CEPS INSTEAD

IGSS Luxembourg

Cross-validating administrative and survey data sets through microsimulation

IJM & EUROSTAT WPs

Frédéric Berger (CEPS/INSTEAD, Luxembourg)
Nizamul Islam (CEPS/INSTEAD, Luxembourg)
PHILIPPE LIÉGEOIS (CEPS/INSTEAD, Luxembourg and DULBEA, ULB, Brussels)
and Raymond Wagener (IGSS, Luxembourg)

(REDIS Project : Coherence of social transfers policies in Luxembourg through the use of microsimulation models / FNR Luxembourg)

INTRODUCTION : THE GENERAL MOTIVATION

- An administrative Data Warehouse built up a few years ago by the Social Security administration in Luxembourg (IGSS)
  - a normalized and exhaustive basis for the generation of statistics serving diversified purposes
  - first operational wave covering 2003
  - complementary to PSELL3/EU-SILC data
- What about a comparison (cross-validation) between survey and administrative data ?
  - No link
  - Comparison at the "macro" level

STRUCTURE OF THE PRESENTATION

- Motivation of the paper
- Preparing the datasets for comparison
- Cross-validating the datasets
- Going ahead : the comparative effects of an important tax reform in Luxembourg (2001-2002) if analyzed through administrative or survey data
- Conclusions

- Comparing what is comparable
- Survey data (PSELL 3/ EU-SILC)
  - Covering the resident population only
- Administrative data
  - First operational wave covering 2003
  - No information from Fiscal administration (mainly through social security admnistration)
  - No information on Capital income and Private transfers
  - Families constructed on a fiscal basis, from links
    - (i) between spouses and
    - (ii) between parents and (socially) dependent children

Topic	EUROMOD survey-based data	EUROMOD administrative-based data	Action / Remarks
Number of individuals before the present adaptation process	443,642 (weighted)	449,025	1 Information about cross-boarders available in administrative data, but not in survey data, hence initially dropped in the latter, leading to 449,025 cases
Unit of analysis	Residence household	2 Fiscal household	All comparisons and actions to be based on fiscal households
Non-private households	Not included	Included but cannot be identified	None (**)
International civil servants	Included	Excluded but may happen that household's members still within the data	(**) Administrative-based data : Drop cases (*) if a married partner announced despite absent from the data (***) Survey-based data : Drop cases (*) if a member of the household not socially insured in GDL (***)
Voluntary insured	Included but cannot be identified	Included and can be identified (but earnings not reliable)	(**) Drop cases (*) in administrative-based data if a member of the household voluntarily insured
Capital income and private transfers	Information collected	3 Unknown	Variables set to "0" in survey-based data
Income from agriculture	Information collected	Information available (but earnings not reliable)	Drop cases (*)
Number of individuals left after the present adaptation process	419,030 (weighted)	4 418,861	Administrative-based data : 7% cases dropped Survey-based data : 5% cases dropped

Table 3 Comparing EUROMOD datasets when unit of analysis is the HOUSEHOLD

Characteristics	Categories	Survey-based EUROMOD data		Administrative-based EUROMOD data (fiscal households only)
		Resident households	Fiscal households	
Number of households	Raw data *(i)*	3,296	4,274	212,578
	Weighted count *(i)*	169,620	205,802	
Number of fiscal households in the resident household	1	80% *(ii)*	Not available	Not available
	2	17%	Not available	Not available
	3 or more	2%	Not available	Not available
Number of persons in the household	1	30%	47%	50%
	2	28%	25%	24%
	3 or 4	33%	23%	21%
	5 or more	9%	5%	5%
Number of workers *(iii)* in the household	0	30%	34%	35%
	1	40%	48%	47%
	2 or more	29%	18%	17%
Type of household	Single (< 65)	19%	35%	37%
	Single (> 65)	11%	12%	14%
	Single with dependent(s) *(iv)*	7%	6%	5%
	Couple – 0 dependent	63%	21%	20%
	Couple – 1-2 dependent(s)		20%	20%
	Couple – 3 dependents or more		5%	5%
	Others		Not relevant	Not relevant

1 2 3

Table 4 Comparing EUROMOD datasets when the unit of analysis is the INDIVIDUAL: Non-monetary characteristics

Characteristics	Categories	Survey-based EUROMOD data	Administrative-based EUROMOD data
Number of persons	Raw data *(i)*	8,657	418,749
	Weighted count *(i)*	419,030	
Gender	Female	50.7%	50.5%
	Male	49.3%	49.5%
Age	Age < 18	22%	22%
	18<= Age < 59	59%	59%
	Age >= 60	19%	20%
Type of household	Single (< 65)	17%	19%
	Single (> 65)	6%	7%
	Single with dependent(s) *(ii)*	7%	6%
	Couple – 0 dependent	21%	21%
	Couple – 1-2 dependent(s)	35%	35%
	Couple – 3 dependents or more	14%	12%
Number of workers *(iii)* in the household	0	25%	26%
	1	45%	45%
	2 or more	30%	29%

- Comparison of monetary characteristics
  - A common reference tool : the equivalent income
    = household disposable income / « weight » of the household
  - Household disposable income
    = primary gross income – taxes – social cont. + benefits
- Using EUROMOD static microsimulation
  - Integration of rules regarding taxes and benefits
  - Given « personal » situation (marital status, household composition), derivation of taxes and benefits (and their changes)
  - Can hold any influence constant while simulating a policy reform, allowing to focus on selected « pure » effects
  - A reduced set of input variables (synthetic basis)
  - Analysis at the individual / household level

Table 5 Comparing EUROMOD datasets when the unit of analysis is the INDIVIDUAL: Monetary characteristics, on average (in EUR / month)

Monetary variables		Survey-based data: Resident households	Survey-based data: Fiscal households	*Ratio: Fiscal/ Resident*	Administrative-based data
Primary income (excluding capital income) (mean)		1,493 [1,416 – 1,570] **1** *(-7.3%)*		Not relevant	1,384
Capital income (mean)		78		Not relevant	Not available in source data
Standard disposable income (excluding capital income) (mean)		1,644 **2** *(-4.0%)*		Not relevant	1,579
Total household primary income (excluding capital income) (mean)		4,489	3,900	0.913	3,561
Total household disposable income (excluding capital income) (mean)		4,715	4,068	0.863	3,822
OECD equivalent weight (mean)		1.96	1.77	0.903	1.74
OECD equivalised income	Mean	2,444	2,314 **3**	0.947	2,200
	Median	2,219	2,095	0.944	1,975
	Poverty line (60% of the median)	1,331	1,257 [1,237 – 1,277]	0.944	**4** 1,185

CEPS INSTEAD

Table 6 Comparing EUROMOD datasets when the unit of analysis is the INDIVIDUAL: Inequality indicators and redistribution effects of the tax system (*)

Inequality indicators	Survey-based EUROMOD data		Administrative-based EUROMOD data	
	Without tax reform (A)	With tax reform (B)	Without tax reform (C)	With tax reform (D)
Gini before tax *(i)* (1)	0.297		0.299	
Gini after tax *(ii)* (2)	0.231	0.245 [0.238 - 0.251] *(iii)*	0.233	0.248
ΔG (3) = (1) - (2) = (4) - (5)	0.067	0.053	0.066	0.051
Reynolds-Smolesnsky index of vertical equity (4) = (6) * ((7) / 1 - (7))	0.068	0.054	0.067	0.052
Re-ranking index of horizontal inequity	0.001	0.001	0.001	0.001
Kakwani index of tax progressivity	0.342	0.411	0.357	0.430
Rate *(iv)* (7)	0.166	0.115	0.158	0.108
P75 / P25	1.721	1.811 [1.772 - 1.850]	1.739	1.823
P90 / P10	2.741	2.917 [2.836 - 2.998]	2.720	2.907
Atkinson index (inequality aversion = 0.5)	0.042	0.047 [0.045 - 0.050]	0.045	0.051
Atkinson index (inequality aversion = 2)	0.161	0.168 [0.160 - 0.177]	0.207	0.226

1
2
3

Characte-ristics	Categories	Data (*)	Share in total population	Poverty rate
All		Adm	100.0%	9.6%
		Survey	100.0%	11.5%
Gender	Female	Adm	50.5%	9.6%
		Survey	50.7%	11.4%
	Male	Adm	49.5%	9.7%
		Survey	49.3%	11.6%
Age	Age < 18	Adm	21.5%	12.1%
		Survey	22.4%	17.0%
	18<= Age < 60	Adm	58.8%	11.0%
		Survey	58.9%	12.1%
	Age >= 60	Adm	19.7%	2.7%
		Survey	18.7%	2.9%
Type of household	Single (< 65)	Adm	18.6%	13.5%
		Survey	17.3%	13.6%
	Single (>= 65)	Adm	6.9%	3.5%
		Survey	6.0%	1.7%
	Single with dependent(s)	Adm	6.4%	24.8%
		Survey	7.5%	26.8%
	Couple - 0 dependent	Adm	20.8%	3.5%
		Survey	20.5%	4.7%
	Couple - 1-2 dependent(s)	Adm	35.2%	9.4%
		Survey	35.2%	11.2%
	Couple - 3 dependents or more	Adm	12.1%	10.2%
		Survey	13.5%	15.8%
Number of workers in the household	0	Adm	26.0%	9.4%
		Survey	24.8%	13.6%
	1	Adm	44.7%	11.9%
		Survey	45.2%	15.0%
	2 or more	Adm	29.3%	6.4%
		Survey	30.0%	4.5%

1 2 3 4 5

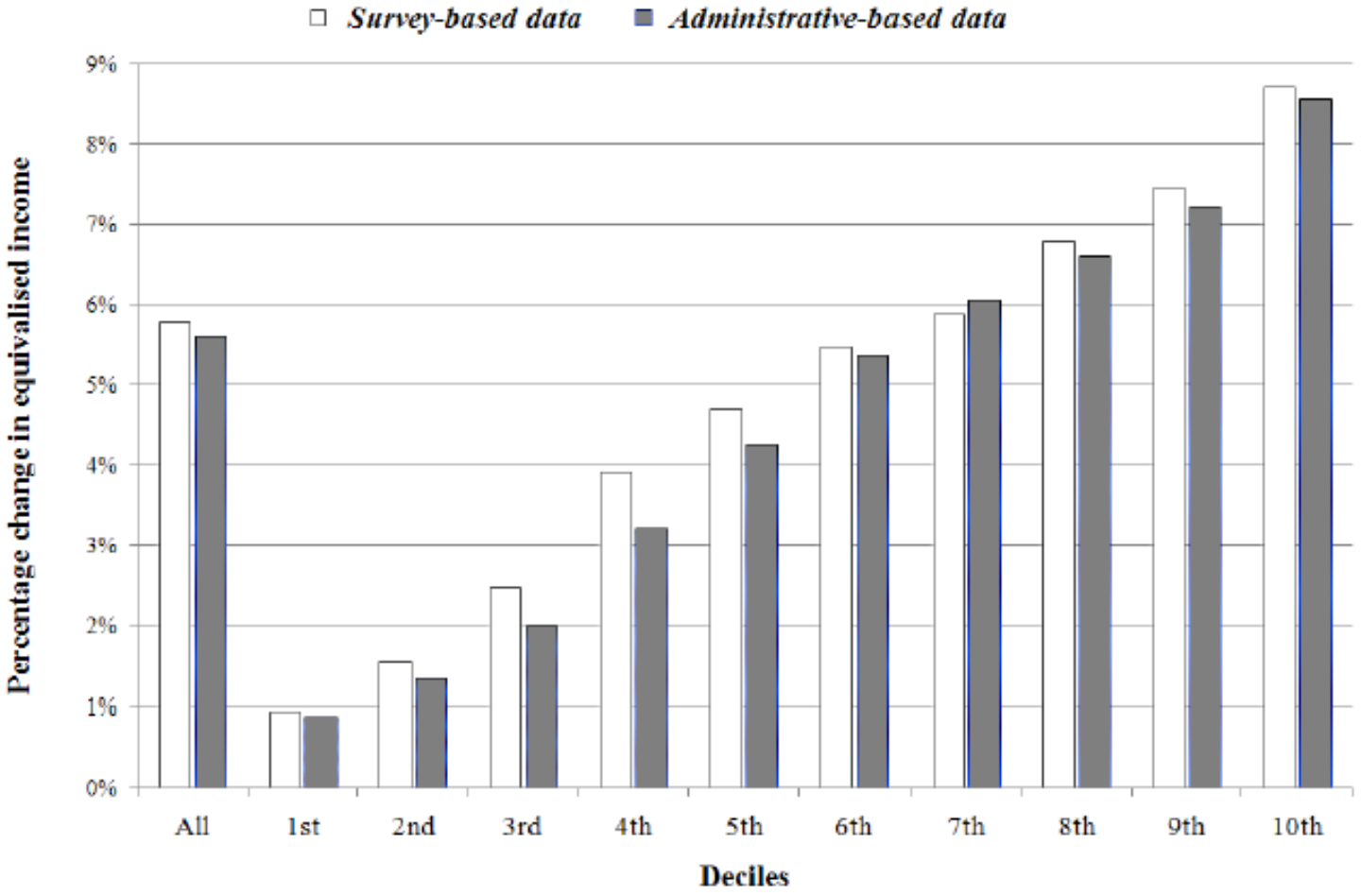
- Going ahead the comparison of raw data
- An application : analysis of the outcome of the important 2001-2002 tax reform
  - First tax bracket (free of tax) enlarged : 6,693 EUR to 9,750 EUR
  - Maximum marginal tax rate reduced from 46% to 38%
- We still target the 2003 population

Inequality indicators	Survey-based EUROMOD data		Administrative-based EUROMOD data	
	Without tax reform (A)	With tax reform (B)	Without tax reform (C)	With tax reform (D)
Gini before tax *(i)* (1)	0.297		0.299	
Gini after tax *(ii)* (2)	0.231	0.245 [0.238 - 0.251] *(iii)*	0.233	0.248
ΔG (3) = (1) - (2) = (4) - (5)	0.067	0.053	0.066	0.051
Reynolds-Smolensnsky index of vertical equity (4) = (6) * ((7) / 1 - (7))	0.068	0.054	0.067	0.052
Re-ranking index of horizontal inequity	0.001	0.001	0.001	0.001
Kakwani index of tax progressivity	0.342	0.411	0.357	0.430
Rate *(iv)* (7)	0.166	0.115	0.158	0.108
P75 / P25	1.721	1.811 [1.772 - 1.850]	1.739	1.823
P90 / P10	2.741	2.917 [2.836 - 2.998]	2.720	2.907
Atkinson index (inequality aversion = 0.5)	0.042	0.047 [0.045 - 0.050]	0.045	0.051
Atkinson index (inequality aversion = 2)	0.151	0.168 [0.160 - 0.177]	0.207	0.226

1

2

Figure 1 Relative change in mean equivalised income due to the tax reform, by decile



- We initiate the cross-validation of administrative data (Luxembourg Social Security Data Warehouse), on the one side, and of the survey data (PSELL3/EU-SILC), on the other side.

- Before validating, control for target populations and make closely related the contents of income variables

→ *6% of initial « populations » dropped ; capital income ignored; the unit of analysis is fiscal household*

- Non-monetary characteristics
  (age classes, types of households, …)

→ *« Strong similarities » between the two data sets*

- Monetary characteristics examined through the EUROMOD framework
  - An initial significant discordance in primary income (7% lower in administrative data) … transferred downstream but reduced through equivalent disposable income
- Inequality most often « similar » in both sets but :
  - Poverty rates lower through administrative data
  - Intensity of poverty higher through administrative data
- Comparing the outcome of a tax reform through both datasets => the impact of the reform on inequality indices are much more important than differences we can observe when comparing administrative and survey datasets