within covariates:	cell `2B`,
between covariates and expenditure categories:	cell `3B`,
within expenditure categories:	cell `4B`.

  - Difference in weighted correlation matrix of HBS and SILC.

    The start position of the matrix is cell `8B`, and the matrix is structured as follows:

difference in correlation	
within covariates	between covariates and expenditure categories
between expenditure categories and covariates	within expenditure categories

    For the covariates part we excluded the square and cube of log income and the categorical variables.

    For the expenditures part, we included entries for total expenditures and saving, but these are not retained in the calculations of the means.

    Cells of the matrix are shaded in darker red the larger the value of the cell (absolute value of difference in estimated correlation for that cell between SILC and HBS).

C Definition of the 20 broad expenditure categories

Table 26 and Table 27 display the composition of the 20 broad aggregates on which we performed the regressions for each country. We indicate the composition by means of the HBS–nomenclature for the commodities and aggregates, but notice that we sometimes overwrote the variables as recorded in the HBS according to the rules described in Section 2.3.

A number of descriptive statistics on the levels and income shares of the expenditures on those broad categories is provided in the sheet `descriptive statistics XX` of the summary file of the imputation for each country (`summary XX`), where `XX` stands for the country code (see Appendix A). *Total expenditure* is defined as the sum of expenditures on these broad categories, and *saving* denotes the difference between household disposable income and total expenditure.

Table 26: Composition of consumption categories

Category	Level of aggregation			
	first	second	third	fourth
1. Food and non–alcoh. beverages	EUR_HE01			
2. Housing: actual rentals		+ EUR_HE041		
3. Utilities		+ EUR_HE045 (Electricity, gas, and other fuels)	+ EUR_HE0441 (Water supply) + EUR_HE0442 (Refuse collection) + EUR_HE0443 (Sewerage collection)	
4. Communication	EUR_HE08			
5. Personal care		+ EUR_HE121		
6. Insurance		+ EUR_HE125		
7. Alcoh. beverages		+ EUR_HE021		
8. Tobacco		+ EUR_HE022		
9. Private transport		+ EUR_HE072 (Operation of personal transport equipment)		
10. Education	EUR_HE10			
11. Clothing and personal items	EUR_HE03 (Clothing and footwear)	+ EUR_HE123 (Personal items)		
12. Health & care	EUR_HE06 (Health products and services)	+ EUR_HE124 (Social protection services)		
13. Restaurants		+ EUR_HE111 (Catering services)		

Table 27: Composition of consumption categories (continued)

Category	Level of aggregation						
	first		second		third		fourth
14. Housing: goods and services	EUR_HE05 (Furnishings, household equipment, and routine maintenance of the house)	+	EUR_HE043 (Maintenance and repair of the dwelling)	+ – – –	EUR_HE0444 (Other services relating to the dwelling) EUR_HE0531 (Large household appliances) EUR_HE0532 (Small electrical household appliances) EUR_HE0551 (Big tools for garden)	– –	EUR_HE05111 (Furniture and furnishings) EUR_HE05112 (Carpets and other floor coverings)
15. Housing: durables				+ + +	EUR_HE0531 (Large household appliances) EUR_HE0532 (Small electrical household appliances) EUR_HE0551 (Big tools for garden)	+ +	EUR_HE05111 (Furniture and furnishings) EUR_HE05112 (Carpets and other floor coverings)
16. Culture and leisure	EUR_HE09 (Recreation and culture)	–	EUR_HE096 (Package holidays)				
17. Public transport		+	EUR_HE073 (Transport services)				
18. Vehicles		+	EUR_HE071 (Purchase of vehicles)				
19. Travelling and holiday		+ +	EUR_HE096 (Package holidays) EUR_HE112 (Accommodation services)				
20. Other		+ +	EUR_HE127 (Other services) EUR_HE126 (Financial services)				

D Covariates used in different countries

As explained in the main text, we use a number of socio–demographic variables as covariates in the regressions. Recall that the covariates for all regressions (20 linear OLS regressions for positive expenditures and 20 probits for estimating the probability of positive expenditures) are the same per country. Table 28 contains a full list of the variables that were considered to be included as explanatory variable. A list of descriptive sample statistics of these variables for both HBS and SILC can be found in the sheet `XX descriptive statistics` of the summary file for each country (`summary XX`).

A variable is excluded from the regressions if the information on the variable is absent or not trustworthy in either the SILC or the HBS dataset. In the next table we indicate by a zero which variables are not included in the regressions for each of the countries.

Table 28: Full list of explanatory variables by country

Variable	AT	BG	EE	HR	LU	LV	MT	NL	SE
HH disposable income (3rd degree polynomial)									
n adult male HH members									
n HH members age $\leq$14									
n HH members 15 – 29									
n HH members 30 – 44									
n HH members 45 – 59									
n HH members age $\geq$ 60									
n employed HH members									0
n unemployed HH members									0
n pensioned HH members							0		0
n disabled HH members							0		0
n student HH members age >14							0		0
n with higher education							0	0	0
n non–EU citizens		0				0	0	0	
reference person farmer		0	0	0	0	0	0	0	0
region dummies			0	0	0	0	0	0	0

E Detailed list of excise goods

1. **02111 Spirits and liquors**
   - eaux-de-vie, liqueurs and other spirits with high alcohol content;
   - mead;
   - pomace brandy, such as pisco, grappa, marc, etc.;
   - aperitifs other than wine-based aperitifs.
2. **02121 Wine**
   - wine from grapes;
   - fortified wines, such as vermouth, sherry, port wine;
   - champagne and other sparkling wines from grapes;
   - ice-wine;
   - low and non-alcoholic wine;
   - wine-based aperitifs.
3. **02122 Other wine**
   - cider and perry, including sake.
4. **02131 Beer**
   - all kinds of beer, such as ale, lager, stout and porter;
   - low-alcoholic beer and non-alcoholic beer.
5. **02311 Cigarettes**
   - cigarettes;
   - cigarettes that are purchased in bars and restaurants, provided that a service charge is not applied.
6. **02312 Cigars**
   - cigars;
   - cigars that are purchased in bars and restaurants, provided that a service charge is not applied.
7. **02313 Other tobacco products**
   - pipe tobacco, chewing tobacco, hookah blends, snus or snuff;
   - cigarette tobacco and tobacco leaf;
   - cigarette papers and single use filters that are consumed with the cigarette;
   - refills for electronic cigarettes, with or without nicotine;
   - tobacco that is consumed with a shisha or hookah pipes if consumed at home.
8. **04511 Electricity**
   - electricity from all sources (coal, solar, hydro, etc.);
   - associated expenditure, such as rental of meters, reading of meters, standing charges, etc;
   - charges for self-produced energy (In some countries, households producing more own electricity than what they consume, are charged storage costs if they feed the surplus electricity back into the electricity supply grid.).
9. **04521 Natural gas through networks**

- natural gas and town gas delivered through gas networks;
- associated expenditure, such as rental of meters, reading of meters, standing charges, etc.

10. **04522 Liquefied hydrocarbons**
    - liquefied hydrocarbons (butane, propane, etc.) delivered in storage containers;
    - associated expenditure, such as rental or purchase of storage containers, standing charges, etc.
11. **04531 Liquid fuels**
    - domestic heating, lighting and cooking fuel oils;
    - biofuels for domestic use;
    - alcohol for fireplaces.
12. **04541 Solid fuels**
    - coal, coke, briquettes, firewood, charcoal, peat and the like, biomass (wheat, nutshell, etc.) and dry animal dung.
13. **07221 Fuels and lubricants for personal transport equipment**
    - petrol and other fuels, such as diesel, liquid petroleum gas, alcohol and two-stroke mixtures;
    - lubricants, brake and transmission fluids, coolants and additives;
    - electricity as fuel for cars when separately priced from other electricity;
    - hydrogen;
    - fuel and lubricants for major tools and equipment covered under (05.5.1) and recreational vehicles covered under (09.1.2).

F Notes on ITT parameters

Specific excise tariffs, a, are expressed as a monetary amount per unit of quantity. The ITT input data however only provide information on expenditures. The general principle to calculate specific excises (T_a) on the basis of expenditure data (e) was laid out in Akoğuz et al. (2020). When one obtains information on consumer prices (denoted by q) in the same units as a, the number of units of quantities, x, is easily obtained from expenditures as follows:

$$x = \frac{e}{q}, \tag{39}$$

and the associated specific excise bill is:

$$T_a = a \cdot x = a\frac{e}{q}. \tag{40}$$

Two problems may arise when implementing this procedure. The first is the aggregation problem. The second is finding a price in the appropriate units. Both problems are somehow related. We start with last mentioned problem: finding a price in the appropriate units. Take for example the case excises on beers. These are usually specified per hectolitre per degree Plato. Moreover, the tariff in some countries varies dependent on the degrees Plato (a higher tariff for stronger beers for example). Of course beer is not sold per degrees Plato per litre, but simply per litre. Now, while it may be straightforward to find a price for beer per litre, there are many types of beer, each having a different degree Plato content, and for one and the same brand and kind of beer, prices vary spatially within one country and over time, within the timespan of one year. Hence a single price (index) for beer should be *constructed*, and, similarly, a general tariff for beer. The aggregation problem is actually the problem of constructing a price (index) for a commodity aggregate. Indeed, generally speaking, expenditures are always registered at a certain level of aggregation. In the specific case of ITT, we observe expenditures at the the COICOP level 4 (5 binary digits) of aggregation. Prices for commodities belonging to such a group are collected from different sources. Sometimes these prices are to be considered as index already. E.g., prices of wine were usually constructed from dividing revenues from sales by volumes of sales (litres in this case), which is one way of constructing an average price (some prefer to call this a unit value rather than a price). In this example, the price of the commodity aggregate wine, say q_w, is the weighted average of the price of all wines sold using the volume share of each bottle of wine sold as weight. Indeed, let x_s be the volume of wine in litres bought by a particular consumer at a particular moment in time during a specific time span, at price q_s. The price of wine, say q_w, is then

$$q_w = \sum_{s \in w} q_s \frac{x_s}{\sum_{t \in w} x_t}. \tag{41}$$

Here, the indices, s and t, run over all sales during a particular period (one year in our case) in a certain geographical unit (a country for ITT).

But wine is not a COICOP level 4 commodity aggregate. The COICOP level 4 classification distinguishes between *wines from grapes* (COICOP **02121**) which comprises still wines, sparkling wines, and fortified wines, and *wine from other fruits* such cider, perry, and rice wine (COICOP **02122**).

The general principle of ITT for constructing price indices at COICOP level 4 aggregates was to look up a number of prices for representative commodities or price indices for commodity aggregates at a lower level of aggregation, and then construct an index by calculating a weighted mean of those prices. To make sense, this requires these prices and price indices at a lower level of aggregation should all be expressed in the same units, and, ideally, in the units in which the specific excise is expressed, or, at least easily convertible in such units.

So, in general, for any COICOP level 4 aggregate g containing commodities subject to a specific excise, we collected consumer price information for commodities or commodity groups belonging to that category g. These price(indice)s are indexed by m and denoted by q_m. We defined a set of weights w_m^q such that the price index of commodity aggregate g is constructed as:

$$q_g = \sum_{m \in g} w_m^q q_m. \tag{42}$$

Ideally these weights should be derived from information on expenditure or volume shares of the different m components in g, but in many cases this information was not readily available, and informed guesses were made. Section F.1 contains some more detail on the source of our price information and general principles of the weighting schemes used.

A similar problem occurs for the excise tariffs. They are usually specified in terms of a higher level of detail than the COICOP level 4 aggregates. In addition, they are sometimes specified in different units for commodities belonging to the same COICOP level 4 category. For example in Luxembourg, LPG is taxed per 1000kg, while petrol per 1000L. Both commodities belong to COICOP level 4 category **07221** (*Fuels and lubricants for personal transport equipment*).

Again, we collected the relevant tariffs of specific excises and defined weighting schemes to aggregate those tariffs to the COICOP level 4 categories. An additional complication was that price(indice)s are usually not available in the same classification as tariffs. Indexing the commodities for which we collected the excise tariffs by m', denoting the associated tariff by $a_{m'}$, and the weights by $w_{m'}^a$, the tariff at COICOP level 4, say a_g, is constructed as:

$$a_g = \sum_{m \in g} w_{m'}^a a_{m'}. \tag{43}$$

Again, ideally the weights should be based on expenditure or volume shares, but, in practice such information was not available to us, so that rules of thumb had to be used. Section F.2 contains some more detail on the source of our excise information and general principles of the weighting schemes used.

Given that the commodity classification of goods subject to specific excises does not necessarily coincide with the good classification for which representative prices or price indices were collected, some bookkeeping rules should be respected. For example, if a set of excise tariffs, m' apply to goods belonging commodity group m, the following equality should hold:

$$w_m^q = \sum_{m' \in m} w_{m'}^a. \tag{44}$$

The following rules should always hold:

$$\sum_{m \in g} w_m^q = 1 \text{ and } \sum_{m' \in g} w_{m'}^a = 1. \quad (45)$$

Full information of prices and excise tariffs with links to sources can be found in the Excel file `ITTv4 parameters.xlsx`. The full information on the weighting schemes for each country can be found in the EM policy systems and is not documented in a separate file.

F.1 ITT consumer prices

Table 29 provides information on the default ITT settings regarding the consumer prices (`$tco_base_xq`). The first column shows the COICOP label of the commodity group including commodities subject to specific excises. The second column refers to the commodities belonging to this COICOP aggregate for which we collected price information, and the third column is the source of the price information. Finally, the last column refers to the weights used to calculate weighted averages for the consumer price indices of the corresponding COICOP level 4 aggregates.[48]

Table 29: Baseline consumer prices of excise goods

	Price Unit	**Price source**	**Weight**
02111 Spirits and liquors	1 L vodka	Statista	
02121 Wine	1 L still wine	Statista	0.90
	1 L sparkling wine	Statista	0.05
	1 L fortified wine	Statista	0.05
02122 Other wine	1 L cider, perry, rice wine	Statista	
02131 Beer	1 L Lager beer	Statista	
02311 Cigarettes	1000 pieces	EDT	
02312 Cigars	1000 pieces	Statista	
02313 Other tobacco products	1000 kg	Statista	
04511 Electricity	1 MWh electricity	Eurostat	
04521 Natural gas	1 Gj natural gas	Eurostat	
04522 Liquefied hydrocarbons	1000 kg LPG	WOB	
04531 Liquid fuels	1000 L gasoil- diesel	WOB	1.00
	1000 L heavy fuel oil	WOB	0.00
	1000 L kerosene	WOB	0.00
04541 Solid fuels	1 Gj antracit coal	Google search	
07221 Fuels and lubricants	1000 L Euro-super 95	WOB	0.50
	1000 L diesel	WOB	0.50
	1000 L kerosene	WOB	0.00
	1 Gj natural gas	Eurostat	0.00

– For some products historical prices were found. These products are cigarettes, electricity, natural gas, LPG, Euro-super 95 and diesel. Cigarettes prices are the MPPC prices per 1000

[48] When the weights are not explicitly given, it means that there is only a single price used to construct the price index for the COICOP level 4 aggregate, and the weight of that price is thus equal to one.

cigarettes reported in the corresponding excise duty tables for tobacco, which can be accessed from the Excel file `ITTv4 parameters.xlsx`.[49] The price of LPG, Euro-super 95, and gas oil is collected from Weekly Oil Bulletins (WOB).

- Statista prices are available in 2019 but not in 2010; 2019 prices were deflated using the country consumer price index.[50]
- For some products, using the prices provided by Statista led to negative producer prices. This was the case for *Other tobacco products* (COICOP level 4 aggregate **02313**) in Croatia for the 2019 policy system. Hence, we adjusted that baseline consumer price *ad hoc*.
- When price information is not available for goods on which specific excises are levied, we assigned a zero weight to that product for constructing the price index (e.g. *heavy fuel oi*l and *kerosene* get zero weight in the price index for COICOP level 4 category **04531** *Liquid fuels*).
- When we had no information at all for the weights, equal weights were used (e.g. *Euro-super 95* and *diesel* in COICOP category **07221** *Fuels and lubricants*, are given equal weight.

F.2 ITT excise parameters

The excise parameters used in ITT are collected from the excise duty tables of the corresponding year.[51] For 2010, the excise tariffs are collected from the excise duty tables of July 2010. For 2019, the excise tariffs are collected from the January 2019 excise duty tables. For Croatia, the July 2013 version of the excise duty tables are used, as we implemented the 2013 policy for that country.

Table 30 lists which specific excise tariffs apply to each of the COICOP level 4 categories.

- The way specific excises are levied on wine consumption varies considerably across countries. Hence, the weighting of specific excise tariffes differs substantially across countires. Details can be found in the country specific `tco_cc` policy systems of each country `cc`.
- COICOP level 4 category **02121** *Wine from grapes* is composed of *still wines*, *sparkling wines* and *fortified wines.* As a rule of thumb, the sum of the weights assigned to various types of specific excises levied on *still wines* was set to 90%; and the sum of the weights assigned to various types of specific excises levied on respectively *sparkling wines* and *fortified wines*, are set to 5% each.
- The excise tariffs for *Intermediate products* are applicable to all alcoholic beverages with an alcohol content of 1.2% to 22%, which cannot be regarded as beer or wine. Generally these are fermented beverages to which ethyl alcohol has been added. Fortified wines such as port and sherry (part of COICOP level 4 category **02121** *Wine from grapes*) belong to this category.

[49] MPPC is the abbreviation for *Most Popular Price Category.*

[50] Consumer price index values are collected from Worldbank databases. Worldbank publishes country level consumer price index on their website at `data.worldbank.org/cpi`. When releasing ITT it deserves recommendation to use a commodity group specific deflator, e.g. `https://ec.europa.eu/eurostat/web/hicp/data/database`.

[51] Historical excise duty rates are also published on the CIRCABC website at `circabc.europa.eu`.

Table 30: Excise categories of excise goods

	Excise category	**weight**
02111 Spirits and liquors	ethyl alcohol	
02121 Wine	**wine, still**, normal rate	**0.90**
	wine, still, reduced rate	
	wine, sparkling, normal rate	**0.05**
	wine, sparkling, reduced rate	
	intermediate products, still, normal rate	**0.05**
	intermediate products, still, reduced rate	
	intermediate products, sparkling, normal rate	
	intermediate products, sparkling, reduced rate	
02122 Other wine	fermented beverages other than wine, still, normal rate	0.900
	fermented beverages other than wine, still, reduced rate	0.025
	fermented beverages other than wine, sparkling, normal rate	0.050
	fermented beverages other than wine, sparkling, reduced rate	0.025
02131 Beer	standard rate	
02311 Cigarettes	cigarettes	
02312 Cigars	cigars	
02313 Other tobacco products	fine cut smoking tobacco	0.50
	other smoking tobacco	0.50
04511 Electricity	electricity	
04521 Natural gas	natural gas, heating non-business	
04522 Liquefied hydrocarbons	LPG, heating non-business	
04531 Liquid fuels	gasoil, heating non-business	1.00
	kerosene, heating non-business	0.00
	heavy fuel oil, heating non-business	0.00
04541 Solid fuels	coal and coke, heating non-business	
07221 Fuels and lubricants	petrol, leaded	0.00
	petrol, unleaded	0.50
	gasoil, propellent	0.50
	kerosene, propellent	0.00
	LPG, propellent	0.00
	natural gas, propellent	0.00

- The COICOP level 4 category **02122** *Other wines* is composed of wines from other fruits than grapes, which cover amongst other things, cider, perry, and sake . These are generally subject to the excise tariffs for *Fermented beverages other than wine*.
- The commodity classification of the input data for Austria, which were imputed from the national version of the HBS survey, not the EUROSTAT one, deviates on some points from the official COICOP classification. For example, all sparkling wines and fortified wines, also those derived from grapes, are classified under COICOP level 4 category **02122** *Other wines*, leaving the COICOP level 4 category **02121** only applicable to ordinary still wines (from grapes).
- There are three specific excises on alcoholic drinks that we did not include in the ITT architecture: i) the excises applied to ethyl alcohol produced by small distilleries, ii) beer produced by small breweries, and iii) excise on low alcohol beer. An exception is the Netherlands where the input data contain expenditure shares for COICOP level 4 aggregate **02133** *Low and non-alcoholic beer*. Hence, contrary to what is the case for other countries, excise on low alcoholic beer is included in the policy system `tco_NL`.[52]
- Austria, for which expenditure information was imputed from the national HBS, and for which the good classification does not always follow the COICOP classification, has two sub-groups in alcoholic beverages that do not occur in other countries:
  - **02141** *wholesale alcoholic beverages*;
  - **02142** *side budget alcoholic beverages*.

  We applied the excise tariff for regular beers to these commodities.
- For many products belonging to COICOP level 4 category **04531** *Liquid fuels*, we could not find price information. We apply the tariff for *gas oil* to the category as a whole.
- The excise tariffs applicable to COICOP level 4 aggregate **07221** *Fuels and lubricants for personal transport equipment* are simplified in Table 30 since there are many excise tariffs for products belonging to this category and they vary immensely across countries. The specific excise module in ITT uses the detailed excise tariffs. However, leaded petrol is no longer used in the EU as a fuel for household vehicles. Hence, the weight of excises for leaded petrol is always set to zero.
- As we had no information on prices or expenditure shares of *heavy fuel oil* and *keroseneon*, also the weights for these tariffs are set to zero.
- Further detailed information on excise tariffs for goods belonging to COICOP level 4 aggregate **07221** can be found the country specific `tco_cc` policy systems.

[52] Recall that for the Netherlands, we imputed data from the EUROSTAT 2015 HBS into SILC. These newer data contain sometimes a more expanded set of the COICOP level 4 commodity groups than what was used for the 2010 EUROSTAT HBS versions.

G Consumer price changes simulated by the CAPRI Model

Table 31: Percentage change in consumer prices

	EU	*BE*	DK	DE	AT	NL	FR	PT	ES	EL	IT	IE	FI	SE	UK	CZ	EE	HU	LT	LV	PL	SI	SK	HR	CY	MT	BG	RO
CERE	0.41%	*0.25%*	0.26%	0.32%	0.33%	0.32%	0.38%	0.61%	0.46%	0.61%	0.46%	0.29%	0.34%	0.25%	0.31%	0.81%	0.61%	1.01%	0.50%	0.79%	0.75%	0.62%	0.77%	0.53%	0.46%	0.67%	1.50%	1.15%
POTA	0.07%	*0.10%*	0.10%	0.12%	0.12%	0.12%	0.14%	0.21%	0.17%	0.21%	0.17%	0.11%	0.13%	0.10%	0.12%	0.15%	0.14%	0.16%	0.14%	0.14%	0.15%	0.12%	0.16%	0.13%	0.11%	0.12%	0.21%	0.23%
TOMA	0.17%	*0.13%*	0.14%	0.17%	0.16%	0.17%	0.20%	0.30%	0.24%	0.30%	0.24%	0.15%	0.18%	0.14%	0.16%	0.40%	0.44%	0.36%	0.43%	0.43%	0.42%	0.35%	0.30%	0.45%	0.33%	0.36%	0.64%	0.63%
OVEG	0.80%	*0.64%*	0.66%	0.73%	0.68%	0.73%	0.70%	0.82%	0.85%	0.94%	0.75%	0.68%	0.75%	0.65%	0.71%	1.02%	1.03%	1.06%	1.02%	1.02%	1.02%	0.96%	1.03%	0.82%	0.94%	0.97%	1.12%	1.12%
APPL	0.26%	*0.19%*	0.20%	0.24%	0.24%	0.24%	0.27%	0.39%	0.32%	0.39%	0.32%	0.21%	0.25%	0.19%	0.23%	0.49%	0.45%	0.56%	0.40%	0.52%	0.51%	0.42%	0.54%	0.48%	0.40%	0.44%	0.64%	0.70%
TAGR	0.17%	*0.14%*	0.15%	0.17%	0.17%	0.17%	0.19%	0.27%	0.22%	0.27%	0.23%	0.15%	0.18%	0.14%	0.14%	0.28%	0.25%	0.34%	0.23%	0.28%	0.30%	0.24%	0.31%	0.35%	0.24%	0.14%	0.42%	0.42%
CITR	0.33%	*0.25%*	0.26%	0.31%	0.31%	0.31%	0.36%	0.53%	0.42%	0.53%	0.43%	0.28%	0.33%	0.25%	0.30%	0.55%	0.50%	0.46%	0.50%	0.58%	0.44%	0.47%	0.47%	0.46%	0.47%	0.51%	0.48%	0.51%
OFRU	0.43%	*0.36%*	0.37%	0.40%	0.41%	0.41%	0.45%	0.55%	0.49%	0.48%	0.49%	0.38%	0.43%	0.36%	0.40%	0.45%	0.54%	0.51%	0.32%	0.53%	0.50%	0.48%	0.38%	0.43%	0.46%	0.49%	0.63%	0.57%
TWIN	0.37%	*0.28%*	0.30%	0.35%	0.36%	0.36%	0.41%	0.59%	0.48%	0.37%	0.49%	0.25%	0.25%	0.28%	0.34%	0.46%	0.37%	0.62%	0.39%	0.36%	0.38%	0.53%	0.45%	0.47%	0.33%	0.45%	0.66%	0.52%
TROP	0.03%	*0.26%*	0.07%	0.13%	0.23%	0.32%	-0.05%	-0.04%	-0.05%	-0.01%	0.16%	0.18%	0.16%	0.16%	-0.10%	-0.16%	-0.20%	-0.50%	0.30%	0.01%	0.05%	-0.04%	0.17%	0.05%	-0.03%	0.10%	-0.04%	-0.09%
COFF	0.29%	*0.29%*	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%	0.29%
TEAS	0.57%	*0.57%*	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.57%	0.58%	0.58%	0.58%	0.58%	0.58%	0.58%	0.00%	0.58%	0.00%	0.00%	0.58%	0.58%	0.58%
COCO	-0.73%	*-0.72%*	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.72%	-0.78%	-0.78%	-0.78%	0.00%	-0.78%	-0.78%	-0.78%	-0.78%	-0.78%	-0.78%	-0.78%	-0.78%	-0.78%
TOBA		*0.00%*	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
BEEF	34.54%	*29.17%*	30.28%	33.61%	33.77%	33.78%	32.43%	36.57%	40.18%	45.02%	38.15%	31.12%	32.11%	29.50%	32.88%	42.48%	38.18%	45.23%	28.38%	30.88%	38.12%	36.24%	40.28%	20.42%	35.97%	38.34%	51.55%	39.71%
PORK	5.15%	*3.58%*	3.77%	4.38%	4.41%	4.41%	4.97%	6.97%	5.76%	6.54%	5.43%	3.92%	4.60%	3.63%	4.24%	7.09%	5.87%	7.80%	4.68%	4.97%	6.08%	5.88%	4.64%	3.18%	5.61%	6.10%	9.63%	6.50%
SGMT	48.99%	*38.78%*	42.58%	46.43%	46.61%	46.62%	46.41%	46.03%	53.63%	58.65%	52.42%	43.56%	38.55%	41.65%	45.59%	55.08%	55.93%	62.73%	42.77%	46.93%	50.80%	48.35%	47.02%	30.71%	52.62%	55.16%	69.22%	54.46%
POUM	4.77%	*3.47%*	4.21%	4.71%	4.74%	4.74%	4.48%	6.47%	5.17%	3.96%	3.82%	4.34%	4.26%	4.10%	4.60%	5.31%	4.74%	7.21%	3.05%	3.47%	4.96%	5.80%	3.48%	2.52%	5.85%	6.18%	8.14%	4.67%
EGGS	1.78%	*1.31%*	1.37%	1.57%	1.58%	1.58%	1.76%	2.36%	2.00%	1.98%	2.03%	1.42%	1.64%	1.33%	1.53%	2.28%	2.30%	2.45%	2.27%	2.27%	2.25%	1.98%	1.95%	1.82%	1.78%	2.00%	2.02%	2.81%
ACQU	1.40%	*1.32%*	1.45%	1.41%	1.47%	1.58%	1.31%	1.51%	1.34%	1.32%	1.30%	1.48%	1.40%	1.31%	1.35%	1.67%	1.75%	1.50%	1.95%	1.89%	1.86%	1.63%	1.87%	1.71%	1.39%	1.79%	1.87%	1.89%
BUTT	1.63%	*1.44%*	1.47%	1.57%	1.58%	1.58%	1.66%	1.73%	1.76%	1.43%	1.74%	1.50%	1.61%	1.45%	1.55%	1.93%	1.95%	1.86%	1.77%	1.93%	1.85%	1.94%	1.68%	1.75%	1.83%	1.84%	1.89%	1.87%
SMIP	2.91%	*2.81%*	0.00%	2.89%	2.90%	2.90%	2.94%	3.00%	2.99%	2.96%	2.97%	2.85%	2.91%	2.81%	2.88%	2.97%	2.98%	2.97%	2.98%	2.92%	2.98%	2.98%	2.89%	2.81%	0.00%	2.92%	2.97%	2.98%
CHES	1.91%	*1.69%*	1.73%	1.85%	1.85%	1.85%	1.94%	2.17%	1.86%	1.84%	1.95%	1.76%	1.88%	1.70%	1.82%	2.20%	2.23%	2.21%	2.18%	2.16%	2.26%	2.22%	2.35%	1.82%	2.18%	2.20%	2.35%	2.35%
FRMI	1.41%	*1.24%*	1.27%	1.38%	1.38%	1.38%	1.47%	1.52%	1.56%	1.33%	1.35%	1.30%	1.41%	1.25%	1.36%	1.60%	1.57%	1.56%	1.48%	1.57%	1.62%	1.57%	1.41%	1.46%	1.30%	1.43%	1.43%	1.19%
CREM	1.25%	*1.13%*	1.16%	1.24%	1.25%	1.25%	1.31%	1.41%	1.39%	1.41%	1.39%	1.18%	1.27%	1.14%	1.23%	1.21%	1.21%	1.21%	1.21%	1.21%	1.21%	1.21%	1.20%	1.21%	1.21%	1.21%	1.21%	1.21%
COCM	1.47%	*1.31%*	1.34%	1.44%	1.44%	1.44%	1.51%	1.62%	1.60%	1.62%	1.61%	1.37%	0.00%	1.32%	1.42%	1.51%	1.51%	1.51%	1.51%	1.51%	1.51%	1.51%	1.48%	1.51%	1.42%	1.50%	1.51%	1.51%
WMIO	2.14%	*2.04%*	2.05%	2.09%	2.09%	2.09%	2.12%	2.16%	2.15%	2.14%	2.15%	2.06%	2.10%	2.04%	2.08%	2.62%	2.64%	2.64%	2.64%	2.63%	2.62%	2.64%	2.57%	2.51%	2.52%	2.57%	2.59%	2.64%
CASE	2.68%	*2.58%*	2.60%	2.66%	2.67%	0.00%	2.71%	2.75%	2.75%	2.77%	2.77%	2.62%	2.68%	0.00%	2.65%	0.00%	0.00%	2.76%	2.76%	0.00%	2.76%	0.00%	2.76%	2.76%	0.00%	2.69%	2.76%	0.00%
WHEP	3.56%	*0.00%*	3.48%	3.55%	0.00%	3.55%	3.60%	3.68%	3.66%	0.00%	3.66%	3.50%	0.00%	3.47%	3.54%	3.34%	3.34%	3.34%	3.31%	0.00%	3.33%	0.00%	3.34%	3.20%	0.00%	0.00%	3.32%	3.34%
OILP	0.37%	*0.21%*	0.19%	0.23%	0.30%	0.24%	0.90%	0.37%	0.47%	0.60%	0.27%	0.25%	0.37%	0.14%	0.28%	0.37%	0.41%	0.33%	0.32%	0.32%	0.27%	0.28%	0.30%	0.37%	0.37%	0.37%	0.46%	0.38%
RICE	1.93%	*1.26%*	1.33%	1.59%	1.61%	1.61%	1.86%	2.88%	2.24%	2.50%	2.28%	1.40%	1.69%	1.28%	1.53%	3.96%	4.04%	4.60%	3.93%	3.93%	3.86%	3.02%	4.10%	2.82%	2.83%	2.56%	5.13%	4.30%
SUGA	0.10%	*0.06%*	0.07%	0.08%	0.08%	0.08%	0.08%	0.16%	0.12%	0.16%	0.12%	0.07%	0.09%	0.06%	0.08%	0.21%	0.22%	0.25%	0.21%	0.21%	0.20%	0.16%	0.22%	0.26%	0.14%	0.16%	0.39%	0.38%

H CAPRI-COICOP mapping of commodity aggregates

Table 32: CAPRI-COICOP mapping of commodity aggregates

CAPRI code		**COICOP Equivalent**
RICE	Rice milled	01111 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 1 Rice
CERE	Cereals	01112 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 2 Bread
		01113 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 3 Pasta products
		01114 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 4 Pastry-cook products
		01115 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 5 Sandwiches
		01116 : 01 Food and nonalcoholic beverages - 1 Food - 1 Bread and cereals - 6 Other products
BEEF	Beef	01121 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 1 Fresh, chilled or frozen meat of bovine animals
		01125 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 5 Dried, salted or smoked meat and edible meat offal
		01126 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 6 Other preserved or processed meat and meat preparations
		01127 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 7 Other fresh, chilled or frozen edible meat
PORK	Pork meat	01122 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 2 Fresh, chilled or frozen meat of swine
SGMT	Sheep and goat meat	01123 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 3 Fresh, chilled or frozen meat of sheep and goat
POUM	Poultry meat	01124 : 01 Food and nonalcoholic beverages - 1 Food - 2 Meat - 4 Fresh, chilled or frozen meat of poultry
ACQU	Fish and other aquatic products	01131 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 1 Fresh, chilled or frozen fish
		01132 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 2 Fresh, chilled or frozen seafood
		01133 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 3 Dried, smoked or salted fish and seafood
		01134 : 01 Food and nonalcoholic beverages - 1 Food - 3 Fish - 4 Other preserved or processed fish and seafood and fish and seafood preparations
FRMI	Fresh milk products	01141 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 1 Whole milk
		01142 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 2 Low fat milk
		01143 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 3 Preserved milk
		01144 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 4 Yoghurt
CHES	Cheese	01145 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 5 Cheese and curd
		01146 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 6 Other milk products
EGGS	Eggs	01147 : 01 Food and nonalcoholic beverages - 1 Food - 4 Milk, cheese and eggs - 7 Eggs
BUTT	Butter	01151 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 1 Butter
OILP	Oils	01152 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 2 Margarine and other vegetable fats
		01153 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 3 Olive oil
		01154 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 4 Edible oils
		01155 : 01 Food and nonalcoholic beverages - 1 Food - 5 Oils and fats - 5 Other edible animal fats
CITR	Citrus fruits	01161 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 1 Citrus fruits (fresh, chilled or frozen)
OFRU	Other fruits	01162 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 2 Bananas (fresh, chilled or frozen)
		01166 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 6 Berries (fresh, chilled or frozen)
		01167 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 7 Other fresh, chilled or frozen fruits
		01168 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 8 Dried fruit
		01169 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 9 Preserved fruit and fruit based products
APPL	Apples pears and peaches	01163 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 3 Apples (fresh, chilled or frozen)
		01164 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 4 Pears (fresh, chilled or frozen)
		01165 : 01 Food and nonalcoholic beverages - 1 Food - 6 Fruit - 5 Stone fruits (fresh, chilled or frozen)
OVEG	Other vegetables	01171 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 1 Leaf and stem vegetables (fresh, chilled or frozen)
		01172 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 2 Cabbages (fresh, chilled or frozen)
		01173 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 3 Vegetables cultivated for their fruit (fresh, chilled or frozen)
		01174 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 4 Root crops, non-starchy bulbs and mushrooms (fresh, chilled or frozen)
		01175 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 5 Dried vegetables
		01176 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 6 Other preserved or processed vegetables
		01178 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 8 Other tubers and products of tuber vegetables
POTA	Potatoes	01177 : 01 Food and nonalcoholic beverages - 1 Food - 7 Vegetables - 7 Potatoes
SUGA	Sugar	01181 : 01 Food and nonalcoholic beverages - 1 Food - 8 Sugar, jam, honey, chocolate and confectionery - 1 Sugar
TROP	Coffee, Teas and Cocoa	
COFF	Coffee	01211 : 01 Food and nonalcoholic beverages - 2 Non alcoholic beverages - 1 Coffee, tea and cocoa - 1 Coffee
TEAS	Tea	01212 : 01 Food and nonalcoholic beverages - 2 Non alcoholic beverages - 1 Coffee, tea and cocoa - 2 Tea
COCO	Cocoa	01213 : 01 Food and nonalcoholic beverages - 2 Non alcoholic beverages - 1 Coffee, tea and cocoa - 3 Cocoa and powdered chocolate

I Income and expenditure shares of CAPRI aggregates

Table 33: Income shares of food commodities: CAPRI Aggregates

Decile	rice	cereals	beef	pork	goat	poultry	fish	milk	cheese	eggs	butter	oils	citrus	other fruit	apple	vegetable	potato	sugar	coffee	tea	cocoa
1	0.11%	4.96%	5.42%	0.32%	0.22%	0.86%	1.39%	0.86%	2.12%	0.28%	0.27%	0.53%	0.37%	1.19%	0.54%	2.87%	0.48%	0.10%	0.44%	0.10%	0.02%
2	0.06%	3.36%	3.96%	0.29%	0.09%	0.50%	1.00%	0.60%	1.43%	0.17%	0.19%	0.30%	0.24%	0.72%	0.36%	1.71%	0.33%	0.06%	0.38%	0.07%	0.02%
3	0.05%	3.28%	3.70%	0.24%	0.12%	0.47%	0.99%	0.56%	1.28%	0.15%	0.19%	0.25%	0.22%	0.59%	0.31%	1.66%	0.26%	0.07%	0.31%	0.04%	0.02%
4	0.04%	3.09%	3.42%	0.24%	0.13%	0.47%	0.88%	0.53%	1.28%	0.13%	0.13%	0.23%	0.19%	0.60%	0.32%	1.52%	0.21%	0.05%	0.24%	0.05%	0.01%
5	0.04%	2.82%	3.16%	0.23%	0.08%	0.40%	0.79%	0.49%	1.14%	0.11%	0.13%	0.18%	0.21%	0.53%	0.26%	1.30%	0.20%	0.05%	0.22%	0.03%	0.02%
6	0.04%	2.73%	2.98%	0.20%	0.07%	0.37%	0.75%	0.45%	1.17%	0.11%	0.12%	0.18%	0.18%	0.52%	0.23%	1.33%	0.17%	0.04%	0.18%	0.03%	0.02%
7	0.03%	2.54%	2.63%	0.17%	0.07%	0.35%	0.71%	0.40%	1.15%	0.09%	0.10%	0.18%	0.19%	0.44%	0.23%	1.14%	0.15%	0.04%	0.16%	0.03%	0.01%
8	0.03%	2.40%	2.47%	0.17%	0.07%	0.36%	0.71%	0.40%	1.00%	0.08%	0.08%	0.14%	0.15%	0.43%	0.21%	1.05%	0.15%	0.03%	0.14%	0.03%	0.01%
9	0.03%	2.15%	2.22%	0.15%	0.07%	0.33%	0.64%	0.38%	0.87%	0.07%	0.07%	0.13%	0.14%	0.39%	0.20%	0.98%	0.11%	0.03%	0.13%	0.03%	0.01%
10	0.02%	1.78%	1.74%	0.11%	0.07%	0.24%	0.59%	0.30%	0.78%	0.07%	0.06%	0.11%	0.14%	0.38%	0.18%	0.88%	0.10%	0.02%	0.12%	0.03%	0.01%
Total	0.04%	2.91%	3.17%	0.21%	0.10%	0.43%	0.84%	0.50%	1.22%	0.13%	0.14%	0.22%	0.20%	0.58%	0.28%	1.44%	0.22%	0.05%	0.23%	0.04%	0.01%

Table 34: Expenditure shares of food commodities: CAPRI Aggregates

Decile	rice	cereals	beef	pork	goat	poultry	fish	milk	cheese	eggs	butter	oils	citrus	other fruit	apple	vegetable	potato	sugar	coffee	tea	cocoa
1	0.07%	3.28%	3.66%	0.24%	0.19%	0.55%	0.94%	0.59%	1.32%	0.20%	0.17%	0.33%	0.22%	0.70%	0.33%	1.79%	0.33%	0.08%	0.33%	0.05%	0.02%
2	0.08%	3.41%	3.96%	0.31%	0.12%	0.54%	0.98%	0.63%	1.44%	0.17%	0.19%	0.30%	0.26%	0.71%	0.36%	1.78%	0.32%	0.07%	0.38%	0.07%	0.02%
3	0.05%	3.48%	3.97%	0.26%	0.13%	0.48%	1.01%	0.59%	1.37%	0.16%	0.21%	0.27%	0.23%	0.62%	0.34%	1.77%	0.28%	0.07%	0.32%	0.04%	0.02%
4	0.04%	3.34%	3.72%	0.28%	0.13%	0.53%	0.99%	0.62%	1.45%	0.14%	0.14%	0.25%	0.21%	0.64%	0.34%	1.64%	0.24%	0.06%	0.26%	0.05%	0.01%
5	0.04%	3.42%	3.79%	0.28%	0.10%	0.50%	0.95%	0.63%	1.39%	0.14%	0.16%	0.22%	0.25%	0.63%	0.30%	1.53%	0.24%	0.06%	0.25%	0.04%	0.02%
6	0.05%	3.35%	3.68%	0.24%	0.07%	0.47%	0.88%	0.56%	1.43%	0.14%	0.15%	0.21%	0.23%	0.62%	0.28%	1.63%	0.21%	0.06%	0.22%	0.03%	0.02%
7	0.03%	3.43%	3.48%	0.23%	0.09%	0.48%	0.94%	0.57%	1.56%	0.12%	0.13%	0.25%	0.28%	0.61%	0.31%	1.56%	0.21%	0.06%	0.21%	0.03%	0.02%
8	0.04%	3.32%	3.45%	0.24%	0.09%	0.50%	0.97%	0.59%	1.38%	0.11%	0.12%	0.20%	0.21%	0.62%	0.30%	1.47%	0.22%	0.04%	0.20%	0.04%	0.02%
9	0.04%	3.49%	3.55%	0.26%	0.10%	0.53%	0.97%	0.60%	1.37%	0.11%	0.12%	0.20%	0.22%	0.63%	0.33%	1.56%	0.18%	0.05%	0.20%	0.04%	0.01%
10	0.04%	3.14%	3.20%	0.20%	0.13%	0.43%	1.01%	0.54%	1.40%	0.12%	0.11%	0.20%	0.25%	0.70%	0.33%	1.58%	0.19%	0.04%	0.20%	0.06%	0.02%
Total	0.05%	3.37%	3.64%	0.26%	0.11%	0.50%	0.96%	0.59%	1.41%	0.14%	0.15%	0.24%	0.24%	0.65%	0.32%	1.63%	0.24%	0.06%	0.26%	0.04%	0.02%

Deciles are constructed with respect to the population of individuals, on the basis of equivalized real disposable incomes. Expenditure shares are household expenditures on a commodity aggregate divided by total expenditures. Entries are means of the shares taken with respect to the population of individuals belonging to a certain cell. Households with negative disposable incomes are included.

J Summary statistics

J.1 Summary statistics for Austria

Table 35: Frequency table of region

region	HBS	SILC
1	0.4488677	0.4490420
2	0.2053666	0.2051908
3	0.3457658	0.3457672

Table 36: Frequency table of household size

HH_size	HBS	SILC
1	0.3574754	0.3603970
2	0.2857688	0.2879672
3	0.1600656	0.1611334
4	0.1291752	0.1220589
5	0.0479270	0.0492728
6	0.0124725	0.0136014
7	0.0051440	0.0040713
8	0.0011418	0.0014106
9	0.0008297	0.0000873

Table 37: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.9042985	0.9110920
1	0.0761292	0.0682225
2	0.0179722	0.0194489
3	0.0014925	0.0012366
4	0.0001076	0.0000000

Table 38: Frequency table of number of household members between 5 and 14 years old

num_HH_mem_btw_5_14	HBS	SILC
0	0.8368685	0.8448845
1	0.1041722	0.0992951
2	0.0506371	0.0466925
3	0.0073775	0.0082657
4	0.0009447	0.0008090
5	0.0000000	0.0000531

Table 39: Frequency table of number of household members between 15 and 24 years old

num_HH_mem_btw_15_24	HBS	SILC
0	0.7796057	0.7853132
1	0.1511333	0.1488924
2	0.0563538	0.0555679
3	0.0113461	0.0093678
4	0.0012892	0.0008587
5	0.0002720	0.0000000

Table 40: Frequency table of number of students in the household between 15 and 24 years old

num_HH_mem_btw_15_24_student	HBS	SILC
0	0.8844503	0.8985027
1	0.0942334	0.0822131
2	0.0179696	0.0180981
3	0.0028957	0.0011157
4	0.0004509	0.0000704

Table 41: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2347832	0.2409708
1	0.3276887	0.3152323
2	0.3970310	0.3953241
3	0.0339669	0.0388545
4	0.0055593	0.0078325
5	0.0008713	0.0017858
6	0.0000995	0.0000000

Table 42: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.7113036	0.7065839
1	0.1894627	0.1977620
2	0.0977512	0.0942928
3	0.0014825	0.0013613

Table 43: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.3453137	0.3736925
1	0.3476808	0.3401240
2	0.2473732	0.2165170
3	0.0447248	0.0509935
4	0.0124874	0.0156752
5	0.0024201	0.0021709
6	0.0000000	0.0008269

Table 44: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.9307146	0.9200003
1	0.0624055	0.0735282
2	0.0067992	0.0051883
3	0.0000807	0.0012831

Table 45: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.3557221	0.3531992
two adults with no dependent children	0.2570966	0.2590941
more than 2 adults with no dependent children	0.0852169	0.0785057
one adult with dependent children	0.0216860	0.0421624
two adults with dependent children	0.1535606	0.1978220
more than 2 adults with dependent children	0.1267178	0.0616193
other	0.0000000	0.0075972

Table 46: Frequency table of whether the household head is a farmer

ref_per_farmer	HBS	SILC
FALSE	0.9638395	0.9911628
TRUE	0.0361605	0.0088372

Table 47: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.9468915	0.9236525
1	0.0320589	0.0443446
2	0.0089492	0.0094528
3	0.0039665	0.0069840
4	0.0046622	0.0099661
5	0.0020985	0.0031290
6	0.0009327	0.0019775
7	0.0003208	0.0004011
8	0.0001196	0.0000924

Table 48: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2465776	0.2422112
1	0.6047917	0.6117444
2	0.1148010	0.1146162
3	0.0282261	0.0261190
4	0.0054788	0.0050736
5	0.0001248	0.0002356

Table 49: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.6191547	0.5992270
1	0.2736576	0.2890095
2	0.1053103	0.1086058
3	0.0016910	0.0028934
4	0.0001865	0.0002644

Table 50: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9877496	0.9836722
1	0.0120714	0.0159374
2	0.0001790	0.0003903

Table 51: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.7661016	0.7437821
1	0.1903508	0.1988955
2	0.0411198	0.0541787
3	0.0019361	0.0028711
4	0.0004917	0.0002726

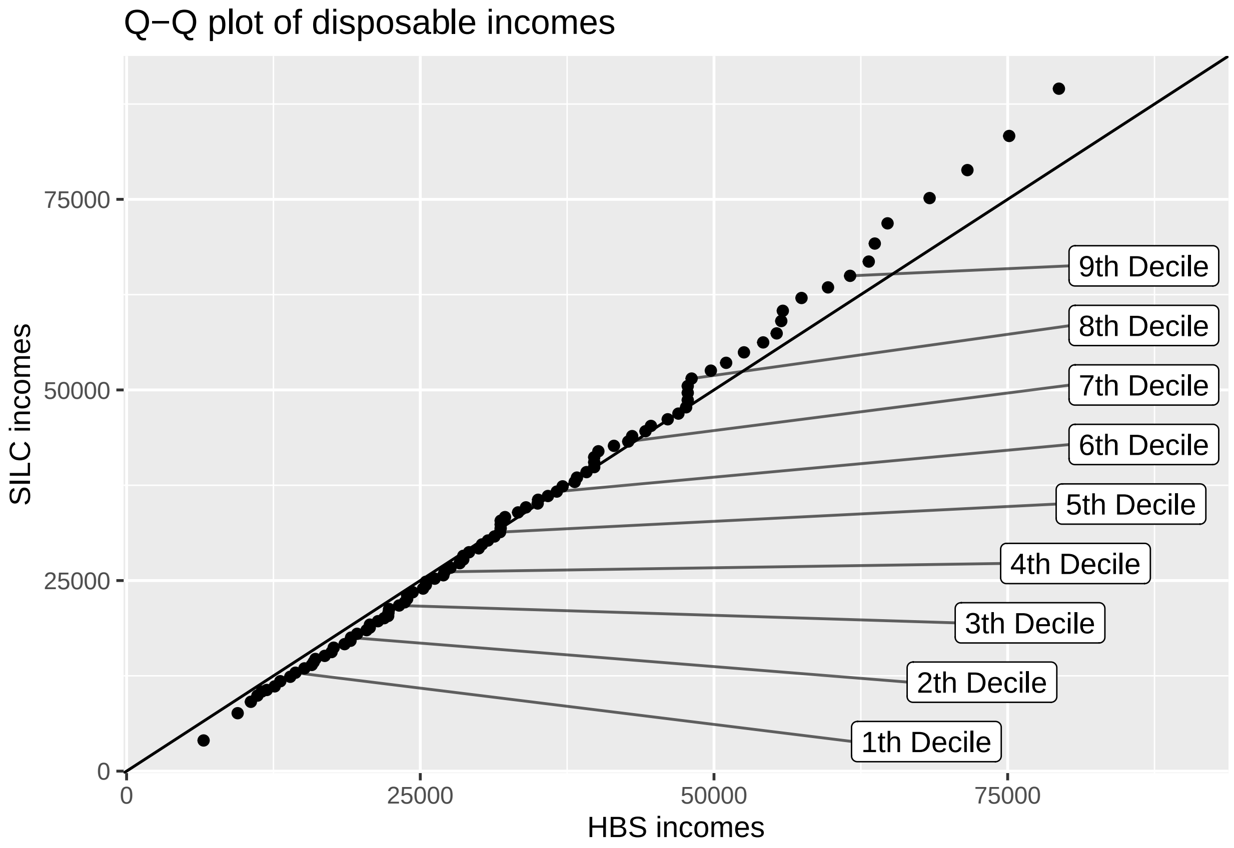


Figure 20: Q-Q plot for disposable income, Austria

J.2 Summary statistics for Bulgaria

Table 52: Frequency table of region

region	HBS	SILC
BG3	0.5187274	0.5256666
BG4	0.4812726	0.4743334

Table 53: Frequency table of household size

HH_size	HBS	SILC
1	0.2548104	0.1948682
2	0.3365722	0.2784111
3	0.2003509	0.2045693
4	0.1431859	0.1933447
5	0.0423616	0.0721219
6	0.0147151	0.0354764
7	0.0049979	0.0110707
8	0.0023453	0.0054658
9	0.0000000	0.0022807
10	0.0006608	0.0005712
11	0.0000000	0.0005452
12	0.0000000	0.0001810
13	0.0000000	0.0003359
14	0.0000000	0.0004040
15	0.0000000	0.0000526
20	0.0000000	0.0000719
25	0.0000000	0.0002293

Table 54: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.9206333	0.8952741
1	0.0735524	0.0925573
2	0.0055769	0.0112430
3	0.0002374	0.0006762
4	0.0000000	0.0002494

Table 55: Frequency table of number of household members between 5 and 15 years old

num_HH_mem_btw_5_15	HBS	SILC
0	0.8558358	0.7978619
1	0.1113443	0.1524289
2	0.0298397	0.0447274
3	0.0020097	0.0020724
4	0.0009706	0.0022160
5	0.0000000	0.0004642
9	0.0000000	0.0002293

Table 56: Frequency table of number of household members between 16 and 24 years old

num_HH_mem_btw_16_24	HBS	SILC
0	0.8160195	0.7298666
1	0.1357650	0.1904420
2	0.0454029	0.0697180
3	0.0028126	0.0079731
4	0.0000000	0.0013633
5	0.0000000	0.0005652
8	0.0000000	0.0000719

Table 57: Frequency table of number of students in the household between 16 and 24 years old

num_HH_mem_btw_16_24_student	HBS	SILC
0	0.8998147	0.8724050
1	0.0830385	0.1077247
2	0.0164789	0.0189984
3	0.0006680	0.0006379
4	0.0000000	0.0002340

Table 58: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.3017554	0.2162993
1	0.2171334	0.1987344
2	0.3706010	0.3996121
3	0.0798937	0.1228269
4	0.0245336	0.0508646
5	0.0048924	0.0082554
6	0.0011906	0.0026513
7	0.0000000	0.0005267
11	0.0000000	0.0002293

Table 59: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.5182906	0.6056456
1	0.3311702	0.2791048
2	0.1497014	0.1143580
3	0.0008378	0.0008916

Table 60: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.4703221	0.3279670
1	0.2683158	0.2451927
2	0.2190325	0.3064372
3	0.0357378	0.0898683
4	0.0052092	0.0267939
5	0.0011453	0.0037410
6	0.0002374	0.0000000

Table 61: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.7671622	0.8070101
1	0.1800492	0.1484111
2	0.0374996	0.0309370
3	0.0114263	0.0098342
4	0.0024579	0.0022908
5	0.0011124	0.0010149
6	0.0002925	0.0004299
11	0.0000000	0.0000719

Table 62: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.2548104	0.1934114
two adults with no dependent children	0.3255183	0.2572986
more than 2 adults with no dependent children	0.1313706	0.1464587
one adult with dependent children	0.0180262	0.0346194
two adults with dependent children	0.1762294	0.1953146
more than 2 adults with dependent children	0.0940450	0.1713539
other	0.0000000	0.0015434

Table 63: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.9991557	0.9945807
1	0.0008443	0.0045953
2	0.0000000	0.0002614
3	0.0000000	0.0005626

Table 64: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2411141	0.1754907
1	0.5533230	0.5240289
2	0.1649936	0.2342889
3	0.0381329	0.0591155
4	0.0024363	0.0059845
5	0.0000000	0.0005788
6	0.0000000	0.0005127

Table 65: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.4172266	0.5098542
1	0.3724093	0.3269331
2	0.2067237	0.1586524
3	0.0036404	0.0044476
4	0.0000000	0.0001127

Table 66: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9249016	0.9533869
1	0.0682219	0.0446542
2	0.0068765	0.0019589

Table 67: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.7225353	0.7020139
1	0.1806152	0.1727870
2	0.0837205	0.0995627
3	0.0118682	0.0207937
4	0.0012608	0.0039971
5	0.0000000	0.0004555
6	0.0000000	0.0003899

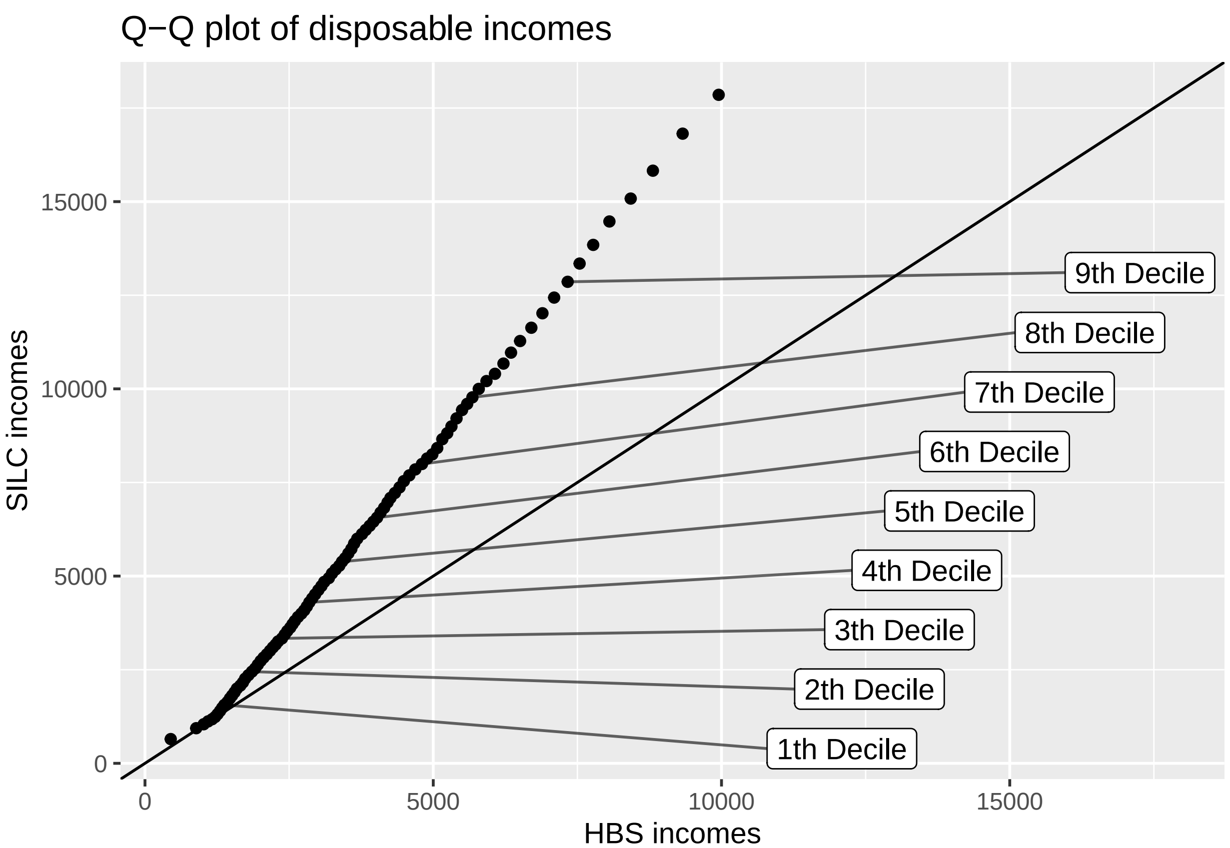


Figure 21: Q-Q plot of disposable income, Bulgaria

J.3 Summary statistics for Estonia

Table 68: Frequency table of household size

HH_size	HBS	SILC
1	0.3701040	0.3450042
2	0.3014597	0.2925835
3	0.1717536	0.1859475
4	0.1096290	0.1239595
5	0.0344129	0.0356611
6	0.0083532	0.0116232
7	0.0030819	0.0031577
8	0.0005407	0.0010042
9	0.0002852	0.0002238
10	0.0002937	0.0003395
11	0.0000000	0.0004505
13	0.0000859	0.0000000
16	0.0000000	0.0000452

Table 69: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.8958592	0.8836221
1	0.0914033	0.0982113
2	0.0123061	0.0171907
3	0.0004314	0.0008584
4	0.0000000	0.0001174

Table 70: Frequency table of number of household members between 5 and 15 years old

num_HH_mem_btw_5_15	HBS	SILC
0	0.8315456	0.8387459
1	0.1215181	0.1219143
2	0.0398583	0.0344086
3	0.0060015	0.0036201
4	0.0007828	0.0007654
5	0.0002937	0.0001831
6	0.0000000	0.0003625

Table 71: Frequency table of number of household members between 16 and 24 years old

num_HH_mem_btw_16_24	HBS	SILC
0	0.7677457	0.7506031
1	0.1828867	0.1879532
2	0.0441674	0.0539870
3	0.0042008	0.0060467
4	0.0009523	0.0011636
5	0.0000000	0.0001686
6	0.0000471	0.0000777

Table 72: Frequency table of number of students in the household between 16 and 24 years old

num_HH_mem_btw_16_24_student	HBS	SILC
0	0.8452235	0.8667610
1	0.1309311	0.1103928
2	0.0218439	0.0208443
3	0.0016461	0.0018497
4	0.0003554	0.0001523

Table 73: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2651016	0.2473337
1	0.3354676	0.3283230
2	0.3676121	0.3818823
3	0.0280054	0.0362997
4	0.0034967	0.0058863
5	0.0002694	0.0002751
6	0.0000473	0.0000000

Table 74: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.6871326	0.6900687
1	0.2442295	0.2365394
2	0.0680288	0.0729134
3	0.0006091	0.0004786

Table 75: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.3705145	0.3486673
1	0.3855519	0.3778351
2	0.2193521	0.2402676
3	0.0208444	0.0291798
4	0.0035634	0.0037238
5	0.0001737	0.0003264

Table 76: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.8385634	0.8262512
1	0.1418403	0.1532283
2	0.0170284	0.0175481
3	0.0025008	0.0024041
4	0.0000670	0.0005684

Table 77: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.3701040	0.3355512
two adults with no dependent children	0.2618396	0.2487112
more than 2 adults with no dependent children	0.0559249	0.0560036
one adult with dependent children	0.0549615	0.0711639
two adults with dependent children	0.2199230	0.2315948
more than 2 adults with dependent children	0.0372470	0.0482988
other	0.0000000	0.0086764

Table 78: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.7971158	0.8040032
1	0.1254494	0.1108647
2	0.0554443	0.0516475
3	0.0163308	0.0208739
4	0.0044009	0.0101575
5	0.0009360	0.0013224
6	0.0000000	0.0003811
7	0.0000000	0.0004941
8	0.0001336	0.0000000
9	0.0001033	0.0001436
10	0.0000000	0.0000668
13	0.0000859	0.0000000
16	0.0000000	0.0000452

Table 79: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.3004712	0.2824730
1	0.5807620	0.5899110
2	0.0995431	0.1062823
3	0.0175220	0.0194380
4	0.0014888	0.0016624
5	0.0001659	0.0001881
6	0.0000471	0.0000452

Table 80: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.6566179	0.6567736
1	0.2663616	0.2635866
2	0.0759568	0.0778603
3	0.0010637	0.0017796

Table 81: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9219744	0.9302691
1	0.0710777	0.0645359
2	0.0066759	0.0050733
3	0.0002720	0.0001217

Table 82: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.5955163	0.5907446
1	0.3061772	0.2898470
2	0.0920689	0.1103917
3	0.0054898	0.0077675
4	0.0007477	0.0012492

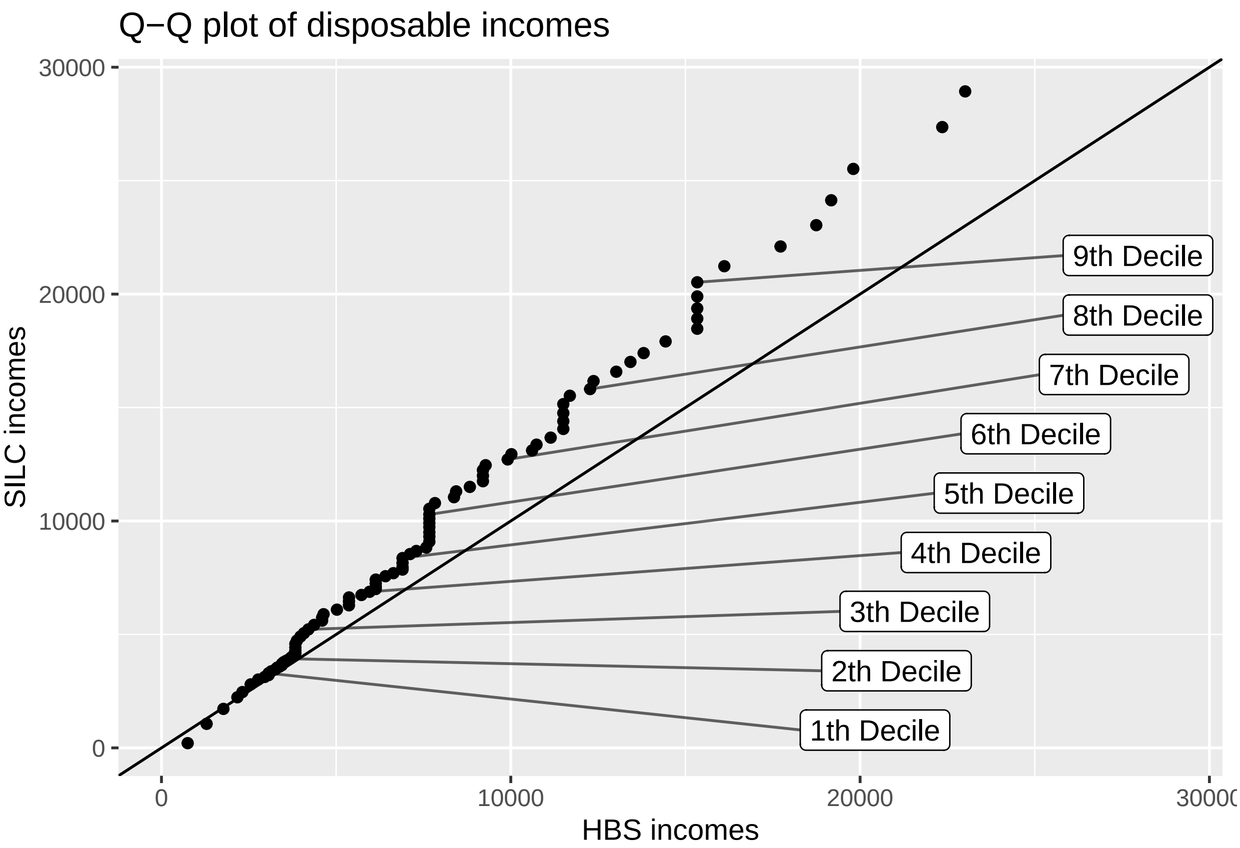


Figure 22: Q-Q plot of disposable income, Estonia

J.4 Summary statistics for Croatia

Table 83: Frequency table of household size

HH_size	HBS	SILC
1	0.2198503	0.2455995
2	0.2960888	0.2564798
3	0.1836309	0.1890877
4	0.1602403	0.1730138
5	0.0866374	0.0789674
6	0.0339321	0.0343716
7	0.0138328	0.0139459
8	0.0037796	0.0057714
9	0.0009488	0.0009166
10	0.0010589	0.0006139
11	0.0000000	0.0005471
12	0.0000000	0.0006853

Table 84: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.9257299	0.8904056
1	0.0624426	0.0950411
2	0.0103605	0.0129762
3	0.0011055	0.0010301
4	0.0003616	0.0000000
5	0.0000000	0.0005471

Table 85: Frequency table of the number of household members between 5 and 14 years old

num_HH_mem_btw_5_14	HBS	SILC
0	0.8096603	0.8006580
1	0.1144850	0.1199235
2	0.0630034	0.0648491
3	0.0103128	0.0127251
4	0.0019378	0.0017455
5	0.0003109	0.0000987
6	0.0002899	0.0000000

Table 86: Frequency table of the number of household members between 15 and 24 years old

num_HH_mem_btw_15_24	HBS	SILC
0	0.7654633	0.7657458
1	0.1455191	0.1543122
2	0.0759168	0.0638883
3	0.0115843	0.0137642
4	0.0011769	0.0014048
5	0.0003396	0.0001994
6	0.0000000	0.0006853

Table 87: Frequency table of number of students in the household between 15 and 24 years old

num_HH_mem_btw_15_24_student	HBS	SILC
0	0.8356308	0.8434306
1	0.1180251	0.1174478
2	0.0427723	0.0333641
3	0.0032322	0.0045636
4	0.0000000	0.0011939
5	0.0003396	0.0000000

Table 88: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2517147	0.2616082
1	0.2024003	0.2042921
2	0.4041130	0.3876730
3	0.1038376	0.1030629
4	0.0357076	0.0350723
5	0.0019318	0.0068913
6	0.0002949	0.0014001

Table 89: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.5528107	0.5585037
1	0.3100673	0.3207999
2	0.1351282	0.1196974
3	0.0019938	0.0009990

Table 90: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.4055538	0.4502255
1	0.2761487	0.2787415
2	0.2363362	0.2153175
3	0.0647786	0.0447519
4	0.0148944	0.0101487
5	0.0016919	0.0008149
6	0.0005964	0.0000000

Table 91: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.7983848	0.7217922
1	0.1624443	0.2125153
2	0.0318850	0.0544074
3	0.0056482	0.0101109
4	0.0007422	0.0006424
5	0.0004931	0.0005318
6	0.0004024	0.0000000

Table 92: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.2198503	0.2425048
two adults with no dependent children	0.2852179	0.2461263
more than 2 adults with no dependent children	0.1519445	0.1350412
one adult with dependent children	0.0196255	0.0226744
two adults with dependent children	0.2038870	0.2219125
more than 2 adults with dependent children	0.1194747	0.1284041
other	0.0000000	0.0033368

Table 93: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.9945540	0.9889667
1	0.0039883	0.0092764
2	0.0012233	0.0003470
3	0.0002344	0.0004617
5	0.0000000	0.0009481

Table 94: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.1944129	0.1958735
1	0.5412005	0.5395827
2	0.2061476	0.2078752
3	0.0510261	0.0501807
4	0.0055969	0.0049720
5	0.0016160	0.0015159

Table 95: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.4201780	0.4131824
1	0.4058706	0.4274121
2	0.1674505	0.1548279
3	0.0061608	0.0045775
4	0.0003400	0.0000000

Table 96: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9669163	0.9815592
1	0.0304441	0.0173425
2	0.0023903	0.0010982
3	0.0002493	0.0000000

Table 97: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.7786998	0.7578441
1	0.1550526	0.1691494
2	0.0569646	0.0622922
3	0.0082988	0.0088262
4	0.0006741	0.0015497
5	0.0003101	0.0003383

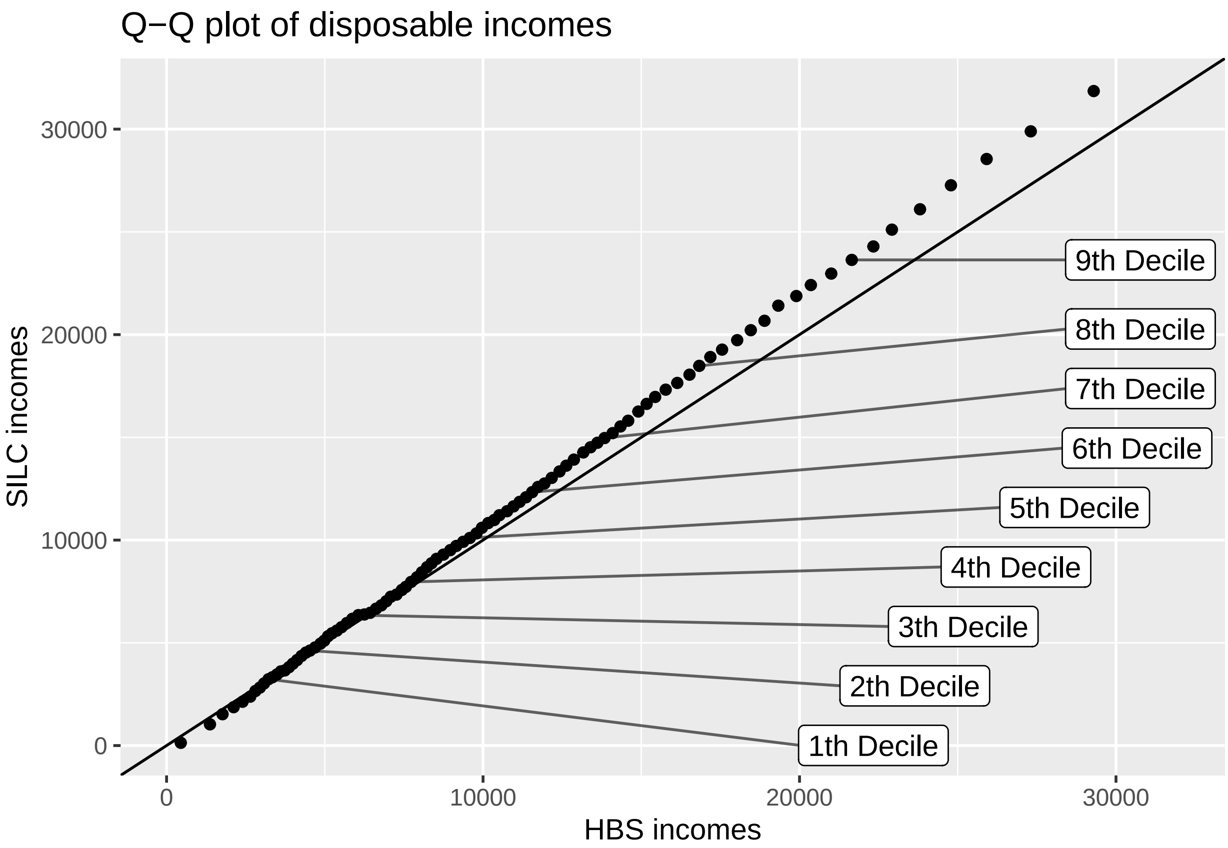


Figure 23: Q-Q plot of disposable income, Croatia

J.5 Summary statistics for Luxembourg

Table 98: Frequency table of household size

HH_size	HBS	SILC
1	0.3330001	0.2940189
2	0.2740000	0.2771210
3	0.1590000	0.1695646
4	0.1505078	0.1907211
5	0.0599603	0.0559190
6	0.0192233	0.0087725
7	0.0026956	0.0021780
8	0.0008270	0.0013721
9	0.0007859	0.0003259
11	0.0000000	0.0000070

Table 99: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.8802730	0.8786468
1	0.0957879	0.0936746
2	0.0224861	0.0273654
3	0.0014531	0.0002687
4	0.0000000	0.0000444

Table 100: Frequency table of number of household members between 5 and 14 years old

num_HH_mem_btw_5_14	HBS	SILC
0	0.8068437	0.8002132
1	0.1181319	0.1260906
2	0.0627516	0.0619379
3	0.0106601	0.0112217
4	0.0014448	0.0002612
5	0.0001679	0.0002754

Table 101: Frequency table of number of household members between 15 and 24 years old

num_HH_mem_btw_15_24	HBS	SILC
0	0.7820093	0.7857585
1	0.1433222	0.1493304
2	0.0590232	0.0540747
3	0.0132619	0.0093537
4	0.0021874	0.0014758
5	0.0001960	0.0000070

Table 102: Frequency table of number of students in the household between 15 and 24 years old

num_HH_mem_btw_15_24_student	HBS	SILC
0	0.8511043	0.8525139
1	0.1046268	0.1091970
2	0.0358514	0.0326088
3	0.0080520	0.0051148
4	0.0003655	0.0005656

Table 103: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.1671478	0.1814416
1	0.3492379	0.2929281
2	0.4371067	0.4627181
3	0.0375947	0.0499366
4	0.0075028	0.0125343
5	0.0012008	0.0003846
6	0.0002093	0.0000567

Table 104: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.7719361	0.7582946
1	0.1609376	0.1511782
2	0.0663292	0.0889543
3	0.0007971	0.0015729

Table 105: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.3200385	0.2953658
1	0.3593599	0.3695782
2	0.2920903	0.2915349
3	0.0235194	0.0368715
4	0.0041314	0.0065764
5	0.0008603	0.0000732

Table 106: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.9013068	0.9237845
1	0.0902585	0.0721166
2	0.0070345	0.0040291
3	0.0014002	0.0000699

Table 107: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.3321232	0.2934959
two adults with no dependent children	0.2524088	0.2562289
more than 2 adults with no dependent children	0.0803417	0.0800631
one adult with dependent children	0.0213930	0.0384494
two adults with dependent children	0.1829432	0.2801408
more than 2 adults with dependent children	0.1307901	0.0507496
other	0.0000000	0.0008722

Table 108: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.9134441	0.9190521
1	0.0522359	0.0441817
2	0.0154233	0.0108362
3	0.0079019	0.0074983
4	0.0071804	0.0111641
5	0.0020971	0.0050030
6	0.0011127	0.0011681
7	0.0004366	0.0003882
8	0.0001679	0.0007081

Table 109: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2086247	0.1930887
1	0.6498096	0.6357192
2	0.1064826	0.1413826
3	0.0305260	0.0279188
4	0.0043710	0.0018908
5	0.0001861	0.0000000

Table 110: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.7517314	0.7249653
1	0.1842777	0.2164540
2	0.0631938	0.0576504
3	0.0007971	0.0003112
4	0.0000000	0.0006190

Table 111: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9325865	0.9484437
1	0.0631063	0.0507798
2	0.0034876	0.0007765
3	0.0008195	0.0000000

Table 112: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.7321886	0.7088085
1	0.1774457	0.1811464
2	0.0859534	0.1042160
3	0.0033874	0.0046855
4	0.0010249	0.0011006
5	0.0000000	0.0000430

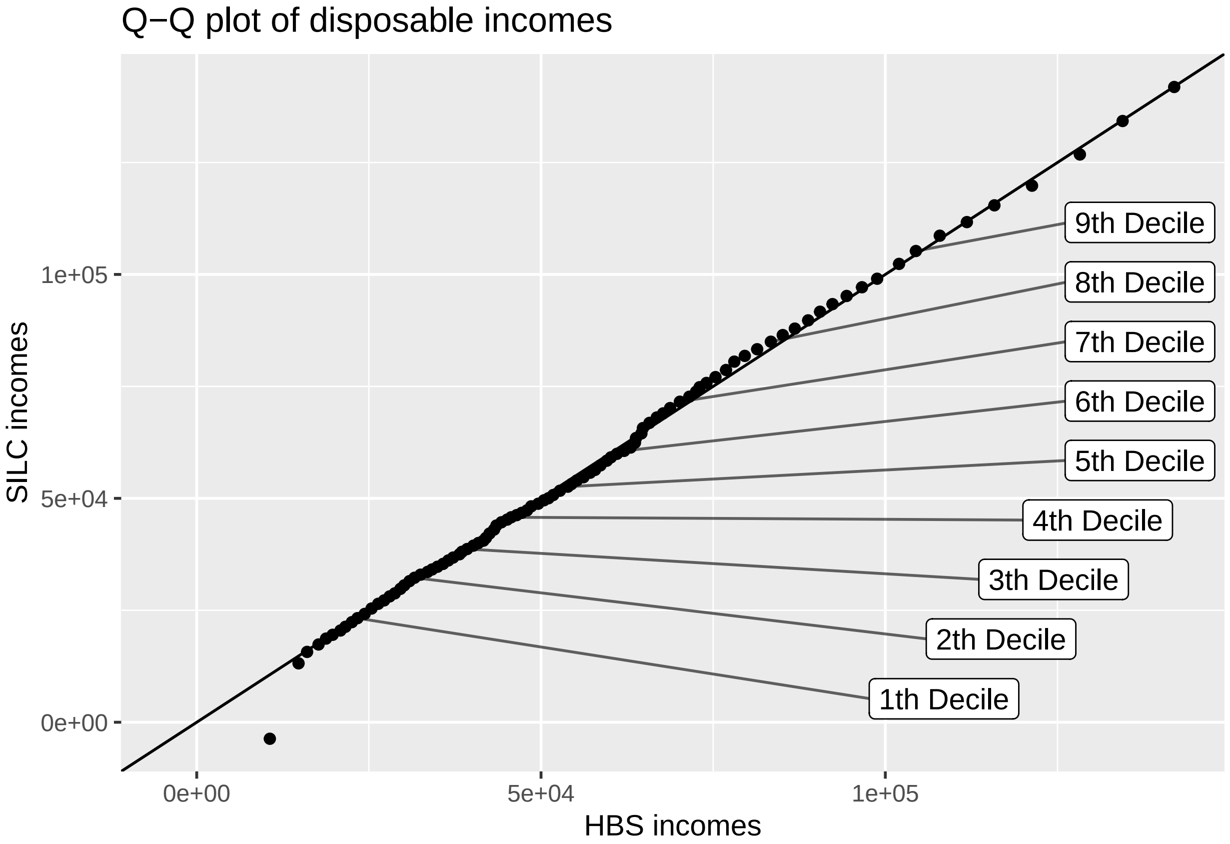


Figure 24: Q-Q plot of disposable income, Luxembourg

J.6 Summary statistics for Latvia

Table 113: Frequency table of household size

HH_size	HBS	SILC
1	0.2571023	0.2743585
2	0.3275271	0.2787133
3	0.2143626	0.2191530
4	0.1286345	0.1356450
5	0.0504186	0.0542445
6	0.0119006	0.0204339
7	0.0055162	0.0101451
8	0.0024909	0.0044206
9	0.0011940	0.0010815
10	0.0006721	0.0008862
11	0.0000000	0.0003829
12	0.0000858	0.0000643
13	0.0000952	0.0000000
15	0.0000000	0.0004712

Table 114: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.8799535	0.8740644
1	0.1037399	0.1101349
2	0.0147372	0.0140094
3	0.0010003	0.0016357
4	0.0005691	0.0000914
5	0.0000000	0.0000643

Table 115: Frequency table of number of household members between 5 and 15 years old

num_HH_mem_btw_5_15	HBS	SILC
0	0.8138104	0.8202563
1	0.1389835	0.1383873
2	0.0397626	0.0347246
3	0.0060315	0.0056792
4	0.0009267	0.0006653
5	0.0003188	0.0002874
6	0.0001666	0.0000000

Table 116: Frequency table of number of household members between 16 and 24 years old

num_HH_mem_btw_16_24	HBS	SILC
0	0.7544724	0.7176864
1	0.1818430	0.2033937
2	0.0543080	0.0673910
3	0.0080451	0.0100177
4	0.0013317	0.0014119
5	0.0000000	0.0000993

Table 117: Frequency table of number of students in the household between 16 and 24 years old

num_HH_mem_btw_16_24_student	HBS	SILC
0	0.8601555	0.8555297
1	0.1131031	0.1199948
2	0.0247373	0.0227215
3	0.0020041	0.0017540

Table 118: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2380439	0.2122492
1	0.2827332	0.3095490
2	0.4133314	0.3858901
3	0.0508709	0.0709332
4	0.0132155	0.0149703
5	0.0012228	0.0045118
6	0.0005823	0.0014348
7	0.0000000	0.0004615

Table 119: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.6451368	0.6532384
1	0.2648928	0.2677089
2	0.0890443	0.0777433
3	0.0009261	0.0012572
4	0.0000000	0.0000522

Table 120: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.3393555	0.3655339
1	0.3498744	0.3444738
2	0.2659935	0.2324240
3	0.0355637	0.0455635
4	0.0076340	0.0102801
5	0.0011862	0.0016073
6	0.0003928	0.0001174

Table 121: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.7835847	0.7497489
1	0.1811531	0.1996526
2	0.0290183	0.0441467
3	0.0049937	0.0050764
4	0.0004666	0.0009042
6	0.0007837	0.0000000
7	0.0000000	0.0004712

Table 122: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.2571023	0.2696261
two adults with no dependent children	0.2985875	0.2254528
more than 2 adults with no dependent children	0.1022696	0.0909659
one adult with dependent children	0.0449871	0.0763982
two adults with dependent children	0.2194116	0.2234148
more than 2 adults with dependent children	0.0776419	0.1084838
other	0.0000000	0.0056582

Table 123: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2669136	0.2609541
1	0.5648055	0.5383123
2	0.1387423	0.1587218
3	0.0239083	0.0357953
4	0.0040781	0.0047997
5	0.0011545	0.0014168
6	0.0003976	0.0000000

Table 124: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.6006052	0.5954363
1	0.2957604	0.3043483
2	0.1027031	0.0976474
3	0.0009313	0.0025680

Table 125: Frequency table of number of disabled household members

num_HH_mem_disab	HBS	SILC
0	0.9468501	0.9344035
1	0.0506018	0.0624176
2	0.0025481	0.0030617
3	0.0000000	0.0001172

Table 126: Frequency table of number of household members with higher education

num_HH_mem_with_higher_edu	HBS	SILC
0	0.6517123	0.6569348
1	0.2416945	0.2393852
2	0.0956813	0.0894832
3	0.0102543	0.0110311
4	0.0006576	0.0024368
5	0.0000000	0.0006682
6	0.0000000	0.0000606

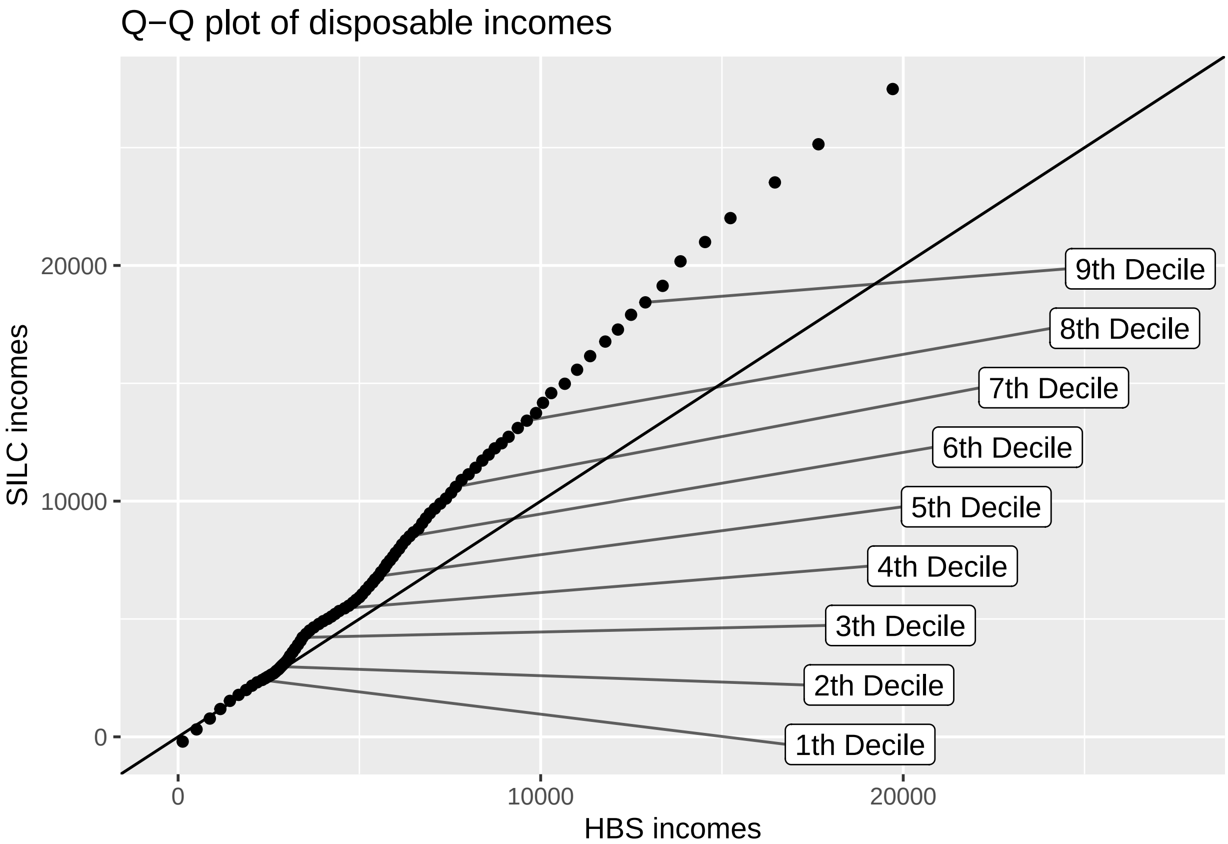


Figure 25: Q-Q plot of disposable income, Latvia

J.7 Summary statistics for Malta

Table 127: Frequency table of household size

HH_size	HBS	SILC
1	0.1883108	0.1878660
2	0.2570275	0.2568041
3	0.2229430	0.2237559
4	0.2202939	0.2202827
5	0.0803801	0.0800895
6	0.0235756	0.0239540
7	0.0046956	0.0055780
8	0.0021162	0.0002780
9	0.0000000	0.0013917
10	0.0006573	0.0000000

Table 128: Frequency table of the number of household members between 0 and 14 years old

num_HH_mem_btw_0_14	HBS	SILC
0	0.6932340	0.6988062
1	0.1698573	0.1718063
2	0.1084320	0.1076525
3	0.0221989	0.0187916
4	0.0050765	0.0015537
5	0.0003769	0.0013898
6	0.0008243	0.0000000

Table 129: Frequency table of the number of household members between 15 and 24 years old

num_HH_mem_btw_15_24	HBS	SILC
0	0.7061607	0.7235025
1	0.1764155	0.1776395
2	0.0939118	0.0823290
3	0.0228124	0.0147177
4	0.0006996	0.0018113

Table 130: Frequency table of number of students in the household between 15 and 24 years old

num_HH_mem_btw_15_24_student	HBS	SILC
0	0.8305072	0.8668362
1	0.1226663	0.1073378
2	0.0413657	0.0232006
3	0.0054608	0.0018679
4	0.0000000	0.0007575

Table 131: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.1656266	0.1864908
1	0.2075779	0.2090870
2	0.5253016	0.4687984
3	0.0877768	0.1093118
4	0.0129082	0.0238884
5	0.0008088	0.0021457
6	0.0000000	0.0002780

Table 132: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.7350707	0.6924001
1	0.1731146	0.2006866
2	0.0895809	0.1032985
3	0.0022338	0.0028170
4	0.0000000	0.0007978

Table 133: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.3179913	0.3339957
1	0.3448090	0.3158142
2	0.2590130	0.2677073
3	0.0622263	0.0666580
4	0.0151932	0.0143626
5	0.0007672	0.0014622

Table 134: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.9339431	0.9252548
1	0.0602913	0.0645777
2	0.0052460	0.0094849
3	0.0005196	0.0006825

Table 135: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.1883108	0.1866820
two adults with no dependent children	0.2425428	0.2383397
more than 2 adults with no dependent children	0.1436653	0.1555460
one adult with dependent children	0.0269853	0.0278636
two adults with dependent children	0.2960436	0.2616611
more than 2 adults with dependent children	0.1024521	0.1285738
other	0.0000000	0.0013337

Table 136: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.1505725	0.1536107
1	0.5941041	0.5903098
2	0.1928751	0.1834819
3	0.0529846	0.0626323
4	0.0087090	0.0099653
5	0.0007547	0.0000000

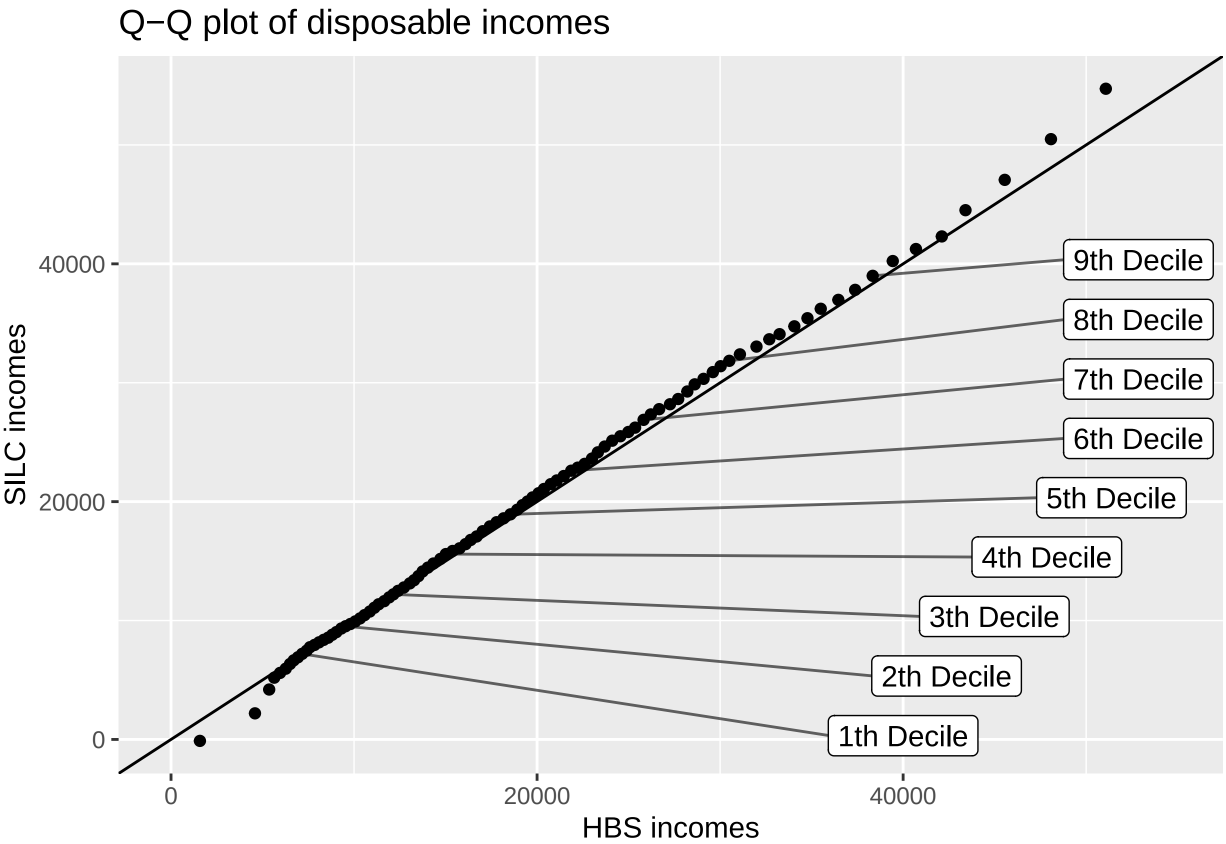


Figure 26: Q-Q plot of disposable income, Malta

J.8 Summary statistics for the Netherlands

Table 137: Frequency table of household size

HH_size	HBS	SILC
1	0.3654075	0.3614221
2	0.3317515	0.3273424
3	0.1222450	0.1229032
4	0.1273838	0.1289068
5	0.0406257	0.0450822
6	0.0090707	0.0112305
7	0.0022532	0.0023281
8	0.0012626	0.0007846

Table 138: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.9083125	0.9056294
1	0.0668542	0.0657670
2	0.0232552	0.0266211
3	0.0015782	0.0018674
4	0.0000000	0.0001152

Table 139: Frequency table of number of household members between 5 and 15 years old

num_HH_mem_btw_5_15	HBS	SILC
0	0.8289611	0.8330083
1	0.0837325	0.0890847
2	0.0674287	0.0579476
3	0.0170266	0.0167540
4	0.0023154	0.0027981
5	0.0004437	0.0004072
6	0.0000920	0.0000000

Table 140: Frequency table of number of household members between 16 and 24 years old

num_HH_mem_btw_16_24	HBS	SILC
0	0.8165383	0.7961839
1	0.1306620	0.1450661
2	0.0453074	0.0497107
3	0.0063374	0.0083271
4	0.0009993	0.0007122
5	0.0001556	0.0000000

Table 141: Frequency table of number of students in the household between 16 and 24 years old

num_HH_mem_btw_16_24_student	HBS	SILC
0	0.8751443	0.8819299
1	0.1016388	0.0956756
2	0.0212696	0.0196758
3	0.0011792	0.0023407
4	0.0007681	0.0003779

Table 142: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2632307	0.2437590
1	0.3317814	0.3065718
2	0.3856485	0.4345055
3	0.0166621	0.0134332
4	0.0024908	0.0015651
5	0.0001866	0.0000000
6	0.0000000	0.0001654

Table 143: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.7206514	0.7639701
1	0.1790420	0.1462485
2	0.1001808	0.0890994
3	0.0001257	0.0003594
4	0.0000000	0.0003226

Table 144: Frequency table of number of employed household members

num_HH_mem_emp	HBS	SILC
0	0.4005166	0.3587543
1	0.2942509	0.3085688
2	0.2799725	0.2984466
3	0.0211973	0.0278753
4	0.0037439	0.0058269
5	0.0003188	0.0003136
6	0.0000000	0.0002146

Table 145: Frequency table of number of unemployed household members

num_HH_mem_unemp	HBS	SILC
0	0.9008280	0.9589202
1	0.0923739	0.0395910
2	0.0067743	0.0014888
3	0.0000238	0.0000000

Table 146: Frequency table of the household types

HH_type	HBS	SILC
one adult with no dependent children	0.3654075	0.3329451
two adults with no dependent children	0.3064413	0.3083479
more than 2 adults with no dependent children	0.0345902	0.0323482
one adult with dependent children	0.0505595	0.0403217
two adults with dependent children	0.2240881	0.2355850
more than 2 adults with dependent children	0.0189133	0.0235361
other	0.0000000	0.0269160

Table 147: Frequency table of whether the household head is a farmer

ref_per_farmer	HBS	SILC
FALSE	1	1

Table 148: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	1	0.9843474
1	0	0.0136029
2	0	0.0004855
3	0	0.0007747
5	0	0.0007895

Table 149: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2338183	0.2215567
1	0.6635823	0.6804395
2	0.0814069	0.0781244
3	0.0184398	0.0181766
4	0.0016915	0.0016358
5	0.0009362	0.0000670
6	0.0001251	0.0000000

Table 150: Frequency table of number of pensioner household members

num_HH_mem_pens	HBS	SILC
0	0.6956562	0.7982515
1	0.1995565	0.1530281
2	0.1046616	0.0486406
3	0.0001257	0.0000798

Table 151: Frequency table of number of disabled household members

num__HH__mem__disab	HBS	SILC
0	0.9419990	0.9389968
1	0.0558203	0.0592558
2	0.0021808	0.0017474

Table 152: Frequency table of number of household members with higher education

num__HH__mem__with__higher__edu	HBS	SILC
0	1	0.6084618
1	0	0.2810357
2	0	0.1086315
3	0	0.0016862
4	0	0.0001848

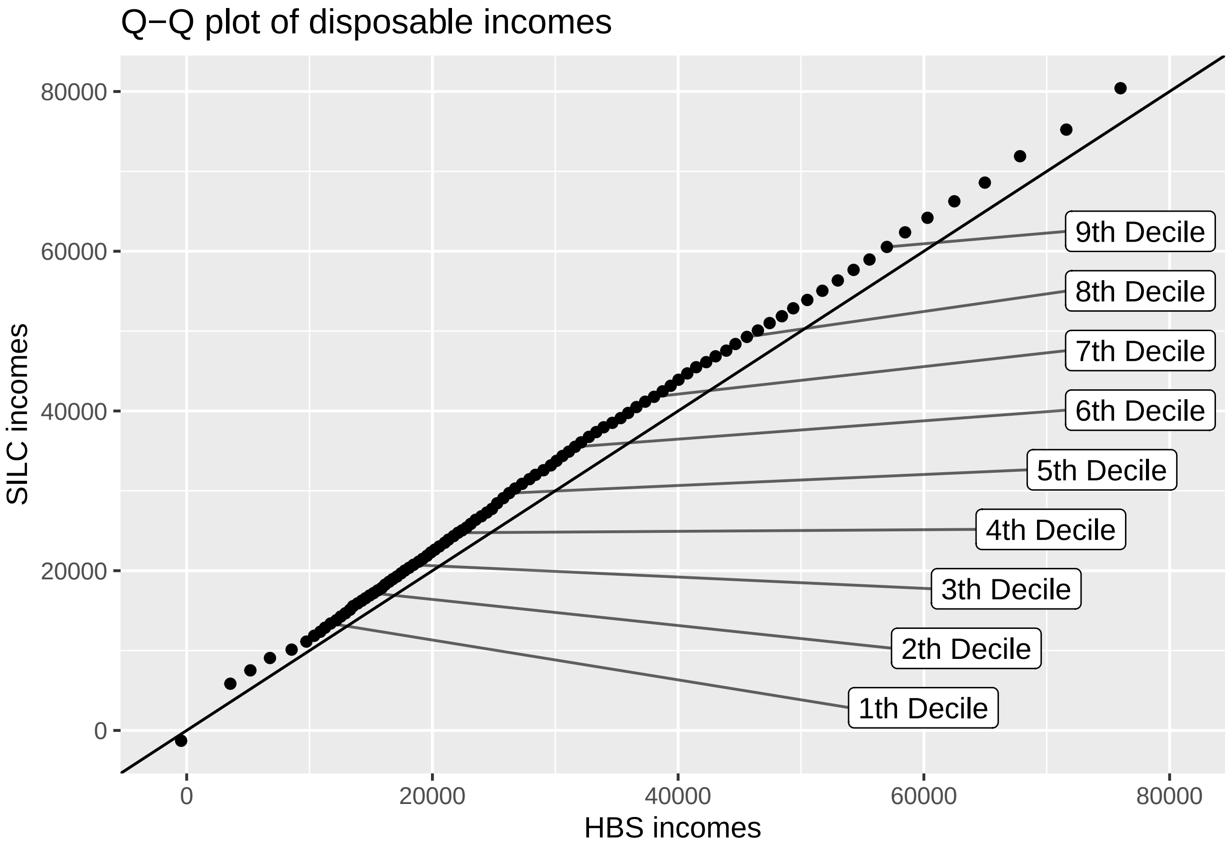


Figure 27: Q-Q plot of disposable income, the Netherlands

J.9 Summary statistics for Sweden

Table 153: Frequency table of household size

HH_size	HBS	SILC
1	0.3785620	0.3952778
2	0.3325951	0.3424640
3	0.1238368	0.1054125
4	0.1157856	0.1131065
5	0.0385967	0.0334136
6	0.0080670	0.0075442
7	0.0016919	0.0022066
8	0.0008648	0.0004526
9	0.0000000	0.0000476
10	0.0000000	0.0000747

Table 154: Frequency table of the number of household members between 0 and 4 years old

num_HH_mem_btw_0_4	HBS	SILC
0	0.9086449	0.8951577
1	0.0752898	0.0798288
2	0.0150217	0.0236844
3	0.0010436	0.0013291

Table 155: Frequency table of number of household members between 5 and 15 years old

num_HH_mem_btw_5_15	HBS	SILC
0	0.8247855	0.8455271
1	0.0950129	0.0841422
2	0.0683043	0.0586879
3	0.0104753	0.0093085
4	0.0014220	0.0017915
5	0.0000000	0.0004681
6	0.0000000	0.0000747

Table 156: Frequency table of number of household members between 16 and 24 years old

num_HH_mem_btw_16_24	HBS	SILC
0	0.7802345	0.7907525
1	0.1673733	0.1562766
2	0.0473510	0.0477831
3	0.0040394	0.0046984
4	0.0010019	0.0004208
5	0.0000000	0.0000687

Table 157: Frequency table of number of students in the household between 16 and 24 years old

num_HH_mem_btw_16_24_student	HBS	SILC
0	0.8713426	0.8859743
1	0.1089400	0.1016667
2	0.0191294	0.0117359
3	0.0005879	0.0006232

Table 158: Frequency table of number of household members between 25 and 64 years old

num_HH_mem_btw_25_64	HBS	SILC
0	0.2435968	0.3174407
1	0.3584419	0.3092310
2	0.3895941	0.3685232
3	0.0077468	0.0045003
4	0.0002886	0.0002227
5	0.0003317	0.0000392
6	0.0000000	0.0000429

Table 159: Frequency table of number of household members older than 64 years old

num_HH_mem_over_65	HBS	SILC
0	0.7730860	0.7125719
1	0.1343588	0.1931360
2	0.0922871	0.0942377
3	0.0002681	0.0000544

Table 160: Frequency table of number of non-EU household members

num_HH_mem_non_EU	HBS	SILC
0	0.9551599	0.9778485
1	0.0366590	0.0161140
2	0.0050640	0.0020448
3	0.0008152	0.0017035
4	0.0013169	0.0009456
5	0.0009849	0.0010351
6	0.0000000	0.0002260
7	0.0000000	0.0000826

Table 161: Frequency table of number of adult male household members

num_HH_mem_adult_male	HBS	SILC
0	0.2426943	0.2446848
1	0.6678224	0.6851284
2	0.0710084	0.0576721
3	0.0159441	0.0112408
4	0.0021655	0.0011930
5	0.0001582	0.0000809
10	0.0002072	0.0000000

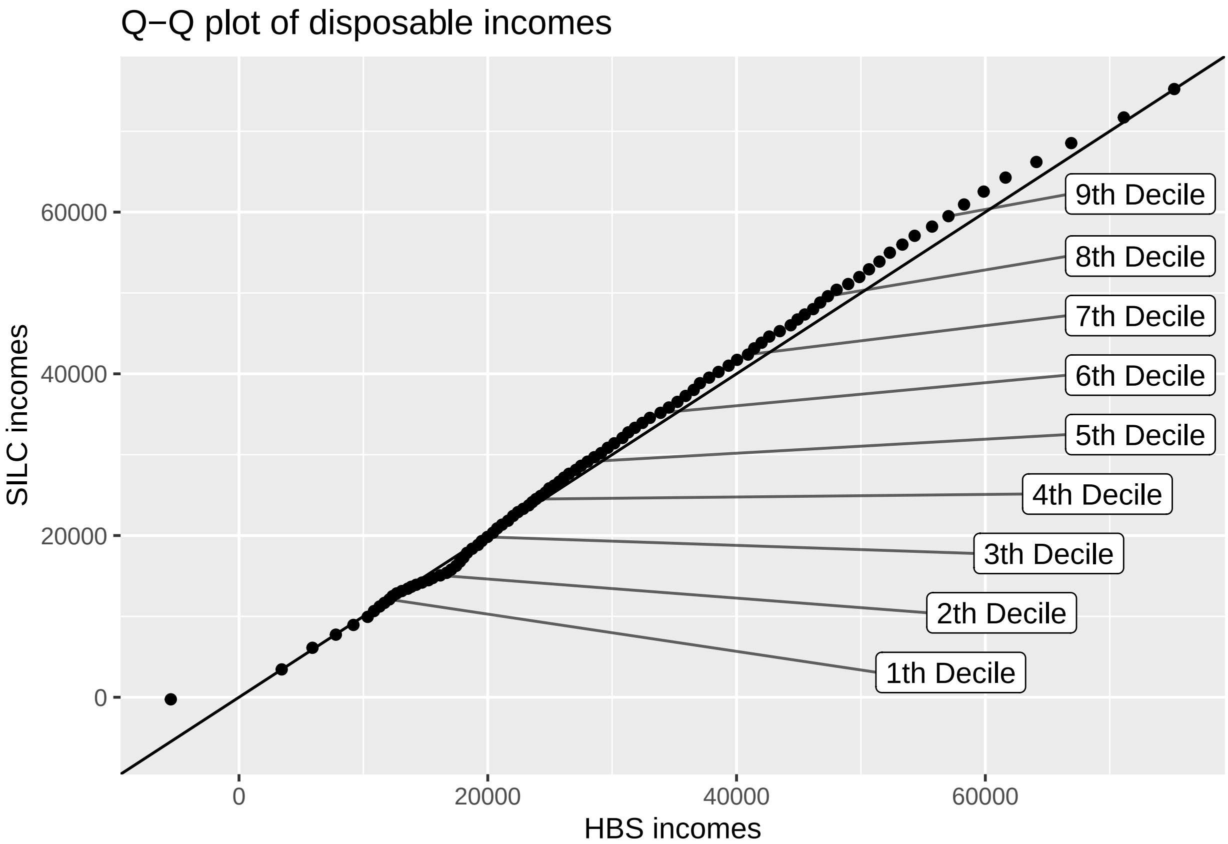


Figure 28: Q-Q plot of disposable income, Sweden