CBR
COUNCIL FOR BUDGET
RESPONSIBILITY

Expenditure imputation and microsimulation of VAT

Zuzana Siebertová, Jana Valachyová
Norbert Švarda, Matúš Senaj

Expenditure imputation and microsimulation of VAT[1]

Zuzana Siebertova[2], Jana Valachyova[3],

Norbert Svarda[4], Matus Senaj[5]

ABSTRACT

In this paper, we document the development process of the microsimulation model for the analysis of the indirect value-added tax liabilities of households in Slovakia. This simulation module can be directly integrated into the framework of SIMTASK, the Slovak microsimulation model of income taxes, health and social security contributions and transfers. In the first step, a combined micro-level dataset that integrates information on disposable income and expenditures of Slovak households has been created. Households' expenditures reported in HBS dataset have been imputed to SK-SILC dataset by estimating parametric Engel curves. Validation of the imputation procedure of households' consumption and simulation of VAT has been discussed.

Keywords: value added tax, tax and transfer system, income distribution, microsimulation

JEL classification: C81, D12, D31, H31

[2] Corresponding author. Address: Council for Budget Responsibility. I. Karvaša 1, 813 25 Bratislava, Slovakia.
E-Mail: siebertova@rrz.sk
[3] Council for Budget Responsibility. I. Karvaša 1, 813 25 Bratislava, Slovakia. E-Mail: valachyova@rrz.sk
[4] Council for Budget Responsibility. I. Karvaša 1, 813 25 Bratislava, Slovakia. E-Mail: svarda@rrz.sk
[5] Council for Budget Responsibility. I. Karvaša 1, 813 25 Bratislava, Slovakia. E-Mail: senaj@rrz.sk

1   Introduction

Microsimulation modelling techniques grew into standard tools for the evaluation of taxes and social transfer system changes. To provide a comprehensive assessment of reform changes at a micro level, a necessary precondition is to have a tool with a detailed representation of the country's tax and transfer system structure. To do so, we employ a microsimulation model SIMTASK developed at the Slovak Council for Budget Responsibility (Siebertova et al., 2016) that simulates the Slovak taxes from labour and social transfers in detail and enables to evaluate the disposable income of individuals as well as households.

In this study, we present a microsimulation module that simulates the value-added tax liabilities of households that can be straightforwardly used within the framework of tax-benefit calculator SIMTASK. The advantage of having this common environment is that reform changes in direct taxes from labour, as well as in social transfers and indirect tax policies can be evaluated simultaneously. In other words, this approach immediately allows to evaluate fiscal and distributional impacts of policy changes by comparing the reform state to the baseline. Moreover, SIMTASK as a tax and transfer calculator enriched with the VAT module can be conveniently integrated into our labour supply and general equilibrium micro-macro estimations and thus to evaluate behavioural and long-run reactions to reform changes (Senaj et al., 2016 and Horvath et al., 2018).

The paper documents several issues. In the first step, a merged micro-level dataset that integrates information on individual disposable income and households' expenditures has been created. SIMTASK runs on the underlying SK-SILC dataset that contains an extensive information on different sources of income, but it does not contain information on consumption. Micro data on households' consumption is available in the Slovak Household Budget Survey (HBS). Using the parametric regression technique of Engel curves estimations developed by Decoster et al. (2013, 2014), we estimated the expenditure allocation model. In this modelling setup, consumption expenditures are estimated separately for durable and non-durable goods. Based on these regressions estimates, the information on households' consumption has been imputed into the underlying SK-SILC data. In the next stage, VAT liabilities of households were simulated based on the estimated households' consumption. Validation of simulations, i.e. the amount of simulated households' consumption expenditures

and VAT revenues, is examined and the results are compared to the official statistics taken from the national accounts. Using a so called "coverage approach" (Barrett et al., 2015), we provide an insight to the sources of variation that result to the substantial gap between household expenditure survey data and expenditure measured in the national accounts.

The rest of the paper is organized as follows. In section 2, theoretical background underlining imputation and simulation techniques that were applied is briefly summarized. In particular, the methodology of the expenditure allocation model is discussed. Section 3 describes the underlying data and documents the results of the imputations. In Section 4, the validation of the simulation results, when compared to the official statistics, is discussed. In the following section, the structure of statutory VAT rates in Slovakia is introduced. The volume of VAT paid by households is calculated by applying the appropriate tax rates to the corresponding expenditure amounts. The section discusses also the distributional impact of the VAT burden.

2 Theoretical background

In the first step, a dataset that combines income and expenditure data needs to be created. Usual technique is to impute or „match" the information on expenditures into income dataset. In our approach[6] we use the estimations of Engel curves (these relate household spending on specific goods and income) to transfer the information on expenditures to income data. The common idea in this class of methods is to find a function, in fact the expenditure allocation model, that links the budget shares in expenditure survey to the common variables in both datasets and then to apply this function to the records in income survey.

We estimate Engel curves of the expenditure aggregates econometrically by means of standard regression tools. Our empirical approach is related to the studies by Decoster et al. (2013) who developed the indirect tax module for EUROMOD, EU wide tax-benefit micro-simulation model, or Abramovsky and Phillips (2015) who enlarged LATAX, the micro-simulation model for Latin American countries, to account for indirect tax simulations. Primary advantage of the

---

[6] Two major categories of matching procedures exist. First are labelled as implicit methods – these use techniques of a statistical matching and do not rely on the assumptions of a theoretical model. In this approach, expenditure information is directly incorporated to income dataset by using a proximity (distance) function that integrates by household cohorts' information from both surveys, conditioning on a set of overlapping variables that are available in both datasets.

applied methodology is that for every value of households' disposable income, expenditures can be simulated.

Once expenditures are imputed, the simulation of VAT liabilities is constructed in two steps. First, reported expenditures in HBS are matched to the corresponding statutory tax rates. A simulation program then computes the amount of VAT paid by every household by applying the statutory tax rates to the estimated expenditure amounts.

2.1 Expenditure allocation model

To estimate the expenditure allocation model, we apply a variant of the two-step procedure of parametric estimation of Engel curves as described by Decoster et al. (2013, 2014) for Belgium and Germany.

The estimation process can be outlined as follows. First, separate models for the total durable and non-durable expenditures are estimated as a polynomial of disposable income and several socio-demographic characteristics at the household level using the expenditure dataset. All durable expenditures are aggregated into one category to reduce the number of zero observations in the estimations. For the non-durable expenditures, 12 categories are formed based on COICOP classification of expenditures.

The estimation of total *durable expenditures* involves two steps. In the first step, a probit model is used to estimate the probability of non-zero households' durable expenditures:

$Pr(D = 1) = \Phi(\beta_{PD}X_{PD})$.

The dummy variable $D = 1$ if household purchased durable goods and $D = 0$ otherwise. The covariates comprised in the vector X_{PD} include functions of households' disposable income and different family and demographic characteristics. For the detailed list of variables used in the estimations, see Table A1 in the Appendix.

Conditional that total durable expenditures are non-zero, in the second step a demand equation is estimated in the semi-logarithmic form $\ln(e_D) = \beta_D X_D + \varepsilon_D$. The dependent variable e_D is the total monthly expenditures for durable goods at the household level. Note that covariates in the first and second equation need not to be the same. The disturbance term ε_D is assumed to be independently and identically distributed.

Total *non-durable expenditures* are estimated directly in one step since every household exhibit non-zero consumption of these goods. Therefore, the total demand equation for non-durable consumption is estimated in the form $\ln(e_N) = \beta_N X_N + \varepsilon_N$.

Using the estimates of durable ($\widehat{e_D}$) and non-durable ($\widehat{e_N}$) expenditures outlined above, savings are defined as a residual between disposable income and the sum of durable and non-durable consumption $\hat{s} = y - (\widehat{e_N} + \widehat{e_D})$.

At the second stage, the allocation of total non-durable expenditures to 12 expenditure categories is formed in terms of *budget share equations*. Certain categories (denoted as $j = 1, \dots, J$) of non-durable expenditures exhibit zero spending in many households. In case of Slovak households, four such classes of expenditures were identified: alcohol, clothing, education and restaurants ($J = 4$).[7] For these categories, a two-step estimation procedure is performed: in the first step a probit model to determine a positive probability of expenditures in each commodity group is estimated as $Pr(D_j = 1) = \Phi(\beta_j X_j)$.

For the households with non-zero expenditures e_j in category j, the share $w_j = {e_j}/{e_N}$ out of total non-durable expenditures is then formed. Conditional budget share demand equations are estimated separately for every commodity group in the form

$$w_j = \gamma_1 \ln e_N + \gamma_2 (\ln e_N)^2 + \gamma_3 (\ln e_N)^3 + \beta_j X_j + \varepsilon_j.$$

Note that this equation specification can be viewed as the *Engel curve* part in QUAIDS model, where only the real income effects are considered and the relative price effects are omitted.

The budget shares for the left-over categories (denoted as $k = 1, \dots, 12 - J$), i.e. those not covered in the commodity groups with many zeros mentioned above, are defined as $w_k = {e_k}/{e_N}$.

Corresponding budget share demand equations are subsequently estimated as

$$w_k = \delta_1 \ln e_R + \delta_2 (\ln e_R)^2 + \delta_3 (\ln e_R)^3 + \beta_k X_k + \varepsilon_k,$$

where e_R denotes the residual amount of non-durable expenditures left from total non-durable expenditures after spending on the commodities defined in the sub-group above has been deducted, i.e. $e_R = e_N - \sum_{j=1}^{J} e_j$.

---

[7] In the Slovak HBS survey from 2015, 20% of households reported no expenditures on alcohol, 18% on clothing, 76% on education and 32% on restaurants.

3 Expenditure imputation

The expenditure allocation model enables to estimate how the households' disposable income is divided into consumption of different commodities and savings. The coefficient estimates of the durable and non-durable consumption as well as the budget shares estimates of non-durables are used to impute households' expenditures into the underlying dataset of the microsimulation model SIMTASK.

3.1 Dataset description

The microsimulation model of the Slovak tax and transfer system SIMTASK uses as an underlying dataset SK-SILC. The SK-SILC survey is conducted by the Slovak Statistical Office on an annual basis and the dataset contains detailed information on income, socio-demographic characteristics of individuals and household composition of the Slovak population. It does not contain information on expenditures. The latest wave available at the time of writing corresponds to income reference period 2015 and contains information on 16,480 individuals and 5,738 households.

On the other hand, information on households' consumption, income and expenditures is delivered in the Slovak Household Budget Survey (HBS). The survey is conducted by the Slovak Statistical Office and the latest wave collects data from 2015. It uses the same sampling frame as it is used for the SK-SILC, but the participating households are drawn independently in the two surveys. The latest cross-section HBS dataset comprises information on 4,785 households. It contains also a wide set of background variables providing information related to socio-demography, income and labour market status of individuals and households. Many characteristics overlap with variables already available in SK-SILC, being exploited also by simulation model SIMTASK.

In both datasets, original sample weights are supplied by the Statistical Office. Weights are calibrated (sample weights are adjusted to match the known population totals in selected categories) and integrated (household weights and personal weights equal) and take into account a number of population characteristics in Slovakia. However, when using SIMTASK, we re-weight the underlying dataset SK-SILC in such a way that new weights replicate the earned income distribution and selected age cohorts directly, as well as household size, number of

employees, self-employed and unemployed. The approach is described in detail in Siebertova et al. (2016) and we showed that the new weights improved the fit between simulated output, underlying data and the official statistics. We applied the equivalent approach to HBS dataset and re-weighted it using the same calibration categories.

A detailed evidence on households' money expenditures is covered in HBS, which is broken down into consumption expenditures and other net expenditures. Information on 310 individual consumption expenditures is aggregated at the 1-digit code level to 12 divisions based on COICOP-5 classification methodology. In our implementation 13 different groups of expenditures are considered, we form 12 commodity groups of non-durables and one group of durable expenditures. Non-durable commodities are defined mostly in line with COICOP classification and include the following classes: (1) food and non-alcoholic beverages, (2) alcoholic beverages and tobacco, (3) clothing and footwear, (4) housing[8] and energies, (5) household services, (6) health, (7) transport, (8) communication, (9) recreation and culture, (10) education, (11) restaurants, and finally (12) miscellaneous goods and services. Durable expenditures cover, among others, commodities like furniture, household appliances, vehicles, or equipment for recreation. In addition, we shifted some expenditures originally included in the category of other net expenditures[9] to miscellaneous goods and services or durables.

For the accuracy of the imputation procedure it is essential that control variables used in demand equations do not exhibit substantial differences in terms of their statistical distribution when comparing the original HBS dataset and the target dataset used by SIMTASK. Table 1 shows that controls related to households' and socio-demographic characteristics are in general comparable (in terms of mean and median). Individual level controls like age, education level or labour market status are related to household head. However, the main variable of interest in the regression equations is the disposable household income. In HBS it is defined such that it includes information on net labour income and all received social transfers as reported by household members. Disposable income used in SIMTASK is based on income information reported in SK-SILC, but the value is simulated. When component variables are available in the

---

[8] Imputed rents have not been so far introduced in the Slovak HBS dataset 2015.
[9] These comprise in total around 14% of net money expenditures of households. It includes capital investments, expenditures on non-commercial insurance, payments of loans and mortgages (moved to miscellaneous goods and services) or reconstruction of a house or dwelling (moved to durables).

original SK-SILC dataset and are simulated by SIMTASK as well (like income taxes, social contributions and selected transfers), the simulated components of disposable income are considered. This constitutes the difference between the two variables of disposable income: in the source dataset HBS it is a reported value, while in a target dataset disposable income is simulated by using SIMTASK.

Table 1 Descriptive statistics of HBS and SIMTASK

	HBS original w.		HBS calibrated w.		SIMTASK	
	mean	median	mean	median	mean	median
Disposable income (eur/month)	1 238.5	1 135.9	1 196.8	1 070.9	1 174.4	1 059.5
Household characteristics						
members	2.93	3	2.93	3.00	2.83	3
economic active members	1.49	2	1.48	2.00	1.55	2
dependent children	0.72	0	0.70	0.00	0.56	0
Household head characteristics						
age	51.0	50	52.1	52	50.3	48
male	0.66	1	0.68	1	0.61	1
education: primary	0.08	0	0.09	0	0.10	0
education: secondary	0.72	1	0.69	1	0.68	1
education: university	0.20	0	0.22	0	0.22	0
employed	0.69	1	0.68	1	0.70	1
unemployed	0.02	0	0.03	0	0.02	0
inactive	0.29	0	0.29	0	0.28	0
Households	4 785		4 785		5 738	

Note: Data in HBS from 2015 are weighted using original or calibrated weights. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015, weights are calibrated and take into account income distribution.
*Disposable income is simulated by SIMTASK.

Both mean and median values of monthly disposable income are well comparable, in HBS 1,196.8 and 1,070.9 € versus 1,174.4 and 1,059.5 € simulated by SIMTASK. Histogram plotted in Figure 1 completes the picture, it shows that disposable income simulated by SIMTASK is only slightly skewed to the left which corresponds to the reported lower mean and median values. The main discrepancy between definition of disposable income in HBS and the simulated disposable income by SIMTASK using SK-SILC is the following. In HBS, data are reported on an annual basis. When the respondent reports a monthly income from labour, this amount is multiplied

by 12 and not by the actual number of months when he was working. On the other hand, in the simulation by SIMTASK, the actual number of months when working is considered. This factor contributes to the overestimation of disposable income in HBS compared to simulation in SIMTASK.

Figure 1 Income distribution in HBS and SIMTASK in 2015

a/ HBS original weights	b/ HBS calibrated weights
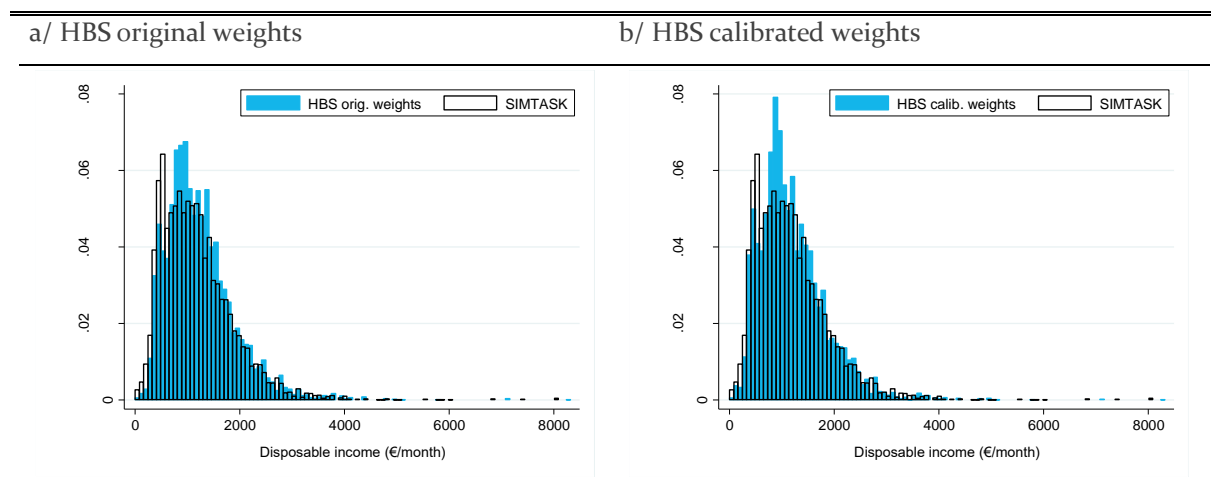
Source: Authors' calculations

3.2 Imputation results

Detailed estimation results of the expenditures allocation model can be found in the Appendix in Table A2. Using these regressions estimates, households' expenditures for durables ($\widehat{e_D}$) and non-durables ($\widehat{e_N}$) are imputed in the target dataset[10].

Based on the estimated non-durable expenditures $\widehat{e_N}$, also the budget shares $\widehat{w_k}$ for 12 non-durable commodity classes for every household can be imputed. However, some predictions (around 1.7 %) of budget shares were estimated as negative. These were set to zero and the remaining imputed budget shares were rescaled such that they sum up to one.

---

[10] Before running regressions on HBS data, some data adjustments were performed. First, to ensure that after taking logs of income the amounts stay positive, extremely small values of disposable income below 5 €/month were dropped (2 observations). Second, reported durable expenditures below 1 €/month were set to zero. Although this adjustment increased the share of zero observations from 13 to 20 %, it considerably improved the fit of the estimated regression. Finally, households with expenditure-to-income ratio over 5 were dropped (1 observation).

The precision of the imputation procedure is documented in Table 2 where the mean and median expenditures of durables as well as non-durables goods and expenditure shares of different categories in the source and target datasets are compared. Total durable and non-durable expenditures were imputed into the underlying SIMTASK dataset based on the estimated regressions. Since the reported disposable income is lower in SIMTASK, this translates into lower values of predicted expenditures.

Table 2 Comparison of expenditures reported by HBS and SIMTASK

	HBS		SIMTASK	
	mean	**median**	**mean**	**median**
Durable expenditures	123	33	110	100
Non-durable expenditures	858	792	847	813
Expenditure shares by categories (%)				
(1) Food and non-alcoholic beverages	23.1	22.0	22.5	21.0
(2) Alcoholic beverages and tobacco	3.0	1.7	2.9	3.4
(3) Garments and shoes	4.2	2.4	4.2	4.6
(4) Electricity and other fuels	23.6	21.5	23.3	21.7
(5) Household services	1.8	1.4	1.8	1.7
(6) Health	3.7	2.6	3.6	3.2
(7) Transport	6.2	4.9	6.2	6.3
(8) Communication	5.5	4.8	5.4	5.4
(9) Recreation and culture	4.0	2.5	4.1	4.0
(10) Education	0.5	0.0	0.4	0.0
(11) Restaurants and hotels	4.3	2.8	4.1	5.1
(12) Other goods and services	11.6	7.9	11.9	12.1
Durables	8.5	3.9	9.7	10.5

Note: Data in HBS from 2015. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015. Data are weighted using population weights. Expenditure variables in euros per month.

Difference between the reported mean expenditure shares in HBS and those imputed in SK-SILC is the largest in categories food and non-alcoholic beverages (0.6 p.p.), followed by other goods and services (0.3 p.p.). In the remaining non-durable categories, the difference is smaller than 0.2 percentage points. Overall, Slovak households allocate almost half of their expenditures in the categories food and non-alcoholic beverages (23.1 %) and electricity and other fuels (23.6 %).

Reported consumption in the remaining categories (except other goods and services) is below 10 % of the total share.

Expenditure patterns across households' disposable income distribution are visualized in Figure 2, where the percentile mean spending in the reported and imputed datasets is plotted in euros per month. Inspection of the plots in panels a/ and b/ shows that durable and non-durable expenditures are distributed differently, but in both cases a lot of variability in the data (compared to the original HBS) vanished. Expenses on durable goods are occasional. This transfers to the fact that when the disposable households' income is low, the spending profile on durables is flat and in the underlying SIMTASK's dataset close to zero. Subsequently, durable expenses increase with income at a slow pace, almost linearly. The graph for mean non-durable expenditures shows that spending increases as a concave function in income and much faster compared to durables. The overall fit of the imputation appears to be very good for households with disposable income below 2,000 €/month.[11] A graphical exposition of the imputation results is favoured, and the scatterplots of non-durable mean expenditure shares are presented in Figure A1 in the Appendix.

Figure 2 Expenditures distribution of households in HBS and SIMTASK in 2015

Note: Data in HBS from 2015. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015. Data are weighted using population weights. Expenditures in €/month.
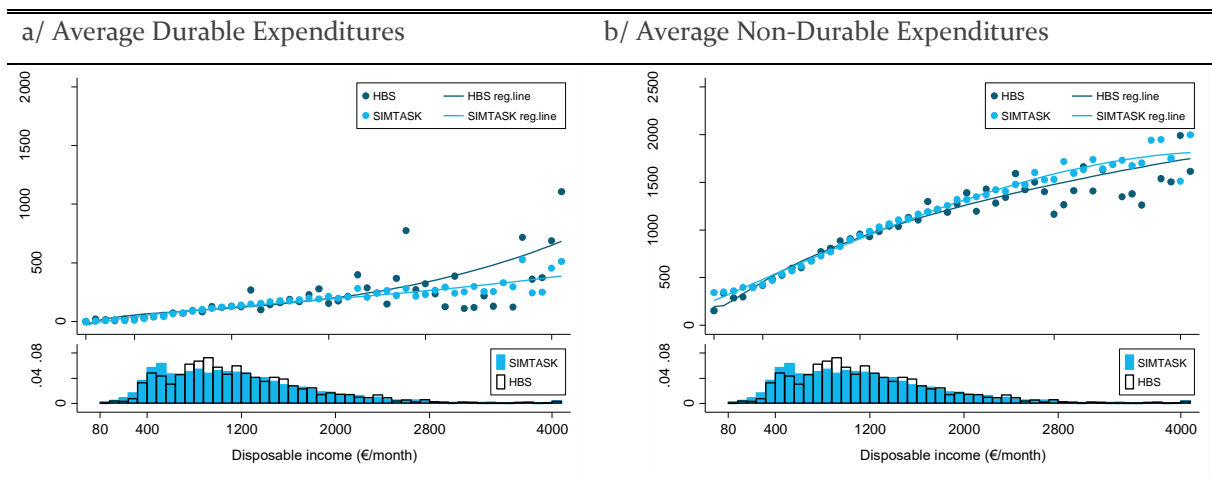Source: Authors' calculations

---

[11] Note that these households constitute around 84 % of the total number of households in the underlying SIMTASK's dataset. Share of households in HBS with the reported income below 2,000 €/month is 90 %.

In Figure 3, the fit of the imputation procedure is inspected in more detail. Expenditure-to-income ratios by disposable income deciles are presented for total consumption as well as for durables and non-durables separately. Detailed categories of non-durables are graphed in Figure A2 in the Appendix. In the lower income groups SIMTASK overestimates total and non-durable expenditure shares (see panels b/ and c/ of Figure 3), whereas in the second half of income distribution the differences are subtle.

Figure 3 Expenditure-to-income ratios by income deciles
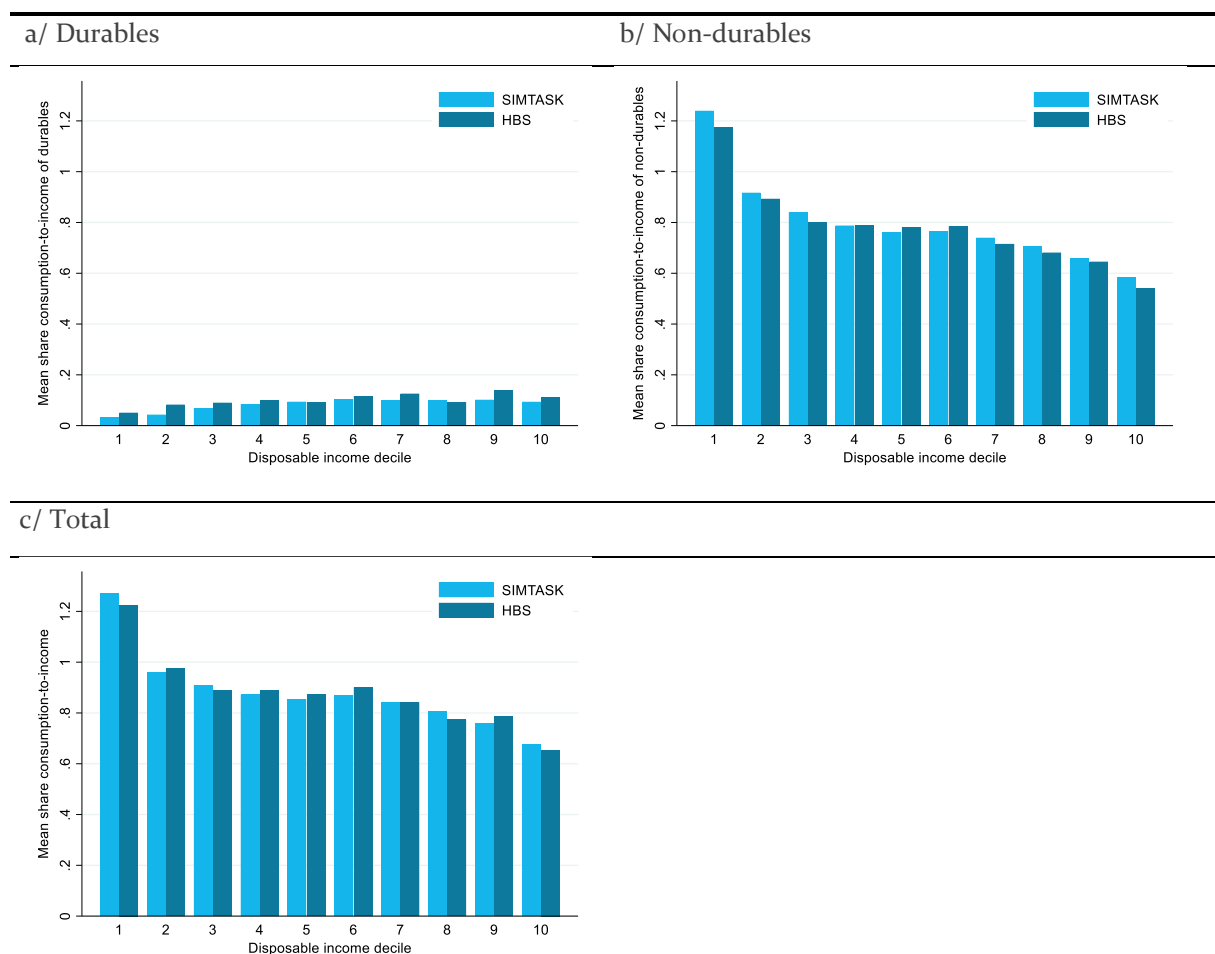
a/ Durables

b/ Non-durables

c/ Total

Note: Data in HBS from 2015. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015. Data are weighted using population weights.
Source: Authors' calculations

This might be attributed to the fact that the disposable income distribution of the underlying dataset of SIMTASK is skewed to the left compared to HBS. The story is different for durables

(see panel a/ of Figure 3), where the imputed shares in SIMTASK are clearly underestimated. Since expenditures on durables account for approximately 10% of the total expenditure-to-income share, the overall picture is driven by the characteristics of non-durables. To sum up, we can conclude that the imputation procedure maintained the main characteristics of households' consumption behaviour for different expenditure categories also when the income distribution is considered. In addition, the graph of total expenditure-to-income ratios in Figure 3 demonstrates that low income households are net borrowers, while high income households tend to be net savers – implying that the savings rate increases with income. This savings pattern has been already documented as a usual trend observed in several OECD countries (OECD, 2014).

4   Validation of the results

The amount of simulated consumption expenditures, when validated against its counterparts in macro statistics, represent only around 50%. This gap is not unique for Slovakia, similar inconsistencies between statistics from micro and macro sources are documented in the literature for other countries as well (see Decoster et al. (2014a) for Belgium or Decoster et al. (2013) for Germany).

In our modelling framework, households' consumption expenditures are simulated based on the information on households' disposable income provided in two households' surveys (SK-SILC and HBS). A consistent picture arises when comparing both, households' consumption expenditures and households' disposable income, to national aggregates. The HBS disposes with the data on 50% of economy wide disposable income reported by national accounts. Disposable income simulated by SIMTASK covers 60% when compared to national accounts. This implies that only aforementioned amounts can be used to model consumption expenditures, which results into 50% coverage rate when compared to national accounts statistics.

The reasons why such gaps exist between the aggregates from households' survey data and national accounts has attracted a considerable attention in the literature and recently was examined also by Fesseau et al. (2013). Authors look at three relevant national accounts aggregates (household adjusted disposable income, household actual final consumption and households' net worth) and examine the extent to which statistical information derived from

micro sources can be aligned to them. As the main sources are identified the differences in population scope, in statistical treatment and in classification.

Household surveys do not provide information on those groups of population which are outside their scope such as households living in non-private dwellings (i.e. prisons, boarding schools, retirement homes, hospitals and nursing homes, religious institutions, hotels, etc.). Also, micro sources do not provide information for several national accounts components such as employers' imputed social contributions, financial intermediation services indirectly measured (FISIM), reinvested earnings on foreign direct investments or social transfers in kind. Also, information on imputed rents for owners-occupied housing and own-account production of goods might be missing. These components are not included in household surveys because they might not be relevant if the focus is on the economic behaviour of households, or because the information is difficult to collect or impute.

Even in those cases when the information is available in both data sources, the differences in how the information is classified and processed, might generate the gaps between micro and macro statistics. Other reasons contributing to the widening of the gap are the differences in valuation of transactions and balance sheet positions and limitation of sources and associated statistical adjustments such as non-response, sampling error, underreporting or time recording. In order to inspect the gap between micro and macro sources in Slovakia in more details, the coverage rates for individual components of gross disposable income are reported in Table 3.

The coverage rates represent the extent to which the statistics reported by HBS, the original dataset SK-SILC and simulated by model SIMTASK[12] correspond to national accounts. A conversion table linking the components of disposable income in households' surveys to their counterparts in national accounts is inspired by Gregorini et al. (2016), and was modified to include the original SK-SILC, the underlying dataset used by SIMTASK and HBS aggregates too (see Table A3 in the Appendix).

Similar coverage rate for gross disposable income is reported for all micro sources, 56% for SK-SILC, 58% the simulation based on SIMTASK and 59% for HBS. In line with the results

---

[12] When working with SIMTASK, the computed statistics might differ from the values reported in the original dataset SK-SILC. Items that are simulated by the model are utilized also for the construction of combined variables. For example, in the computation of households' disposable income we are using as components simulated personal income tax, paid social security contributions and selected transfers - rather than the reported original values.

documented in the literature, the household survey data cover relatively well employee income (coverage rate well above 80% in SK-SILC and HBS and more than 86% when simulated with SIMTASK) and social benefits other than social transfers in kind (coverage rate 64% in SK-SILC, 72% in HBS and 63% with SIMTASK).

Table 3 Coverage rates of the components of disposable income, Slovakia 2015

	National Accounts (mil. €)	HBS (mil. €)	Cover. rate (%)	SK-SILC (mil. €)	Cover. rate (%)	SIM-TASK (mil. €)	Cover. rate (%)
Employee income (wages and salaries)	24 476	19 906	81	20 394	83	21 090	86
Employers' contributions	7 337	n.a.	n.a.	n.a.	n.a.	6 208	85
Income from self-employment & Property income/Resources	17 940	3 199	18	3 507	20	3 064	17
Income from household production of services for own consumption	802	7	1	124	15	129	16
Social benefits other than social transfers in kind/Resources	11 358	8 153	72	7 277	64	7 114	63
Other current transfers and other resources received	2 615	430	16	177	7	125	5
Current income total	**64 528**	**31 695**	**49**	**31 479**	**49**	**37 731**	**58**
Social contributions and taxes on income paid/Use	14 807	n.a.	n.a.	n.a.	n.a.	11 639	79
Taxes on income	2 626	1 706	65			2 070	79
Social contributions paid by employees	4 827	2 480	51	5 196	70	3 361	70
Social contributions paid by employers	7 337	n.a.	n.a.	n.a.	n.a.	6 208	85
Taxes on wealth paid/use	152	57	38	49	32	50	33
Property income paid	312	n.a.	n.a.	274	88	327	105
Other current transfers paid	2 964	155	5	62	2	45	2
Current expenditures total	**18 218**	**4 398**	**24**	**5 581**	**31**	**12 061**	**66**
Gross Disposable Income	**46 310**	**27 297**	**59**	**25 898**	**56**	**27 042**	**58**

Note: Gross Disposable Income in NA, HBS and SK-SILC is the difference between current income total and expenditures. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015, weights are calibrated. When using SIMTASK, several items are simulated. Cover. rate (coverage) represents perecentage of the statistics to national accounts.
Source: National Accounts (SO SR), SK-SILC, HBS and SIMTASK

A significant gap between micro sources and national accounts is reported for income from self-employment and property income that are aggregated into one category. The coverage rate is

less than 20%. The item most closely corresponding to income from self-employment in national accounts is the mixed income. In Slovakia, the Statistical Office aggregates into mixed income in the sector of households the imputed rents and the operating surplus from self-employment. As it is documented in Table 3, almost 18 mld. of euros reported for this category in national accounts constitutes around 28% of the total revenues. However, in micro sources, the counterpart category aggregates only 3 mld. of euros and this discrepancy forms the major contribution to the overall gap. The reasons for a gap mentioned in Fesseau et al. (2013) include different treatment of property income received/paid by the enterprises or explicit adjustment for hidden activities in national accounts item. The reasons for poor alignment of interest and dividends received incorporate the fact that wealthy households are underrepresented in household surveys, as well as the effect of under-declaration of income by households. Exclusion of non-private dwellings from household surveys contributes to the gap in this item as well. Very low coverage rates are observed for other current transfers, both received and paid in both micro sources. Almost missing in HBS is the information on income from household production of services for own consumption, while SK-SILC and SIMTASK cover around 15% of this component.

On the expenditure side, similar picture (like on the side of revenues) arises when micro and macro sources are compared. Labour taxes and social insurance contributions paid are represented well (coverage rate 65% in HBS, 70% in SK-SILC and more than 70% with SIMTASK). A significant gap is documented for taxes on wealth with the coverage rate around 30%.

Table 4 reports the coverage rates for final consumption expenditure of households, relating the consumption statistics both from HBS and the imputed values by SIMTASK to national accounts statistics. The highest coverage rate reaches 80% for transport and communication. Relatively well aligned are expenditures on clothing and footwear, health and food and non-alcoholic beverages. On the other side, significant gaps for alcohol and tobacco, recreation and culture, education and miscellaneous goods and services exist. The reasons for a gap might be due to underreporting by households, often documented in the case of alcohol and tobacco in a number of countries (see the discussion in Fesseau et al., 2013). Also, the impact of territorial classification might contribute to the gap.

Table 4 Final consumption expenditure of households by consumption purpose, Slovakia 2015

	National Accounts (mil. €)	HBS (mil. €)	Cover. rate (%)	SIM-TASK (mil. €)	Cover. rate (%)
(1) Food and non-alcoholic beverages	7 548	4 383	58	4 353	58
(2) Alcoholic beverages and tobacco	2 171	630	29	659	30
(3) Garments and shoes	1 678	1 111	66	1 041	62
(4) Electricity and other fuels	10 557	4 353	41	4 284	41
(5) Household services	2 503	1 118	45	342	14
(6) Health	1 073	656	61	643	60
(7) Transport	3 186	2 553	80	1 448	45
(8) Communication	1 451	1 146	79	1 159	80
(9) Recreation and culture	4 209	1 422	34	897	21
(10) Education	696	162	23	104	15
(11) Restaurants and hotels	2 444	1 085	44	999	41
(12) Other goods and services	4 916	1 581	32	2 903	59
Durables				2 447	
Total	**42 433**	**20 200**	**48**	**21 280**	**50**

Note: Final consumption expenditures in national accounts use a domestic concept (excludes consumption of residents in the rest of the world and includes consumption of non-residents on the economic territory), while in the micro data these adjustments are by construction not available.
Source: NA 2015 (Eurostat), HBS, SIMTASK

5 Simulation of VAT liabilities

5.1 Representation of VAT rates across expenditure categories

Two VAT rates were applied in Slovakia in 2015. The standard rate was set to 20 %, the reduced rate to 10 % and several goods were exempted and taxed at 0 rate. Exempted categories include usual categories like rents, health care, post services, books, education and financial services. Table 5 summarizes how are the statutory VAT rates represented in the respective expenditure categories listed in HBS. Overall, 84 % of expenditures in 2015 were taxed at a standard rate, 3 % at a reduced rate and the remaining 13 % belong to the category of exempted. Most items in two

major categories were taxed at the standard rate – all goods in the category food and more than 80 % in the category electricity and other fuels.

Table 5 VAT shares by COICOP expenditure categories in HBS in 2015

Expenditure categories	VAT 0%	VAT 10%	VAT 20%
Food and non-alcoholic beverages	-	-	100
Alcoholic beverages and tobacco	-	-	100
Garments and shoes	-	-	100
Electricity and other fuels	19.23	-	80.77
Household services	-	-	100
Health	46.67	40.00	13.33
Transport	-	-	100
Communication	28.57	-	71.43
Recreation and culture	8.57	11.43	80
Education	100	-	-
Restaurants and hotels	-	-	100
Miscellaneous goods and services	51.72	-	48.28
Durables	-	-	100
Total	**12.70**	**3.26**	**84.04**

Source: Authors' computations

After expenditures are imputed in the target dataset, total volume of VAT paid by households can be estimated. In the first step, expenditure items as recorded in HBS are matched with the corresponding tax rates. Expenditures are then grouped to COICOP categories and the VAT shares by categories are determined (see Table 5). Given these shares of statutory VAT rates reported for every expenditure category and using the simulated consumption from the expenditure allocation model, the aggregate revenues from VAT paid by households can be evaluated. The amount of VAT duty paid by every household is estimated for every expenditure category separately and then summed up. Simulated expenditures and the corresponding VAT liabilities broke down by different commodity categories are summarized in Table 6. The last column in Table 6 summarizes the effective VAT rates in the corresponding categories computed as a share of revenues from VAT to total expenditures.

Table 6 Simulated expenditures and VAT liabilities paid by households in 2015

Expenditure categories	Expenditures (€/mil)	Revenues from VAT (€/mil)			Effective VAT (%)
		10%	20%	Total	
(1) Food and non-alcoholic beverages	4 353	0	726	726	16.7
(2) Alcoholic beverages and tobacco	659	0	110	110	16.7
(3) Garments and shoes	1 041	0	174	174	16.7
(4) Electricity and other fuels	4 284	0	577	577	13.5
(5) Household services	342	0	57	57	16.7
(6) Health	643	23	14	38	5.9
(7) Transport	1 448	0	241	241	16.7
(8) Communication	1 159	0	138	138	11.9
(9) Recreation and culture	897	10	117	127	14.2
(10) Education	104	0	0	0	0
(11) Restaurants and hotels	999	0	167	167	16.7
(12) Other goods and services	2 903	0	221	221	7.6
Durables	2 447	0	401	401	16.4
Total	**21 280**	**34**	**2942**	**2976**	**14.0**

Note: Effective VAT rate is given as the ratio of revenues from VAT to expenditures (including VAT).
Source: Authors' computations

At this point, we can discuss the validation of VAT revenues. Total VAT revenues in Slovakia in 2015, as reported by the national accounts statistics, aggregate to 5,420 million €. As opposed to items discussed above (disposable income and consumption), no official statistics in national accounts on the aggregate amount of VAT payments by the sector of households is reported. However, SIMTASK simulates VAT revenues for the population of private households covered in the survey. In 2015 simulation of VAT revenues in SIMTASK sum up to 2,976 million €, making up 55% of the total amount reported by national accounts. This ratio is in line with coverage rates already observed for disposable income and consumption expenditures.

5.2 Distributional effects of VAT

The base of the incidence analysis – VAT liabilities decomposed by disposable income deciles are graphed in the left panel of Figure 4. The VAT paid by households grows with disposable
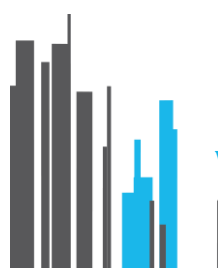
income faster than linearly. While households in the 1st decile pay on average 940 € for VAT yearly, for households in the top decile 10 it is more than 2,518 €.

Figure 4 VAT liability in 2015
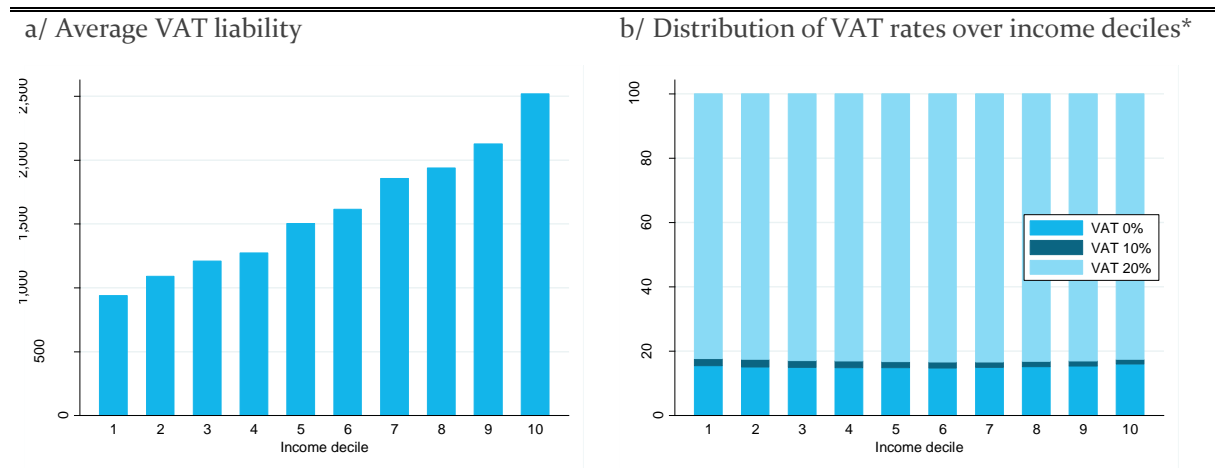
a/ Average VAT liability	b/ Distribution of VAT rates over income deciles*

Note Data in HBS from 2015. SIMTASK uses as an underlying dataset SK-SILC with income reference period 2015. Data are weighted using population weights. Income deciles are based on equivalence weighted disposable household income. Data are weighted using population weights.
*Bars show mean values of expenditure shares by the corresponding VAT rates.
Source: Authors' calculations

The distribution of VAT rates over disposable income deciles does not change substantially (right panel of Figure 4). It illustrates that mean expenditure share values of reduced rate commodities only slightly decrease with increasing income – from 2.2 to 1.5 % for the reduced rate. Correspondingly, expenditure share of commodities taxed at standard rate increase with income also only marginally from 82.3 to 82.5 %.

Looking at the VAT payments as a percentage of disposable income (left panel in Figure 5), a regressive pattern of VAT in Slovakia is confirmed. With increasing income, the average VAT burden decreases from 16 % in the lowest decile to 8 % in the top decile. Using this analytical approach, it can be concluded that VAT is a highly regressive tax. The same finding has been observed in numerous European country studies. Nevertheless, the picture is different when focus is on the expenditure-base results (right panel in Figure 5). In this case, the VAT burden exceeds 14 % and the profile is relatively proportional over consumption deciles. Again, this

result is in line with outcomes presented in the literature, showing that VAT burden is almost flat or slightly progressive over consumption distribution (Decoster et al., 2010).

Figure 5 VAT incidence in 2015

a/ Average VAT as a percentage of disposable income and consumption across income deciles
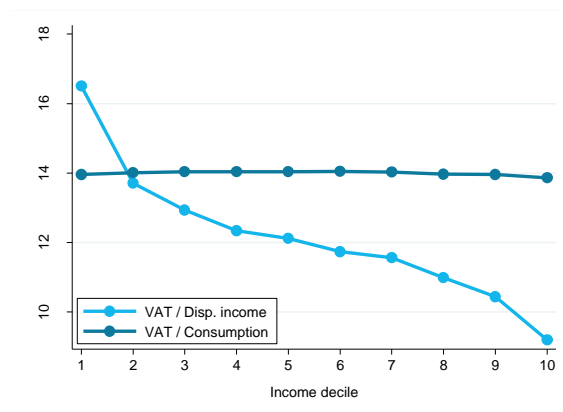
b/ Average VAT as a percentage of disposable income and consumption across consumption deciles
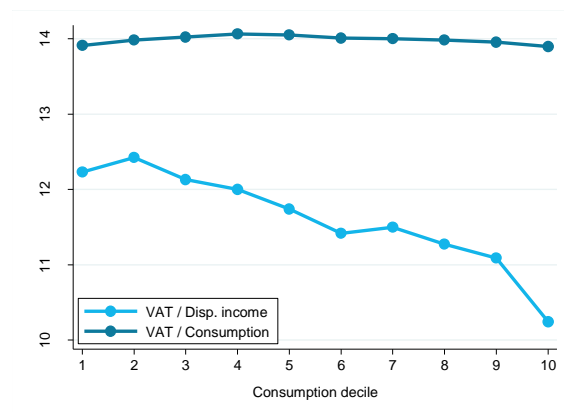
Note: Income deciles are based on equivalence weighted disposable household income. Consumption deciles are computed from equivalence weighted total household consumption. Data are weighted using population weights. Source: Authors' calculations

6   Conclusion

We constructed a VAT simulation module that can be directly integrated into SIMTASK, a Slovak microsimulation model of direct taxes and social transfers. To do so, we matched two survey datasets such that we transferred expenditure information on the household level from the household budget survey to the underlying dataset of SIMTASK that uses an income survey SK-SILC. We found that the goodness-of-fit of the imputation procedure is reasonable enough to work with this combined dataset in the framework of the microsimulation model. Finally, the validation of the simulation results shows the common pattern: simulated consumption expenditures and resultant VAT revenues aggregate only around 50% compared to national accounts, a typical gap observed when micro and macro sources are compared.

7 References

Abramovsky, L. and Phillips, D. (2015): „LATAX: A multi-country flexible tax micro-simulation model", mimeo, Institute for Fiscal Studies

Barrett, G., Levell, P., and Milligan, K. (2015): „A Comparison of Micro and Macro Expenditure Measures across Countries Using Differing Survey Methods", in Carroll, Ch., Crossley, T. and Sabelhaus, J. (eds.): „Improving the Measurement of Consumer Expenditures", Studies in Income and Wealth, vol. 74, University of Chicago Press

Decoster, A., Loughrey, J., O'Donoghue, C. and Verwerft, D. (2010): "How regressive are indirect taxes? A microsimulation analysis for five European countries", Journal of Policy Analysis Management, 29: 326–350

Decoster, A., Ochmann, R. and Spiritus, K. (2013): "Integrating Indirect Taxation into EUROMOD - Documentation and Results for Germany", EUROMOD Working Paper EM 20/13. Essex: EUROMOD at the Institute for Social and Economic Research

Decoster, A., Ochmann, R. and Spiritus, K. (2014a): "Integrating VAT into EUROMOD. Documentation and results for Belgium", EUROMOD Working Paper EM 12/14. Essex: EUROMOD at the Institute for Social and Economic Research

Decoster, A. and Spiritus, K. (2014b): "SInTax: Microsimulation of VAT and Excises in the Context of EUROMOD. User Manual". FLEMOSI Technical Note 16. 1, 13

Fesseau, M.F., Wolff and Mattonetti, M. (2013): "A Cross-country Comparison of Household Income, Consumption and Wealth between Micro Sources and National Accounts Aggregates", OECD Statistics Working Papers, No. 2013/03, OECD Publishing, Paris

Horvath, M., Senaj, M., Siebertova, Z., Svarda, N. and Valachyova, J. (2018): "Evaluating the Aggregate Effects of Tax and Benefit Reforms", CBR working paper, forthcoming

Gregorini, F., Grundiza, S. and Lamarche, P. (2016): "Distributional Income Indicators in a Micro-Macro Data Integration Perspective", Paper prepared for the 34th IARIW General Conference

OECD/Korea Institute of Public Finance (2014): „The Distributional Effects of Consumption Taxes in OECD Countries", OECD Tax Policy Studies No. 22, OECD Publishing

Senaj, M., Siebertova, Z., Svarda, N., and Valachyova, J. (2016): „Labour force participation elasticities and the move away from a flat tax: the case of Slovakia", IZA Journal of European Labor Studies, Volume 5

Siebertova, Z., Svarda, N. and Valachyova, J. (2016): "Improving the Validity of Microsimulation Results: Lessons from Slovakia", International Journal of Microsimulation, 9(2), 77-105

Appendix

Table A1 List of variables used in the estimations

log(income)	Logarithm of households' disposable income.
r_log(income)^2, r_log(income)^3	Prefix r_ denotes residuals from regressions of higher order variables on all lower order ones. Specification used to avoid multicollinearity.
Male	Binary variable that equals 1 if the household head is man, 0 if woman.
Age	Variable indicating the household head's age.
Education group dummies	3 binary variables are created based on ISCED classification (**EDU: Primary** [reference cat.], **EDU: Secondary**, **EDU: Tertiary**). If the household head belongs to a group according to his highest degree awarded, the corresponding binary variable equals 1, otherwise 0.
Economic inactive	Binary indicator that equals to 1 if the household head is economic inactive, 0 otherwise.
Dependent children	Variable indicating number of dependent children in household.
Household members	Variable indicating number of household members.
Economic active members in HH	Variable indicating number of economic active household members.
HH type dummies	3 binary variables are created based on declared household type of dwelling (**HH type: house** [reference cat.], **HH type: multiple dwelling**, **HH type: other**). If the household belongs to a group according to its dwelling type, the corresponding binary variable equals 1, otherwise 0.
HH ownership dummies	2 binary variables are created based on declared household ownership type (**HH ownership: owner** [reference cat.], **HH ownership: tenant**). If the household belongs to a group according to its ownership type, the corresponding binary variable equals 1, otherwise 0.
Region dummies	8 binary variables are created based on the region where the household resides (**REG: Bratislava** [reference category], **REG: Trnava, REG: Trencin, REG: Nitra, REG: Zilina, REG: Banska Bystrica, REG: Presov, REG: Kosice**). If the person belongs to a group according to the degree of urbanization of his residence, the corresponding dummy variable equals 1, otherwise 0.
Degree of urbanisation dummies	3 binary variables are created based on number of inhabitants of the area where the household resides (**Density: Dense** [reference category], **Density: Average**, **Density: Sparse**). If the household belongs to a group according to the degree of urbanization of its residence, the corresponding dummy variable equals 1, otherwise 0.

Note: Variable names in bold.

Table A2 Estimates of the expenditure allocation model

Model	Probit Durables			WLS Durables			WLS Non-Durables		
Dependent var	Positive dur expend			log(expend_dur)			log(expend_nondur)		
log(income)	0.627	***	(0.004)	1.103	***	(0.005)	0.476	***	(0.001)
r_log(income)^2	0.411	***	(0.007)	-0.316	***	(0.011)	0.06	***	(0.001)
r_log(income)^3	-0.118	***	(0.002)	0.104	***	(0.003)	-0.026	***	(0.000)
Characteristics of HH head									
Male	0.1	***	(0.003)	0.165	***	(0.003)	0.033	***	(0.001)
Age	0.001	***	(0.000)	-0.015	***	(0.000)	-0.002	***	(0.000)
EDU: Secondary	0.376	***	(0.004)	0.029	***	(0.005)	0.118	***	(0.001)
EDU: Tertiary	0.552	***	(0.005)	0.235	***	(0.006)	0.201	***	(0.001)
Economic inactive	-0.364	***	(0.005)	0.028	***	(0.005)	-0.039	***	(0.001)
HH characteristics									
Dependent children	0.164	***	(0.003)	0.114	***	(0.003)	0.079	***	(0.000)
Household members	-0.149	***	(0.002)	-0.086	***	(0.002)	0		(0.000)
Ec. active members in HH	-0.046	***	(0.002)	0.032	***	(0.002)	0.051	***	(0.000)
HH type: multiple dwelling	-0.114	***	(0.003)	-0.335	***	(0.003)	-0.009	***	(0.001)
HH type: other	0.53	***	(0.021)	1.779	***	(0.026)	0.13	***	(0.005)
HH ownership: tenant	-0.285	***	(0.004)	-0.214	***	(0.004)	-0.044	***	(0.001)
Region: Trnava	0.132	***	(0.005)	-0.023	***	(0.005)	0.049	***	(0.001)
Region: Trencin	0.621	***	(0.006)	0.334	***	(0.005)	0.076	***	(0.001)
Region: Nitra	0.105	***	(0.005)	0.208	***	(0.005)	-0.007	***	(0.001)
Region: Zilina	0.228	***	(0.005)	0.176	***	(0.005)	0.021	***	(0.001)
Region: Banska Bystrica	0.204	***	(0.005)	0.28	***	(0.005)	-0.098	***	(0.001)
Region: Presov	0.346	***	(0.005)	-0.004		(0.005)	-0.001		(0.001)
Region: Kosice	0.202	***	(0.005)	0.189	***	(0.005)	-0.053	***	(0.001)
Density: Average	0.051	***	(0.004)	0.138	***	(0.003)	-0.013	***	(0.001)
Density: Sparse	0.061	***	(0.004)	0.093	***	(0.004)	-0.046	***	(0.001)
Constant	-3.425	***	(0.026)	-3.261	***	(0.032)	3.22	***	(0.005)
Pseudo / Adj R-squared	0.143			0.181			0.635		

Note: Standard errors in parentheses. *p < 0.05, **p < 0.01, ***p < 0.001.
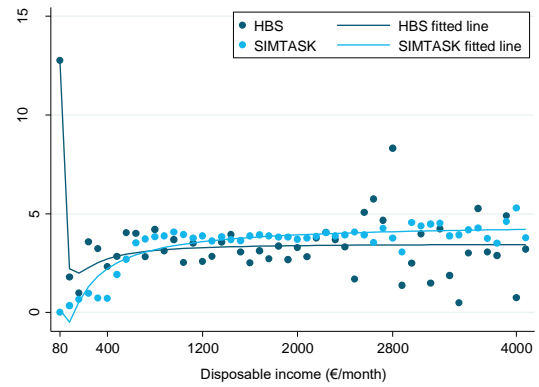Estimations are weighted using population weights.

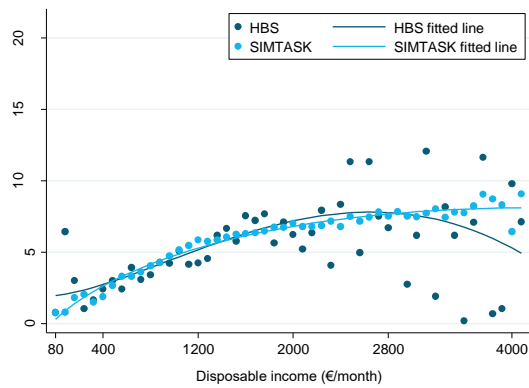Figure A1 Scatterplots of non-durable mean expenditure shares
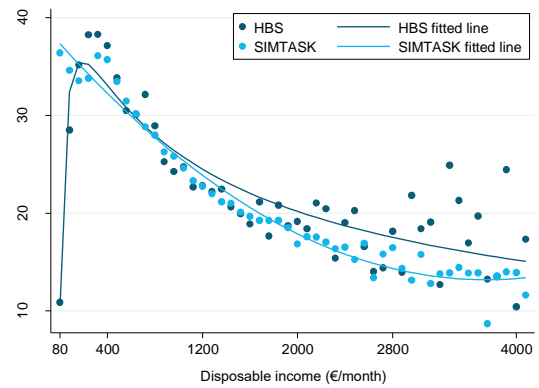
(1) Food and non-alcoholic beverages
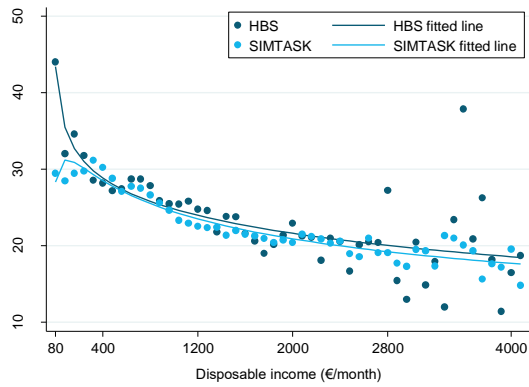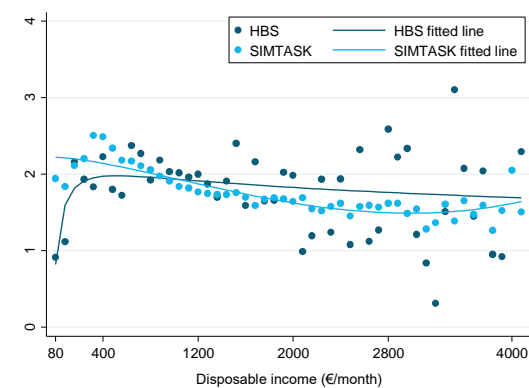
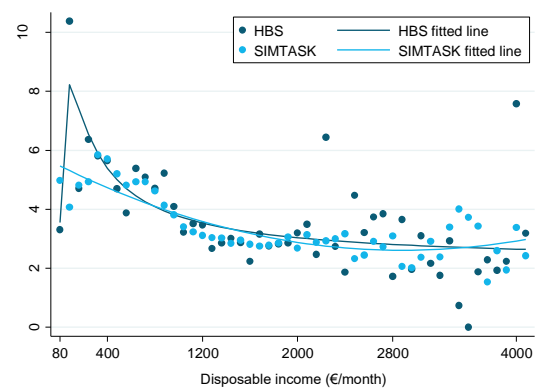(2) Alcoholic beverages and tobacco

(3) Garments and shoes

(4) Electricity and other fuels

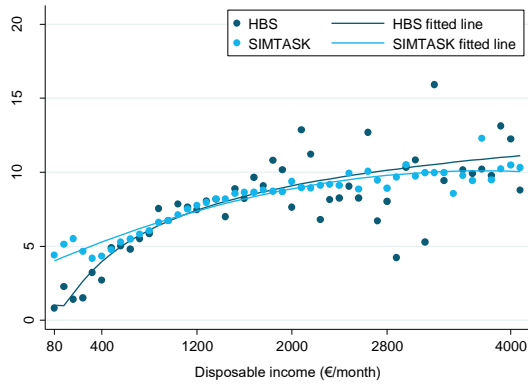(5) Household services

(6) Health

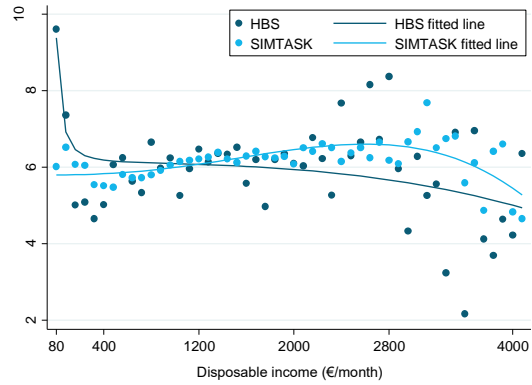Source: Authors' calculations

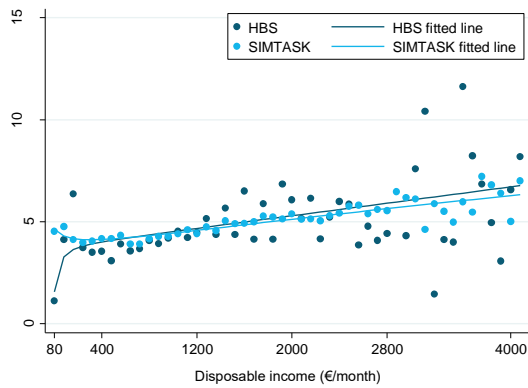Figure A1 Scatterplots of non-durable mean expenditure shares (continued)
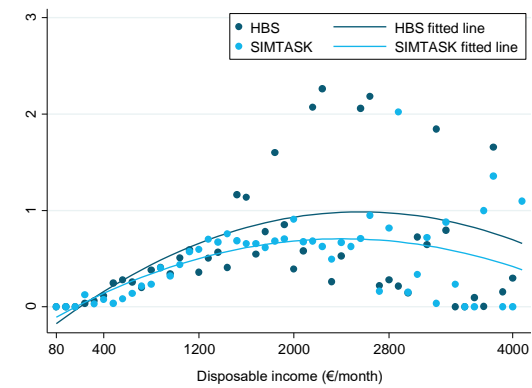
(7) Transport
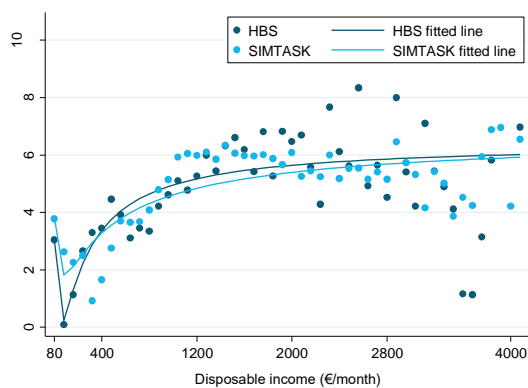

(8) Communication



(9) Recreation and culture

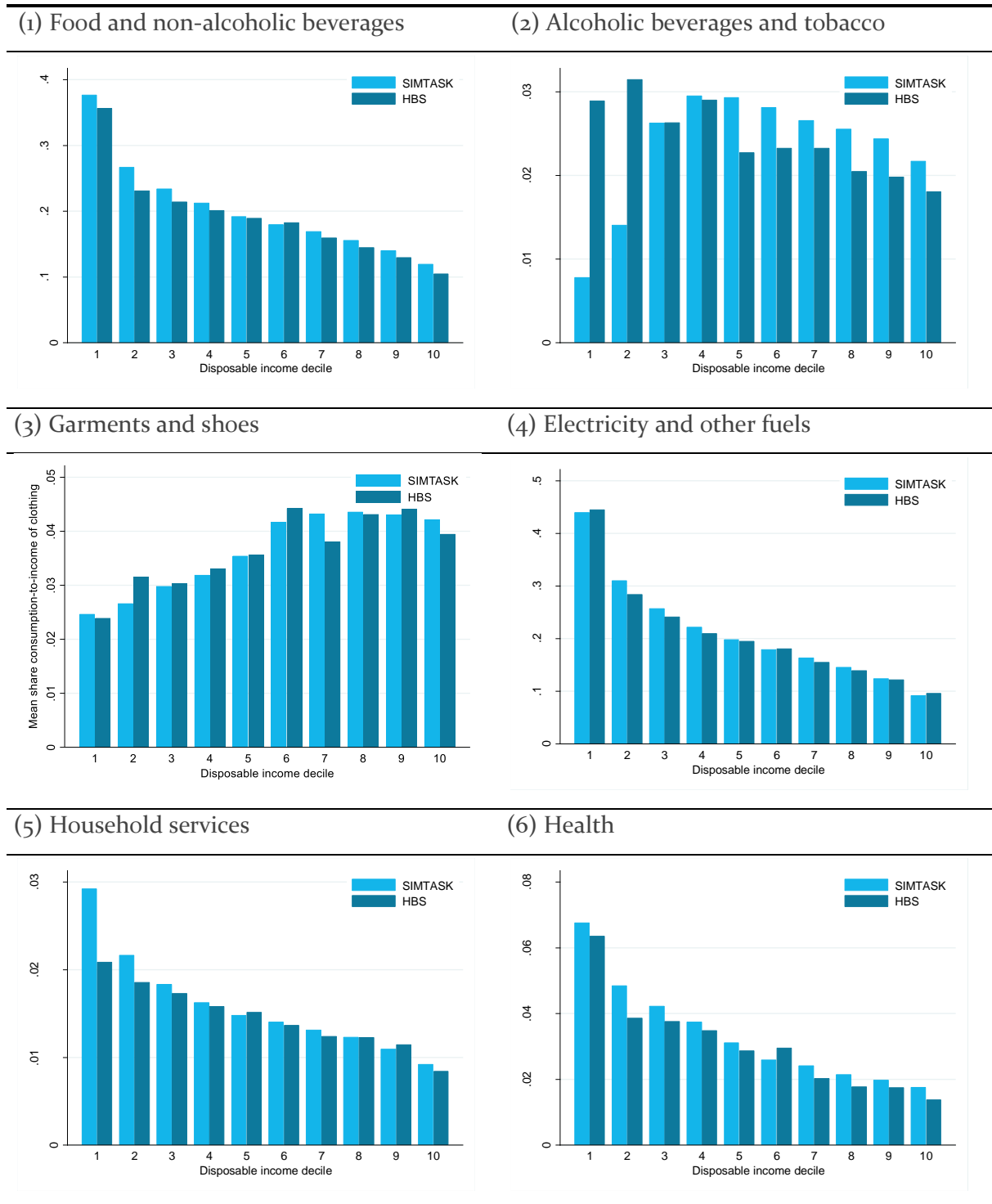

(10) Education

(11) Restaurants and hotels

(12) Other goods and services

Source: Authors' calculations

Figure A2 Mean expenditure-to-income ratios in selected categories of non-durables by income deciles

(1) Food and non-alcoholic beverages

(2) Alcoholic beverages and tobacco

(3) Garments and shoes

(4) Electricity and other fuels

(5) Household services

(6) Health
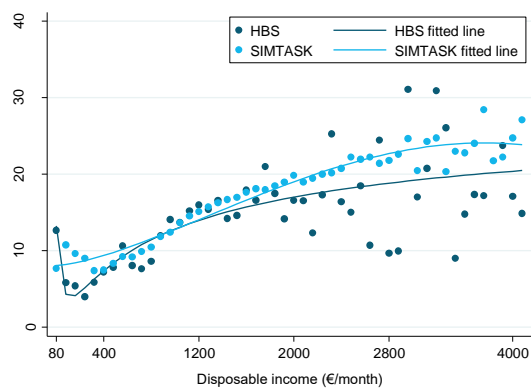
Source: Authors' calculations

Figure A2 Mean expenditure-to-income ratios in selected categories of non-durables by income deciles (continued)

(7) Transport

(8) Communication

(9) Recreation and culture

(10) Education

(11) Restaurants and hotels

(12) Other goods and services

Source: Authors' calculations

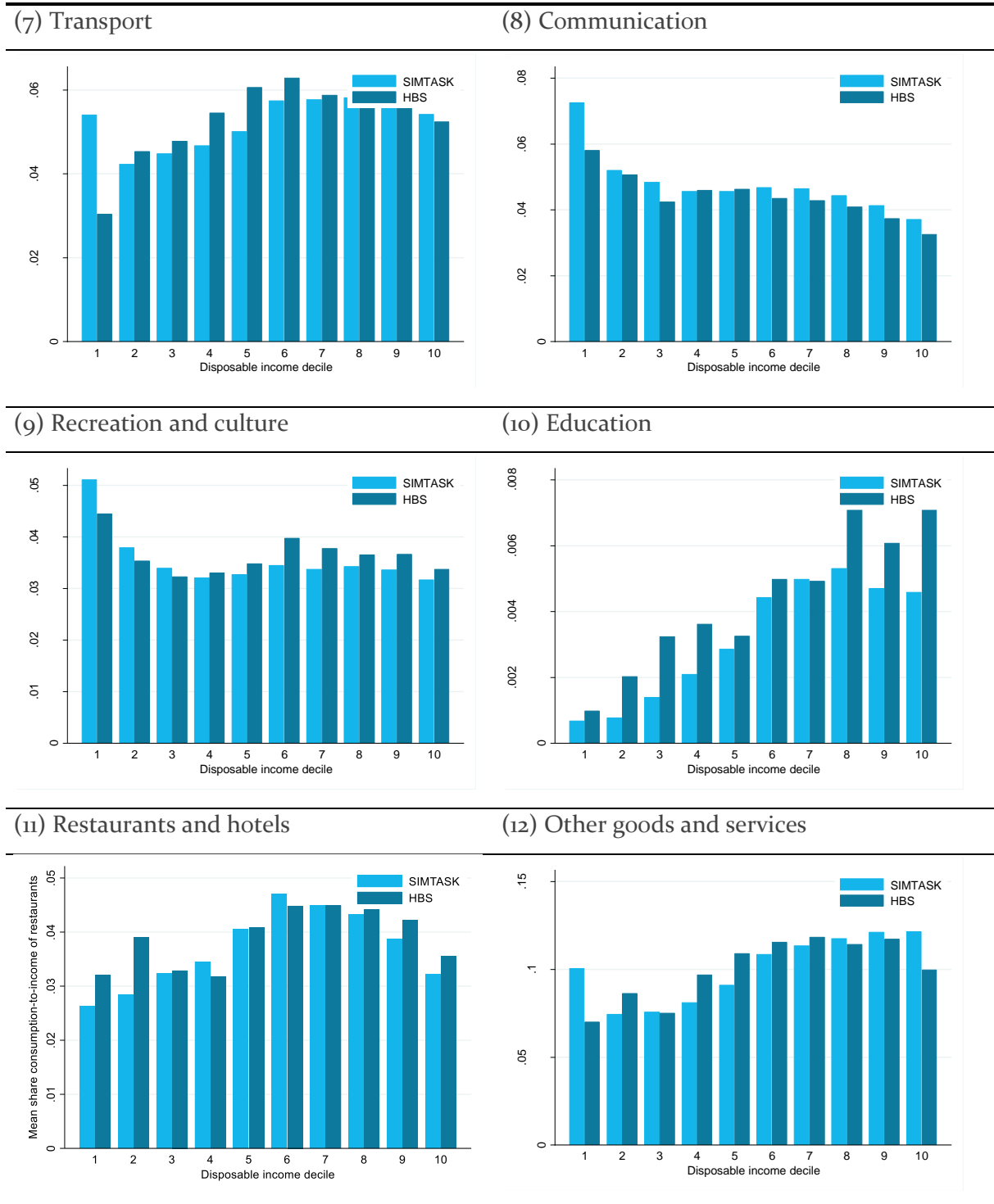Table A3 Conversion table linking the components of disposable income in households' surveys to their counterparts in national

accounts

Disposable income	EU-SILC DI components	HBS 2015	NA GDI components
Employee income (wages and salaries)/ Resources	PY010G: Employee cash or near cash income PY021G: Company car PY020G*: Non-cash employee income	PrPrZ_3: Gross income from employment PZStr: Meal vouchers	D11/rec: Wages and salaries (in kind and cash)
Income from self-employment /Resources	PY050G: Cash benefits or losses from self-employment	CPrPC: Net income from self-employment	B3G: Mixed income, gross
Property income/Resources	HY090G: Interest, dividends, profit from capital investments in unincorporated business HY040G: Income from rental of a property or land	AC2012: Interest, dividends, profit from capital investments in unincorporated business AC2022: Income from rental of a property or land	D4/rec: Property income
Income from household production of services for own consumption	HY170G*: Value of goods produced for own consumption	AC2032: Income from internal production of agricultural goods AC2042: Income from internal production: selling used goods	B2G Operating surplus, gross
Social benefits other than social transfers in kind/Resources	HY050G: Family/children related allowances HY060G: Social exclusion not elsewhere classified PY090G: Unemployment benefits PY100G: Old-age benefits PY110G: Survivor' benefits PY120G: Sickness benefits PY130G: Disability benefits PY140G: Education-related allowances HY070G: Housing allowances	StarD: Old-age benefits InvD: Disability benefits PozosD: Survivors' benefits NemDav: Sickness benefits MatRodP: Family/children allowances DavNez: Unemployment benefits PrOpZTP: Benefits for caregivers of disabled individuals AC112, AC2082: Family related transfers at the household level AC122: Material need benefit AC132: Housing allowances AC142: Education-related allowances AC152: Other regular allowances AC2092, AC2102: Other transfers	D62/rec: Social benefits, other than social transfers in kind

*Note: Not covered in disposable income definition used in SK-SILC
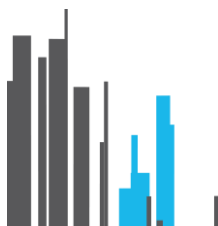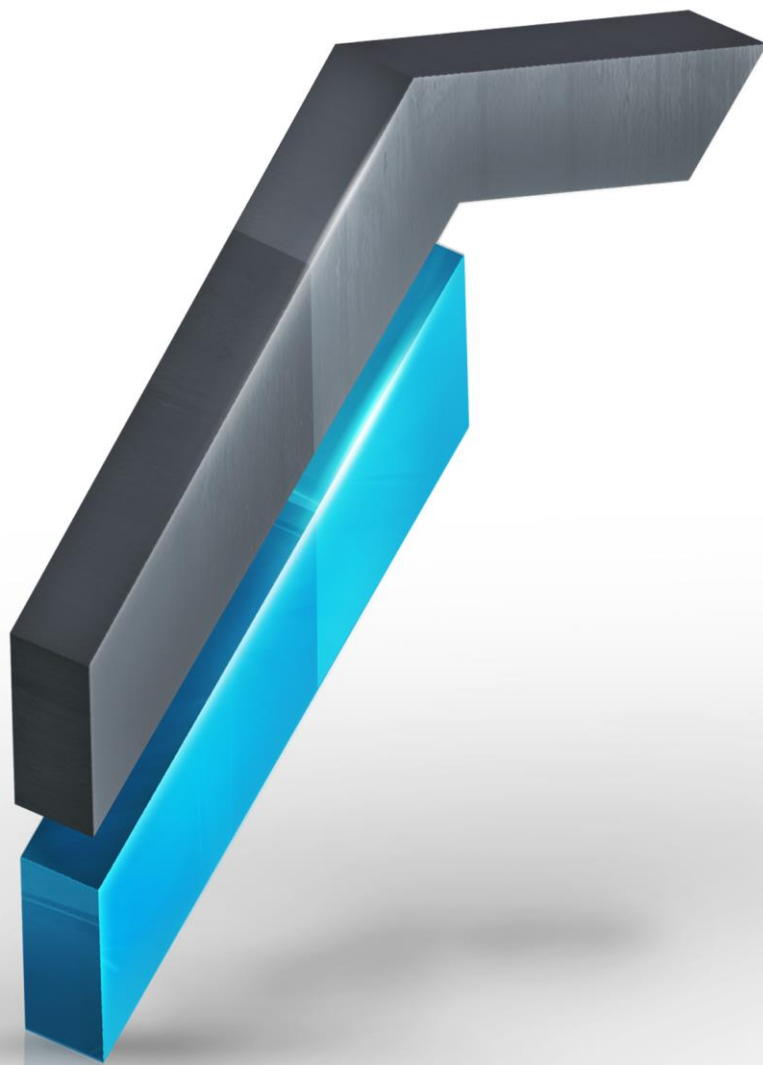Source: Eurostat, SK-SILC 2015, HBS 2015, National Accounts 2015

Table A3 Conversion table linking the components of disposable income in households' surveys to their counterparts in national

accounts (continued)

Disposable income	EU-SILC DI components	HBS 2015	NA GDI components
Other current transfers and other resources received	HY080G: Regular inter-household cash transfer received HY110G: Income received by people aged under 16 PY080G: Pension from individual private plans	PrDDPS: Pension from individual private plans AC162: Regular inter-household cash transfer received AC2052: Other inter-household cash transfer received AC2062: Returned loans	D7/rec: Other current transfers
Social contributions and taxes on income paid/Use	HY140G: Tax on income and social contributions	DanPrij_3: Tax on income DochNemP_3: Social insurance contributions PovZPoi_3: Health insurance contributions	D61/use Net social contributions D51/use: Taxes on income
Taxes on wealth paid/use	HY120G Regular taxes on wealth	x201021: Regular taxes on wealth paid x201022: Irregular taxes on wealth paid	D59/use: Other current taxes
Property income paid	HY100G*: Interest repayments on mortgage		D4/use: Property income
Other current transfers paid	HY130G: Regular inter-household cash transfer paid	AC162: Regular inter-household cash transfer paid	D7/use: Other current transfers
Total disposable income	**Total disposable income, EU-SILC**	**Total disposable income, HBS**	**Gross disposable income, NA**

*Note: Not covered in disposable income definition used in SK-SILC
Source: Eurostat, SK-SILC 2015, HBS 2015, National Accounts 2015