BANCA D'ITALIA
EUROSISTEMA

COMBINING SURVEY AND ADMINISTRATIVE DATA TO ESTIMATE THE DISTRIBUTION OF HOUSEHOLDS' DEPOSITS

Andrea Neri
Matteo Spuri
Francesco Vercelli

IARIW-Bank of Italy Conference
Naples, Italy
March 31, 2023

Motivation

- Aggregates on **household wealth** obtained from surveys are generally lower than the corresponding macroeconomic figures.
- Broad literature on this topic – e.g. D'Alessio 1990, D'Aurizio 2008, Neri 2012.
- **ESCB** methodology - understand, quantify and explain these differences (Distributional Wealth Accounts, **DWA**).
- Concerning **deposits**, an **alternative methodology** exploiting administrative data available in Italy is proposed.

- ESCB shared methodology
- BoI extensions to ESCB methodology
  - Highly reliable detection and imputation
  - Calibration correction
- Conclusion

- ESCB shared methodology
- BoI extensions to ESCB methodology
  - Highly reliable detection and imputation
  - Calibration correction
- Conclusion

DWA: data sources and main goals

- Compile **quarterly distributional results** on **household wealth**.

- Based on microeconomic survey data and macroeconomic data:

Survey data	Macro-aggregates
• Household Finance and Consumption Survey (HFCS)	• Quarterly Sector Financial Accounts (QSA) • Annual Sector Balance Sheets (Non-financial assets)

- Provide **distributional information** on households wealth:
  - Median and mean of net wealth and its components
  - Share of wealth of top 5%, 10% and bottom 50%
  - **Gini Index** (net wealth)

DWA: compilation steps

1. Reconcile definitions and concepts between the **HFCS** and National Accounts (**NAs**).

2. Align HFCS and NAs totals when the survey is available:
   1. Correction of households' **deposits**
   2. Coverage of **wealthiest households** - addition of **Forbes' rich list** and **synthetic households** from the estimated **Pareto** distribution
   3. **Proportional allocation** to account for the remaining gap

3. **Interpolate** distributional results between surveys and **extrapolate** to the most recent QSAs.

Each country can improve such methodology exploiting available data.

DWA: compilation steps

1. Reconcile definitions and concepts between the **HFCS** and National Accounts (**NAs**).

2. Align HFCS and NAs totals when the survey is available:
   1. Correction of households' **deposits** ← **Extension 1: Detection and imputation**
   2. Coverage of **wealthiest households** - addition of **Forbes' rich list** and **synthetic households** from the estimated **Pareto** distribution
   3. **Proportional allocation** to account for the remaining gap

   ← **Extension 2: Calibration correction**

3. **Interpolate** distributional results between surveys and **extrapolate** to the most recent QSAs.

These extensions are not mutually exclusive nor required jointly

- ESCB shared methodology

- BoI extensions to ESCB methodology
  - Highly reliable detection and imputation
  - Calibration correction

- Conclusion

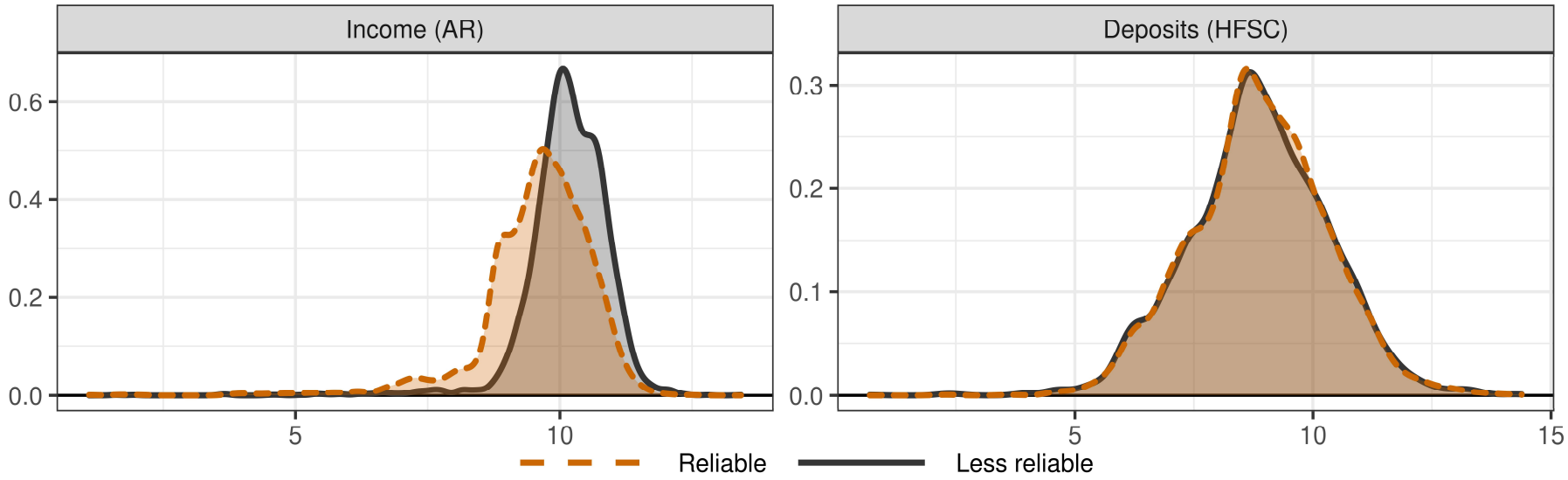
Comparison with ESCB base method and data sources

	ESCB Base methodology	Alternative BoI methodology
Highly reliable detection	Income criterion and/or asset criterion	Survey responses vs administrative records
Deposits Imputation	Average values (by income groups)	Model based (linear regression)

- **Administrative Records (AR)** data linked to **HFCS** through fiscal identifiers:
  - Fiscal income from the tax register
  - Housing wealth from the cadastral records
  - Debts from Bank of Italy's Credit Register

Highly reliable households detection

- **Highly reliable households** identified as: HFCS income > 0.95*AR income.



- **Robustness checks** using 6 alternative definitions of highly reliable households.

Regression based imputation

- Regression on the **sample of highly reliable respondents**.
- Models valuated through the **out-of-sample prediction**.
- Estimate **deposit** values for **less reliable respondents**.
- **Robustness checks** comparing linear regression with **hurdle models**.

Available variables	Model 6
income (AR) (log)	0.160***
real estate (AR) (log)	0.026***
financial assets (excl. deposits) (log)	0.019*
loans (AR) (log)	0.007
wages (AR) (log)	-
pensions (AR) (log)	-
self-employed income, profits, rents (AR) (log)	-
expenditures using banknotes (log)	0.212***
durables (log)	0.035***
non-durable consumption (log)	0.562***
overdraft credit (log)	-0.073**
credit card debt (log)	0.007
savings (log)	-
age of the head of the household	0.013
age of the head of the household (squared)	-0.000
Adjusted R^2	0.209

This model also includes occupation, household composition and residence

Impact on final statistics (net wealth)

Model	Deposits Coverage		Top 5%	Top 10%	Top 20%	Bottom 50%	Gini
	Before deposits adjustment	After deposits adjustment					
ESCB Base model	**34.2%**	**44.7%**	50.7%	60.7%	73.3%	7.1%	**72.3%**
Linear Regression imputation	**34.2%**	**49.8%**	50.1%	60.0%	72.6%	7.5%	**71.6%**

- **Deposits coverage** increases significantly.
- **Gini index** reduces, but not substantially.
- Distribution across percentiles is slightly different.

Calibration data sources

- Exploitation of Italian **Bank Supervisory Reports (BSR)** data.
- Includes **outstanding amounts** of deposits by asset range.

< 12.5k	12.5-50k	50-250k	250-500k	>500k

- Comparison between **BSR** and **HFCS** data is not straightforward (different unit of observation) and requires some adjustments.
- Households with **more than one deposit** account are attributed with **66%** of their total deposits on the **main one** (evidence from 2020 survey).

Calibration methodology

- Deposits are adjusted solving the following **multivariate calibration**:

$$\min_{\boldsymbol{a}} \sum_{i=1}^{n} \frac{(w_i a_i - w_i)^2}{w_i}$$

$$s.t. \sum_{i=1}^{n} w_i \cdot (\boldsymbol{a_i} \cdot \boldsymbol{x_i}) \cdot I_{i,c} = X_c$$

$$a_i \in [min_a, max_a] \; \forall i \in \{1, \dots, n\}$$

a_i: **adjustment parameter** for account *i*

x_i: **deposits** in account *i*

w_i: **weights** of household holding account *i*

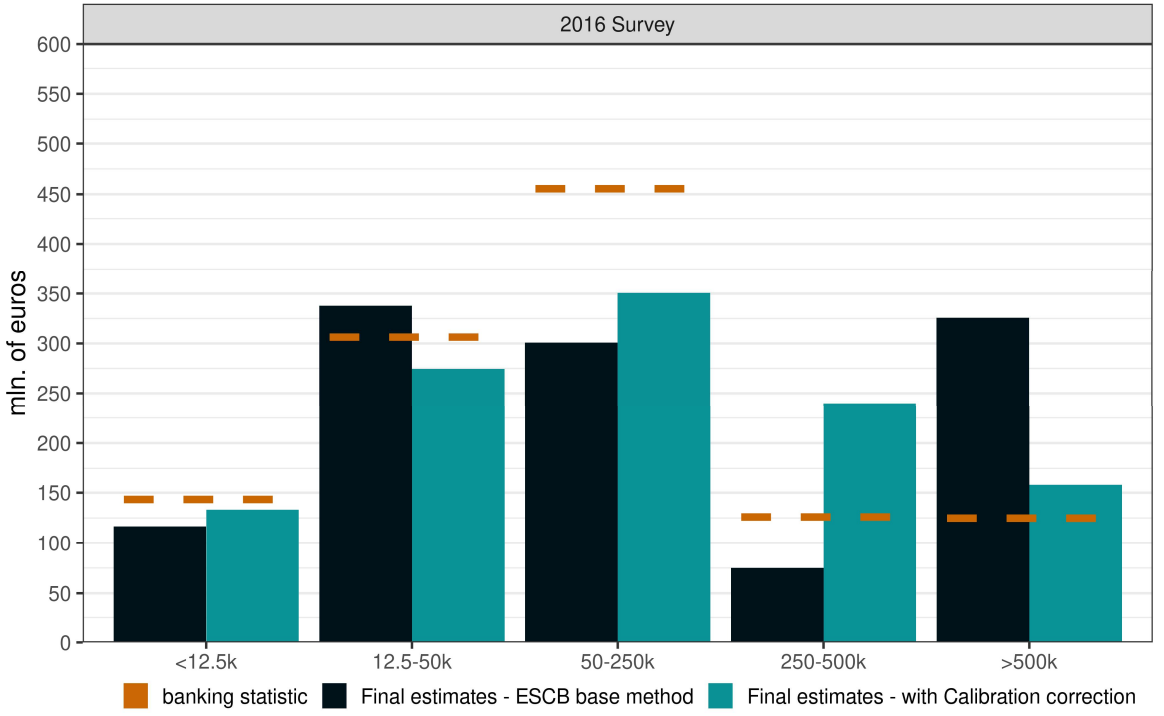
X_c: **aggregate amount** in class *C* from BSR

$I_{i,c}$: indicator function for account *i* belonging to class *C*

Calibration impact

Model	Top 5%	Top 10%	Bottom 50%	Gini
ESCB Base methodology	50.7%	60.7%	7.1%	**72.3%**
Alternative methodology	49.6%	60.0%	6.8%	**72.4%**

- The calibration reduces the gap with respect to aggregate BSR data.
- The **wealth share** decreases for both top 5% and bottom 50%.
- The **Gini index** is almost unaffected.



- ESCB shared methodology
- BoI extensions to ESCB methodology
  - Highly reliable detection and imputation
  - Calibration correction
- Conclusion

Summary of our proposal

Outlier detection and imputation

- **Administrative Records** microdata
- Different definition of **reliable** households
- Deposits imputation through **linear regression model**
- Impact on **deposits coverage** and **Gini Index**

Calibration correction

- **Bank Supervisory Reports (BSR)** aggregated data
- **Adjustment parameter** applied to the deposits value
- Deposits distribution more in line with BSR information

Final remarks

- Usage of administrative data is not straightforward and requires several **adjustments** and **assumptions**.

- Main results of the proposed methodology:
  - Increase of coverage
  - Overall **distribution** more in line with the administrative data one.
  - Change in **inequality statistics** (non substantially with respect to ESCB method).

Next steps

- Further robustness checks on the identification of **highly reliable households**.
- **Explore alternative penalization methods** on the calibration technique – e.g. Ridge, Lasso.
- Extend the calibration correction to the **quarterly microdata**.
- Explore the usage of BSR on **other instruments** – e.g. debt securities, listed shares, mutual fund shares.

THANK YOU FOR YOUR ATTENTION

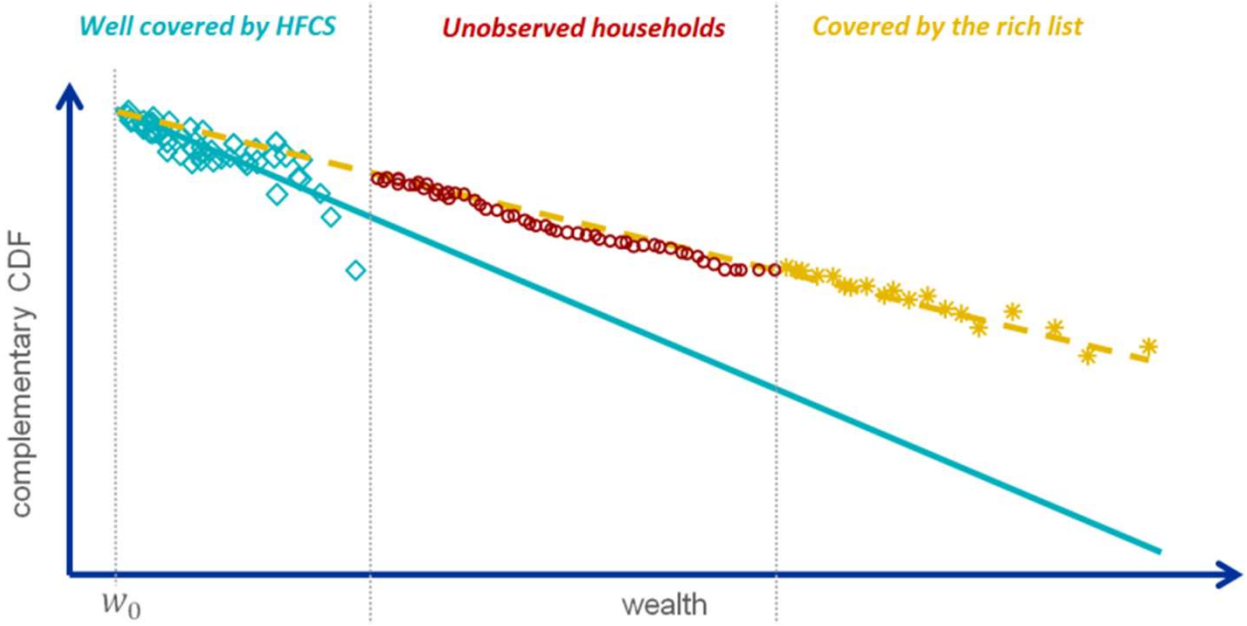
BANCA D'ITALIA

EUROSISTEMA

- ESCB shared methodology
- BoI extensions to ESCB methodology
  - Highly reliable detection and imputation
  - Calibration correction
- Conclusion
- Supplementary material

Rich list and Pareto curve estimation

- Complementing **HFCS** sample with “rich lists” improves estimation.
- **Large gap** remains between richest HFCS household and poorest in rich list.
- **Synthetic households** sampled from the estimated **Pareto tail**.



Highly reliable detection – alternative definition

No.	Definition	No. of obs.
1	**SHIW income > 0.95*AR income**	**3,069**
2	SHIW income > 0.90*AR income	3,756
3	SHIW no. of properties > (AR no. of properties) - 100 (perc. points)	3,226
4	SHIW income > 0.90*AR income and meeting the ECB income criterion	3,524
5	SHIW income > 0.90*AR income and meeting the ECB income and asset criterion	3,121
6	meeting both the ECB income and asset criteria	5,663
7	meeting at least 2 criteria: income (2); no. of properties (3); not an ECB's outlier (6)	4,403

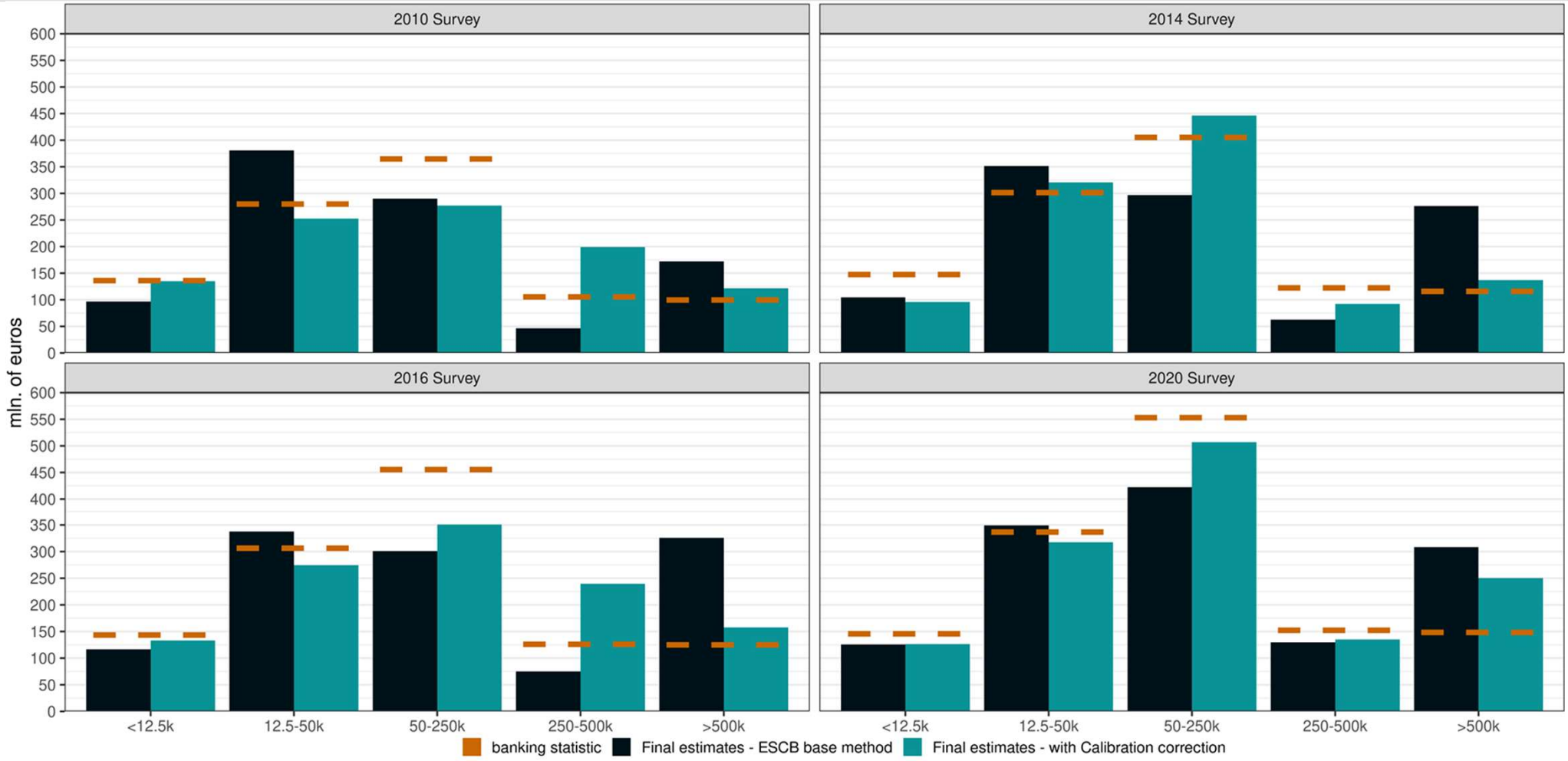
Regression coefficient estimates

	(1) deposits (log) b/se	(2) deposits (log) b/se	(3) deposits (log) b/se	(4) deposits (log) b/se	(5) deposits (log) b/se	(6) deposits (log) b/se
income (AR) (log)	0.273***		0.147***		0.229***	**0.160*** **
real estate (AR) (log)	0.036***	0.041***	0.030***	0.031***	0.033***	**0.026*** **
financial assets (excl. deposits) (log)	0.037***	0.046***	0.024**	0.024**	0.036***	**0.019* **
loans (AR) (log)	0.004	0.004	0.002	0.002	0.003	**0.007**
wages (AR) (log)		0.016		-0.004		
pensions (AR) (log)		0.045***		0.034***		
self-employed income, profits, rents (AR) (log)		0.047***		0.014		
expenditures using banknotes (log)			0.218***	0.188***		**0.212*** **
durables (log)			0.033***	0.040***		**0.035*** **
non-durable consumption (log)			0.432***	0.588***		**0.562*** **
Overdraft credit (log)			-0.066**	-0.060**	-0.053*	**-0.073** **
Credit card debt (log)			0.005	0.008	0.009	**0.007**
savings (log)					0.042***	
age of the head of the household						**0.013**
age of the head of the household (squared)						**-0.000**
Constant	5.832***	7.862***	1.381*	1.204	6.000***	**-0.680**
10-fold CV RMSE (ave)	**1.322**	**1.326**	**1.304**	**1.296**	**1.313**	**1.285**
Adjusted R^2	**0.106**	**0.091**	**0.151**	**0.156**	**0.119**	**0.209**

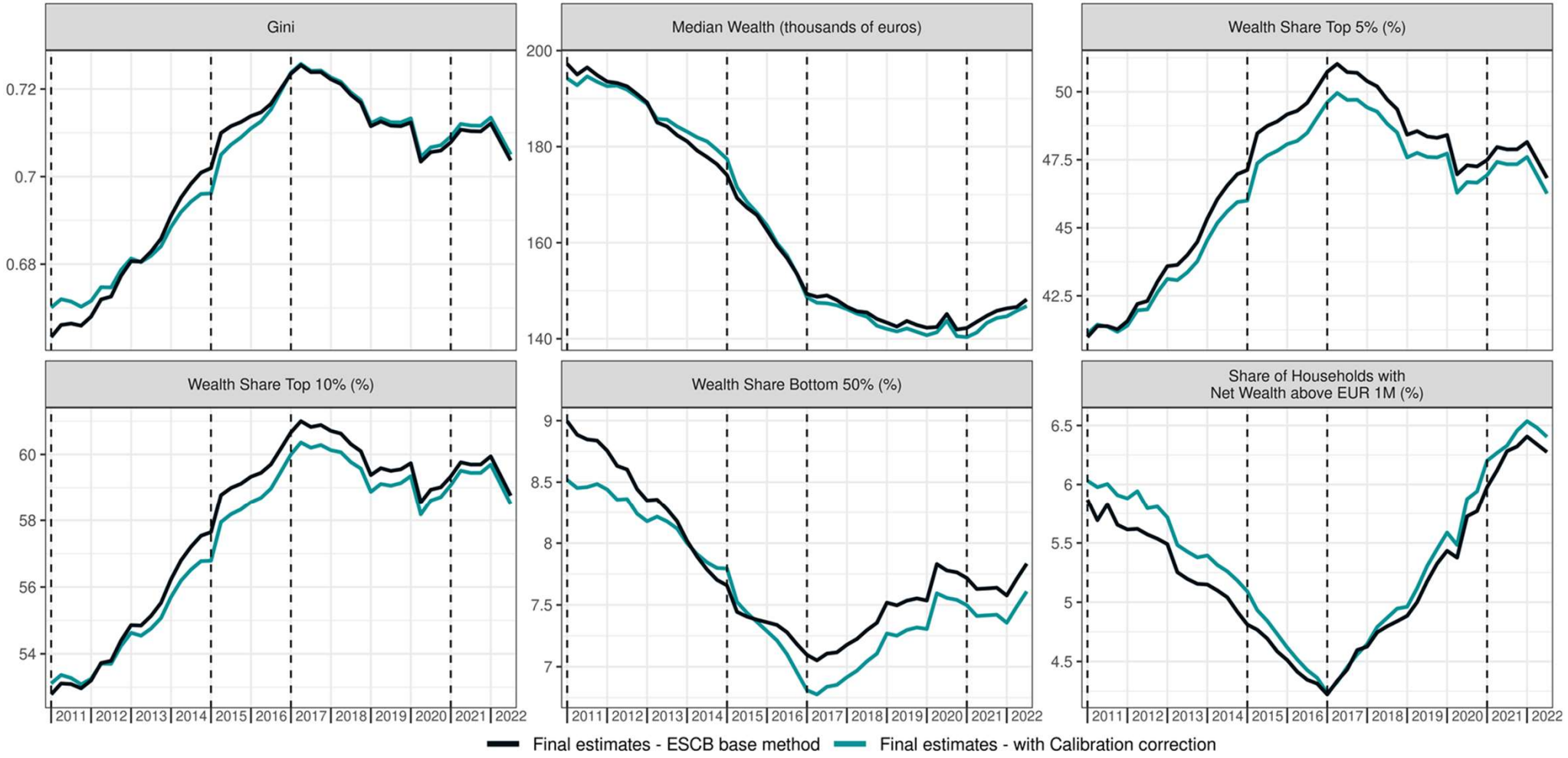
Amounts and percentage of deposits by asset range

Year	<12.5k	12.5-50k	50-250k	250-500k	>500k	Total	<12.5 (%)	12.5-50k (%)	50-250k (%)	250-500k (%)	>500k (%)	Total (%)
2013	132,736	272,734	359,55	77,075	74,114	916,209	14.5	29.8	39.2	8.4	8.1	100.0
2014	133,876	274,093	373,384	84,263	81,400	947,016	14.1	28.9	39.4	8.9	8.6	100.0
2015	132,967	274,802	388,707	88,438	86,820	971,734	13.7	28.3	40.0	9.1	8.9	100.0
2016	131,213	280,656	426,778	91,742	92,444	1,022,833	12.8	27.4	41.7	9.0	9.0	100.0
2017	132,199	283,860	440,306	94,304	95,544	1,046,213	12.6	27.1	42.1	9.0	9.1	100.0
2018	132,166	287,876	454,984	98,629	99,779	1,073,434	12.3	26.8	42.4	9.2	9.3	100.0
2019	130,400	292,579	485,909	109,835	112,125	1,130,848	11.5	25.9	43.0	9.7	9.9	100.0
2020	137,054	319,605	528,516	116,808	114,740	1,216,723	11.3	26.3	43.4	9.6	9.4	100.0

Calibration impact – BSR data comparison



Calibration impact – Inequality indicators



Overall impact