

# What's Wrong with Survey-based Top Wealth Shares: Evidence from Housing Wealth of French Households

---

Olivier MESLIN<sup>1</sup>

<sup>1</sup>Insee, Statistical Innovation Unit - SSP Lab

November 21, 2025

# The limitations of wealth surveys

Wealth surveys are one of the main statistical sources on wealth, but are unlikely to reflect accurately the distribution and the concentration of wealth due to multiple biases ([Kennickell, 2017], [Eckerstorfer et al., 2016], [Vermeulen, 2018]). Two main categories of hypothesis:

- Underrepresentation of wealthy households;
- Underreporting of assets.

# The limitations of wealth surveys

Wealth surveys are one of the main statistical sources on wealth, but are unlikely to reflect accurately the distribution and the concentration of wealth due to multiple biases ([Kennickell, 2017], [Eckerstorfer et al., 2016], [Vermeulen, 2018]). Two main categories of hypothesis:

- Underrepresentation of wealthy households;
- Underreporting of assets.

The true causes of biases in wealth surveys remain unclear, often because of a lack of a benchmark measure ([Meyer et al., 2015]).

Insee has been developing a new administrative database describing the housing wealth of French households ([André and Meslin, 2021] and [André and Meslin, 2025]).

Insee has been developing a new administrative database describing the housing wealth of French households ([André and Meslin, 2021] and [André and Meslin, 2025]).

In this paper, I combine the 2017 French wealth survey with this database and other administrative sources to shed light on the causes of biases in wealth surveys.

# What is exactly the problem?

Data source	Share in the population					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
Administrative database	50.0%	25.0%	15.0%	5.0%	4.0%	1.00%
Survey data (true quantiles)	45.2%	29.1%	16.8%	4.9%	<b>3.5%</b>	<b>0.53%</b>

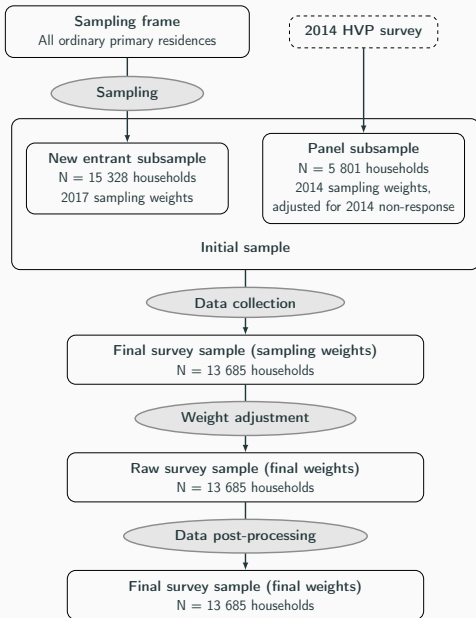
Source	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
Administrative database	6.5%	24.6%	26.0%	14.1%	18.2%	10.8%
Survey data (survey quantiles)	10.8%	25.5%	25.6%	13.3%	<b>16.1%</b>	<b>8.8%</b>
Survey data (true quantiles)	6.2%	29.1%	28.9%	14.2%	<b>15.6%</b>	<b>6.0%</b>

With comparable definitions, almost half of the top 1% is missing in the survey, both in population share and wealth share.

Three steps:

- I first link sampled households with their counterpart in the administrative database;
- I decompose discrepancies between survey-based estimates and estimates based on administrative data to unveil the sources of bias;
- I finally compare the housing units reported by respondents with the housing units they actually own.

# The French wealth survey



# The French wealth survey: questions on housing wealth

- Households are asked **how many housing units they own, either in full ownership or in bare ownership, either directly or through a SCI** (*société civile immobilière*);  
⇒ housing units owned in usufruct must not be reported (except main residences).
- Collected information:
  - Nature of the housing unit (house or flat);
  - Location (*commune*);
  - Share owned by the household;
  - Year and price of purchase;
  - Use of the housing unit (owner-occupied, leased to tenants, vacant);
  - Estimate of the current market value.
- Several flats can be described jointly under two conditions: they are located in the same building and used for the same purpose (eg, all are leased to tenants).  
⇒ Households actually report housing assets, not individual housing units.

Database built for 2017 following a two-step methodology:

- Reconstituting the list of all real estate assets owned by resident households.  
⇒ Systematic linkage of cadastral data, commercial register data and real estate transactions data with the Fideli administrative database (describing all residents households and individuals)  
[André and Meslin, 2021]

Database built for 2017 following a two-step methodology:

- Reconstituting the list of all real estate assets owned by resident households.  
⇒ Systematic linkage of cadastral data, commercial register data and real estate transactions data with the Fideli administrative database (describing all residents households and individuals)  
[André and Meslin, 2021]
- Estimating the market value of all housing units located in France.  
⇒ Statistical model based on machine learning algorithms.  
[André and Meslin, 2025]

## Linking survey data with the administrative database

The linking procedure is highly accurate and the composition of households is very similar in both sources.

# Linking survey data with the administrative database

The linking procedure is highly accurate and the composition of households is very similar in both sources.

The linked data contains:

- Information on assets and wealth reported by respondents;  
⇒ **Reported outcomes**
- Information on assets and wealth from the administrative database for all sampled households (including non-respondents).  
⇒ **"True" outcomes** (although administrative data is far from perfect!)

# Linking survey data with the administrative database

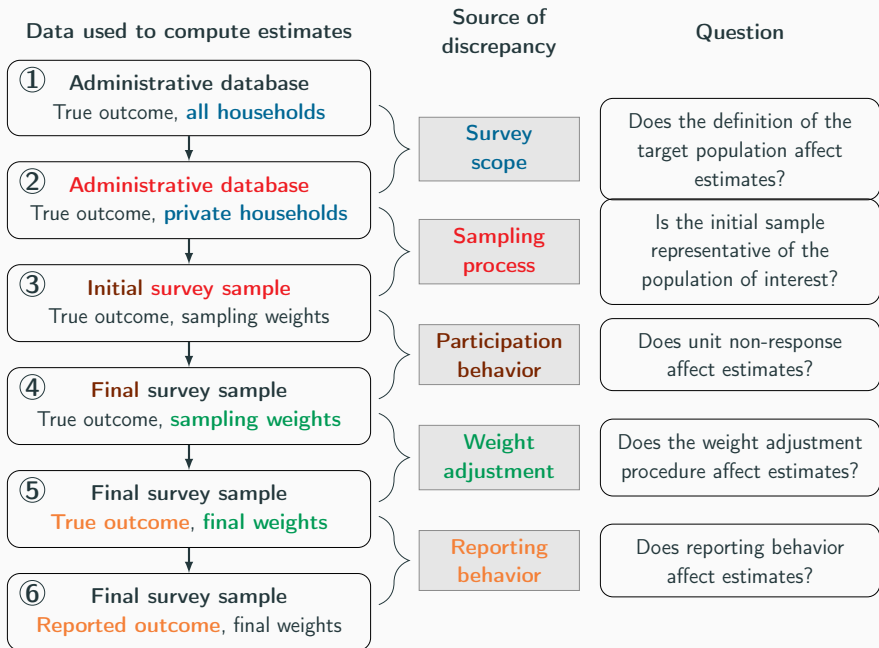
The linking procedure is highly accurate and the composition of households is very similar in both sources.

The linked data contains:

- Information on assets and wealth reported by respondents;  
⇒ **Reported outcomes**
- Information on assets and wealth from the administrative database for all sampled households (including non-respondents).  
⇒ **"True" outcomes** (although administrative data is far from perfect!)

Key advantages:

- the administrative database is close to the sampling frame of the survey;
- the definitions of housing assets and wealth are exactly the same;
- the definition of households is the same.



# Decomposing discrepancies: wealth concentration

	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
① Administrative data	6.5%	24.6%	26.0%	14.1%	18.2%	10.8%
⑥ Survey data	6.2%	29.1%	28.9%	14.2%	15.6%	6.0%
<b>Total discrepancy</b>	<b>-0.2</b>	<b>4.5</b>	<b>2.9</b>	<b>0.1</b>	<b>-2.5</b>	<b>-4.8</b>
Scope of survey	-0.1	0.0	0.1	0.0	0.0	0.0
Sampling process	-0.2	0.0	-0.3	0.7	-0.2	0.0
Participation behavior	0.1	1.8	1.1	0.7	-0.8	-2.9
Weight adjustment	0.2	-0.4	-0.8	-0.1	-0.4	1.6
Reporting behavior	-0.2	3.2	2.8	-1.3	-1.1	-3.4

Two main problems: underrepresentation of wealthy households and underreporting of assets.

The discrepancy due to unit non-response may be due to three causes:

- The household/housing unit is **out of the survey scope** (eg: vacant housing units);  
⇒ This is usually not considered as unit non-response.
- The household is in the survey scope but **could not be reached**;
- The household could be reached but **refused to participate**.

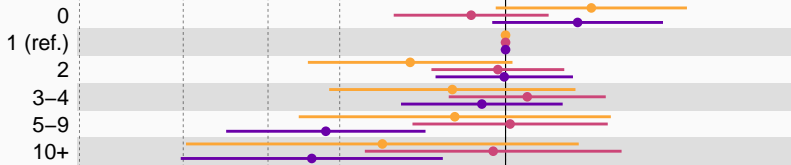
# Investigating households' participation behavior

	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
③ Initial sample	6.2%	24.5%	25.7%	14.8%	18.0%	10.8%
④ Final sample	6.3%	26.3%	26.9%	15.5%	17.1%	7.8%
<b>Effect of participation behavior</b>	<b>0.1</b>	<b>1.8</b>	<b>1.1</b>	<b>0.7</b>	<b>-0.8</b>	<b>-2.9</b>
Out of survey scope	-0.1	0.2	0.0	0.1	0.0	-0.1
Could not be reached	0.1	0.8	0.6	0.1	-0.7	-0.8
Refused to participate	0.2	0.8	0.6	0.5	-0.1	-2.0

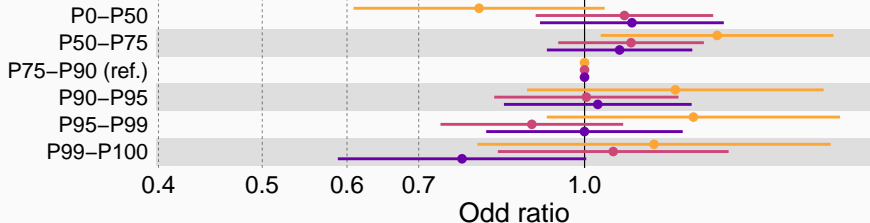
## Type of household



## Number of housing units owned by the household



## Position in the housing wealth distribution



● In-scope rate ● Contact rate ● Participation rate

Findings on unit non-response:

- **The contact rate decreases with total wealth.**  
⇒ Wealthier households seem more difficult to reach.
- **The participation rate decreases with total wealth, housing wealth and the number of housing units owned.**  
⇒ Wealthy households are specifically reluctant to participate.  
⇒ The non-response is not ignorable, even conditionally.
- **Consequence: small underrepresentation of wealthy households, but significant downward bias in top wealth shares.**  
⇒ Confirmation of [Kennickell, 2019]'s hypothesis.

# Investigating respondents' reporting behavior

I answer three questions on the reporting behavior:

- Where do the reporting biases come from?
  - ⇒ Mostly from underreporting in the number of housing assets.
  - ⇒ The top 1% cumulates all flavours of underreporting.
- What housing units are reported?
  - ⇒ Households report housing units they have daily economic control upon.
- What happens at the top of the distribution?
  - ⇒ I compare housing wealth reported in the survey with housing wealth tax returns (IFI).
  - ⇒ Underreporting increases with housing wealth.

# Investigating the reporting behavior

The reporting behavior modifies estimates through four channels:

- **Price effect:** the market value of housing assets reported by households may be inaccurate;
- **Quantity effect:** the number of reported housing units may be inconsistent with administrative data;
  - **Intensive margin:** the average number of **housing units per housing asset** may be inconsistent with administrative data;
  - **Extensive margin:** the number of reported **housing assets** may be inconsistent with the number recorded in administrative data;
- **Households reshuffling effect:** a household can be re-classified in a different group based on reported wealth.

# Investigating the reporting behavior

	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
⑤ Final sample, true wealth	6.4%	25.9%	26.1%	15.5%	16.7%	9.4%
⑥ Final sample, reported wealth	6.2%	29.1%	28.9%	14.2%	15.6%	6.0%
<b>Effect of reporting behavior</b>	-0.2	3.2	2.8	-1.3	-1.1	-3.4
Average price of housing unit	0.1	0.4	-0.3	0.2	-0.1	-0.4
Number of housing units	8.9	0.9	-1.3	-1.8	-3.5	-3.0
<i>Number of h.h./h.a.</i>	-0.1	-0.1	0.9	0.0	-0.1	-0.6
<i>Number of housing assets</i>	9.0	0.9	-2.3	-1.8	-3.4	-2.5
Reshuffling of households	-9.2	1.9	4.4	0.4	2.5	0.0

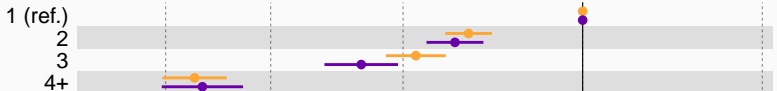
### Ownership right held by the household



### Ownership intermediation through an SCI



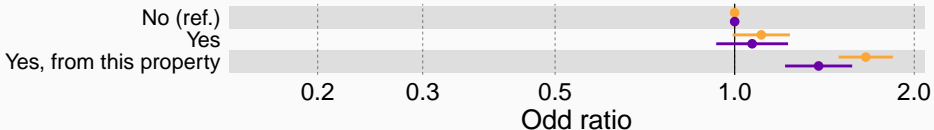
### Number of households owning the property



### The recipient of property tax belongs to the household



### The household earns rental income



● All housing units ● All housing assets

## Type of housing unit

House (ref.)  
Flat

## When did the household become the owner of the housing unit

1979 and before  
1980–1989  
1990–1999  
2000–2009 (ref.)  
2010 and after

## Usage of housing unit

Main residence (ref.)  
Secondary  
Rental  
Vacant  
Professional

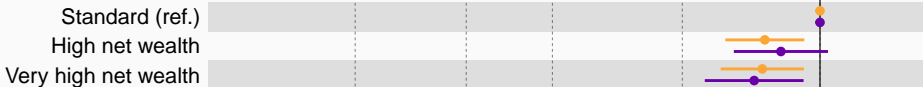
## Number of housing units owned by the household in the built-up area

1 (ref.)  
2  
3  
4  
5–9  
10+

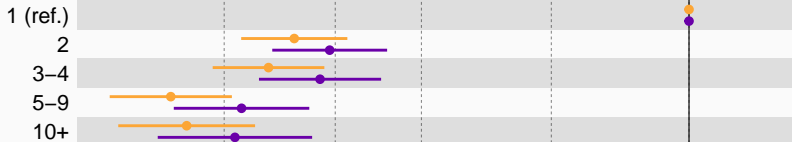
0.2 0.3 0.5 1.0 2.0  
Odd ratio

● All housing units ● All housing assets

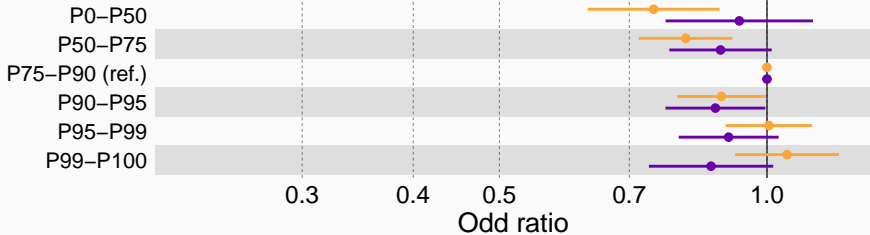
## Type of household



## Number of housing units owned by the household



## Position in the housing wealth distribution



● All housing units ● All housing assets

## What properties are reported, and why?

- ⇒ Households are more likely to report housing units they have daily economic control upon.
- ⇒ Reporting is better among households with better information and higher educational level.
- ⇒ Housing assets/housing units are less likely to be reported when the household owns a large number of housing units and is high-net-wealth.

## What happens at the top of the distribution?

Idea: Let's compare the housing wealth reported by respondents with housing wealth tax returns.

# What happens at the top of the distribution?

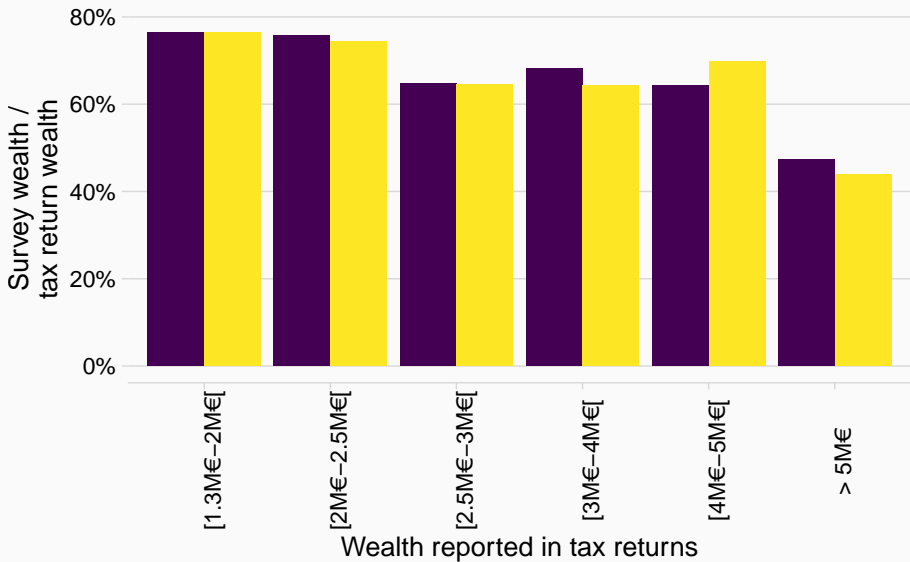
Idea: Let's compare the housing wealth reported by respondents with housing wealth tax returns.

Let's compare the two measures reported by households:

- I match respondents with their housing wealth tax returns (IFI) in 2018.  
⇒ Small subsample of wealthy households ( $N = 877$ ).
- I compute the ratio between the housing wealth reported in the survey and the housing wealth reported to the tax administration.

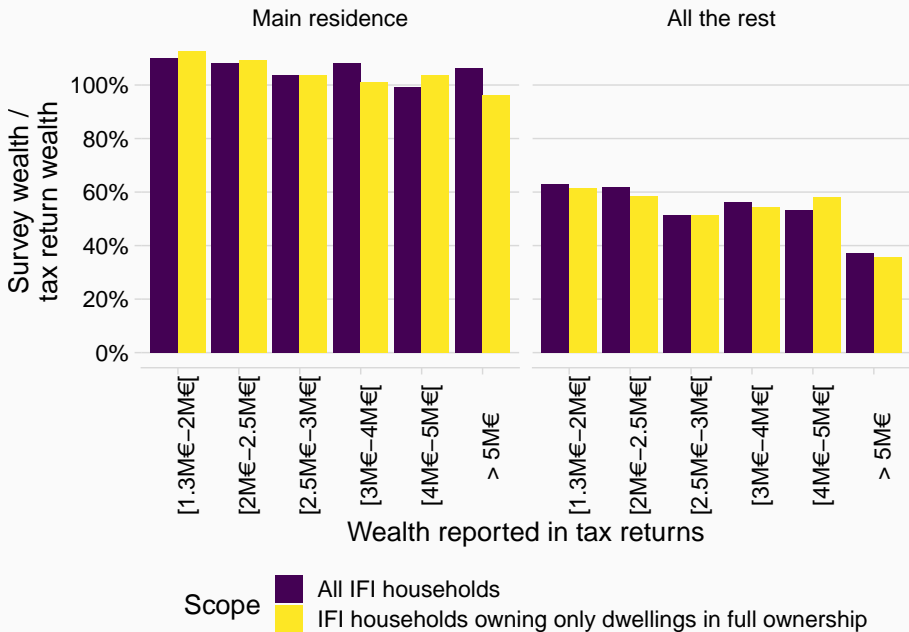
Key advantage: both estimates of housing wealth are reported by households, roughly at the same date, so that two sources of discrepancies can be safely assumed away:

- Unobserved asset quality;
- Mismeasurement of the share owned by the households.



Scope

- All IFI households
- IFI households owning only dwellings in full ownership



# Conclusion

- The downward bias in housing wealth concentration estimates comes in equal parts from underrepresentation of wealthy households and underreporting of assets;
- Underrepresentation of wealthy households:
  - Households with large housing wealth are specifically reluctant to participate to the survey  $\Rightarrow$  Non-ignorable non-response;
- Underreporting of assets:
  - Households report dwellings they have daily economic control upon;
  - Large downward bias in the reported number of dwellings;
  - Underreporting is specifically strong at the top of the distribution.



André, M. and Meslin, O. (2025).

**Le bonheur est dans le prix : Estimation du patrimoine immobilier brut des ménages sur données administratives.**

document de travail 2025-04, Insee.



André, M. and Meslin, O. (2021).

**Et pour quelques appartements de plus: Étude de la propriété immobilière des ménages et du profil redistributif de la taxe foncière.**

document de travail 2021-04, Insee.



Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F., and Wildauer, R. (2016).

**Correcting for the missing rich: An application to wealth survey data.**

*Review of Income and Wealth*, 62(4):605–627