Matching between SILC and HBS to compute direct and indirect taxes

Speaker: **Carmen Marín** (FEDEA)

Julio López-Laborda: Universidad de Zaragoza & FEDEA
Jorge Onrubia: Universidad Complutense de Madrid & FEDEA

JRC Seville. European Commission

12 February 2020

Summary

1 Introduction

Motivation

Motivation

The aim of this paper is to estimate the tax burden of Spanish households including direct and indirect taxation.

It is necessary a dataset which combines both income and expenditure data.

Motivation

Motivation

The aim of this paper is to estimate the tax burden of Spanish households including direct and indirect taxation.

It is necessary a dataset which combines both income and expenditure data.

Not available

Exceptions

- USA: *Consumer Expenditure Survey*
- UK: *Living Costs and Food Survey*

Household Surveys

The surveys in Spain are:

- **Household Budget Survey (HBS)**
  - Provides annual information about household expenditure
  - Expenditure File (COICOP/ HBS: 12 rubrics, 262 products).

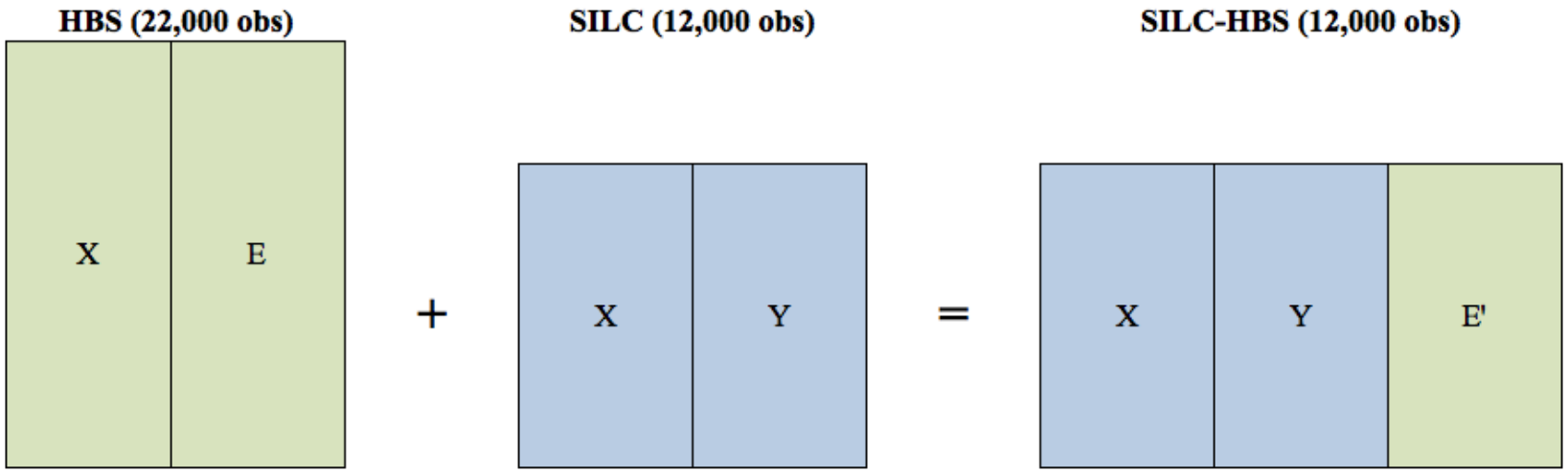
- **Survey of Income and Living Conditions (EU-SILC)**
  - Provides information about household income and direct taxation

Solution

SILC-HBS matching
Regression imputation techniques (parametric methods)

Matching Process

O'Donoghue *et al.* (2004), Decoster *et al.* (2014), De Agostini *et al.* (2017), Savage (2017).



- X = Specific household characteristics
- Y = Household Income and Direct Taxation (SILC)
- E = Household Expenditure (HBS)
- E' = Imputed Household Expenditure (SILC)

Variables included

Depedent variable: monetary expenditure (E)

Independent variable: disposable income (y).
It is rescaled in the SILC

Independent variable: household specific dummy variables (x).
Hellinger Distance

- Population density
- Household size
- Household type
- Household tenure
- Householder labour status

Difficulties of the estimation process

- The dependent variable: Household Expenditure presents a high skewness and kurtosis
- Heteroskedasticity inherent to the estimation process caused by:
  - Missespecification due to the exclusion of some important variables
  - Survey measurement error
- Treatment of zero expenditures
- Low variance of imputed expenditure

Literature Review

Article	Country	Model	Treatment Zero Expenditure	Treatment Low variance
O'Donoghue et al. (2004)	10 countries	log OLS	No	Yes (Normal)
Decoster et al. (2014)	Belgium	Log OLS	Yes	No
De Agostini et al. (2017)	10 countries	log OLS	Yes	No
Savage (2017)	Ireland	Log OLS	Yes	Yes (Normal)

Literature Review

Article	Country	Model	Treatment Zero Expenditure	Treatment Low variance
O'Donoghue et al. (2004)	10 countries	log OLS	No	Yes (Normal)
Decoster et al. (2014)	Belgium	Log OLS	Yes	No
De Agostini et al. (2017)	10 countries	log OLS	Yes	No
Savage (2017)	Ireland	Log OLS	Yes	Yes (Normal)

We propose:

- GLM log Gamma Model
- zero expenditure treatment to all goods

- low variance treatment: error term χ^2

log OLS vs. GLM log Gamma

Log OLS

We are interested in household expenditure in levels and not in logarithms. The total expenditure estimates must be corrected for retransformation bias using smearing estimates. Under heteroskedasticity, the smearing estimates produce a bias in the estimation.

GLM

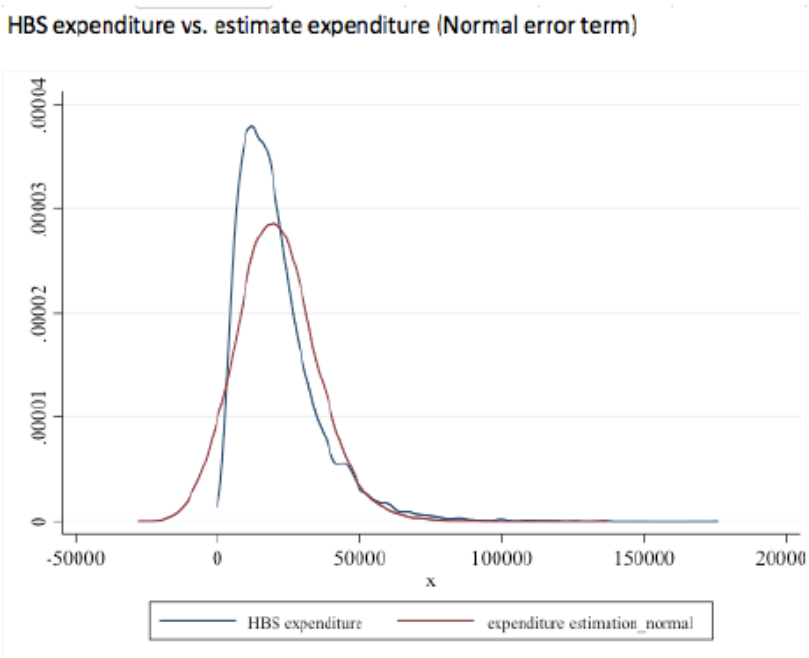
- GLMs do not suffer from the retransformation problem
- GLMs can allow heteroskedasticity through the distributional families

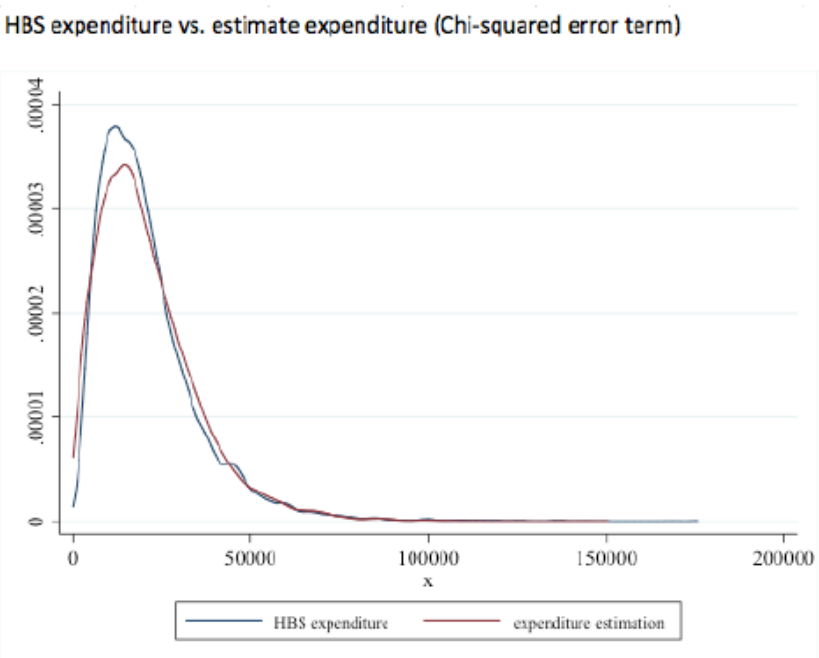
Main disadvantage: It is necessary to use the appropiate link function and distributional family to have more accurate results

Stochastic regression imputation

The SILC imputed expenditure has a lower standard deviation than the HBS expenditure ($R^2 \simeq 0.5$).

The skewness and kurtosis of a χ^2_1 is similar to HBS expenditure





2 Methodology

Statistical Matching Process

1. Determine SILC household expenditure

Estimate expenditure by GLM (log-gamma) in the HBS and impute the results in the SILC.

2. Divide the SILC total expenditure into different goods categories

2.1. HBS estimation and SILC imputation of the demand for each goods category using a Probit
2.2. HBS estimation and SILC imputation of goods shares for the households with positive demand (OLS)

3. Simulate the indirect taxation from SILC imputed expenditures

Estimation and Imputation of total expenditure

Equation 1. HBS Expenditure Estimation. GLM Log Gamma

$$ln(E_i^B) = \alpha + \gamma_1 ln(y_i^B) + \gamma_2 ln(y_i^B)^2 + \gamma_3 ln(y_i^B)^3 + x_i^{B\prime}\beta + \epsilon_i$$

Estimation and Imputation of total expenditure

Equation 1. HBS Expenditure Estimation. GLM Log Gamma

$$ln(E_i^B) = \alpha + \gamma_1 ln(y_i^B) + \gamma_2 ln(y_i^B)^2 + \gamma_3 ln(y_i^B)^3 + x_i^{B\prime}\beta + \epsilon_i$$

Equation 1'. Stochastic Expenditure Imputation in the SILC

$$ln(\hat{E}_i^I) = \hat{\alpha} + \hat{\gamma}_1 ln(y_i^I) + \hat{\gamma}_2 ln(y_i^I)^2 + \hat{\gamma}_3 ln(y_i^I)^3 + x_i^{I\prime}\hat{\beta}$$

$$E_i^I = \hat{E}_i^I + \mu_i$$

$$\mu_i \sim \chi_1^2(0, d) \text{ where } d \text{ is } \sqrt{Var(E_i^B) - Var(\hat{E}_i^I)}$$

Estimation and Imputation of total expenditure

HBS expenditure vs SILC imputed expenditure

Year	Mean		Standard Deviation		Skewness		Kurtosis	
	HBS expenditure	SILC imputation	HBS expenditure	SILC imputation	HBS expenditure	SILC imputation	HBS expenditure	SILC imputation
2013	20,979	20,960	14,490	14,458	2.05	1.41	10.95	6.38
2014	21,032	21,173	14,590	14,909	1.95	1.56	10.65	6.83
2015	21,439	21,627	14,973	15,246	2.06	1.61	11.55	7.94
2016	22,330	22,358	15,320	15,451	2.08	1.50	13.21	6.91
2017	23,354	23,505	16,037	16,056	1,93	1,39	10,12	6,29

Estimation and Imputation of total expenditure

Percentile comparison HBS expenditure and SILC imputation

Year	Percentile 90th		Percentile 95th		Percentile 99th	
	HBS expenditure	SILC imputation	HBS expenditure	SILC imputation	HBS expenditure	SILC imputation
2013	38,191	38,756	46,769	47,007	65,788	64,468
2014	38,821	38,817	47,505	48,270	66,216	65,021
2015	39,328	39,734	48,315	48,823	68,293	66,857
2016	40,647	40,832	49,603	49,285	69,379	69,841
2017	42,704	43,443	52,184	52,984	72,769	70,992

Division of HBS expenditure in different categories

The expenditure is divided in 43 subgroups

1. Food and non-alcoholic drinks (2)
2. Alcoholic beverages (4)
3. Tobacco (3)
4. Clothing (1)
5. Renting (1)
6. Home expenditures and fuels, electricity, water (8)
7. Health and pharmacy (4)
8. Private Transport (3)
9. Public Transport (1)
10. Communication (1)
11. Recreation and Holidays (4)
12. Education and Culture (4)
13. Durable Goods (3)
14. Other Expenditures (4)

Estimation and Imputation of expenditure shares

For k=1,..., 43

Equation 2: Probability of positive demand (HBS). Probit

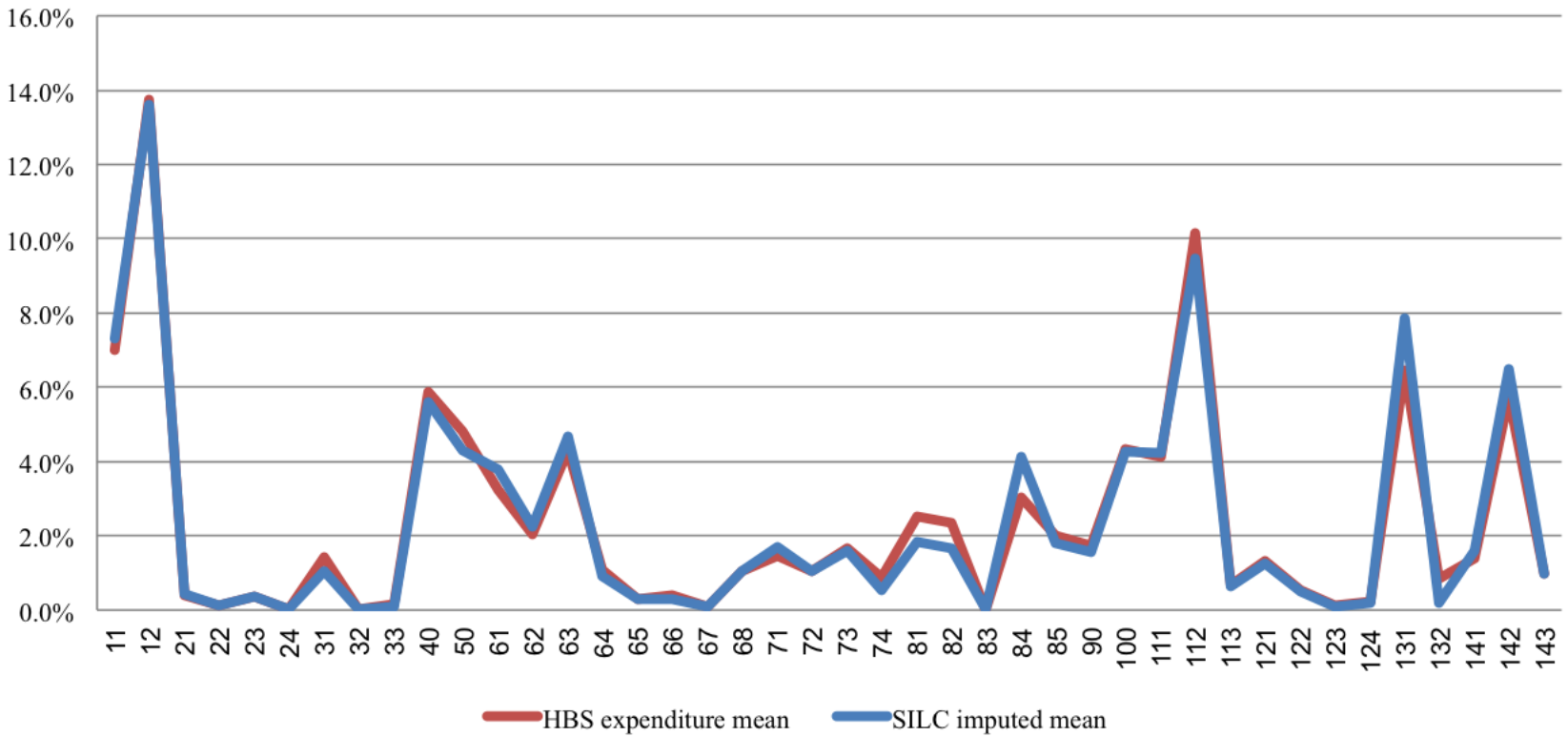
$$\Pr(D_i^k = 1) = \Phi\left(\alpha^k + \delta_1^k ln(y_i) + \delta_2^k ln(y_i)^2 + x_i'\beta^k + v_i^k\right)$$

Ecuación 3. HBS Estimation of expenditure shares. OLS

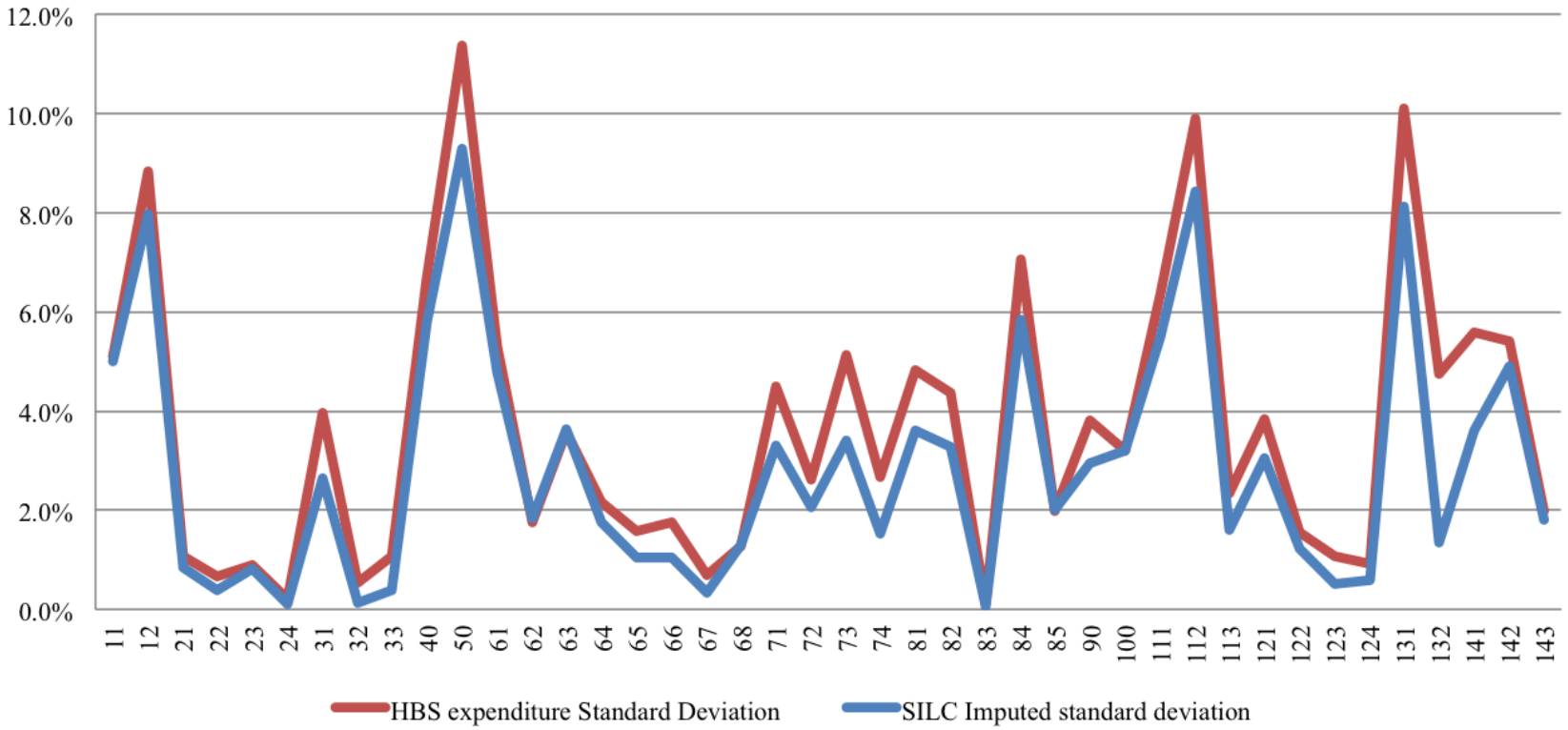
$$w_i^k = \alpha^k + \rho_1^k ln(E_i) + \rho_2^k ln(E_i)^2 + x_i'\beta^k + \epsilon_i^k \text{ if } \hat{D}_i^k = 1$$

Chi-Squared error term is added with zero mean and a variance st the new variable has the same standard deviation as the original one.

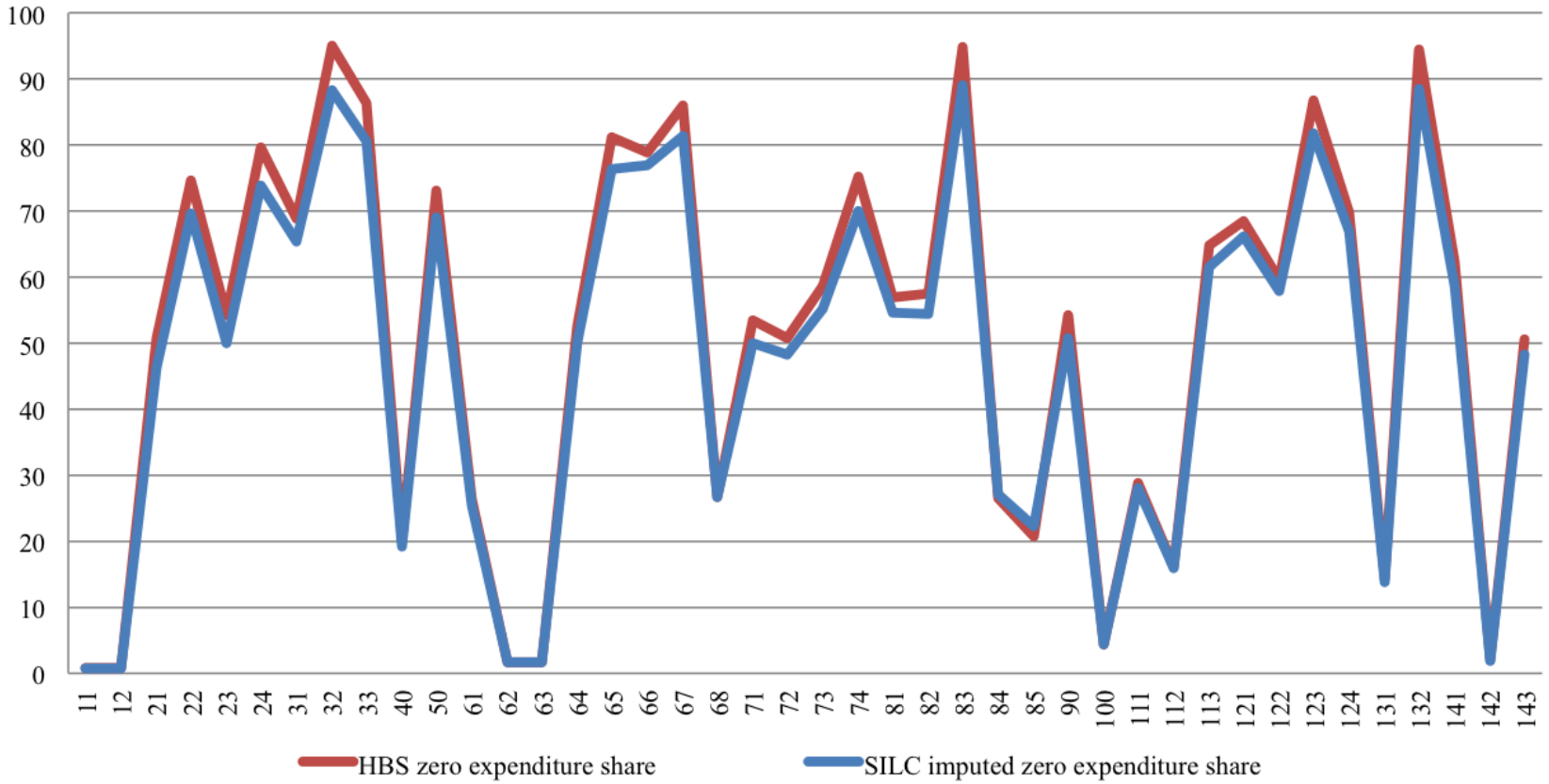
HBS expenditure and SILC-imputed mean. Year 2017



HBS and SILC-imputed standard deviation. Year 2017



HBS and SILC-imputed zero expenditure shares. Year 2017



3 Tax burden and redistributive impact analysis (2013-2017)

Household Gross Income

SILC Average household gross income (2013-2017) in euros

Quintile	2013	2014	2015	2016	2017	2017-2013 rate of change	% 2017
Quintile 1	7,249	7,267	7,662	7,441	8,248	14%	4.8%
Quintile 2	15,636	15,907	15,911	16,498	17,284	11%	10.1%
Quintile 3	24,396	24,430	24,739	25,843	26,627	9%	15.6%
Quintile 4	36,869	36,602	37,190	38,875	39,900	8%	23.4%
Decile 9	52,953	52,975	53,904	55,846	57,986	10%	17.0%
Centile 91-99	83,541	83,025	83,956	84,305	88,619	6%	23.4%
Centile 100	166,107	179,428	189,461	195,049	194,385	17%	5.7%
Total	31,305	31,405	31,941	32,854	34,130	9%	100%

Taxes Included

- **Direct Taxation** (Directly Computed from the SILC)
  - Personal Income Tax (PIT)
  - Wealth Tax (WT)
  - Employee Social Security Contributions (SSC)
- **Indirect Taxation** (Simulated Taxes from SILC Imputed Expenditures)
  - Value Added Tax and Transfer Tax (VAT&TT)
  - Excise Duties (Tobacco, Alcoholic Drinks, Fuels, Electricity) and Insurance Premium Tax (ED&IP)

Social Cash Transfers included

- Old age pension
- Survivor pension
- Unemployment subsidies
- Permanent disability
- Other: sickness, social assistance, housing, family and children, grants.

Household Tax Burden per taxation figure

Average Tax Rate per taxation figure (2013-2017)(%)

Tax	2013	2014	2015	2016	2017
Direct Taxes	16.7	17.2	16.5	16.4	16.8
Personal income tax	11.5	11.9	11.4	11.3	11.7
Wealth tax	0.06	0.04	0.05	0.05	0.07
Social Security	5.1	5.2	5.1	5.1	5.0
Indirect taxes	9.9	9.9	9.8	9.9	10.0
VAT & Transfer Tax	7.8	7.9	7.9	8.0	8.1
Excise & Insurance Premium	2.1	2.0	2.0	1.9	1.9
Total Tax Burden	26.6	27.1	26.4	26.3	26.8

Household Tax Burden per gross income groups

Tax Burden. Average Tax rate (2013-2017) (%)

Quintile	2013	2014	2015	2016	2017
Quintile 1	29.4	29.4	26.8	27.7	25.3
Quintile 2	22.5	22.6	23.0	22.7	22.4
Quintile 3	23.3	23.2	22.7	23.0	23.9
Quintile 4	25.3	25.6	25.1	24.6	25.5
Decile 9	26.8	27.4	26.7	26.8	27.2
Centile 91-99	29.5	29.9	29.1	29.2	30.0
Centile 100	33.3	37.4	34.4	34.2	35.3
Total Tax Burden	26.6	27.1	26.4	26.3	26.8

Household Tax Burden per gross income groups

Direct Taxes. Average Tax Rate (2013-2017) (%)

Quintile	2013	2014	2015	2016	2017
Quintile 1	9.8	9.0	8.1	7.7	7.1
Quintile 2	8.3	9.1	8.8	8.6	8.5
Quintile 3	11.8	11.8	11.3	11.4	12.3
Quintile 4	15.4	15.6	15.2	15.0	15.5
Decile 9	18.1	18.6	18.0	18.1	18.4
Centile 91-99	22.4	22.6	21.9	21.8	22.5
Centile 100	27.8	32.2	29.5	28.9	29.7
Direct Tax Burden	16.7	17.2	16.5	16.4	16.8

Household Tax Burden per gross income groups

Indirect taxes. Average Tax Rate (2013-2017) (%)

Quintile	2013	2014	2015	2016	2017
Quintile 1	19.6	20.3	18.6	20.0	18.2
Quintile 2	14.2	13.5	14.3	14.1	13.8
Quintile 3	11.5	11.4	11.4	11.6	11.7
Quintile 4	9.9	10.0	10.0	9.6	10.1
Decile 9	8.7	8.8	8.6	8.7	8.8
Centile 91-99	7.1	7.3	7.2	7.3	7.5
Centile 100	5.6	5.2	4.9	5.3	5.6
Indirect Tax Burden	9.9	9.9	9.8	9.9	10.0

Social cash transfers per figure

Social cash transfers. Average Subsidy (2013-2017) (%)

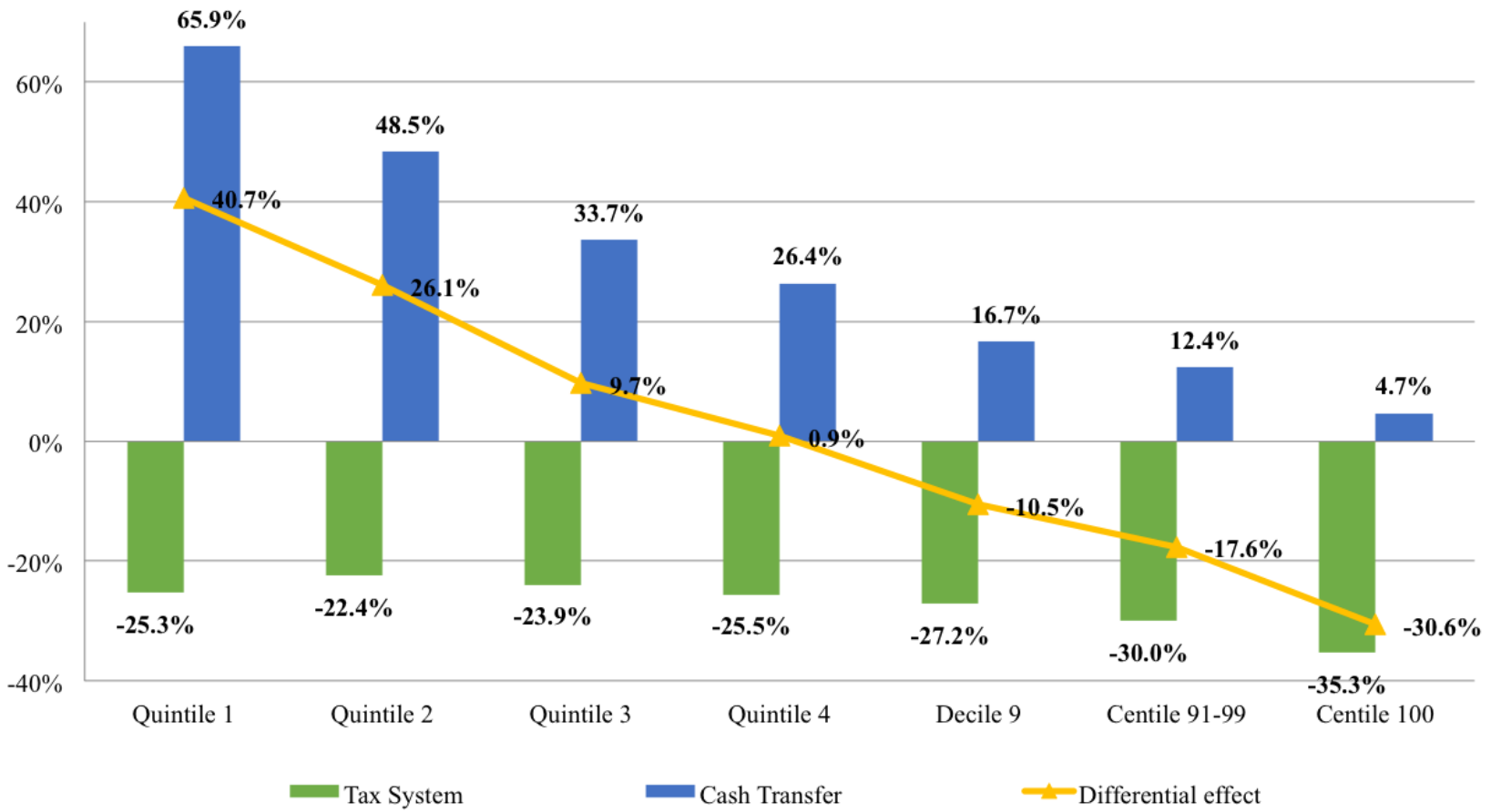
Cash Transfer	2013	2014	2015	2016	2017
Old age pension	15.6	15.8	15.9	15.8	15.8
Survivor pension	4.0	4.0	4.0	3.9	3.8
Unemployment subsidy	4.8	3.8	3.1	2.7	2.5
Permanent disability	2.4	2.3	2.4	2.2	2.2
Other	1.3	1.2	1.4	1.4	1.3
Total	28.0	27.2	26.8	26.0	25.5

Social cash transfers per gross income groups

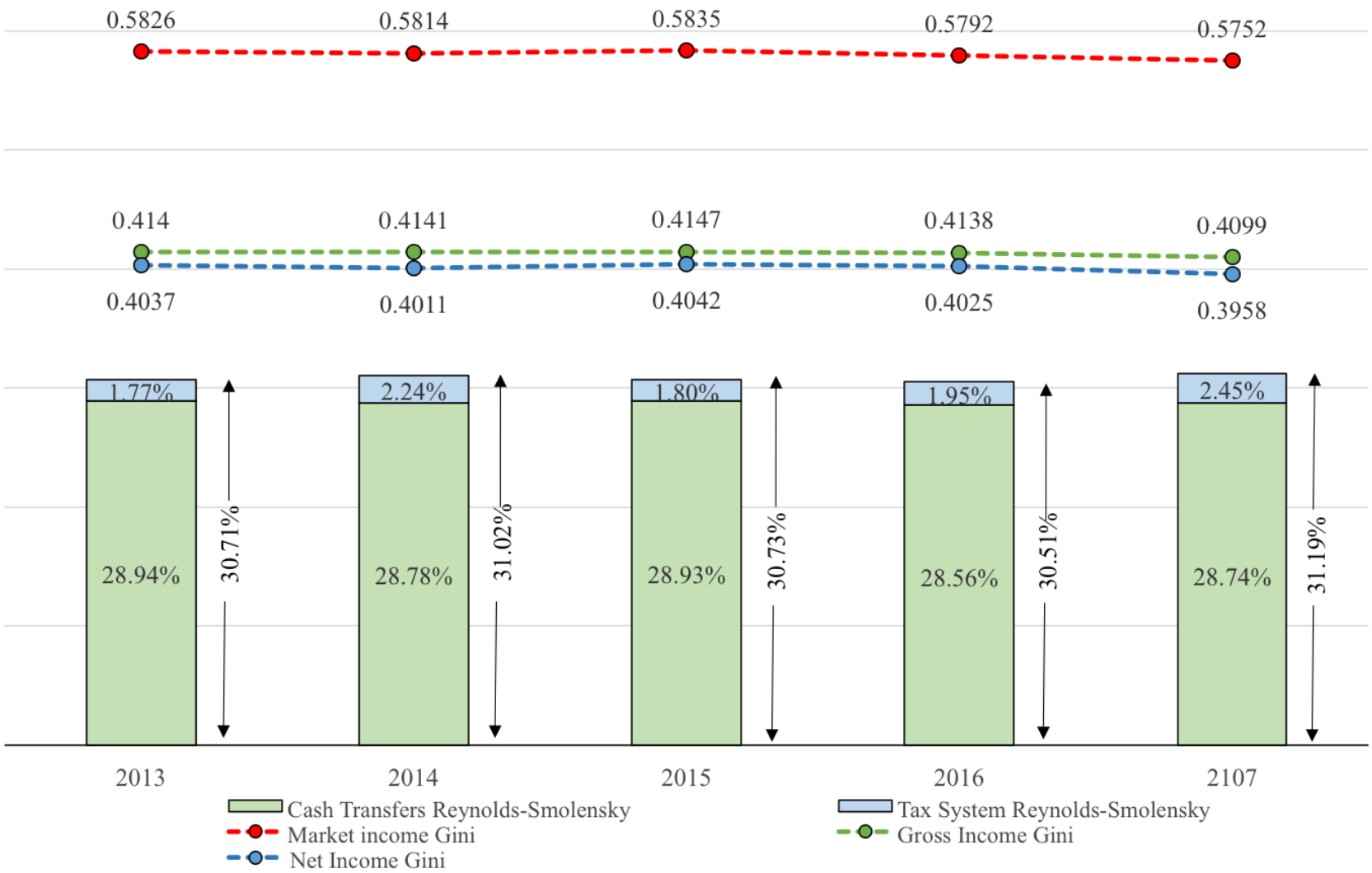
Social cash transfers. Average Subsidy (2013-2017) (%)

Quintile	2013	2014	2015	2016	2017
Quintile 1	67.3	67.8	67.3	67.0	65.9
Quintile 2	52.2	51.1	49.2	49.7	48.5
Quintile 3	38.8	36.8	36.9	36.0	33.7
Quintile 4	28.6	27.5	27.2	25.7	26.4
Decile 9	19.2	18.4	18.0	18.0	16.7
Centile 91-99	13.6	14.2	14.0	12.4	12.4
Centile 100	7.4	5.3	4.1	4.6	4.7
Total	28.0	27.2	26.8	26.0	25.5

Distribution of Cash Transfers and Taxes. Year 2017



Redistributive Impact Analysis



Redistributive Impact Analysis

Onrubia, Picos & Rodado (2014)

Equation 5. Reynolds-Smolensky Decomposition

$$\Pi^{RS} = \sum_{i=1}^{5} \frac{\overline{Y - S_i}}{\overline{Y - S}} (\hat{\Pi}^{RS}_{Y,Y-S_i} - R_i) - R$$

$$R_i = G_{Y-S_i} - C_{Y-S_i}$$

$$\hat{\Pi}^{RS}_{Y,Y-S_i} = C_Y - C_{Y-S_i}$$

Equation 6. Kakwani Decomposition

$$\Pi^{K} = \frac{\bar{Y}}{\bar{S}} \sum_{i=1}^{5} \frac{\bar{S}_i}{\bar{Y}} \Pi^{K}_{Y,Y-S_i}$$

Redistributive Impact Analysis

Decomposition of Reynolds-Smolensky Index over gross income Gini (2013-2017) (%)

Tax	2013	2014	2015	2016	2017
Personal income tax	8.96	9.56	9.06	8.72	9.09
Wealth tax	0.04	0.03	0.01	0.02	0.06
Social Security	-1.29	-1.25	-1.14	-0.75	-0.75
VAT & Transfer Tax	-4.11	-4.19	-4.37	-4.46	-4.08
Excise & Insurance Premium	-0.88	-0.79	-0.76	-0.65	-0.66
Reranking	0.24	0.21	0.27	0.17	0.20
Total Tax System	2.49	3.16	2.54	2.71	3.46

Redistributive Impact Analysis

Decomposition of Kakwani Index (2013-2017) (%)

Tax	2013	2014	2015	2016	2017
Personal income tax	0.1092	0.1136	0.1112	0.1072	0.1087
Wealth tax	0.0005	0.0004	0.0002	0.0003	0.0007
Social Security	-0.0111	-0.0093	-0.0092	-0.0050	-0.0050
VAT & Transfer Tax	-0.0439	-0.0436	-0.0468	-0.0480	-0.0421
Excise & Insurance Premium	-0.0092	-0.0081	-0.0080	-0.0068	-0.0067
Total Tax System	0.0456	0.0531	0.0474	0.0478	0.0557

4 Conclusions

Conclusions

- **GLM log Gamma and the Chi-squared procedure** is the model chosen to estimate the HBS expenditure and impute the results in the SILC.
- **The household tax burden** was on average nearly 26-27% for the period 2013-2017. The lowest taxation ocurred in 2016 (26.31%), whereas the highest one in 2014 (27.10%).
- **First quintile households** support a high indirect tax burden and social security contributions (high level of consumption and the existence of minimum social security basis).
- **Personal Income Tax** causes the income redistribution and progressiveness of the Spanish Tax System.

Conclusions

- **The Spanish tax system** is progressive and reduces the gross income inequality, as a whole. The greatest reduction in inequality was produced in 2017 (3.46%) and the smallest one in 2013 (2.49%).
- **The cash transfers** reduce the gross income inequality in almost 29%, jointly with the taxation system, the reduction in inequality amounts nearly 31%.
- **Pensions** causes the majority of the income redistribution.

References

- Agostini, P.D., B. Capau, A. Decoster, F. Figari, J. Kneeshaw, C. Leventi, K. Manios, A. Paulus, H. Sutherland and T. Vanheukelom (2017), "Euromod extension to indirect taxation: Final report", EUROMOD Technical Note Series EMTN-3.0, EUROMOD.
- Decoster A., R. Ochmann and K. Spiritus (2014), "Integrating VAT into EUROMOD. Documentation and results for Belgium", Euromod Working Paper Series, EM12/14.
- O'Donoghue C., M. Baldini and D. Mantovani (2004), "Modelling the redistributive impact of indirect taxes in Europe: an application of Euromod", Euromod Working Paper Series, EM7/01.
- Onrubia J., F. Picos and M.C. Rodado (2014), "Rethinking the Pfähler-Lambert decomposition to analyse real-world personal income taxes", International tax and public finance, 21:796-812.
- Savage M. (2017), "Integrated modelling of the impact of direct and indirect taxes using complementary datasets", The Economic and Social Review, 48(2), 171-205.

THANK YOU