Understanding the Biases of Wealth Surveys: Evidence from Housing Wealth of French Households

Olivier MESLIN[1]

[1]Insee, Statistical Innovation Unit - SSP Lab

June 5, 2024

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]).

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.

This paper

Insee has been developing a new administrative database describing the housing wealth of French households ([André and Meslin, 2021] and [André and Meslin, 2024]).

$\Rightarrow$ **We have a benchmark measure to assess the quality of wealth surveys!**

This paper

Insee has been developing a new administrative database describing the housing wealth of French households ([André and Meslin, 2021] and [André and Meslin, 2024]).

$\Rightarrow$ **We have a benchmark measure to assess the quality of wealth surveys!**

In this paper, I combine the 2017 French wealth survey with this database to understand 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	45.2%	29.1%	16.8%	4.9%	**3.5%**	**0.53%**

Source	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
Administrative database	6.5%	24.5%	26.0%	14.1%	18.2%	10.8%
Survey data	6.3%	29.0%	28.9%	14.2%	**15.6%**	**6.0%**

What is exactly the problem?

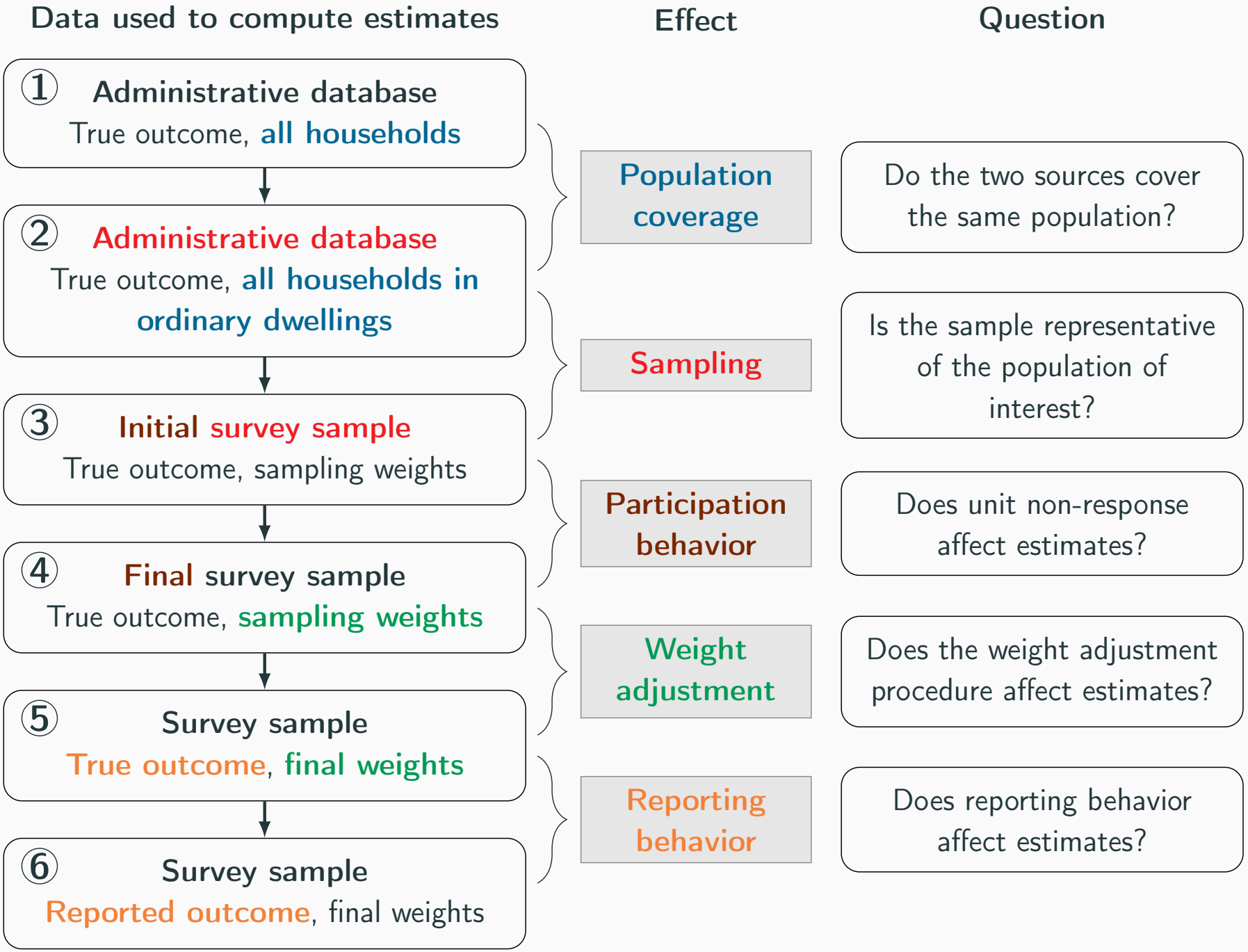
Data source	Share in the population					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
Administrative database	50.0%	25.0%	15.0%	5.0%	4.0%	1.00%
Survey data	45.2%	29.1%	16.8%	4.9%	**3.5%**	**0.53%**

Source	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
Administrative database	6.5%	24.5%	26.0%	14.1%	18.2%	10.8%
Survey data	6.3%	29.0%	28.9%	14.2%	**15.6%**	**6.0%**

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

Three-step methodology

- Linking households sampled for the survey with their counterpart in the administrative database;
  $\Rightarrow$ **Dataset with reported assets and true assets**
- Systematic decomposition of discrepancies between survey-based estimates and estimates based on administrative data;
  $\Rightarrow$ **Precise measurement of each source of bias**
- Comparing the dwellings reported by respondents with the dwellings they actually own.
  $\Rightarrow$ **Understanding the determinants of reporting behavior**


Data used to compute estimates
Effect
Question
① Administrative database
True outcome, all households
② Administrative database
True outcome, all households in ordinary dwellings
③ Initial survey sample
True outcome, sampling weights
④ Final survey sample
True outcome, sampling weights
⑤ Survey sample
True outcome, final weights
⑥ Survey sample
Reported outcome, final weights
Population coverage
Sampling
Participation behavior
Weight adjustment
Reporting behavior
Do the two sources cover the same population?
Is the sample representative of the population of interest?
Does unit non-response affect estimates?
Does the weight adjustment procedure affect estimates?
Does reporting behavior affect estimates?

Decomposing discrepancies: wealth concentration

	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-100
Administrative data	**6.5%**	**24.5%**	**26.0%**	**14.1%**	**18.2%**	**10.8%**
Survey data	**6.3%**	**29.0%**	**28.9%**	**14.2%**	**15.6%**	**6.0%**
Total discrepancy	**-0.2**	**4.5**	**2.9**	**0.1**	**-2.5**	**-4.8**
Scope of survey	-0.1	0.0	0.0	0.0	0.0	0.0
Sampling	-0.1	0.1	0.0	0.9	-0.5	-0.4
Unit non-response	0.0	1.7	1.1	0.6	**-0.5**	**-2.9**
Weight adjustment	**0.2**	**-0.4**	**-0.8**	**0.0**	**-0.6**	**1.5**
Reporting behavior	-0.2	3.1	2.5	**-1.4**	**-1.0**	**-3.0**

Decomposing discrepancies: wealth concentration

	Share in gross housing wealth					
	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-100
Administrative data	**6.5%**	**24.5%**	**26.0%**	**14.1%**	**18.2%**	**10.8%**
Survey data	**6.3%**	**29.0%**	**28.9%**	**14.2%**	**15.6%**	**6.0%**
Total discrepancy	**-0.2**	**4.5**	**2.9**	**0.1**	**-2.5**	**-4.8**
Scope of survey	-0.1	0.0	0.0	0.0	0.0	0.0
Sampling	-0.1	0.1	0.0	0.9	-0.5	-0.4
Unit non-response	0.0	1.7	1.1	0.6	**-0.5**	**-2.9**
Weight adjustment	**0.2**	**-0.4**	**-0.8**	**0.0**	**-0.6**	**1.5**
Reporting behavior	-0.2	3.1	2.5	**-1.4**	**-1.0**	**-3.0**

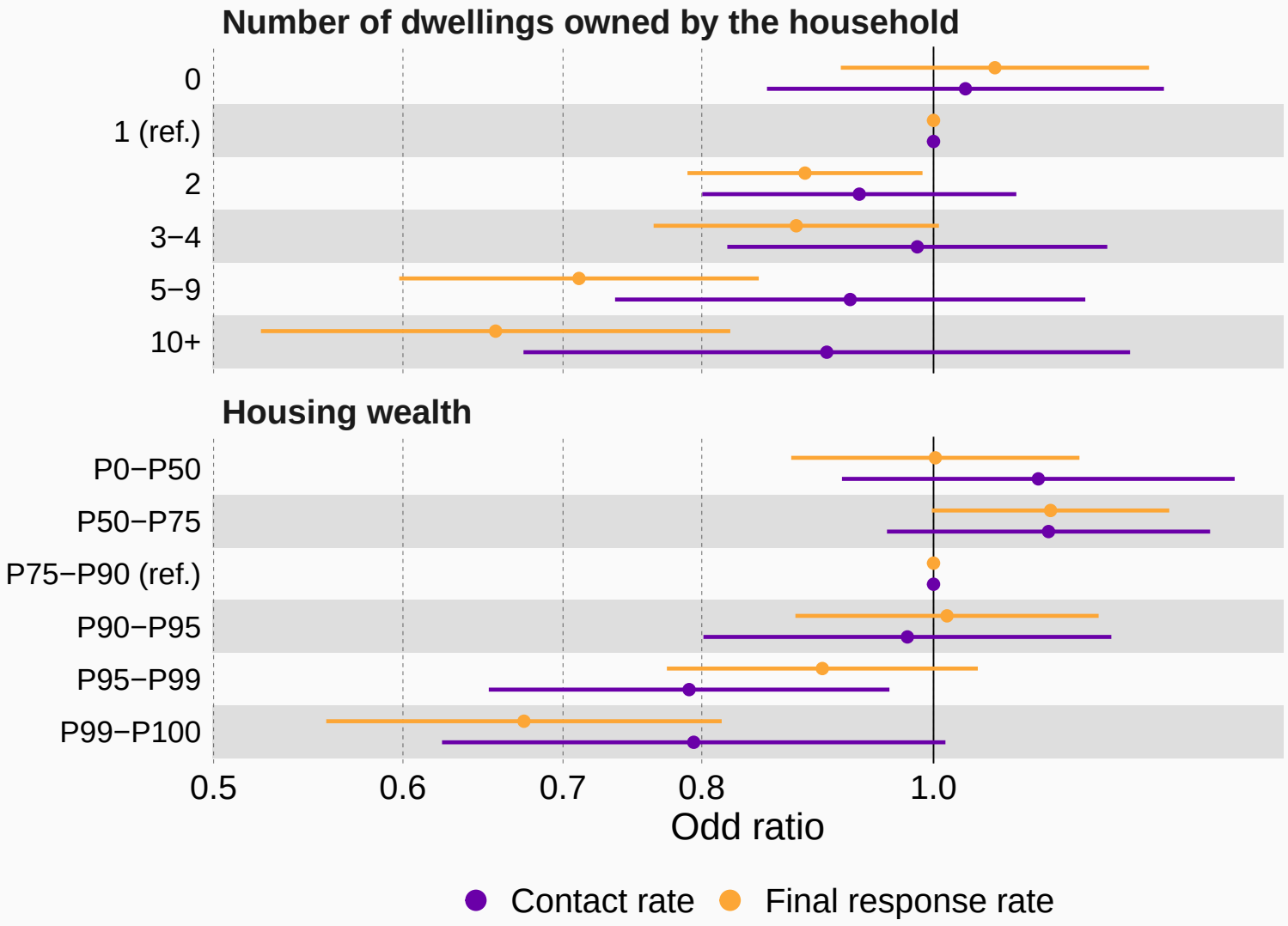
$\Rightarrow$ The downward bias in housing wealth concentration estimates comes in equal parts from underrepresentation of wealthy households and underreporting of assets.

Investigating unit non-response

Econometric approach: large landlords and households of the top 1% are slightly more difficult to contact and specifically reluctant to participate.

Investigating unit non-response

Econometric approach: large landlords and households of the top 1% are slightly more difficult to contact and specifically reluctant to participate.



What properties are reported, and why?

I compare the list of dwellings reported by respondents with the list of dwellings they own to understand the reporting behavior.

What properties are reported, and why?

I compare the list of dwellings reported by respondents with the list of dwellings they own to understand the reporting behavior.

1. **Asset misreporting is ubiquitous: underreporting and overreporting.**
    - Respondents' answers are often inconsistent with the precise question asked about their real estate assets.

What properties are reported, and why?

I compare the list of dwellings reported by respondents with the list of dwellings they own to understand the reporting behavior.

1. **Asset misreporting is ubiquitous: underreporting and overreporting.**
    - Respondents' answers are often inconsistent with the precise question asked about their real estate assets.

2. **Households are more likely to report dwellings they have actual economic control upon**:
    - the dwellings are owned in full ownership (or usufruct);
    - the dwellings are not owned jointly with other households;
    - the building is used as the household's main residence;
    - the household derives property income from them;
    - the property tax recipient belongs to the household.

What properties are reported, and why?

3. **Reporting is better among households with better information and higher educational level**:
    - the respondent is the owner of the dwelling/building;
    - the respondent uses documents and can easily answer questions;
    - the respondent is not older than 70 and has a higher educational level.

What properties are reported, and why?

3. **Reporting is better among households with better information and higher educational level**:
   - the respondent is the owner of the dwelling/building;
   - the respondent uses documents and can easily answer questions;
   - the respondent is not older than 70 and has a higher educational level.

   I finally compare housing wealth reported by households in the survey with the housing wealth they reported to the French tax administration (housing wealth tax).
4. **Clear signs of intentional underreporting at the top of the distribution.**

Conclusion

Comparing survey data with administrative data really helps identifying opportunities to improve survey quality:

- The weight adjustment procedure does not correct all of the selective nonresponse problem;
  $\Rightarrow$ Opportunity to test improvements of the weight adjustment procedure.

Conclusion

Comparing survey data with administrative data really helps identifying opportunities to improve survey quality:

- The weight adjustment procedure does not correct all of the selective nonresponse problem;
  $\Rightarrow$ Opportunity to test improvements of the weight adjustment procedure.
- Asset misreporting is ubiquitous (underreporting and overreporting) because the question asked is not consistent with respondents' conceptual framework;
  $\Rightarrow$ Opportunity to improve the questionnaire.
- Clear signs of intentional underreporting at the top of the distribution.
  $\Rightarrow$ Opportunity to leverage administrative data to improve the measurement of the top tail of wealth distribution.

Reference

André, M. and Meslin, O. (2024).
Le bonheur est dans le prix : Estimation du patrimoine immobilier brut des ménages sur données administratives.
document de travail 2024-04, 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.