# » Using territorial data to define sampling of HFCS

Catarina Valente

Francisco Vala

João Lopes

DMSI / ME
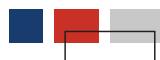
Outubro de 2020

# Motivation

## » Increase in Inequality

- Capital income becoming increasingly important

- Inheritances are returning to be major factor [Alvaredo et al. 2013]

## » Household Finance and Consumption Survey

- Harmonized survey across euro zone, at household-level

- Collect information on income, assets and liabilities

- Establishment of monetary and financial stability policies

- Understand macroeconomic shocks [HFCN 2009, 2013, 2016]

» HFCS in Portugal

- Collected in person (costly; large burden on respondents)

- Stratified sample (PSU: territorial areas; SSU: households)

- Over-sample wealthy classes [INE 2017]

» Over-sampling

- Distributions are skewed-right and difficult to characterize

- Non-response rates are biased towards wealthy

Underestimation of inequality [Kennickell 2005, 2008; HFCN 2016]

» Sampling scheme in Portugal

Use Useful Area information at household-level:

- Area-probability subsample (4,000 households)

- Wealthy subsample (4,000 households)

» Alternative sampling scheme

- Use economic information at regional-level [Valente 2019]

Decrease sampling effort to statistical authorities

Decrease statistical burden to respondents

## » Economic information at regional-level

- Personal Income Tax (2016)
- Real-estate prices (2017)

## » Variables [Valente 2019]

- value $P_{80}$ income at regional-level
- proportion ≥ $P_{80}$ income at country-level
- value $P_{80}$ real-estate at regional-level
- proportion ≥ $P_{80}$ real-estate at country-level
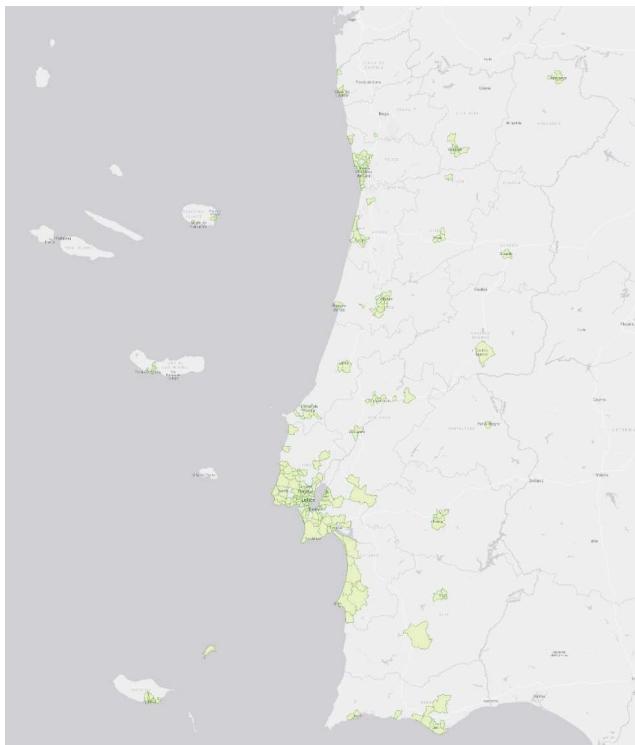
## » Regional-level data

- wealthy (clusters **A** and **B**)
- non-wealthy (**C** and NA)

[Valente 2019]

## » Universe reduction

- 31.4% (95,978 in 306,129)
- 2.5% (95,978 in 3,946,593)

## » Sub-samples

- Base-case: "wealthy household"

- Regional+: "wealthy household" and "wealthy region"

- Regional–: "wealthy household" and "non-wealthy region"

(on ISFF 2017 sample)

| | n | Income | Financial | Real | Liabilities |
|---|---|---|---|---|---|
| Regional+ | 1,303 | 45k | 60k | 454k | 41k |
| Regional– | 1,739 | 28k | 36k | 262k | 31k |
| Base-case | 3,042 | 35k | 46k | 344k | 36k |
| Total | 5,924 | 31k | 36k | 260k | 29k |

## » Pairwise comparison

- Base-case vs Regional+

- Regional+ vs Regional–

(on Gross income, Financial and Real Assets, and Liabilities)

## » Summary Statistics and Statistical Tests

- Central Tendency and Dispersion

- Inequality, Concentration indexes and $S_x$ ratios

- Lorenz Curves comparison

⌄

» Expectation

$\bar{x}$, $\tilde{x}$: Regional+ ≥ Base-case ≥ Regional–;

$s$: Regional+ ≈ Base-case ≈ Regional–;

$A_\varepsilon$: Regional+ ≥ Base-case ≥ Regional–;

$G$, $LAC$: Regional+ ≈ Base-case ≈ Regional–;

$S_x$:$S_{1-x}$: Regional+ ≥ Base-case ≥ Regional–.

Overall, Regional+ capture wealth, while preserving variation.

⌄
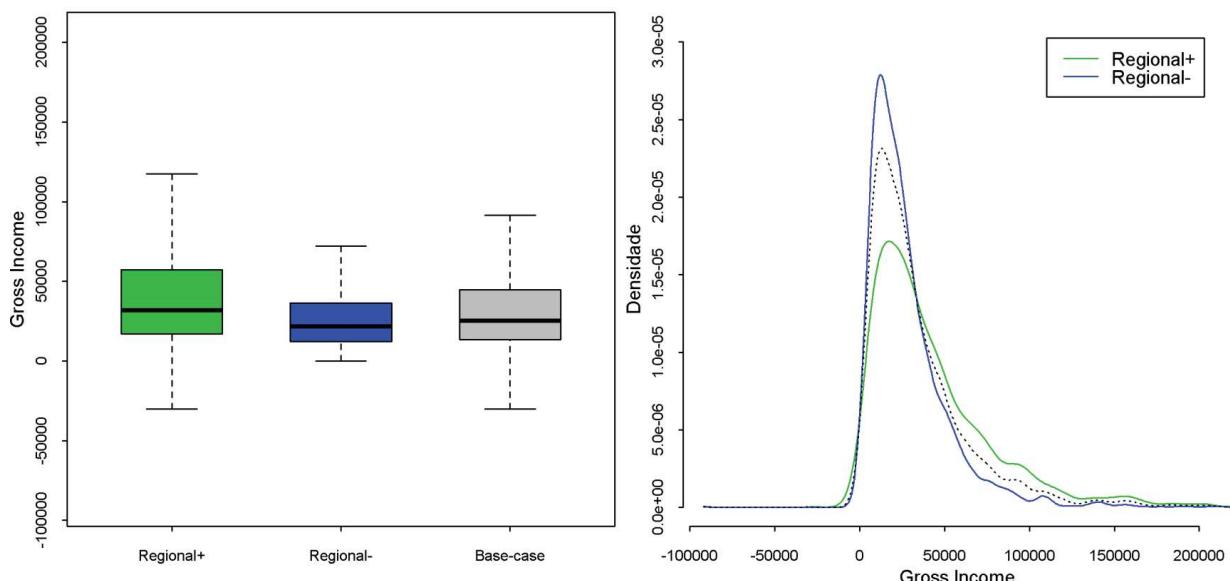
» Gross income
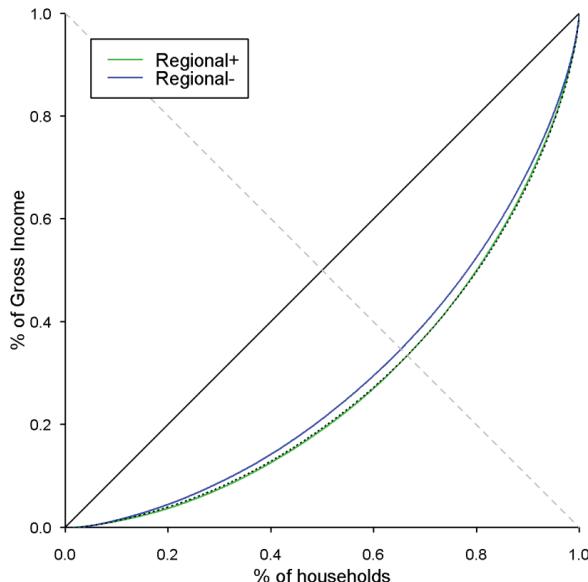
» Gross income



| | $A_1$ | $G$ | $LAC$ |
|---|---|---|---|
| Reg+ | 0.33 | 0.46 | 0.95 |
| Reg− | 0.29 | 0.43 | 0.98 |
| Base | 0.32 | 0.46 | 0.98 |

| | $S80{:}S20$ | $S90{:}S10$ | $P$ |
|---|---|---|---|
| Reg+ | 15 | 28.6 | 10.4 |
| Reg− | 11 | 21.8 | 8.7 |
| Base | 13 | 26.5 | 10.3 |

» Gross income

- retaining households with high incomes;

- keeping the range of households;

- inequality indexes with no substantial difference;

- concentration indexes with somewhat higher values.

» Financial assets
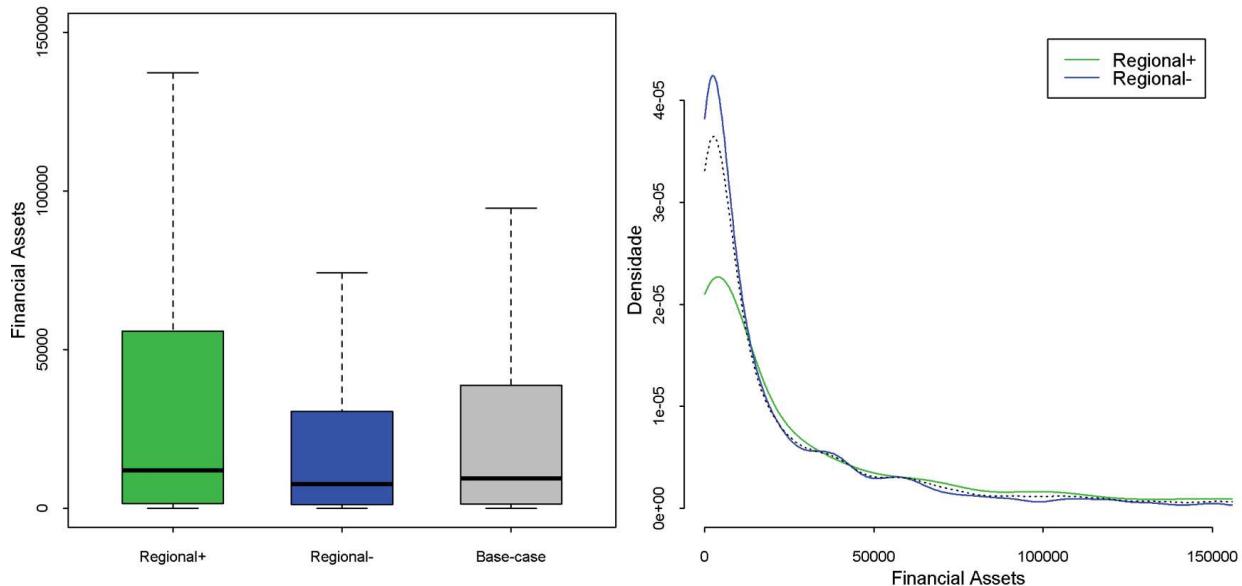
» Financial assets
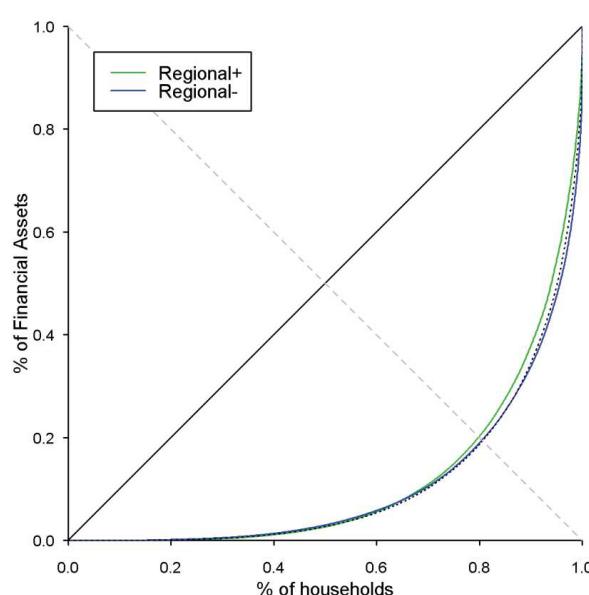


|       | $A_1$ | $G$  | $LAC$ |
|-------|-------|------|-------|
| Reg+  | 0.85  | 0.77 | 0.92  |
| Reg−  | 0.83  | 0.79 | 0.95  |
| Base  | 0.84  | 0.79 | 0.96  |

|       | $S80{:}S20$ | $S90{:}S10$ | $P$   |
|-------|-------------|-------------|-------|
| Reg+  | 697.9       | 5,274.8     | 218.0 |
| Reg−  | 432.2       | 2,321.0     | 192.5 |
| Base  | 589.1       | 3,432.2     | 212.3 |

» Financial assets

- retaining wealthy households;

- keeping the range of households;

- inequality indexes with no substantial difference;

- concentration indexes with considerable higher values.
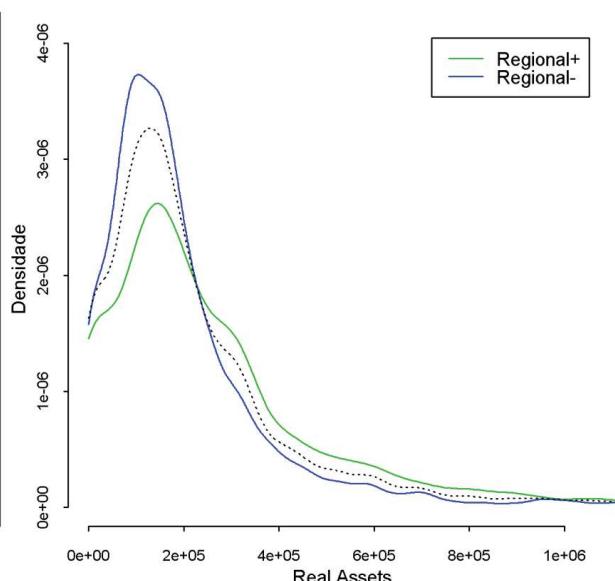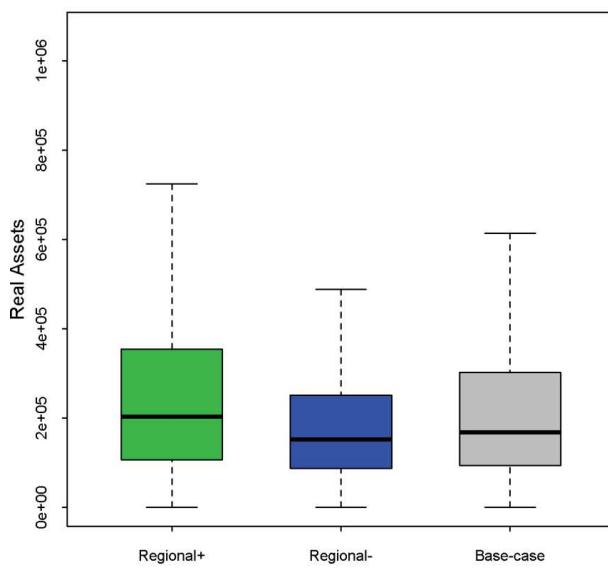
» Real assets
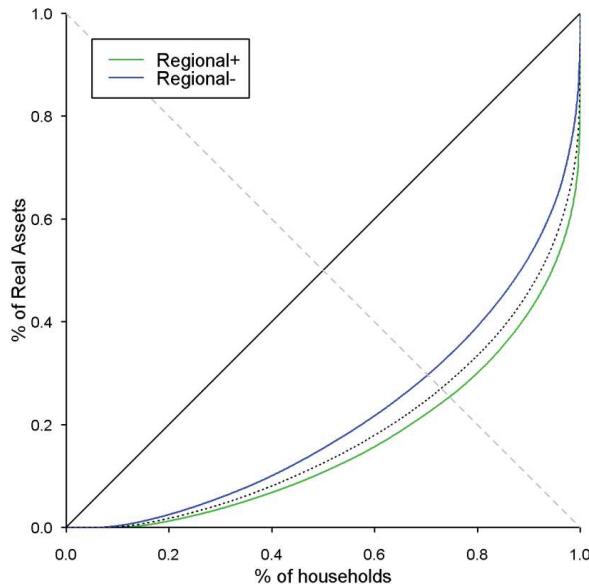
» Real assets



| | $A_1$ | $G$ | $LAC$ |
|---|---|---|---|
| Reg+ | 0.62 | 0.67 | 1.13 |
| Reg− | 0.48 | 0.57 | 1.12 |
| Base | 0.56 | 0.63 | 1.14 |

| | $S80{:}S20$ | $S90{:}S10$ | $P$ |
|---|---|---|---|
| Reg+ | 55.0 | 1,356.5 | 33.9 |
| Reg− | 24.1 | 137.6 | 18.9 |
| Base | 37.2 | 365.6 | 26.5 |

» Real assets

- retaining wealthy households;
- keeping the range of households;
- $A_1$ and $G$ with higher values;
- concentration indexes with considerable higher values.

In Portugal, real-estate account for > 50% of real assets [INE 2019]
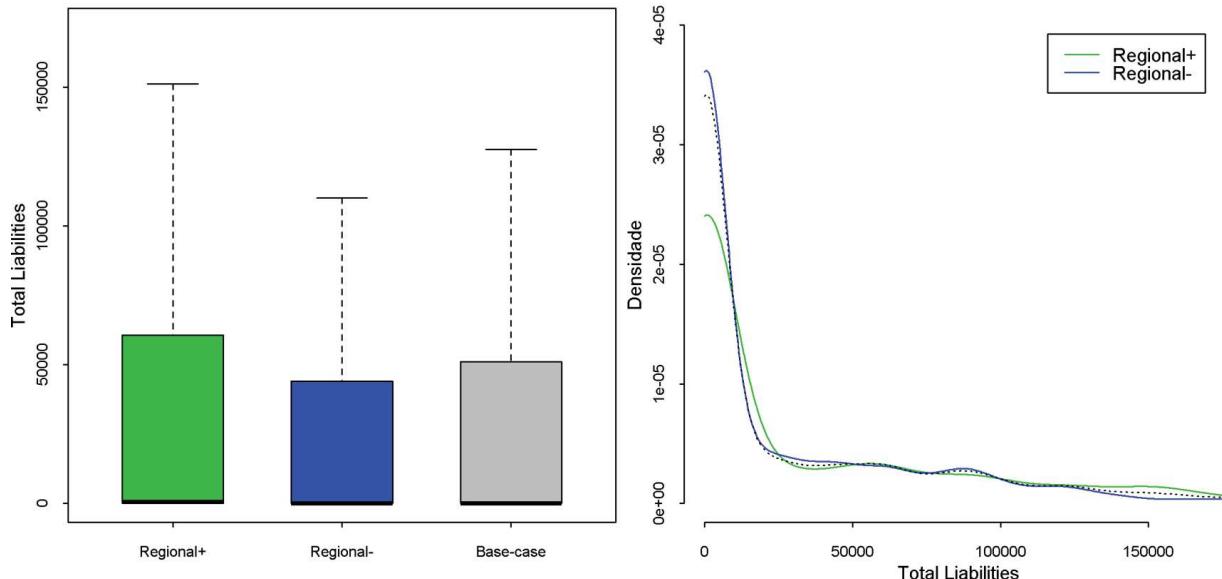
» Liabilities
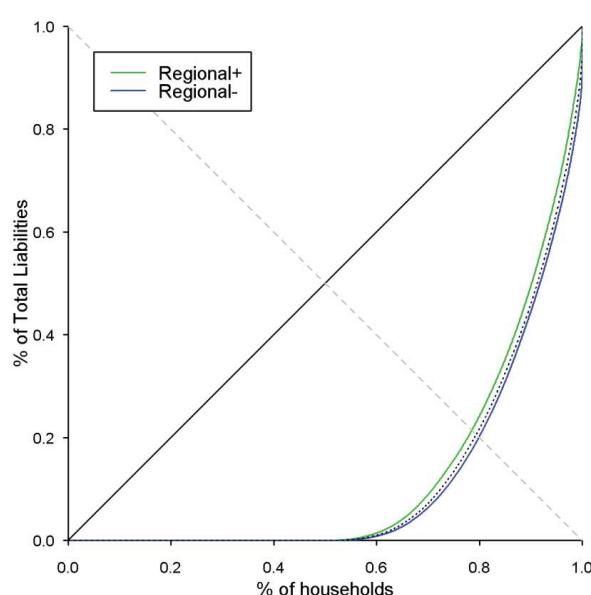
» Liabilities



|  | $A_1$ | $G$ | $LAC$ |
|---|---|---|---|
| Reg+ | 0.54 | 0.75 | 0.75 |
| Reg− | 0.56 | 0.78 | 0.78 |
| Base | 0.56 | 0.77 | 0.77 |

|  | $S80{:}S20$ | $S90{:}S10$ | $P$ |
|---|---|---|---|
| Reg+ | – | – | – |
| Reg− | – | – | – |
| Base | – | – | – |

## » Liabilities

- retaining households with high liabilities;

- keeping the range of households;

- inequality indexes with no substantial difference;

- concentration indexes not considered.

… but relation between wealth and liabilities is complex!

## » Main bullet-points

- Use information on Income and Real-estate (regional level);

- Take conservative approach (keep wide sample universe);

- Retaining wealthy households;

- Consistent shift-to-the-right while keeping wide range;

- Better representation of wealthier households (Real assets);

- Reduce sample effort and decrease cost (2.5% of universe).

Alvaredo F, Atkinson AB, Piketty T and Saez E. 2013. "The Top 1 Percent in International and Historical Perspective." *J Econom Persp* 27(3): 3–20.

HFCN. 2009. "Survey Data on Household Finance and Consumption: Research Summary and Policy Use." *ECB Occas Pap Ser* 100.

HFCN. 2013. "The Household Finance and Consumption Survey: methodological report for the first wave." *ECB Stat Pap Ser* 1.

HFCN. 2016 "The Household Finance and Consumption Survey: methodological report for the second wave." *ECB Stat Pap Ser* 17.

INE. 2017. "Inquerito a Situacao Financeira das Familias." *INE Doc Met* 130 (1.2).

Kennickell AB. 2005. "Darkness made visible: Field management and nonresponse in the 2004 SCF." In Annual Joint Statistical Meetings, August, 2005. Minneapolis, MN: American Statistical Association.

Kennickell AB. 2008. "The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances." In 56th International Statistical Institute Session, August 22-29, 2008. Lisbon, PT: International Statistical Institute.

Valente CF. 2019. *Identificação e delimitação de territórios homogéneos de residentes com níveis de rendimento e património mais elevados.* Lisbon: ISEG. (DM-CIFV No. 10400.5/19065).

## » Clusters of regions [Valente 2019]

- Hierarchical clustering (Ward method, Euclidean distance)
- 3 clusters of 907 regional units

|  | A | B | C |
|---|---|---|---|
| income (€) | 32k | 29k | 19k |
| income (%) | 27.89 | 26.19 | 13.56 |
| real estate (€) | 2.8k | 1.3k | 0.9k |
| real estate (%) | 66.41 | 8.61 | 3.27 |
| cluster size | 54 | 153 | 700 |

## » Clusters of regions [Valente 2019]

- Hierarchical clustering (Ward method, Euclidean distance)
- 3 clusters of 907 regional units

|  | A | B | C |
|---|---|---|---|
| $n$ | 484,387 | 1,068,422 | 2,375,183 |
| $n_{wealthy}$ | 26,917 | 69,061 | 208,373 |
| cumsum | 484,387 | 1,552,809 | 3,927,992 |
| $cumsum_{wealthy}$ | 26,917 | 95,978 | 304,351 |

NA: 18,601 households (1,778 wealthy households)

**» Defining wealthy regions**

- Robust definition of wealthy

- Robust sampling with enough available households

**» Sampling procedure concerns**

- Pool of PSU: other surveys deplete available households;

- PSU: low density units can be problematic;

- SSU: overestimation of availability (outdated information);

- Longitudinal studies: assure availability for several waves.