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

Document de travail

N°M2025-08-Décembre 2025

Insee
Mesurer pour comprendre

Olivier MESLIN

Institut national de la statistique et des études économiques

2025/08

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

Olivier MESLIN

Décembre 2025

Remerciements :

Cet article est issu d'une suggestion initiale de Pierre Lamarche, que je remercie tout particulièrement. Je remercie sincèrement mes collègues de la Division Logement et Patrimoine de l'Insee de m'avoir donné accès aux données de production de l'enquête Histoire de vie et Patrimoine et d'avoir répondu à mes nombreuses questions, en particulier Marie-Cécile Cazenave-Lacrouts, Aliette Cheptitsky, Pierre Cheloudko et Julie Labarthe. Je remercie Odran Bonnet, Antoine Bozio, Christel Colin, Michel Duée, Bertrand Garbinti, Mélina Hillion, Pierre Lamarche, Eric Lesage, Romain Lesur, Thomas Piketty et Corinne Prost pour leurs commentaires et suggestions utiles, ainsi que les participants aux séminaires de la Paris School of Economics, à la conférence Qualité Eurostat 2024 et à divers séminaires internes de l'Insee.

Direction de la méthodologie et de la coordination statistique et internationale
Timbre L001
88 Avenue Verdier - CS 70058 - 92541 Montrouge Cedex - France -
Tél. : 33 (1) 87 69 55 00 - E-mail : DG75-L001@insee.fr - Site Web Insee : http://www.insee.fr

Résumé

Cet article étudie les causes des biais affectant le haut de la distribution de patrimoine immobilier dans les enquêtes sur le patrimoine, en comparant l'enquête Histoire de vie et Patrimoine 2017 à une nouvelle base de données décrivant le patrimoine immobilier des ménages français. Cet article s'inscrit dans le cadre des travaux préparatoires à la refonte en profondeur de l'enquête Histoire de vie et Patrimoine qui sera conduite par l'Insee au cours des prochaines années.

Une comparaison avec cette nouvelle source montre qu'il ne manque dans l'enquête qu'entre 10 % et 20 % du dernier décile de la distribution de patrimoine immobilier (tant en part de population qu'en part dans le patrimoine total), mais qu'il manque en revanche entre 40 % et 50 % du dernier centile de patrimoine immobilier.

Afin d'éclairer les causes de ce phénomène, les ménages échantillonnés (répondants et non-répondants) sont appariés à la base de données de référence, puis les données appariées sont systématiquement exploitées de façon à mesurer les écarts induits par chacune des étapes du processus d'enquête. Cette analyse conclut que le biais à la baisse affectant le haut de la distribution de patrimoine immobilier dans l'enquête est dû à parts égales à deux mécanismes. Premièrement, les ménages appartenant au dernier centile sont fortement sous-représentés dans l'enquête, principalement parce qu''ils sont un peu plus difficiles à contacter et nettement plus réticents à participer à l'enquête que le reste de la population. La correction de la non-réponse par repondération ne compense que partiellement cette sous-représentation et le calage sur marges déforme, de manière inattendue, la distribution de patrimoine immobilier. Deuxièmement, la sous-déclaration des actifs est nettement plus marquée parmi les ménages les plus aisés.

Le comportement de déclaration des répondants est étudié en comparant les actifs immobiliers déclarés par les répondants aux actifs qu'ils détiennent effectivement. Il apparaît que les ménages tendent à ne déclarer que les actifs sur lesquels ils exercent un contrôle juridique total et un contrôle économique quotidien. La sous-déclaration des actifs s'accroît fortement avec le nombre de logements détenus et parmi les ménages à très haut patrimoine. Ensuite, les valeurs de marché déclarées par les répondants pour leur résidence principale est comparée à la fois à des estimations statistiques et à des transactions immobilières effectivement observées. Cette comparaison conclut que les ménages situés dans le bas de la distribution de patrimoine immobilier tendent à surestimer la valeur de leur logement, tandis que ceux situés dans le haut ont tendance à la sous-estimer.

Enfin, le patrimoine immobilier total déclaré dans l'enquête par les ménages les plus fortunés est comparé à la fois aux montants estimés dans la base de données de référence et aux déclarations d'impôt sur la fortune immobilière de ces mêmes ménages. Cette comparaison montre que la sous-déclaration dans l'enquête atteint environ 40 % pour les ménages les plus fortunés (patrimoine immobilier brut supérieur à 5 millions d'euros). L'article propose finalement des pistes d'améliorations de la méthodologie des enquêtes sur le patrimoine.

JEL : D31, E01, C81, C83

Mots-clés : enquête sur le patrimoine, méthodologie d'enquête, données administratives, patrimoine immobilier, inégalités de patrimoine, décomposition de biais, sous-déclaration, sous-représentation.

Abstract

This paper investigates the causes of biases in survey-based top wealth shares by comparing the 2017 French wealth survey with a new benchmark database on housing wealth of French households. It contributes to preliminary work toward a major overhaul of the French wealth survey currently conducted by the French National Statistical Institute.

Compared to this benchmark, only 10% to 20% of the top 10% of the housing wealth distribution is missing in the survey (both in population share and in wealth share), but this proportion increases to 40%-50% for the top 1%. I link all sampled households to the benchmark database, respondents and non-respondents alike, and use this linked data to measure the discrepancy induced by each step of the survey process, based on an innovative decomposition approach. I conclude that the downward bias in survey-based top wealth shares comes in equal parts from two causes. First, households belonging to the top 1% are strongly underrepresented in the survey mostly because they are somewhat more difficult to contact and much more reluctant to participate than the rest of the population. The weight adjustment procedure does not fully compensate this underrepresentation and calibration unexpectedly distorts the wealth distribution. Second, wealth underreporting is more intense among wealthy households.

I then compare reported housing assets with the assets actually owned by respondents and show that households tend to report assets they have full legal and daily economic control upon, and that asset underreporting increases sharply with the number of housing units owned by households and among high net wealth households. I compare market values reported by respondents for their primary residence with both statistical estimates and prices observed in real estate transaction data and conclude that households at the bottom of the housing wealth distribution tend to overestimate the value of their home, whereas households at the top tend to underestimate it.

I finally compare reported housing wealth with both the benchmark database and wealth tax returns and show that wealth underreporting among the very wealthiest households (with an estimated housing wealth above 5 meuro) amounts to approximately 40%. Based on these findings I suggest potential improvements to wealth survey methodology.

JEL: D31, E01, C81, C83

Keywords: wealth surveys, survey methodology, administrative data, housing wealth, wealth inequality, underreporting, underrepresentation.

Contents

1 Introduction

In a context of rising income and wealth inequality, obtaining reliable measures of wealth concentration is of considerable importance for researchers, policymakers, and also the general public (Piketty (2014); Piketty et Zucman (2015)). Household wealth surveys are one of the most important data sources on wealth, but are known to fail to reflect accurately the upper tail of the wealth distribution, resulting in wealth concentration estimates markedly lower than in other sources. A large literature has explored multiple paths to understand and overcome these limitations of surveys: investigating survey methodology to unveil the nature and causes of potential biases (Juster, Smith, et Stafford (1999); Vermeulen (2018); Kennickell (2019)), reconciliating survey data and administrative data (Bricker, Henriques, Krimmel, et Sabelhaus (2016a,b); Kennickell (2017b); Meriküll et Rõõm (2021)), correcting survey data to better reflect the true distribution of wealth (Vermeulen (2016, 2018); Bach, Thiemann, et Zucco (2019); Blanchet, Flores, et Morgan (2022); Cantarella, Neri, et Ranalli (2024)) and finding alternative, supposedly more accurate, data sources (Kopczuk et Saez (2004); Saez et Zucman (2016)).

Although underrepresentation of wealthy households and underreporting of assets are the usual suspects, the literature provides limited reliable empirical evidence on the magnitude and causes of the biases in survey based top wealth shares because of three limitations. First and most importantly, "the fundamental problem in assessing survey bias [...] is the lack of a benchmark measure of the true outcome" (Meyer, Mok, et Sullivan (2015)), even more so for wealth than for other outcomes such as income or consumption. Second, even when a benchmark measure is available, wealth surveys typically include only a small number of wealthy households, making it challenging to obtain compelling empirical evidence on whether their participation or reporting behaviors differ from the general population. Third, the existing literature lacks systematicity by focusing on *some* sources of biases, without accounting for *all* sources of discrepancy between survey-based outcomes and true outcomes. An additional limitation of a somewhat different nature is that the literature tends to focus mostly on financial wealth, based on two implicit hypotheses: housing wealth reported in wealth surveys is rather reliable (at least more than reported financial wealth), and biases in housing wealth concentration are moderate.

This paper overcomes these limitations and investigates the biases in survey-based top wealth shares by linking the French wealth survey with a new benchmark database on housing wealth of French households, and by systematically measuring all sources of discrepancy in housing wealth between the two data sources. I take advantage of the strong oversampling of wealthy households in the survey to focus on the specific behaviors of the top of the housing wealth distribution. I conclude that survey-based estimates of housing wealth concentration are affected by severe biases that come in equal parts from two causes: although they are very accurately represented in the initial survey sample, wealthy households are strongly underrepresented in the survey final sample mostly because they are specifically reluctant to participate, and asset underreporting among wealthy households is much more intense than in the rest of the population. I estimate that housing wealth underreporting among the wealthiest households (with an estimated housing wealth above 5 m€) amounts to at least 40%. From an institutional perspective, this paper belongs to preliminary steps toward a major overhaul of the French wealth survey currently conducted by the French National Statistical Institute.

This paper analyzes the shortcomings of wealth surveys through a careful comparison of two data sources: the 2017 French Household Finance and Consumption Survey (HFCS) and a new statistical database built on administrative data describing the housing wealth of all resident households as of January 1st, 2017, along with rich socio-economic data on households (André et Meslin (2021) and André et Meslin (2025)). I first describe the discrepancies between the two data sources and show that between 40% and 50% of the top 1% of the housing wealth distribution is missing in the survey (both in population share and in wealth share). The rest of the paper aims at explaining why. I link all sampled households to the benchmark database, respondents and non-respondents alike, and use this linked data to measure the bias induced by each step of the survey process in estimates of two key outcomes: the share of the population belonging to the top 1% of the gross housing wealth distribution, and the share of total gross housing wealth owned by the top 1%. I finally investigate how the participation behavior of sampled households and the reporting behavior of respondents relate to housing assets and housing wealth.

Comparisons between surveys and administrative data sources are often obfuscated by the fact that discrepancies may arise from three distinct series of issues: shortcomings of surveys, limitations of administrative data, and comparability issues between the two sets of sources (see for instance Bricker, Henriques, Krimmel, et Sabelhaus (2016b)). It turns out that I can focus almost exclusively on the shortcomings of wealth surveys, because the benchmark database is particularly reliable and highly comparable with the survey: the two data sources use the same definitions of assets and of households, and cover almost the same populations. The benchmark database has nevertheless two limitations: the share owned by each household in each asset is not available in administrative data, and market values of housing units are not observed but estimated using a machine learning algorithm. When necessary I leverage other administrative data sources to prove that the results are not sensitive to these two limitations.

The contribution of this paper is threefold. First, I introduce a general, assumption-free, methodology to decompose biases in survey-based estimates when the survey can be linked to a benchmark dataset. Existing studies (Vermeulen (2018); Johansson-Tormod et Klevmarken (2022); Meriküll et Rõõm (2021)) analyze only a few specific problems encountered in surveys and fail to offer consistent sets of quantitative estimates of *all biases* affecting survey-based concentration estimates, so that it is difficult to assess their relative importance. Contrasting with this literature, I offer a comprehensive approach where discrepancies between survey based estimates and estimates based on the benchmark database are additively decomposed, each source of discrepancy being precisely defined and measured. This methodology can easily be adapted to other settings and other research questions.

Second, this paper sheds light on why wealth surveys fail to accurately reflect the upper tail of the housing wealth distribution (the top 1%). Existing studies that linked wealth surveys with an administrative benchmark could not focus on the upper tail because it was insufficiently covered in the survey (Johansson-Tormod et Klevmarken (2022) and Meriküll et Rõõm (2021)). Fortunately, the sample of the French wealth survey was drawn using a state-of-the-art sampling plan that strongly oversamples wealthy households. I take advantage of this sampling plan to focus specifically on the top 1%. Applying the decomposition methodology mentioned above, I conclude that the two main causes of the underestimation of the top 1% wealth share are the underrepresentation of wealthy households and the underreporting of assets at the top of the distribution and that most of these two biases is due to specific behaviors of wealthy households, offering clear empirical evidence in support of hypotheses formulated in Kennickell (2019).

Third, I bring forward a rich set of results on the determinants of participation and reporting behaviors. Following Meriküll et Rõõm (2021), I distinguish three causes of non-participation: the household is out of the survey scope, could not be contacted by the interviewer or refused to participate. Consistent with most of the existing literature (Kennickell et Woodburn (1999), Johansson-Tormod et Klevmarken (2022), Alvargonzález, Barcelo, Bover, Cobreros, Crespo, El Amrani, García-Uribe, Gento, Gómez, Villanueva, et al. (2024)), Using logistic regressions, I conclude that wealthy households are significantly more difficult to contact and that wealthy households and households owning a large number of housing units are particularly reluctant to participate in the survey. I then analyze the determinants of asset reporting by comparing reported assets with assets actually owned by respondents. I reach three main conclusions: households tend to report assets they have full legal and daily economic control upon, underreporting decreases with the information available to the respondent, and underreporting increases sharply with the number of housing units owned by the household. I investigate the asset evaluation behavior by comparing market values reported for the household's primary residence with statistical market value estimates and prices observed in real estate transaction data. Consistent with Johansson-Tormod et Klevmarken (2022), I conclude that households at the bottom of the housing wealth distribution tend to overestimate slightly the value of their home, whereas households at the top tend to underestimate it slightly. I finally focus on the tip of the distribution and compare reported housing wealth with both the benchmark database and the housing wealth reported by households in their wealth tax returns. I conclude that the housing wealth of the wealthiest households is underreported by roughly 40%.

The remainder of the paper proceeds as follows. Section 2 presents the survey and the benchmark database, describes the key discrepancies between them in the measurement of the top tail of the housing wealth distribution and reviews the potential causes of these discrepancies. Section 3 introduces the decomposition approach and identifies the causes of the discrepancies between the benchmark and survey data. Section 4 investigates the causes of the underrepresentation of wealthy households in the survey. Section 5 investigates the causes of asset and wealth underreporting in the survey. Section 6 discusses the results and suggests methodological improvements in wealth survey methodology. Section 7 concludes.

2 Two diverging data sources

In this section, I first describe the relevant features of the French wealth survey (2.1) and of the benchmark administrative database (2.2). I then compare estimates based on the two data sources and identify three large discrepancies related to the measurement of the top tail of the housing wealth distribution (2.3). Finally I review the potential causes of these discrepancies, based on the literature addressing limitations of wealth surveys (2.4).

2.1 Overview of the French wealth survey

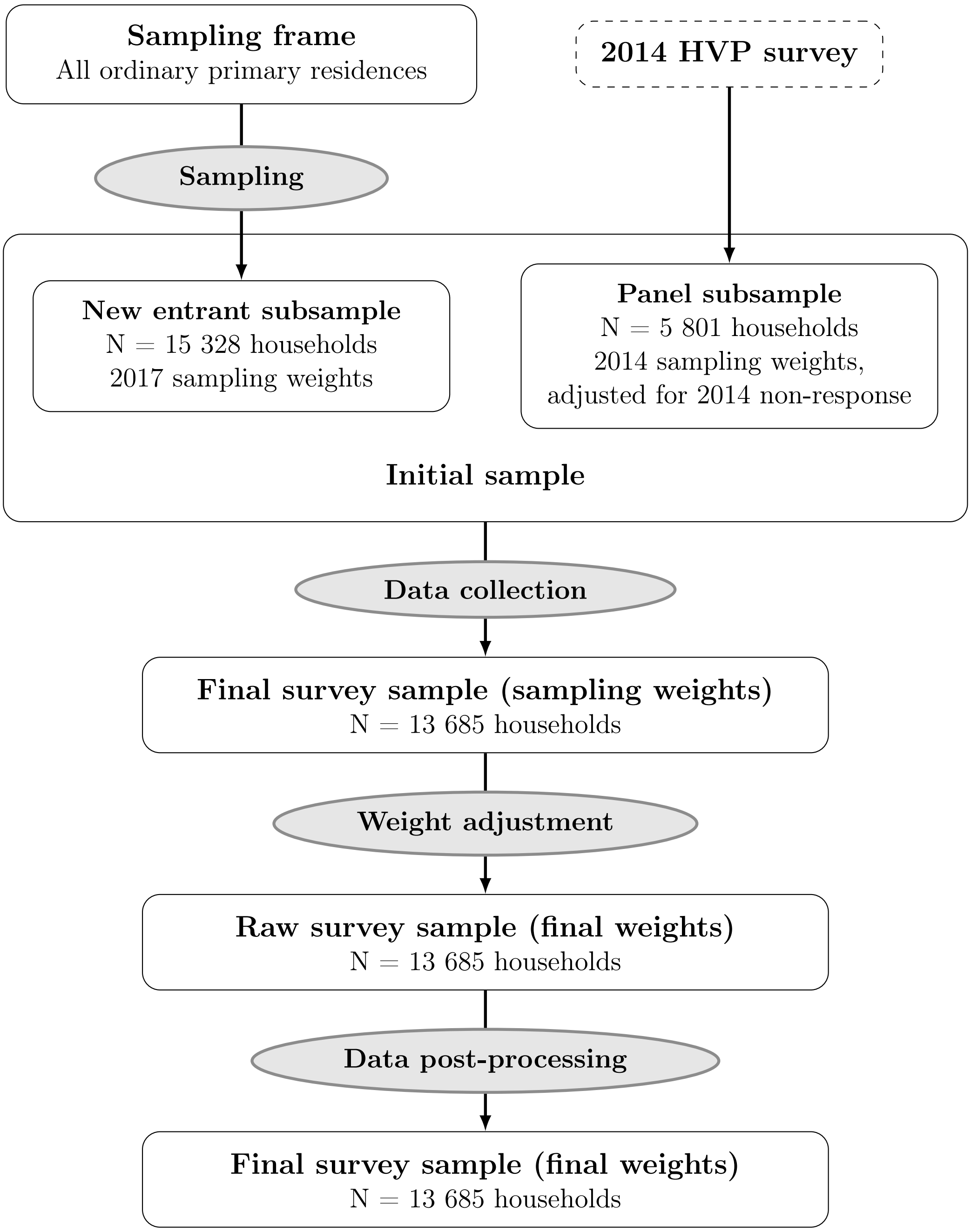
The French wealth survey (*enquête Histoire de vie et Patrimoine*, HVP hereafter) was introduced in 1986 by the French national statistical institute (Insee) and takes place every third year from 2014 onwards. This survey collects detailed information on the real estate, financial and professional assets of households and the associated debt, as well as on the factors related to asset accumulation: personal and professional biography, inheritances and gifts, income and financial situation. It is the source of the French component of the Household Finances and Consumption Survey (HFCS), a European survey on wealth coordinated by the European Central Bank. Two important methodological improvements were introduced in the last fifteen years. First, from the 2010 survey onwards the sampling plan has been improved to strongly oversample wealthy households. This sampling plan is a major advantage of the French wealth survey, as it makes it possible to zoom on the upper tail much more than in most other wealth surveys. Second, starting in 2014, a subsample of individuals (called panel individuals) are surveyed four times over a period of nine years[1] (four consecutive surveys), turning the survey into a rotating panel. The following paragraphs describe the four main steps of the production process of the survey summarized in figure 1: sampling (2.1.1), data collection (2.1.2), weight adjustment (2.1.3) and data editing and imputation (2.1.4).

2.1.1 Sampling plan

The initial sample of the 2017 HVP survey contains approximately 21 100 households and consists in two subsamples: a *new entrant subsample* surveyed for the first time in 2017, and a *panel subsample* already surveyed in 2014.

[1] This longitudinal dimension has been introduced by most countries participating in the Household Finance and Consumption Survey (13 countries as of 2021, see Network (2023)).

Figure 1. Production process of the 2017 HVP survey



The new entrant subsample The sampling frame of the new entrant subsample[2] is a modified version of the 2016 Fidéli database produced by the French statistical institute (Insee). This database relies on fiscal data and describes all housing units located in France and all households living in these housing units: housing units' characteristics, composition and location of households, and detailed data on income (wages, pensions, capital income, social transfers). The sampling frame for the survey is built by modi-

[2]This subsample is called *refreshment sample* in the official methodology of the Household Finance and Consumption Survey. Following Lynn (2012), I use the arguably more intuitive expression of *new entrant subsample*.

fying this database in two ways. First, it is restricted *private households* (roughly 94% of resident households accounting for 96% of the resident population), excluding all individuals living in institutions. Second, the Fidéli database is complemented with data on wealth drawn from wealth tax returns (*Impôt de solidarité sur la fortune*). Within this sampling frame, high net wealth households are defined as tax units subject to the French wealth tax (*impôt de solidarité sur la fortune*); this group accounts for roughly 1% of French tax units. Wealth reported in wealth tax returns is considered as a proxy for true wealth and is used to oversample high net wealth households, quite similarly to the methodology used in the US Survey of Consumer Finances (see Kennickell (2017b)).

The new entrant subsample (15 300 fiscal households) is drawn from this sampling frame using a two-stage sampling plan: local areas are drawn in a first step, then housing units are drawn within the selected areas[3]. The second step of the sampling plan is stratified with respect to geography, age of household head, income composition and wealth. In particular, high net wealth households are isolated in three specific strata and are heavily oversampled: the sampling rate is 0.053% on average, but close to 2.7% for the top strata of very wealthy households. The strata are described in table 1. All in all, high net wealth households account for approximately 18.6% of the new entrant subsample (but only for 1% of the sampling frame).

Table 1. Descriptive statistics on the sampling plan of the new entrants subsample

	Description	Share in population	Number of sampled households	Sampling rate	Share in the sample
1	Very high net wealth households, urban (> 3M€)	0.09%	730	2.7%	4.8%
2	Very high net wealth households, rural (> 3M€)	0.11%	1 075	2.6%	7.0%
3	Other high net wealth households	0.82%	1 039	0.40%	6.8%
	All high net wealth households	*1.0%*	*2 844*	*0.87%*	*18.6%*
4	Older households (reference person 60 or older)	41.1%	3 459	0.03%	22.6%
5	Households with high business income	2.8%	1 174	0.15%	7.7%
6	Households with high capital income	2.3%	523	0.07%	3.4%
7	Households with high wages	4.9%	2 302	0.17%	15.0%
8	All other households	47.8%	5 026	0.04%	32.8%
	All standard households	*99.0%*	*12 484*	*0.04%*	*81.4%*
	All households	**100.0%**	**15 328**	**0.05%**	**100.0%**

Source: production data of the HVP 2017 survey, author's computation.

This sampling design has three consequences. First, individuals living in institutions (prisons, hospitals, student residences, nursing homes...) are excluded from the sampling frame and are not surveyed. Second, the reference units in the sampling frame are not households but housing units, implying that the new entrant subsample is a set of housing units that may be home to more than one household (eg, flatmates), whereas the panel subsample is a sample of individuals. Third, the high heterogeneity of sampling rates

[3]This two-stage sampling plan aims at limiting the budgetary cost of the survey by reducing distances travelled by interviewers.

induces a large dispersion of sampling weights and eventually of final household weights. For instance, the 99th percentile of final weights is close to 10 000, whereas the 1st percentile is close to 6.7.

The panel subsample Half of the sample already surveyed in the 2014 survey was included in the panel subsample. This subsample contains 5 800 households and derives entirely from the sample of respondents of the 2014 survey; the 2014 sample was drawn using the same methodology as the 2017 new entrant subsample (including the same 8 sampling strata). Importantly, the sampling weights used for this subsample in the 2017 survey are the sampling weights of housing units sampled in 2014, after the correction for non-response operated in 2014 (but before the 2014 calibration, see section 2.1.3 for details).

2.1.2 Data collection

Data collection took place between September 2017 and January 2018 and followed the standard procedure of French household surveys. Interviewers start by identifying the individuals living in each housing unit of the new entrant subsample and by verifying that the housing unit is used as a primary residence. Households belonging to the panel subsample are contacted directly by phone and/or email thanks to information collected in 2014 and in a subsequent follow-up survey. Around 1 650 housing units or households are excluded at this stage for various reasons (e.g. the housing unit is vacant or not used as a primary residence, the interviewer could not locate the housing unit, all household members are deceased or moved abroad...), leaving 19 450 valid households in the sample. A letter announcing the survey is then sent to the individuals living in the housing unit and an appointment is made for the interview. A second letter insisting on the importance of the survey is sent to households who initially refused to participate. Finally, the interview takes place face-to-face at the respondent's home.

Lots of efforts are devoted to achieving a high response rate: interviewers undergo specific training focused on convincing households to participate, and particular attention is paid to the high net wealth sampled households. All in all, 13 742 households living in 13 685 housing units were interviewed. The overall response rate of the 2017 survey amounts to 63.5%[4]; it is higher in the panel subsample (75.8%) and lower in the new entrant subsample (58.5%). Response rates are above 50% in the top strata (table 2); such response rates among wealthy households are remarkably high when compared to other similar surveys in Europe and in the United States (Kennickell et Woodburn (1999), Alvargonzález, Barcelo, Bover, Cobreros, Crespo, El Amrani, García-Uribe, Gento, Gómez, Villanueva, et al. (2024)), where response rates at the top are often comprised between 20% and 40%.

The interview can take place in the presence of several members of the household,

[4] This overall response rate is computed as ratio between the number of respondents and the number of sampled housing units and households, including out-of-scope units.

Table 2. Response rate by stratum and subsample

Stratum	All households	Subsample: Entrant	Panel
Very high net wealth households, urban	52.8%	48.8%	71.2%
Very high net wealth households, rural	51.1%	47.4%	67.6%
Other high net wealth households	58.1%	53.6%	71.6%
All high net wealth households	*54.2%*	*50.0%*	*70.3%*
Older households	64.4%	60.4%	73.8%
High business income	66.1%	59.9%	79.7%
High capital income	69.4%	62.1%	82.9%
High wages	66.3%	62.3%	75.2%
All other households	64.8%	59.7%	77.6%
All standard households	*65.3%*	*60.5%*	*76.6%*
All households	**63.5%**	**58.5%**	**75.8%**

Source: production data of the HVP 2017 survey, author's computation.

but in principle at least with the reference person (the individual bringing the largest resources to the household) or his/her spouse. Interviewers are asked to interview the person most knowledgeable about the household's assets. Respondents may use personal documents to answer questions (bank statements, legal documents related to real estate transactions...). The first part of the questionnaire contains general questions on the household's members (age, gender, highest degree, family links...). Importantly, interviewers are asked to collect detailed personal information on all households members (first and last names, date of birth). Respondents are then asked questions on all their assets (primary residence, other real estate assets, financial assets, durable goods, liabilities), on their consumption level, on gifts and inheritances and on their personal and professional history (childhood, marital status, employment history). The whole interview is typically quite long: it lasts for more than one hour on average for standard households, and between one hour and a half and two hours on average for high net worth households. For some households with diversified assets the interview may even take place over two visits.

Regarding housing wealth, households are asked how many housing units they own, either in full ownership or in bare ownership (more on this below), either directly or through a French real estate holding company (*Société civile immobilière*, SCI), on the day of the survey. The only exception to this definition is the primary residence of the household: it must also be reported if it is held in usufruct. In addition, households are asked a few more questions on each housing unit[5]: the type of housing unit (house or flat), the municipality (*commune*) where it is located, the share owned by the reference individual, his/her spouse, other members of the households and other households, the year and price of purchase of the housing unit and how it is used (owner-occupied, leased

[5]Respondents are not asked whether the housing unit is owned directly or through an SCI.

to tenants, vacant...). Moreover, respondents are asked to give lower and upper bounds for the current market value of the housing unit. Importantly, households are allowed to describe several flats jointly if two conditions are met: the flats must be located in the same building and must be used in the same way (for instance, all are leased to tenants). This implies that in practice some assets reported by large landlords are more akin to buildings or fraction of buildings (eg, 5 flats located in the same building and leased to tenants) than to housing units. In the rest of this article, I refer to these groups of housing units as *housing assets*.[6]

The mention of bare ownership in the definition of housing wealth used in the survey is an important methodological choice, as roughly 15% of privately-held housing units in France are subject to separation of ownership rights between usufruct and bare ownership[7]. Under French law, the individual holding usufruct of a housing unit may use it (eg, live in it) and derive an income from it, is liable for the property tax on the housing unit, but does not have the right to sell it. Conversely, the individual holding the bare ownership may sell the (bare ownership of the) housing unit, but cannot use it or derive an income from it. The bare owner becomes automatically full owner when the usufructuary dies.

Finally, the interviewer completes an interview quality report immediately after the interview, outside the presence of the surveyed household. This report contains qualitative measures of the attitude of the respondent(s): whether the respondent was mistrustful before/after the interview, whether she understood the questions, whether she was interested in the survey, whether she used documents, etc.

The data collection phase is followed by a long phase of post-processing lasting more than a year. This phase include many steps such as linking the survey sample with income tax returns and other administrative data to retrieve information on income and social benefits, anonymization and writing the documentation. For the sake of brevity, I describe in detail only the two steps that impact survey-based estimates of wealth distribution and concentration: weight adjustment (2.1.3) and data editing and imputation (2.1.4).

2.1.3 Weight adjustment

The sampling weights of the HVP survey are adjusted through a standard two step procedure (see Haziza et Beaumont (2017) for a detailed presentation). This procedure

[6] Arguably, they could have been called "buildings", but this might have suggested that households owned *all* housing units in the building, which is not always the case.

[7] There are two reasons for the widespread use of separation of ownership rights in France. First, giving the bare ownership of an asset is tax-favoured with respect to giving the full ownership. As a consequence, French parents frequently give the bare ownership of their real estate assets to their children, as a way to reduce their tax burden on intergenerational wealth transfers while keeping control of the assets. Second, separating ownership rights is one of the default settings of inheritance process when a surviving spouse is present. For instance, imagine a married couple with three children owning their primary residence (with equal shares). Under some legal conditions that are frequently met in practice, if one of the spouses dies, the surviving spouse receives the usufruct of its late spouse's share in the house, and the children receive the bare ownership of their deceased parent's share.

is carried out separately for the two subsamples. First, sampling weights are modified to account for unit non-response (defined as a complete lack of information on a given sampled unit) using the homogeneous response groups method (Caron (2005)). In this approach, the response probability of each household is estimated using a logit model, based on the information available in the sampling frame. Households are then clustered in a small number of homogeneous response groups, where all households in a group have a similar predicted response probability. Finally, the sampling weights of respondents of each group is divided by the average weighted response rate of the group. For instance, if one specific subgroup of households has on average a 33% response probability, the sampling weight of the households of this subgroup that actually answered the survey is multiplied by 3 (1/0.33) to correct for unit non-response. Second, weights are adjusted through a calibration approach to ensure consistency between survey estimates and known population totals available from external benchmark sources (such as the census and income tax returns). This step is an optimization problem where one looks for the smallest relative modification of weights that ensures consistency between survey estimates and known totals. Auxiliary data used for calibration include the type of household, the place of residence, the age, highest degree and socio-economic status of the reference person, earned income, capital income and net wealth (as measured by tax returns). Importantly, this auxiliary data does not come from the survey sampling frame but from other sources such as the census and fiscal data, and is used in the calibration of other household surveys so as to ensure aggregate consistency between surveys[8].

2.1.4 Data editing and imputation

Regarding housing assets and housing wealth, the data editing process takes place in three steps. First, raw survey data undergoes a thorough manual inspection, aiming at detecting and correcting errors and inconsistencies, and at imputing some missing values. Edited variables include homeownership status, the share owned in each asset, the number of housing units owned and reported market value intervals. All in all, corrections are limited: the number of housing units owned is corrected for roughly 3.2% of households, (most of the time by one unit) and less than 0.2% of reported market value intervals are corrected. Second, final reported asset values are imputed using an econometric model trained on the market value intervals reported by respondents. Third, housing wealth is then computed as the sum of all market values (weighted by the share owned by each household).

The first step of this postprocessing happens to be quite difficult and labour-intensive for very wealthy households, for two reasons: no external benchmark was available at the time of the survey, and the subsample of very wealthy households is too sparse to test the plausibility of one household's answers by comparing them with the answers of other, similar households. As a consequence, statisticians have to decide whether and how to

[8]For instance two surveys collected in the same year and calibrated with the same auxiliary data will have the same distribution of household type.

correct respondents' answers based only on the internal consistency of these answers and on short written comments by interviewers. Although few in number, these corrections are important for the survey as these few wealthy households have a significant impact on concentration estimates, in line with the simulations by Kennickell (2019).

2.2 Overview of the benchmark database on housing wealth

The benchmark database on housing wealth is a statistical database built on administrative data. It describes all real estate assets owned by all resident households in France and estimates their gross housing wealth on January 1st, 2017. The following paragraphs describe the data sources (2.2.1) and the methodology (2.2.2) underlying this database.

2.2.1 Administrative data sources

The benchmark database on housing wealth relies on four administrative data sources:

– **Households database**: The *Fidéli database* produced from 2011 onwards by the French statistical institute (Insee) describes all housing units located in France and all resident individuals living in these housing units. Importantly, this database contains a dataset identifying *uniquely* all adult individuals known to the tax administration (the *taxpayers dataset*), along with detailed personal information (first names, birth name, married name, date and place of birth), the household they belong to and the housing unit they live in. The construction of this database relies on two assumptions. First, individuals are supposed to be living in the housing unit they reported to the tax administration as their primary residence for tax purposes, although they might actually live somewhere else[9]. Second, individuals living in the same housing units are supposed to be members of the same households (called *fiscal households*). The benchmark database on housing wealth relies on the 2017 edition of the Fideli database.

– **Cadastral data**: the French cadastral data produced by the French tax administration describes the universe of buildings and land plots located in France and subject to property tax. It contains a detailed description of housing units (floor area, level for flats, number of rooms, bedrooms and bathrooms, whether the housing unit has a garage, a terrace, a swimming pool...) and personal information on their owners (first name, birth name, married name, date and place of birth, address, type of ownership rights). The benchmark database on housing wealth relies on the 2017 edition of French cadastral data.

– **Commercial register data**: The *registre du commerce et des sociétés* (RCS) is a legal register managed by the commercial courts' registries. All firms located in France must be registered by at least one commercial court registry. This data contains information on companies (name, legal form, address of the head office) and

[9]For instance, many students live in student residences although their primary residence for tax purposes remains their parents' primary residence.

their owners and managers (natural persons or legal entities). It is used to unveil the shareholders of real estate holding companies (*Sociétés civiles immobilières*, SCI).

- **Real estate transaction data**: this dataset contains information on all real estate transactions subject to stamp duties between 2015 and 2019 (roughly 3 million transactions). All transactions are geocoded.

2.2.2 Methodology

The benchmark database on housing wealth is built through a three step methodology. First, data sources are systematically linked to reconstitute the list of all real estate asset owned by each household. Second, the market value of each housing unit is estimated using a statistical model. Third, the housing wealth of households is computed as the sum of the market values of all housing units owned by each household. The complete methodology is described in detail in André et Meslin (2021) and André et Meslin (2025).

Who owns what? The list of all real estate asset owned by each household is reconstituted in three steps:

- Unveiling SCIs' shareholders: when a housing unit belongs to a SCI, cadastral data contains information on the company, but not on its shareholders. To overcome this data limitation, cadastral data is linked with commercial register data to retrieve personal information on SCIs' shareholders.
- Linking owners with taxpayers: housing units' owners described in cadastral data are linked with the taxpayers dataset using personal information[10] (first names, birth name, married name, date and place of birth, address). This step yields a dataset linking each individual to the real estate asset she or he owns, along with the precise ownership right held by the individual.
- Reconstituting households' assets: the 2017 Fidéli database is used to gather individuals into households. This step yields a dataset linking each household to the real estate asset owned by the households' members.

What is the market value of each housing unit? The market value of all housing units is estimated in two steps. First, a machine learning algorithm is trained on a large dataset of real estate transactions to predict the market value of housing units. Second, this model is used to predict the market value of all housing units located in France in the first quarter of 2017. This statistical model has two prominent advantages discussed in André et Meslin (2025): the spatial structure of housing prices is very accurately accounted for in a non-parametric way, and it has a small bias, meaning that the average predicted price is close to the average observed price for any sufficiently large group of housing units.

[10] First and last names collected in surveys are systematically deleted at the end of the survey process. The French wealth survey is an exception to this rule because of the panel component: personal information is necessary to follow respondents over time.

What is the housing wealth of each household? The housing wealth of resident households can be readily estimated as the sum of market value of all housing assets owned by each household. Two caveats apply nonetheless. First, one must choose how to allocate housing units jointly owned by two or more households. Given that cadastral data does not contain any information on the share owned by each individual, equal split between owners is assumed. This assumption is inconsequential most of the time (as between 85% and 90% of privately-owned housing units are owned by only one household), but may induce serious biases for large landlords at the very top of the distribution, where several households jointly own very expensive assets (eg, large buildings with many rental flats located in city centers). Second, when property rights on a housing unit are separated between bare ownership and usufruct, one must allocate the asset to either the bare owners or the usufructuary. These housing units are assumed to be owned by the bare owners[11], to keep the definition of housing wealth consistent with the HVP definition.

2.3 Three puzzling discrepancies

In this section, I compare estimates based on the survey and on the benchmark database. Three points are worth clarifying before diving in the comparison. First, for the sake of concision, the terms "wealth" and "housing wealth" designate gross housing wealth throughout the paper, unless stated otherwise. Second, although I will frequently refer to the outcomes measured in the benchmark database as the *true* outcomes, I do not suggest that these outcomes are flawless or perfectly reliable. Third, wealth groups are *always defined with respect to the distribution of gross housing wealth in the benchmark database* (the only exception being table 4b). For instance, a household belongs to the top 1% if its gross housing wealth is higher than 1.512 M€, because the 99$^{\text{th}}$ quantile of the gross housing wealth distribution as measured by the benchmark database amounts to 1.512 M€. As a consequence, a household can belong to different groups, depending on the definition of housing wealth used to determine its position in the distribution. For instance, a household whose true housing wealth and reported housing wealth amounts respectively to 2 M€ and 1.1 M€ will be classified in the top 1% based on true wealth, and in a lower group based on reported wealth. What changes is the *measure of wealth* used to allocate households to wealth groups, not the *definition* of wealth groups.

Based on these conventions, I identify three large discrepancies related to the measurement of the top tail of the housing wealth distribution. The rest of the paper will then focus on explaining carefully the causes of these discrepancies.

First, *the share of large landlords is much lower in the survey.* At first sight, figure 2 may suggest that the two sources yield similar estimates, as the share of resident households owning at least one housing unit is close to 60% in both sources: 58.2% in the benchmark database and 60.5% in the survey. However, the two sources yield diverging estimates on the share of large landlords: whereas 3.5% of resident households own five

[11] This information is available in cadastral data.

housing units or more in the benchmark database, this share amounts to only 1.5% in the survey, roughly 60% less.

Second, *the upper tail of the housing wealth distribution is partly missing in the survey.* The two distributions of housing wealth are quite close on the central part of the distribution (figure 3), roughly between the 60$^{\text{th}}$ and the 95$^{\text{th}}$ percentiles. However, the quantiles are increasingly divergent in the upper tail of the distribution. At the very top of the distribution, survey-based quantiles are roughly 20% lower than quantiles computed using administrative data. For instance, the 99th percentile amounts to 1,512k€ in the benchmark database, but only to 1,244k€ in the survey. As a consequence, the top tail of the housing wealth distribution is increasingly underrepresented in the survey: the population share of households belonging to the true top 10% amounts to 8.9% in the survey, 4% for the true top 5%, and only 0.53% for the true top 1% (first panel of table 3). In others words, with comparable definitions, roughly half of the true top 1% population is missing in the survey.[12]

Third, *the concentration of housing wealth is lower in the survey.* At first sight, the top housing wealth shares seem to be slightly lower in the survey (table 3): the top 1% would own 10.8% of total housing wealth according to the benchmark database and 8.8% in the survey (29% and 25% respectively for the top 5%). However, this comparison is misleading because housing wealth groups are defined using quantiles from two different distributions. When housing wealth groups are defined using the quantiles from the benchmark database (last row of the second panel in table 3), the discrepancy becomes much larger: the wealth share of the true top 1% amounts to only 6% in the survey as opposed to 10.8% in the benchmark database. In others words, with comparable definitions, roughly 40% of the true top 1% wealth share is missing in the survey.

2.4 Where could these discrepancies come from?

In this section, I briefly review the potential causes of these discrepancies. The literature addressing limitations of wealth surveys identify three broad sources of bias (in particular Kennickell (2017b), Kennickell (2017a) and Vermeulen (2018)): mis-coverage of the population of interest (some subgroups are unintentionally oversampled and/or undersampled), non-ignorable unit non-response along with inadequate weight adjustment, and reporting errors.

Coverage errors may be due to the imperfections and age of the sampling frame, resulting in samples that are not perfectly representative of the population of interest. For instance, the oversampling of wealthy households relies on wealth tax describing

[12] Conversely, survey quantiles are larger than quantiles based on administrative data at the bottom of the distribution (below P70). This discrepancy has two causes. First, restricting the benchmark database to private households (from ① to ② in figure 4) induces an upward shift of the housing wealth distribution, because individuals living in institutions are generally less wealthy than private households. This shift is also visible in table 4: restricting the benchmark database to private households increases the share of households owning at least one housing unit by 2.6 pp. Second, households at the bottom of the distribution tend to overestimate the value of their assets (see Johansson-Tormod et Klevmarken (2022), section 5.3 and figure 10).

Figure 2. Distribution of households by number of housing units owned

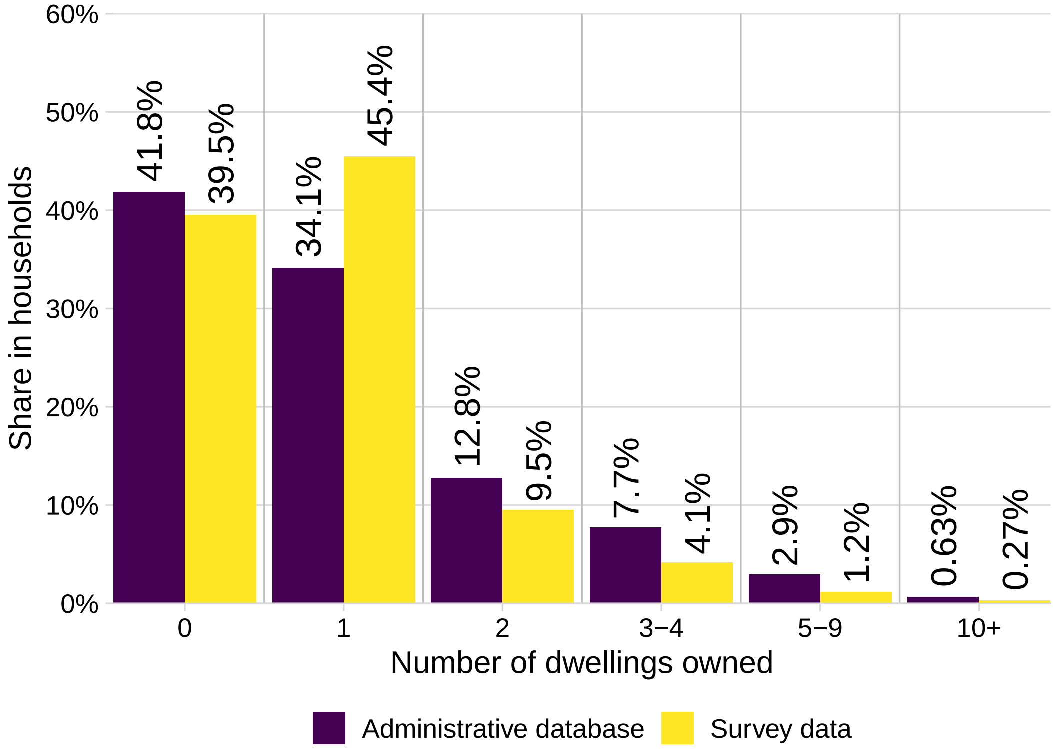


Figure 3. Quantile-quantile plot of the housing wealth distribution

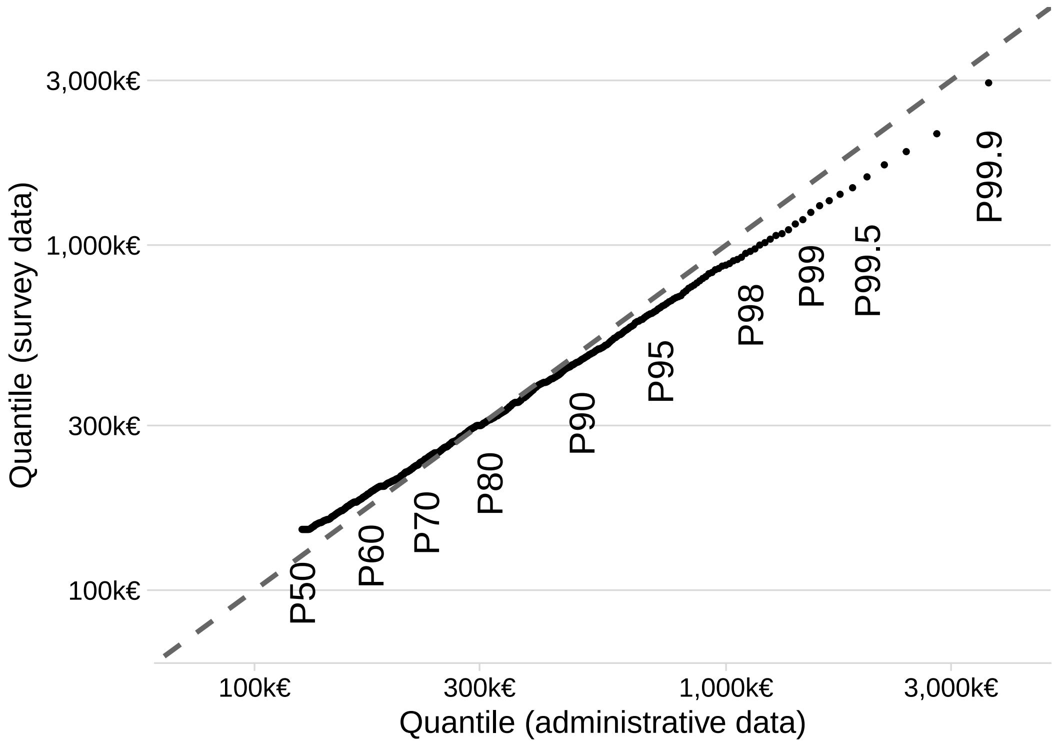


TABLE 3. The concentration of housing wealth is lower in the survey

(a) Share in population

Data source	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%	3.5%	0.53%

(b) Share in housing wealth

Source	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%	16.1%	8.8%
Survey data (true quantiles)	6.2%	29.1%	28.9%	14.2%	15.6%	6.0%

Notes: survey quantiles are computed using the distribution of reported housing wealth (weighted with the final survey weights). True quantiles are computed using the benchmark database. The share in population is a share of individuals, not a share of households.

wealth as of January 1st, 2016, so that the targeting of high net wealth households in year 2017 may be imperfect. In addition, wealth tax returns are an imperfect proxy for true wealth, for at least two reasons: taxpayers may have underreported their assets to reduce their tax liability, and some tax-exempt assets are not reported in wealth tax returns (most notably professional assets), inducing a downward bias for some of the high net wealth households. Another source of uncertainty pertains to the fact that extremely wealthy households are rare and as such may be under- or overrepresented in the realized sample, resulting in an imperfect coverage of the upper tail. This being said, this source of variance in the coverage of the upper tail in the initial sample is unlikely to be a major problem in the present case, because of the strong oversampling of wealthy households in the survey and because the distribution of housing wealth is undoubtedly less skewed than the distribution of financial wealth.

Compared to other household surveys, *unit non-response* is an acute problem in wealth surveys as "wealthy households appear generally to be far more difficult to contact and to persuade to participate" (Kennickell (2017a)): the participation rate decreases sharply with wealth, and can be as low as 25% at the top of the distribution (Kennickell (2017b); Alvargonzález, Barcelo, Bover, Cobreros, Crespo, El Amrani, García-Uribe, Gento, Gómez, Villanueva, et al. (2024); Osier (2016)), inducing large biases in wealth distribution estimates. The standard methodological remedy to this differential unit non-response bias consists in a weight adjustment procedure based on available information on sampled households (see Haziza et Beaumont (2017)). However, if non-responding wealthy households differ systematically from responding ones even conditionally on available controls (for instance by owning more professional assets that go unreported in wealth tax returns), then the weight adjustment procedure does not compensate adequately unit non-response biases. Moreover, even small imperfections in the weight adjustment procedure may induce large and unpredictable changes in wealth concentration estimates, because these estimates are heavily influenced by the small number of very wealthy households included in the sample (see Kennickell (2019) for compelling examples).

Following Kennickell (2017b), three different kinds of *reporting errors* can be distinguished: mistaken answers due to lack of information or to inattention to the questions, deliberately incorrect answers, and answers based on a "conceptual framework that differs in important ways from that intended in the survey design." The standard methodological remedy to reporting errors consists in imputing plausible values and other data editing methods. In the context of my study, reporting errors may concern the number and market value of housing assets, as well as the share owned in each asset, the way it is used and the date of acquisition. An important remark is that deliberately incorrect answers are difficult to tell apart from the two other flavors in practice, as the information available to the respondent and his or her intentions are typically unobserved. This point matters when it comes to finding ways to improve wealth surveys: unlike unintentional mistakes, deliberate mistakes cannot be addressed by modifying the survey design (for instance by asking more precise questions) and requires data imputation based on external information such as administrative data.

3 Where do the discrepancies between administrative data and survey data come from?

In this section, I explain the methodology used to combine the two data sources (3.1), before introducing the decomposition methodology (3.2). The results of this decomposition are presented in subsection 3.3.

3.1 Linking the survey with the benchmark database

Linking sampled households data with the benchmark database I link the initial HVP sample with the benchmark database, using confidential datasets from the production process of the survey. This record linkage procedure is carried out in slightly different ways depending on the subsample and the participation status. For surveyed households, I use personal information collected during the survey (first and last names, date of birth, address) to link all household members with the taxpayers dataset of the 2017 Fidéli database. For non-responding panel households, I apply the same record linkage procedure, using this time personal information collected in the 2014 survey. Finally, for non-responding new entrant households, I simply retrieve all individuals belonging to the sampled fiscal household (see 2.1.1). By doing so, I assume that the composition of the household is exactly the one described in the 2017 Fidéli database. This record linkage procedure is very accurate thanks to the high quality of personal information in both sources: at least one adult household member was identified in the benchmark database for 99.6% of responding households, and all adult household members were identified for 96.3% of responding households[13]. Moreover, the composition of households is very close in both sources: 91% of adult members of surveyed households are actually living in the same housing unit they were supposed to be living in in the benchmark database; this share amounts to 95% among homeowners.

Reconstituting the housing wealth of sampled households Once the record linkage procedure has been carried out, retrieving the list of real estate assets owned by the individuals belonging to sampled households is straightforward. I then reconstitute the real estate assets and the housing wealth of all sampled households by pooling the assets owned by the members of each household. In doing so, I apply the exact HVP survey definition of housing wealth: all housing units owned either in full ownership or in bare ownership, either directly or through a SCI. Once again, equal split between owners is assumed when a housing unit is jointly owned by two or more households (see 2.2.2).

Content of the final database The final database contains two datasets: the household dataset and the housing unit dataset. The household dataset contains two types of information: housing assets and wealth actually owned by sampled households (*true outcomes*), and housing assets and wealth as reported by respondents (*reported outcomes*). The housing unit dataset specifically documents all housing units actually owned by respondents.

This linked data offers several key advantages for investigating the specific limitations of wealth surveys. First, the benchmark dataset closely matches the survey's sampling frame, meaning both the survey and benchmark data cover approximately the same population. Second, the definitions of housing assets and wealth are identical across both sources, implying that the comparison should not be plagued by subtle conceptual differences. Third, the definition of households is identical in both datasets, eliminating usual concerns about inconsistencies of observation units across sources (e.g., tax units versus households).

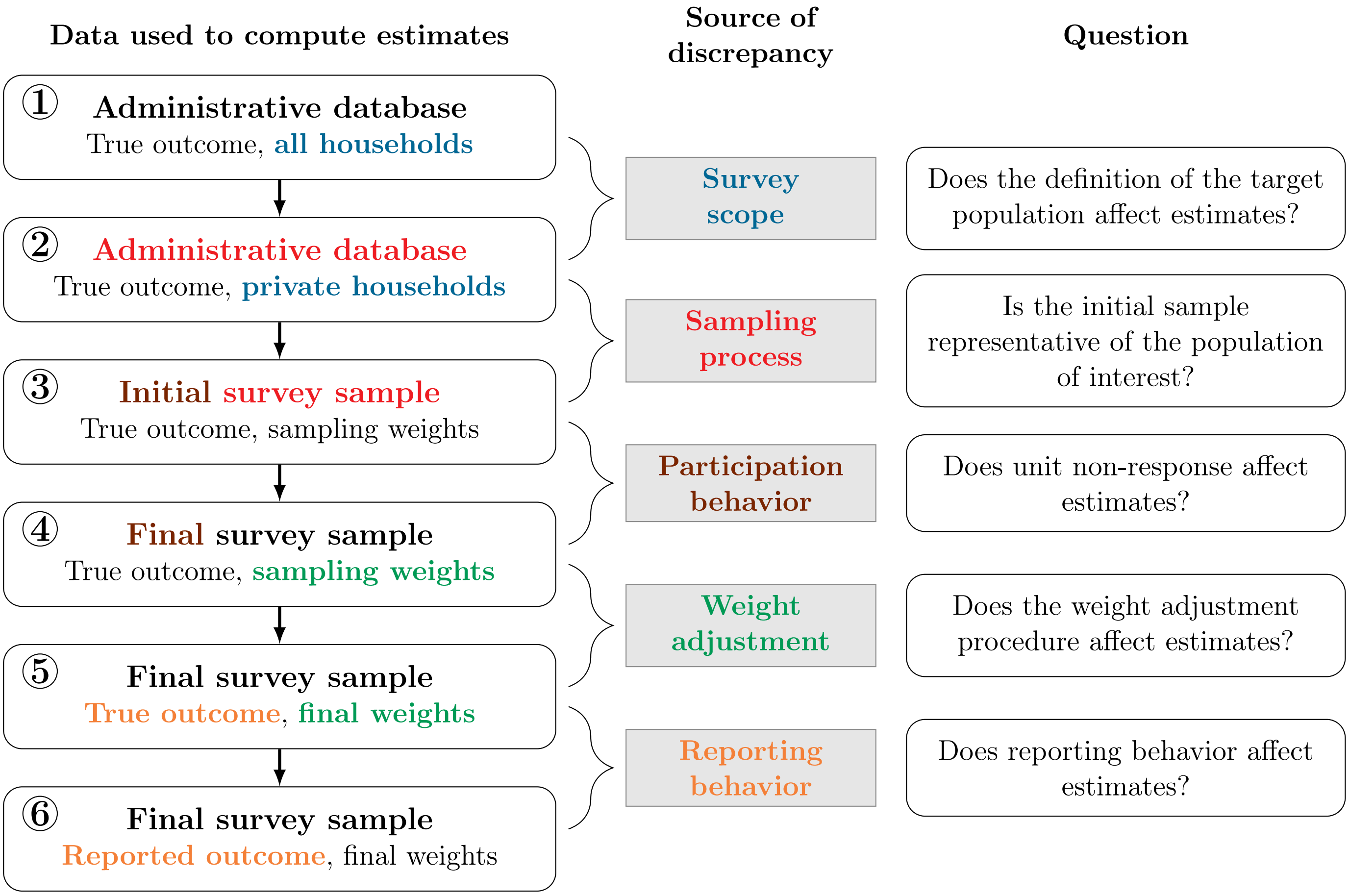
3.2 Decomposing discrepancies

In this section, I introduce a decomposition methodology to measure each source of discrepancy between survey-based estimates and benchmark-based estimates. This approach inspired by Lustig et al. (2020) and Johansson-Tormod et Klevmarken (2022) consists in two simple steps: computing the same outcome using slightly different datasets, and computing pairwise differences. The two key advantages of this decomposition are additivity and systematicity: any discrepancy between survey-based estimates and

[13] Adults are defined are individuals aged 18 and more. Most children belonging to responding households could not be identified in the benchmark database, because the benchmark database does not contain personal information on children. This may induce a downward bias in the reconstituted housing wealth of sampled households (if the children own some real estate assets). However, this bias is likely to be negligible, given that individuals aged 17 or less account for less than 0.5% of all ownership links between housing units and individuals in French cadastral data.

benchmark-based estimates is decomposed into a series of effects that always adds up to the total discrepancy, and no large source of discrepancy can go unnoticed.

Figure 4. Decomposing discrepancies



The first step of this decomposition consists in computing the same outcome of interest (say, the share of total housing wealth owned by the top 1%) using an ordered series of slightly different definitions. What varies across definitions is the precise data used to compute the outcome, the crucial point being that each pair of consecutive definitions differs by exactly one element. The left part of figure 4 entails six definitions; more can (and will later) be introduced, depending on the exact number of sources of discrepancy one wants to isolate. The first and last definitions are exactly identical to the benchmark-based and survey-based estimates presented in section 2.3; the four intermediate definitions are hybrid estimates mixing features of benchmark-based and survey-based estimates (for instance, using survey data but replacing the wealth reported by respondents with wealth estimates from the benchmark database). The second measure is identical to the first measure, except that it restricts the benchmark database to the sampling frame (i.e., all private households). Compared to the second definition, the third definition substitutes the restricted benchmark database with the initial survey sample including both respondents and non-respondents. The fourth measure substitutes the initial sample with the final sample (respondents). The fifth measure substitutes sampling weights with final weights. Finally, the sixth measure substitutes the true outcome with the reported outcome.

The second step of the decomposition consists in computing pairwise differences between consecutive estimates, so as to measure the effect of each source of discrepancy. Comparing ① and ② measures the effect of definition of the target population. Comparing ② and ③ measures the effects of the sampling process. Comparing ③ and ④ measures the effects of participation behavior. Comparing ④ and ⑤ measures the effects of weight adjustment. Comparing ⑤ and ⑥ measures the effects of reporting behavior.[14]

Interpreting these estimates is the most delicate part of this decomposition approach insofar as identifying and measuring a statistical discrepancy does not imply the existence of a measurement bias in the survey. For instance, comparing ① and ② shows that the definition of the target population explains a (small) part of the *discrepancy* between survey-based outcomes and true outcomes. But this finding does not imply that the target population definition used in the survey introduces a *bias* in outcome measurement: the two data sources simply rely on different definitions. As a consequence, I will consider that *a statistical discrepancy is indeed a bias only when its effects clearly contradict the intentions of the survey designers.*

As an illustration, this decomposition is applied to the share of households owning at least one housing units (table 4). The first panel presents the estimates of this share according to the six definitions of figure 4. The second panel presents pairwise differences along with the total discrepancy. These results can be interpreted as follows: the share of households owning at least one housing units is 2.3 percentage points higher in the survey than in the benchmark database. This discrepancy is mostly due to the effect of population coverage (+2.6 pp), reflecting the fact that private households are more frequently homeowners than individuals living in institutions. The participation behavior induces a strong upward bias in this share (reflecting the fact that tenants are generally less likely to participate to the survey), but this bias is almost perfectly corrected by the weight adjustment procedure. Finally, the sampling process and reporting behavior are not major factors explaining the total discrepancy.

3.3 Results

I apply the decomposition approach described above to the distribution and to the concentration of housing wealth. Results are presented in table 5. First, the two main causes

[14]Ideally, it would have been preferable to isolate the effect of data post-processing, by adding one more definition and by distinguishing the raw reported outcome and the final reported outcome (see Kennickell (2015) in the case of the US wealth survey). Unfortunately, reconstituting raw reported housing wealth is not straightforward for two reasons. First, reported information is often incomplete (eg, asset shares are missing). Second, there is no raw reported asset value as respondents are asked to report an interval rather than a point estimate. As a consequence, reconstituting what raw reported housing wealth could have been would require a set of assumptions, in particular regarding how to turn asset value intervals into point estimates. Tests showed that any estimation of the effect of data post-processing would be very sensitive to these assumptions, and might reflect discrepancies between the analyst's assumptions and those made by statisticians, rather than potential biases introduced by data post-processing. I thus decided not to isolate the specific effect of data post-processing.

Table 4. Decomposing the discrepancy in the share of households owning at least one housing unit

Data used to compute the estimate	Estimate
① **Administrative database**	**58.2%**
② Administrative database, true outcome, private households	60.8%
③ Sampled households, true outcome, sampling weights	60.4%
④ Surveyed households, true outcome, sampling weights	66.5%
⑤ Surveyed households, true outcome, final weights	59.8%
⑥ **Survey data**	**60.5%**

	Estimate
Administrative database	**58.2%**
Survey data	**60.5%**
Total discrepancy	**2.3**
Population coverage	2.6
Sampling process	-0.3
Participation behavior	6.1
Weight adjustment	-6.7
Reporting behavior	0.6

of the discrepancies in the top 1% population and wealth shares are unit non-response and reporting behavior, with remarkably similar magnitudes. Second, the effect of unit non-response seems to affect specifically the top 5%, whereas the reporting behavior affects the whole top 10%. Moreover, the reporting behavior effect is particularly strong at the top of the distribution: whereas this source of discrepancy reduces the P90-P99 wealth share by 8% (2.6 pp out of 32.3%), it shrinks the top 1% wealth share by 30% (3.3 pp out of 10.8%). Third, the weight adjustment procedure compensates only partially the unit non-response bias for the top 1% shares, and has an unexpected negative sign for the P95-P99 group. Fourth, the sampling process contributes to the underrepresentation of wealthy households; this is unexpected given that the sampling plan has been designed to ensure an accurate coverage of wealthy households. Fifth, the scope of the survey plays almost no role in explaining the discrepancies between top population and wealth shares.

These findings suggest two general conclusions: the downward bias in housing wealth concentration estimates in the 2017 French wealth comes in equal parts from underrepresentation of wealthy households and underreporting of assets, and the underrepresentation of wealthy households seems to be due to three distinct causes: sample selection in the sampling process, unit non-response and imperfect weight adjustment. The rest of this paper will shed light on the precise mechanisms underlying the underrepresentation of wealthy households (section 4) and the underreporting of assets (section 5).

Table 5. Results of the decomposition

(a) Share in population

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
① Administrative data	**50.0%**	**25.0%**	**15.0%**	**5.0%**	**4.0%**	**1.00%**
⑥ Survey data	**45.2%**	**29.1%**	**16.8%**	**4.9%**	**3.5%**	**0.53%**
Total discrepancy	**-4.8**	**4.1**	**1.8**	**-0.1**	**-0.5**	**-0.47**
Scope of survey	-1.6	0.8	0.5	0.2	0.1	0.03
Sampling process	-0.4	0.3	-0.1	0.3	-0.1	-0.06
Participation behavior	-2.7	2.1	1.0	0.1	-0.2	-0.25
Weight adjustment	3.9	-2.0	-1.3	-0.4	-0.3	0.03
Reporting behavior	-4.0	2.8	1.8	-0.3	-0.1	-0.22

(b) Share in 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%**
Total discrepancy	**-0.2**	**4.5**	**2.9**	**0.1**	**-2.5**	**-4.8**
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

Notes: see text and figure 4 for the methodology.

4 Investigating the wealthy households' underrepresentation

In this section, I investigate the mechanisms underlying the underrepresentation of wealthy households. I first prove that the effect of participation behavior on top population and wealth shares is due to two mechanisms: wealthy households are more difficult to contact, and wealthy households and households owning a large number of housing units are specifically reluctant to participate in the survey (4.1). I then show that the weight adjustment procedure does not fully correct this non-response bias, and that the calibration procedure induces a significant distorsion in the wealth distribution in the survey (4.2). I finally show that the sample selection effect highlighted previously pertains to the lack of representativeness of panel subsample (4.3), induced by both non-ignorable non-response and imperfect weight adjustment in the 2014 survey.

4.1 Where does the unit non-response bias come from?

This paragraph isolates the causes of specific unit non-response among wealthy households and investigates the determinants of this behavior.

Decomposing participation behavior Non-participation in the HVP survey may be due to three unrelated causes intervening at different stages of the data collection process. The household can be *out of the survey scope*: the sampled housing units is vacant, destroyed or not used as a primary residence, all members of the household are dead, the panel household moved abroad... This case is usually not considered as unit non-response. The household can also be in the survey scope but *could not be reached*: the household did not answer to Insee letters and the interviewer could not contact any of its member. Finally, the household could be reached but explicitly *refused to participate*. Households may use a variety of motives to justify their refusal: lack of financial literacy, lack of time, protection of privacy, fear that reported information may be used for tax auditing purposes...

Following Meriküll et Rõõm (2021), I decompose the average response rate of households as the product of three rates: the average in-survey-scope rate, the average contact rate and the average participation rate, with respect to the number of housing units owned by the households (first panel of table 6) and to the position in the housing wealth distribution (second panel of table 6). This decomposition leads to three findings. First, the overall response rate of very large landlords (owning ten housing units and more) and of households belonging to the top 1% is approximately 10 pp lower than the sample average. Second, this lower response rate is driven mostly by a much lower participation rate. This suggests that large landlords and households belonging to the top 1% are specifically reluctant to participate to the survey. Third, the average contact rate decreases with housing wealth and with the number of housing units owned, suggesting that wealthier households and large landlords are somewhat more difficult to reach. Additional results presented in table A1 in appendix 8.1 show that these mechanisms are even stronger among the wealthiest households (strata 1 and 2) than among slightly less wealthy high net wealth households. Overall, these findings are entirely consistent with the literature (Kennickell (2017a), Alvargonzález, Barcelo, Bover, Cobreros, Crespo, El Amrani, García-Uribe, Gento, Gómez, Villanueva, et al. (2024)).

Table 6. Descriptive statistics on participation behavior

(a) By number of housing units owned

Number of housing units owned	Overall response rate	Decomposition of the overall response rate: In-survey-scope rate	Contact rate	Participation rate
0	**55.3%**	85.7%	78.9%	81.8%
1	**69.4%**	94.8%	86.1%	85.1%
2	**68.8%**	94.7%	85.5%	85.0%
3-4	**67.6%**	95.1%	85.0%	83.7%
5-9	**60.0%**	95.0%	82.6%	76.5%
10+	**54.3%**	93.1%	80.8%	72.2%
Whole sample	**63.5%**	**91.9%**	**83.3%**	**82.9%**

(b) By position in the housing wealth distribution

Position in the housing wealth distribution	Overall response rate	Decomposition of the overall response rate: In-survey-scope rate	Contact rate	Participation rate
P0-P50	**58.6%**	87.4%	80.9%	82.9%
P50-P75	**72.1%**	96.2%	87.7%	85.5%
P75-P90	**69.8%**	95.2%	86.8%	84.5%
P90-P95	**69.3%**	95.8%	86.0%	84.1%
P95-P99	**61.9%**	94.9%	80.5%	81.1%
P99-P100	**51.8%**	93.7%	77.6%	71.3%
Whole sample	**63.5%**	**91.9%**	**83.3%**	**82.9%**

Notes: all figures are unweighted. The overall response rate is the product of the three rates on the right side of the table.

The specific effect of each of the three non-response causes on population and wealth shares can be measured by decomposing the discrepancy between steps ③ and ④ of figure 4 in three substeps. The first substep substitutes the initial sample with the initial sample restricted to in-survey-scope households. The second substep substitutes the initial sample restricted to in-survey-scope households with the initial sample restricted to contacted households. The third substep substitutes the initial sample restricted to contacted households with the final survey sample (the households that were actually surveyed). Results are presented in table 7, with two findings. First, difficulty to contact wealthy households reduces the top 5% population and wealth shares (by -0.27 pp and -1.5 pp respectively). Second, refusal to participate has the largest effect and affects specifically the top 1% population and wealth shares (by -0.21 pp and -2.1 pp respectively).

Table 7. Effects of participation behavior on population and wealth shares

(a) Share in population

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
③ Initial sample	**48.0%**	**26.1%**	**15.4%**	**5.45%**	**4.07%**	**0.97%**
④ Final sample	**45.2%**	**28.2%**	**16.4%**	**5.54%**	**3.86%**	**0.72%**
Effect of participation behavior	**-2.7**	**2.1**	**1.0**	**0.09**	**-0.21**	**-0.25**
Out of survey scope	-1.5	0.8	0.4	0.14	0.11	0.02
Could not be reached	-1.0	0.9	0.4	0.02	-0.22	-0.05
Refused to participate	-0.1	0.4	0.2	-0.07	-0.10	-0.22

(b) Share in 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%**
Effect of participation behavior	**0.1**	**1.8**	**1.1**	**0.7**	**-0.8**	**-2.9**
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

Notes: see text and figure 4 for the methodology. All shares are computed using sampling weights.

Is unit non-response (conditionally) ignorable? Lower contact and participation rates behavior among wealthier households and large landlords may be explained by many factors other than wealth and assets. For instance, households with a higher standard of living might be more difficult to contact and specifically reluctant to participate, inducing the observed patterns because standard of living, housing assets and housing wealth have a strong positive correlation.

In order to isolate the specific effect of assets and wealth on the participation behavior, I run a logistic regression on three binary outcomes: whether the housing units or household is in the survey scope, whether the interviewer reached the household (conditionally on being in the survey scope), and whether the household participated to the survey (conditionally on having been reached). Covariates include: subsample (panel or new entrant), size of the household, standard of living, region, high net wealth status (standard/high net wealth/very high net wealth, see table 1), age, number of housing units owned by the household, number of owner-occupied housing units owned by the household, and the household's position in the housing wealth distribution. Results are presented in figure 5 in the form of odd-ratio relative to the reference category, along with 95% confidence intervals. An odd ratio larger than 1 means that the related group has a larger probability to be in the survey scope, be reached, or participate to the survey than the reference group. Four findings can be drawn from this figure. First, households of the panel subsample are easier to reach and more likely to participate than new entrants households, suggesting that there might be systematic differences between the two subsamples (see 4.3). Second, the contact rate and the participation rate vary with respect to households' socio-economic characteristics: households with a higher standard of living are easier to reach and more likely to participate; two-person households (mostly couples without children) are also more prone to participate; households older than 80 are more often out of scope and more difficult to contact (maybe because they actually live in nursing homes). Third, and most importantly, the contact rate and the participation rate vary with respect to households' assets and wealth *all other things being equal*: high net wealth households are significantly more difficult to reach and less likely to participate, and large landlords and households belonging to the top 1% are significantly less likely to participate. Fourth, the in-scope rate is not highly correlated with any covariate, except the subsample[15], the occupancy status[16] and the high net wealth status. Based on these findings, I conclude that unit non-response among wealthy households is conditionally non-ignorable based on the information available at the time the survey is conducted[17] (as it is directly correlated to the outcomes of interest such as the number of housing units owned), and that its main causes lies in the specific difficulty to contact wealthy households and in their reluctance to participate to the survey.

4.2 Does the reweighting procedure mitigate the non-response bias?

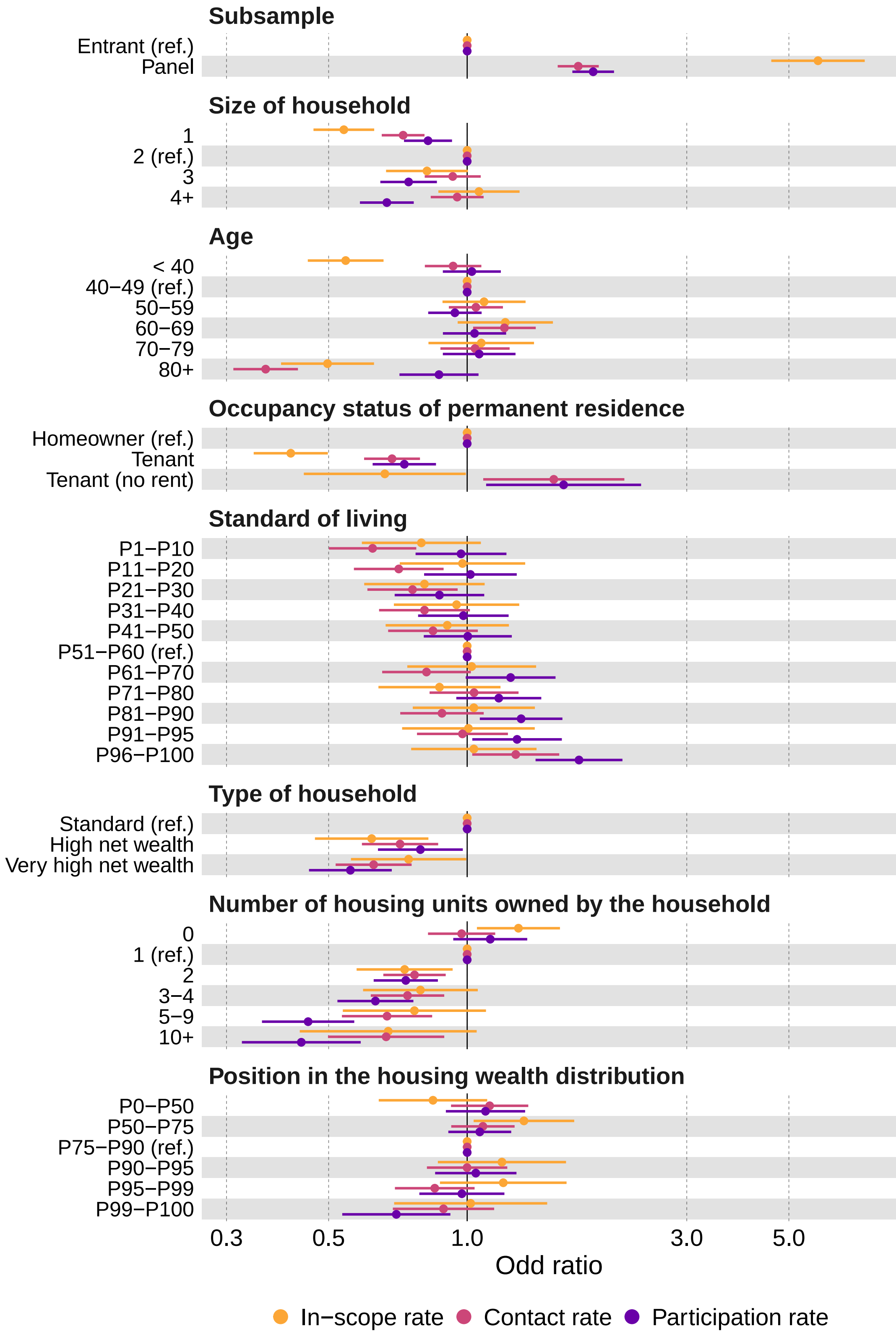
This paragraph investigates the effects of the reweighting procedure that consists in two stages. First, sampling weights are modified to account for unit non-response using the homogeneous response groups method (Haziza et Beaumont (2017); Caron (2005)). In this approach, the response probability of each household is estimated using a logit model, based on the information available in the sampling frame. Households are then clustered in homogeneous response groups (where all households have a similar response probability). Finally, the sampling weights of respondents of each group is divided by the average weighted response rate of the group. Second, weights are adjusted through a calibration approach so that the sample of respondents is representative of French households with respect to several variables such as the type of household, the place of residence, the age, highest degree and socio-economic status of the reference person. Importantly, earned income, capital income and net taxable wealth as measured by

[15] By definition, households included in the panel subsample were in the survey scope in 2014, and may be out of it in 2017 only for a limited number of reasons: all members are dead, moved abroad or to an institution such as a nursing home. Conversely, a housing units sampled in the new entrant subsample may be out of the survey scope for many reasons (the housing units is vacant, not used as a primary residence, impossible to find, destroyed...). As a consequence, panel households are mechanically much more likely to be in the scope of the survey.

[16] This reflects the fact that rental housing units are more frequently vacant than owner-occupied housing units.

[17] Of course, using the administrative database in subsequent survey waves may help mitigating this problem.

FIGURE 5. **Determinants of participation behavior**



income and wealth tax returns (only for the three high net wealth strata) are included in these marginal calibration variables.

Table 8. Effects of weight adjustment on population and wealth shares

(a) Share in population

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
④ Final sample, sampling weights	**45.2%**	**28.2%**	**16.4%**	**5.54%**	**3.86%**	**0.72%**
⑤ Final sample, final weights	**49.2%**	**26.3%**	**15.1%**	**5.17%**	**3.59%**	**0.75%**
Total discrepancy	**3.9**	**-2.0**	**-1.3**	**-0.37**	**-0.27**	**0.03**
Correction of non-response	1.6	-1.2	-0.5	-0.01	0.10	0.11
Marginal calibration	2.4	-0.7	-0.8	-0.36	-0.37	-0.08

(b) Share in housing wealth

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
④ Final sample, sampling weights	**6.3%**	**26.3%**	**26.9%**	**15.5%**	**17.1%**	**7.8%**
⑤ Final sample, final weights	**6.4%**	**25.9%**	**26.1%**	**15.5%**	**16.7%**	**9.4%**
Total discrepancy	**0.2**	**-0.4**	**-0.8**	**-0.1**	**-0.4**	**1.6**
Correction of non-response	-0.1	-1.0	-0.7	0.0	0.5	1.3
Marginal calibration	0.3	0.6	-0.1	-0.1	-0.9	0.3

Notes: see text and figure 4.

The specific effect of each of these reweighting stages can be precisely measured by decomposing the discrepancy between steps ④ and ⑤ of figure 4 in two substeps. The first substep substitutes initial sampling weights with sampling weights corrected for unit non-response. The second substep substitutes sampling weights corrected for unit non-response with final weights (ie after calibration). Results are presented in table 8, with two findings. First, the correction for non-response works as expected and corrects almost half of the unit non-response biases in the top 1% population and wealth shares: this correction increases the top 1% population and wealth shares by 0.11 pp and 1.3 pp respectively, compared with unit non-response biases of -0.23 pp and -2.9 pp respectively. Second, the calibration procedure distorts the distribution in unexpected ways, increasing slightly the weight of the top 1% but decreasing the weight of the P95-P99 wealth group. This is due to distortions in the structure of the final sample (table 9): whereas calibration does not modify the sample share of the two strata of very wealthy households, it reduces the sample share of other wealthy households by 13% (from 0.83% to 0.72%). These distortions are hardly surprising from a theoretical standpoint: calibration methods ensure that the survey-based weighted sum of calibration variables are close or identical to the sum measured on the whole population, but may modify the distribution of other, non-calibrated variables in unexpected ways. These changes in distributions are difficult to predict, apart from two general considerations: a larger number of calibrated variables tend to increase the dispersion of calibrated weights, and the changes in distributions tend to be larger for variables that are poorly related to the calibration variables (see Haziza et Beaumont (2017)). These distortions may be mitigated by including information on sampling strata in calibration variables (see section 6).

Table 9. Descriptive statistics on the reweighting of the new entrants subsample

Description	Share in population	Share within respondents	Share after correction for non-response	Share in the final sample
Very HNW households, urban	0.09%	0.05%	0.08%	0.08%
Very HNW households, rural	0.11%	0.09%	0.11%	0.11%
Other HNW households	0.82%	0.71%	0.83%	0.72%
All HNW households	*1.02%*	*0.85%*	*1.03%*	*0.91%*
Older households	41.2%	41.8%	41.6%	40.0%
High business income	2.8%	2.8%	2.8%	3.0%
High capital income	2.3%	2.3%	2.3%	2.1%
High wages	4.9%	4.7%	4.8%	4.5%
All other households	47.8%	47.6%	47.4%	49.4%
All standard households	*99.0%*	*99.1%*	*99.0%*	*99.1%*

Notes: HNW stands for "high net wealth". This table reads like this: high net wealth households account for 1.02% of the total population, but only for 0.85% among respondents. The correction for unit non-response increases this share to 1.03%. Finally, calibration reduces this share to 0.91%.

4.3 Is something wrong with the survey sample?

Finding that the sampling process contributes to the underrepresentation of wealthy households is surprising, given that the sampling plan has been carefully designed to ensure an accurate representation of wealthy households. I argue that this discrepancy is indeed a bias, entirely due to the lack of representativeness of the panel subsample. This can be shown by recomputing outcomes of interest using definition ③, using only the new entrant subsample, and then only the panel subsample, and comparing these estimates with outcomes of interest computed using definition ②. Results are presented in table 10. The new entrant subsample appears to be almost perfectly representative of the sampling frame (all private households), suggesting no major problem in the sampling plan. For instance, the top 1% wealth share computed on this subsample is very close to the top 1% wealth share computed on the sampling frame (11.1% and 10.8% respectively). Conversely, the panel subsample lacks representativeness: the P50-P90 household group is overrepresented (population share of 44.7%, instead of 40%) and the top 1% wealth share is 0.6 pp below the top 1% wealth share computed on all private households.

Given that the HVP survey methodology has remained mostly unchanged between 2014 and 2017 (except for the introduction of the panel subsample), the most likely explanation of this bias is that the 2014 HVP survey was affected by the two problems described in sections 4.1 and 4.2: non-ignorable non-response and imperfect weight adjustment by the homogeneous response group method. This led to a lack of representativeness of the final 2014 survey sample, that simply carried over to the 2017 survey through the panel subsample.

Table 10. Effects of the sampling process on population and wealth shares

(a) Share in population

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-100
② Sampling frame	**48.4%**	**25.8%**	**15.5%**	**5.2%**	**4.1%**	**1.03%**
③ Initial sample	**48.0%**	**26.1%**	**15.4%**	**5.4%**	**4.1%**	**0.97%**
New entrants subsample	49.8%	25.1%	15.0%	5.2%	4.0%	0.99%
Panel subsample	44.2%	28.3%	16.3%	6.0%	4.2%	0.94%

(b) Share in housing wealth

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-100
② Sampling frame	**6.4%**	**24.6%**	**26.0%**	**14.1%**	**18.2%**	**10.8%**
③ Initial sample	**6.2%**	**24.5%**	**25.7%**	**14.8%**	**18.0%**	**10.8%**
New entrants subsample	6.3%	24.3%	25.7%	14.4%	18.2%	11.1%
Panel subsample	5.9%	25.0%	25.9%	15.6%	17.4%	10.2%

Notes: see text and figure 4. All shares are computed using sampling weights.

This finding is far from anecdotal for two reasons. First, this sample selection problem may become significantly larger if the weight of the panel subsample increases in subsequent waves of the survey. This is already the case in the 2020 wave of the survey, where panel households account for roughly 50% of the initial sample (and new entrants for 50%), as opposed to only roughly 25% in the 2017 wave. Second, the longitudinal dimension has been introduced by most countries participating in the Household Finance and Consumption Survey (13 countries as of 2021, see Network (2023)), meaning that this problem is likely to affect most recent wealth surveys in Europe.

5 Investigating the respondents' reporting behavior

In this section, I investigate the mechanisms underlying the reporting behavior of respondents. I first show that housing wealth underreporting is almost entirely caused by underreporting in the number of housing assets owned by respondents rather than asset value underestimation (5.1). I then compare reported assets with assets actually owned by respondents to uncover the determinants of asset reporting (5.2). I then study the asset value estimation behavior by comparing reported asset values with statistical market value estimates and real estate transaction data (5.3). I finally focus on the top of the distribution and compare wealth reported in the survey with housing wealth estimates available in the benchmark database and with housing wealth tax returns (5.4).

5.1 Where does the reporting bias come from?

The reporting behavior affects survey-based population and wealth shares through three channels. First, the market value of housing assets reported by households may be inaccurate or inconsistent with market conditions (*underestimation*). Second, the number of housing assets reported by households may be inconsistent with administrative data (*underreporting*). This underreporting may take two forms: either a housing asset goes entirely unreported (*underreporting at the extensive margin*) or the number of flats per housing asset may be underreported (*underreporting at the intensive margin*) because flats may be reported jointly (see 2.1.2). For instance, a wealthy household owning a 20-flats building may report the housing asset in the survey but with 10 flats only. Third, a household may be re-classified in a different wealth group if final reported wealth differs from true wealth (*household reshuffling effect*). For instance, imagine that a household with a 2 M€ true housing wealth reports only 1 M€ in the survey. Based on its true wealth, this household is classified in the top 1% wealth group (because the 99$^{\text{th}}$ quantile of the housing wealth distribution amounts to 1.512 M€), but based on its reported wealth this household is reclassified to a lower wealth group.

The effect of these three mechanisms on population and wealth shares can be measured by decomposing the discrepancy between steps ⑤ and ⑥ of figure 4 in four substeps. The first substep substitutes the average estimated housing units value with the average housing units value reported by respondents, measuring the effect of asset value estimation. The second substep substitutes the true number of housing units owned by respondents with the number of housing units reported by respondents, distinguishing what comes from the number of housing assets and from the number of housing units per housing asset. In these two substeps, the wealth groups household belong to are kept unchanged (and are defined based on their true housing wealth). Finally, the fourth substep substitutes true wealth with final reported wealth when allocating households to wealth groups. Results are presented in table 11.

First, housing wealth underreporting reduces significantly the apparent share of wealthy households in the survey (first panel of table 11): whereas 0.75% of surveyed households do actually belong to the true top 1%, this share decreases to only 0.53% when classifying households based on their reported wealth. In other words, at least 30% of interviewed households belonging to the true top 1% are misclassified in a lower wealth group because they underreported their housing wealth. Second, underreporting in the number of housing units owned by households is the major cause of the underreporting bias in housing wealth concentration. It affects all of the top 25% of the housing wealth distribution, and its magnitude increases with the position in the top tail: this source of underreporting reduces the P90-P95 wealth share by 12% (1.8 pp out of 15.5%), the P95-P99 wealth share by 21% (3.5 pp out of 16.9%) and the top 1% wealth share by 32% (3 pp out of 9.3%). Third, most of the underreporting of housing units happens at the extensive margin (underreporting in the number of housing assets) rather than at the intensive margin (underreporting in the number of housing units by housing asset).

TABLE 11. **Effects of asset reporting on population and wealth shares**

(a) Share in population

	P0-P50	P50-P75	P75-P90	P90-P95	P95-P99	P99-P100
⑤ Final sample, true wealth	**49.2%**	**26.3%**	**15.1%**	**5.2%**	**3.6%**	**0.75%**
⑥ Final sample, reported wealth	**45.2%**	**29.1%**	**16.8%**	**4.9%**	**3.5%**	**0.53%**
Effect of reporting behavior	**-4.0**	**2.8**	**1.8**	**-0.3**	**-0.1**	**-0.22**

(b) Share in 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%**
Effect of reporting behavior	**-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
Ratio housing units/location	*-0.1*	*-0.1*	*0.9*	*0.0*	*-0.1*	*-0.6*
Number of locations	*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

Notes: see text and figure 4.

Fourth, the asset valuation behavior does not play a significant role, except by reducing marginally the top 1% wealth share. Fifth, the true top 1% is the only wealth group cumulating all three flavours of underreporting.

5.2 What properties get reported, and why?

In this subsection, I investigate the determinants of asset reporting and answer two different questions. First, why are some assets more likely to be reported than others? Second, are wealthy households less likely to report their assets, other things being equal?

Detecting reported housing units Understanding the determinants of asset reporting based on the linked data may seem straightforward: one should build a one-to-one mapping between reported assets and assets owned by respondents, and then investigate why some assets went unreported. Unfortunately, such a one-to-one mapping happens to be unfeasible for the whole set of housing units[18], for two different reasons. First, information on housing units reported in the survey is scarce and often somewhat inaccurate. For instance, the owner of a housing units located in the close suburbs of Paris may report to own a housing units in Paris, or the year of purchase may be off by a few years. Second, as explained in section 2.1.2, households may report several flats jointly if two conditions are met: the flats must be located in the same building and must be

[18]However, this one-to-one mapping is feasible for a subset of housing units, see section 5.3.

used in the same way. This makes a one-to-one mapping a meaningless exercise when the reported number of flats is inaccurate. Imagine for instance that a wealthy household owning 20 flats in a building in Paris reports in the survey that it owns exactly 8 flats in this building: it makes no sense to try to detect which among the 20 flats are the reported ones, and which are the unreported ones.

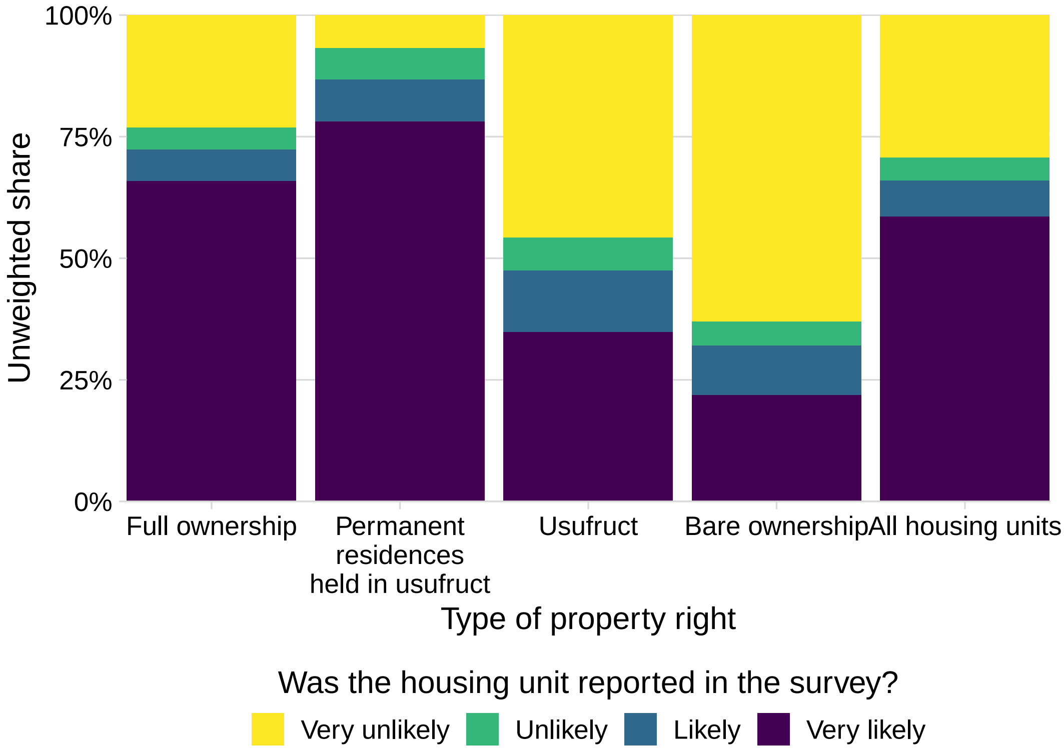
Table 12. Reporting probability: definition and summary statistics

Reporting likelihood	Definition	Share of housing units	Share of housing assets
Very likely	Same municipality AND Same type of housing unit AND Similar years of purchase OR the household owns housing units at only one location in the municipality	58.7%	61.8%
Likely	Same municipality AND Same type of housing unit	7.3%	7.3%
Unlikely	Same municipality	4.7%	3.5%
Very unlikely	All the rest	29.3%	27.4%

As a consequence, I followed a different path: I approximate the likelihood that housing units owned by a household were reported by comparing them to all the housing units reported by this household, based on available information (municipality, type of housing units, year of purchase). The more similar a housing units is to one of the reported housing units, the higher the reporting likelihood. I define a four-degree scale of reporting likelihood detailed in table 12. Although such a likelihood scale is somewhat imprecise by definition, it is worth stressing that a housing units is classified as very unlikely to have been reported only if the household reported no housing units located in the same municipality. A significant limitation is that all housing units located at the same address and reported jointly have the same degree of reporting likelihood even if the number of housing units at this address has been underreported. As a robustness check, I apply the same methodology to housing assets. Summary statistics reported in the two last columns of table 12 show that this approach successfully distinguishes two clear sets of housing units (or housing assets), as most housing units are either very likely or very unlikely to have been reported. Only 12% of housing units and 11% of housing assets owned by respondents are in intermediate, ambiguous categories.

What question do households actually answer? A striking finding is that answers given by households are frequently inconsistent with the question asked on housing assets: housing units held in usufruct are more likely to be reported than housing units held in bare ownership, although households are explicitly asked to report housing units held in bare ownership but not those held in usufruct (except for the primary residence). Among households owning housing units in bare ownership, 70% are likely to have forgotten to report at least one of those housing units. Conversely, among households owning housing units in usufruct other than its primary residence, 41% are likely to have erroneously reported at least one of those housing units. And among the small number of households owning simultaneously housing units in bare ownership and housing units in usufruct, 32% are likely to have made a double mistake: reporting at least one housing units held in usufruct and forgetting to report at least one housing units held in bare ownership.

FIGURE 6. Reporting likelihood and ownership rights



Notes: all shares are unweighted.

As a consequence, around 70% of all housing units held in bare ownership by respondents are unlikely or very unlikely to have been reported whereas almost half of the housing units held in usufruct other than the primary residence are likely or very likely to have been reported (figure 6). Moreover, housing units held in full ownership are more likely to be reported than those held in usufruct and bare ownership. These ubiquitous misreporting patterns suggest that respondents report assets based on a definition of housing wealth "that differs in important ways from that intended in the survey design" (Kennickell (2017b)). Given that the main difference between usufruct and bare ownership lies in the usufructuary having daily economic control (using the housing units, deriving an income from it and paying the associated property tax), I hypothesize that households tend to report assets they have full legal and daily economic control upon. The next paragraph provides compelling empirical evidence supporting this hypothesis.

Understanding asset reporting In order to investigate the determinants of asset reporting, I estimate a logistic regression on a dataset describing all housing units actually owned by respondents. The dependent variable is a dummy indicating whether a housing units is very likely to have been reported. As a robustness check, I also estimate the same model at the housing asset level[19]. All covariates are binary or categorical and cover five different topics: the features of the housing units, the precise ownership right held by the household, variables describing the socio-economic situation of the household, variables describing the respondents and subjective evaluation by the interviewer. All variables are described in table A2 in Appendix 8.2. Results are presented in figures 7 to 9 in the form of odd-ratio relative to the reference category, along with 95% confidence intervals. An odd ratio larger than 1 means that housing units (or housing assets) of the considered category have a larger probability to be reported in the survey than housing units (or housing assets) belonging to the reference category.

Five conclusions can be drawn from the results. First, households are more likely to report housing units that they have full legal control upon, i.e. housing units that they own in full ownership, without intermediation through an SCI, and whose ownership they do not share with other households. Conversely, housing units owned in bare ownership, through an SCI or jointly with other households are less likely to be reported. Second, households are more likely to declare large assets (housing assets where they own a large number of housing units) and those housing units they use as a primary residence, pay property taxes on, or derive rental income from. This pattern suggests that households predominantly report properties over which they exercise daily economic control and contributes to explain why housing units held in usufruct are declared more frequently than those held in bare ownership. Third, the probability of reporting varies with the characteristics of respondents: it increases with their educational level and decreases with their age. Importantly, a housing units is significantly more likely to be reported when at least one respondent has a ownership right on it, suggesting either that information on assets may not be perfectly shared between household members, or that it is easier for respondents to remember their own assets rather than the assets of other household members. Fourth, the reporting likelihood of housing units varies with the attitude of respondents during the interview: housing units are less likely to be reported when respondents are mistrustful, do not use documents and seem to be hiding some information according to the interviewer. Fifth and most importantly, housing units owned by large landlords are much less likely to be reported in the survey *all other things being equal*, as well as housing units owned by high net wealth households, whereas reporting likelihood is not significantly correlated with standard of living or the position in the housing wealth distribution. This suggests that asset underreporting is a specific feature of large landlords and wealthy households, more than a specific feature of households at the top of the housing wealth distribution.

[19] In this case the dependent variable is a dummy indicating whether at least one housing unit located at the housing asset is very likely to have been reported.

Figure 7. Determinants of reporting behavior (1)

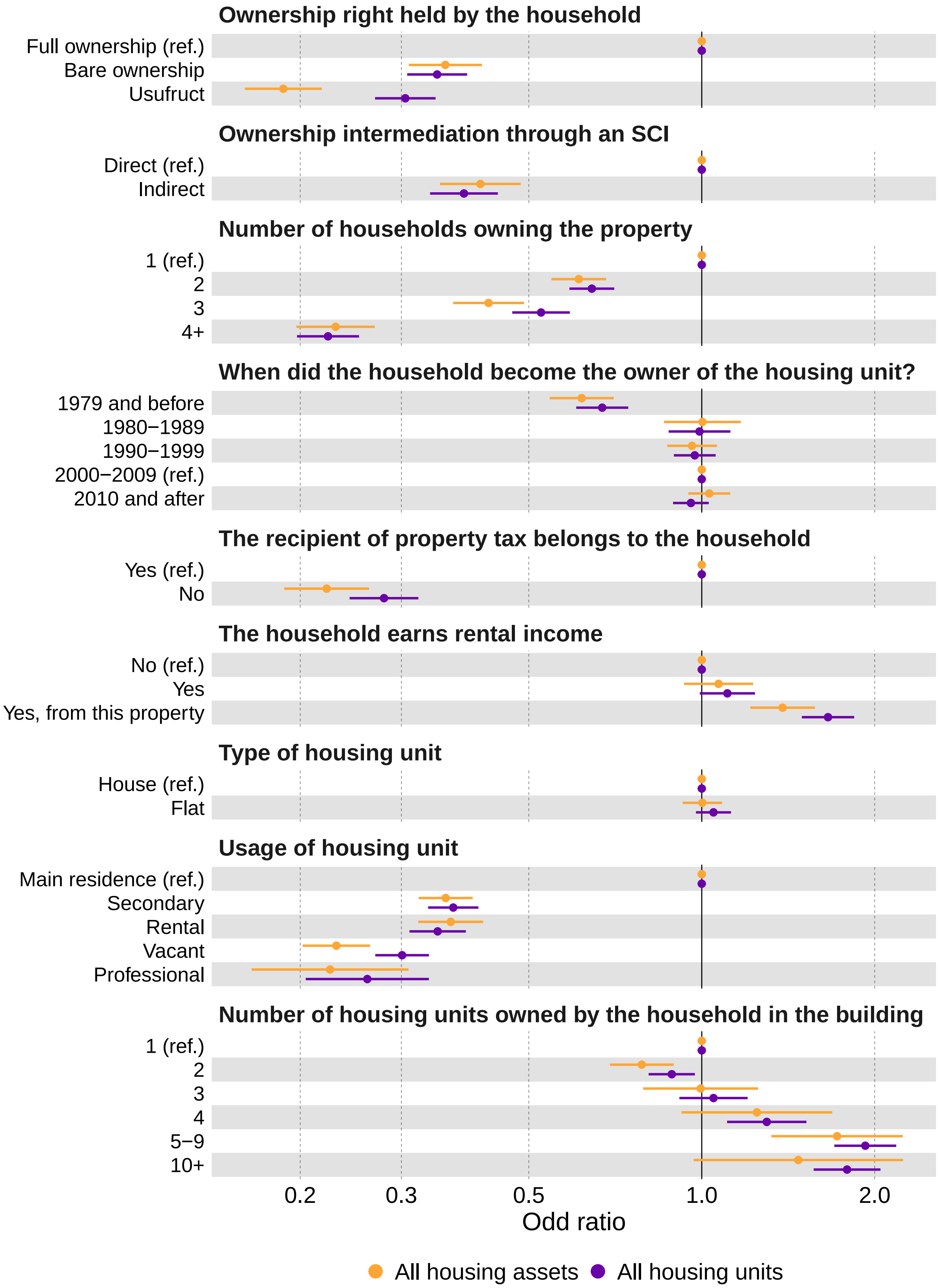


FIGURE 8. Determinants of reporting behavior (2)

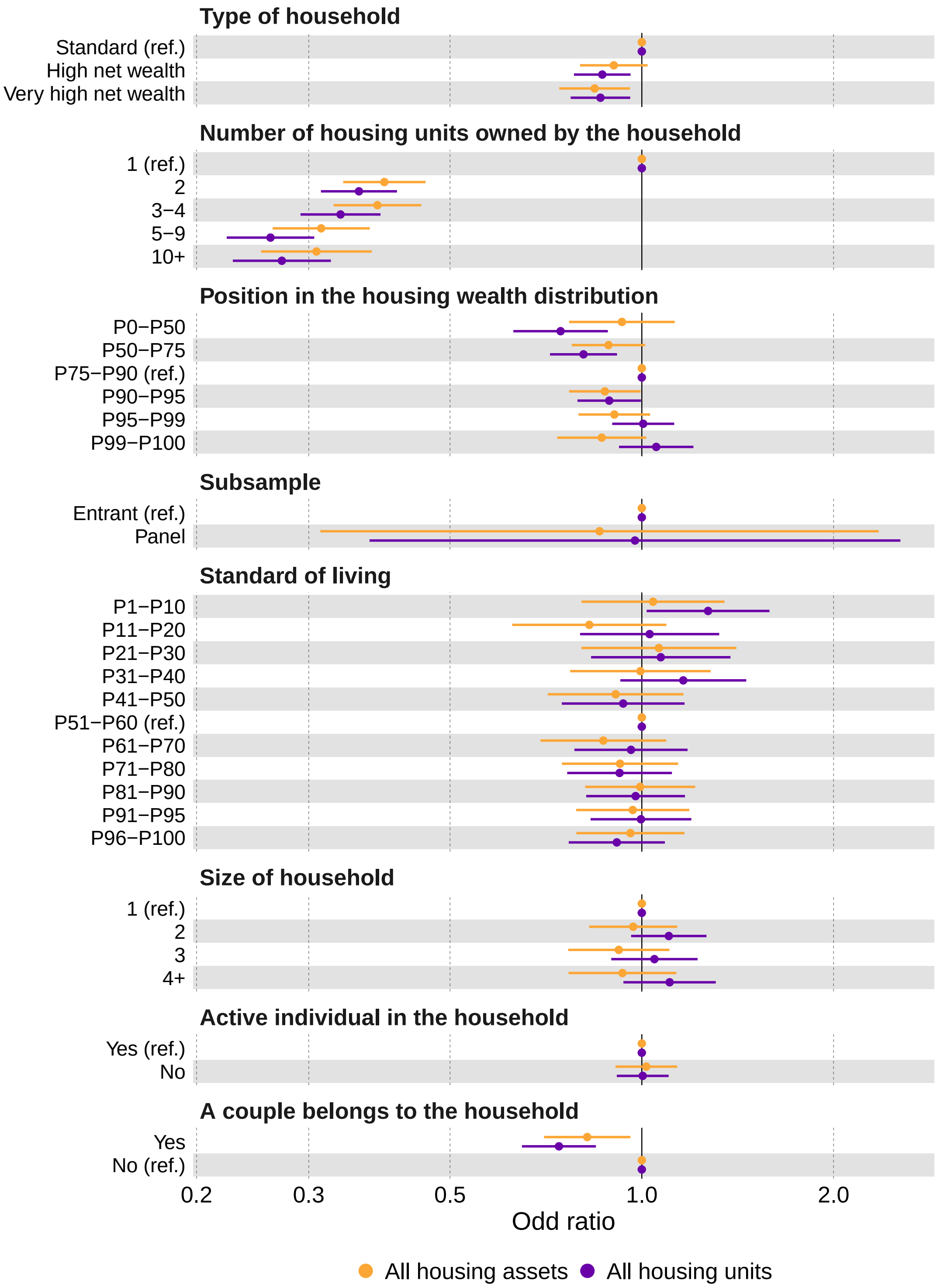
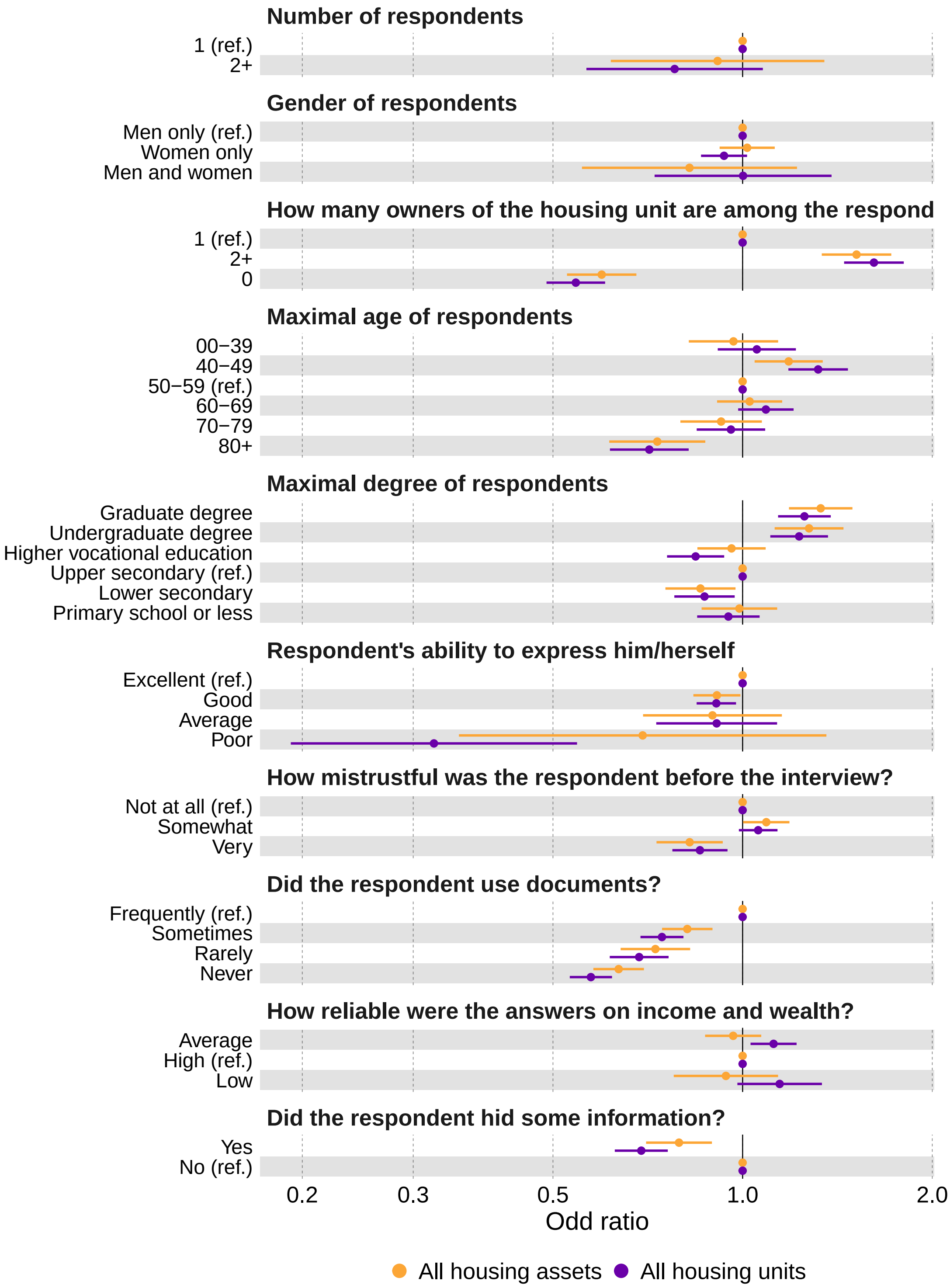


FIGURE 9. Determinants of reporting behavior (3)



5.3 How well do households estimate the value of their assets?

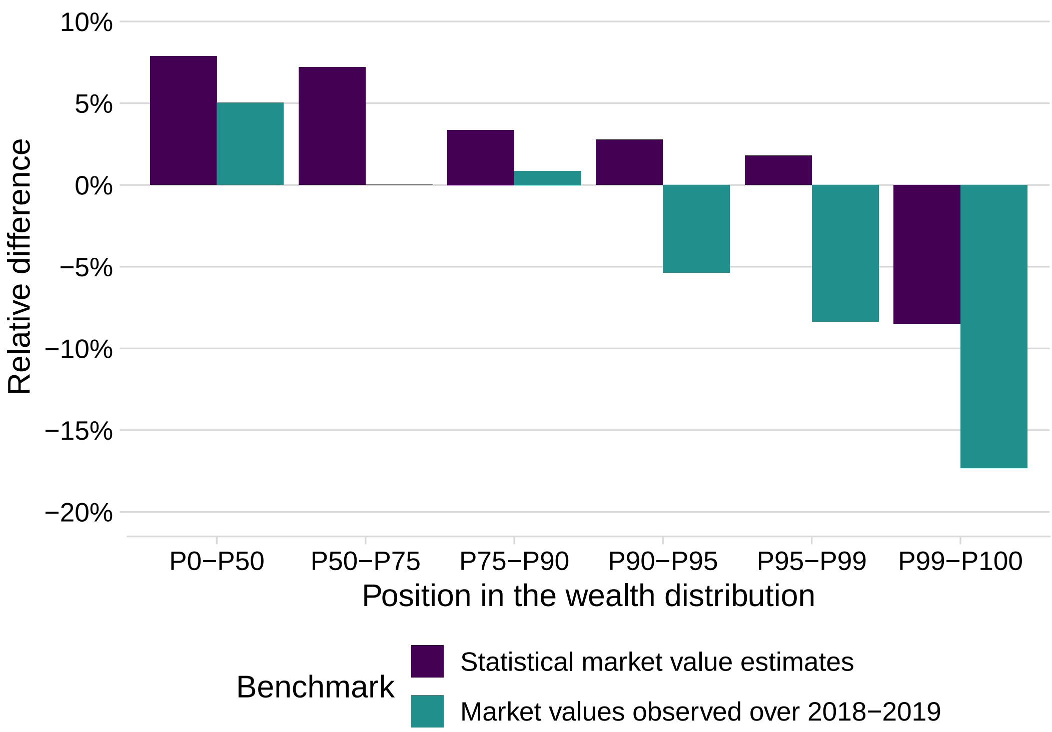
In this section I analyze the asset valuation behavior of respondents with two objectives: investigating the causes of asset under- and overvaluation, and determining whether some groups of households tend to over- or underestimate the market value of their real estate assets, particularly at the top of the housing wealth distribution. To do so, I perform a one-to-one mapping between reported assets and assets owned by respondents on the subset of primary residences, and then compare reported asset values with two benchmarks: the statistical estimates available in the benchmark database, and observed market values available in real estate transaction data.

Linking primary residences with the benchmark database and real estate transaction data Though unfeasible for all housing units (see 5.2), a one-to-one mapping between reported assets and assets owned by respondents is possible for the subset of primary residences, for three reasons. First, the new entrant subsample is a sample of housing units used as primary residences (either by tenants or by homeowners) and already contains cadastral identifiers of sampled housing units, so the mapping is straightforward as soon as the household living in the sampled housing units reports to be a homeowner. Second, for the panel subsample, information reported on the primary residence is often more precise and reliable than on other housing units, allowing for a one-to-one mapping. Third, the joint reporting of flats described in section 2.1.2 is not an issue for primary residences, because most households do not use multiple flats located at the same housing asset as their primary residence.

The mapping procedure works as follows. For new entrant households, a housing units described in the initial sample is mapped to the primary residence reported by the household if two conditions are met: members of the sampled household own the housing units described in the initial sample and the household reports to be homeowner. The procedure is slightly more involved for panel households: the reported primary residence is mapped to one of the housing units owned by a panel household only if the household owns only one housing units used as an owner-occupied housing units and located in the same municipality as the reported one. This procedure maps 8 862 reported primary residences to housing units owned by respondents, out of 9 613 homeowner households (92.2%).

Comparing reported asset values with external benchmarks Asset values reported for primary residences in the survey can be confronted with two benchmarks: *statistical market value estimates* and *observed market values*. Statistical estimates are computed using the valuation methodology described in section 2.2.2, and are identical to asset values available in the benchmark database, except for one thing: in the present section market values are estimated for the fourth quarter of 2017, so that they are perfectly contemporaneous with the survey. Observed market values come from real estate transaction data (the *Demande de Valeurs Foncières* database covering the 2017-2023 period).[20] I compare reported market values to these two benchmarks because each of them has significant limitations. The benchmark database gives statistical estimates for almost all primary residences (meaning no selection bias) and at the survey data collection date (meaning no bias due to housing price inflation), but the reliability of the underlying statistical model might be questioned (see Johansson-Tormod et Klevmarken (2022)), although André et Meslin (2025) strongly suggest that these flaws are small and unsystematic as soon as sufficiently large groups of housing units are considered. Conversely, real estate transaction data gives reliable market values observed in actual transactions, but only for the potentially selected subset of 1 159 primary residences sold in the years following the survey. Moreover, the comparison between asset values reported in the 2017 survey and market values observed in the subsequent years is obfuscated by the strong housing price inflation observed over the 2017-2022 period (+6.5% between 2017 and 2019, +28% between 2017 and 2022). For this reason, I focus only on housing units sold within two years after the survey[21] (in 2018 and 2019). I argue that comparing reported asset values with these two benchmark datasets may lead to conclusive cumulative evidence because the two benchmarks do not share the same limitations.

FIGURE 10. Comparing reported asset values with external benchmarks



[20] Once reported primary residences have been mapped to the benchmark database, finding transactions on these housing units in real estate transaction data is straightforward because both the benchmark database and the transaction dataset contain cadastral identifiers.

[21] Figure A1 in appendix 8.3 presents the results for all primary residences sold after the survey, with qualitatively unchanged results.

For each benchmark, I compute the average relative difference between reported asset values and benchmark values, expressed as a share of benchmark values. For instance, a 12% relative difference between reported values and statistical estimates means that on average reported values are 12% higher than statistical estimates. Results are presented in figure 10, with two findings. First, the comparison with statistical market value estimates suggests a clear pattern: households belonging to the bottom 75% of the housing wealth distribution seem to overestimate slightly the value of their primary residences on average, the values reported by households between the 75$^{\text{th}}$ and the 99$^{\text{th}}$ percentiles are roughly in line with estimated market values, and households of the top 1% seem to underestimate the value of their primary residences by approximately 10% on average. Second, the comparison with observed market values gives similar results: in spite of the significant inflation in housing price observed after 2017, reported values of primary residences sold less than two years after the survey are on average 5% higher than observed market values for households belonging to the bottom 50%, and just in line with observed market values for households between the 50$^{\text{th}}$ and the 75$^{\text{th}}$ percentiles, suggesting again that these households tend to overestimate the value of their primary residence. Conversely, reported values of primary residences sold by households of the top 1% less than two years after the survey are on average 17% lower than observed market values. Although part of this discrepancy might come from the primary residences of top 1% households living in places with stronger housing price inflation, regional housing price indexes demonstrate that regional differences in trends are not a sufficient explanation.[22] All in all, this comparison exercise yields an unambiguous message: the market value of the main residence is on average overestimated by households located at the bottom of the wealth distribution, and underestimated by households located at the very top of the distribution, and the magnitude of this mis-evaluation amounts to roughly ± 10%. However, it is unclear whether this conclusion applies to other real estate assets.

Investigating the asset valuation behavior I investigate the determinants of asset valuation by a linear regression. The dependent variable is the relative difference between the reported market value [23] and the benchmark value (either the statistical market value estimate or the observed market value). All covariates are binary or categorical and are described in table A3 in Appendix 8.3. The main results are presented in figure 11. Complete results are presented in figures A2 and A3 in Appendix 8.3.

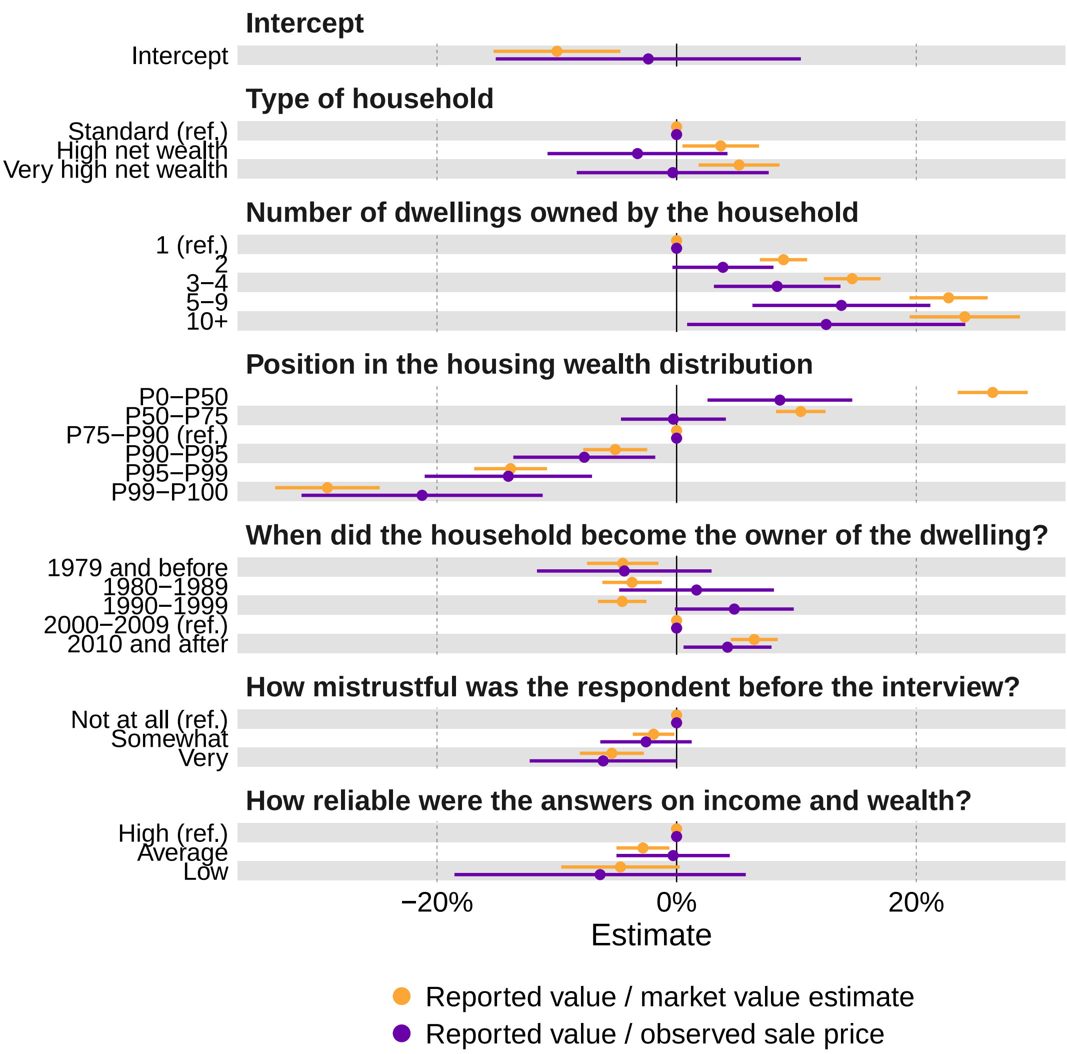
Although point estimates differ significantly depending on the benchmark used, four

[22] For instance, prices of flats located in Paris were on average 6.7% higher in 2018-2019 than in 2017 Q4, as opposed to +3.8% for all housing units in France.

[23] As explained in section 2.1, respondents are asked to report an interval for asset values rather than point estimates, and final asset values are imputed by statisticians, so that the conclusions of this section could be due to biases in the data post-processing. To invalidate this explanation, I compared the two benchmarks with an hypothetical unedited reported asset value, reconstituted by computing the geometric mean of the lower and upper bounds reported by respondents. Figure A4 in appendix 8.3 shows that the results presented in figure 11 are qualitatively unchanged if the geometrical mean of reported interval bounds is used instead of the final reported value.

main findings stand out in both settings. Consistent with figure 10, households located at the top of the wealth distribution tend to come up with a markedly lower evaluation of their home than households in the middle of the distribution, whereas households located at the bottom of the distribution tend to report a higher evaluation. This finding stands in stark contrast with the fact that large landlords tend to report higher market values. A potential interpretation is that large landlords have a better knowledge of the housing market, so that they are able to come up with more accurate market value estimates for their primary residence compared to other wealthy households owning only a few housing units. Moreover, reported market values are higher when the primary residence was acquired more recently, suggesting again that the accuracy of market value estimates varies with the respondent's knowledge of the housing market. Finally, the asset evaluation behavior varies with the attitude of respondents during the interview: reported asset values tend to be lower when respondents are mistrustful and are perceived as unreliable by the interviewer.

FIGURE 11. Determinants of asset valuation behavior



5.4 What happens at the top of the distribution?

In this section I measure the precise extent of wealth underreporting at the very top of the housing wealth distribution by comparing housing wealth reported in the survey with two benchmarks: statistical estimates available in the benchmark database and respondents' housing wealth tax returns.

Linking wealthy households with their housing wealth tax returns Introduced in 2018 as a replacement for the net wealth tax (*Impôt de Solidarité sur la Fortune* or ISF), the French housing wealth tax (*Impôt sur la Fortune Immobilière* or IFI), is an annual tax levied on tax units households whose net housing wealth exceeds 1.3 M€. The tax base is total housing wealth net of liabilities, and tax rates are progressive, starting at 0.5% and rising up to 1.5% for assets above 10 million euros. Taxpayers are required to report all their real estate assets (primary and secondary residences, rental properties and land), along with liabilities related to real estate assets, as of January 1$^{\text{st}}$ of the year in which the tax will be due. Importantly, taxpayers subject to IFI must report the market value of each asset and their share in the asset. However, housing wealth tax returns are not currently available at the asset level, but only at the tax unit level with a rough breakdown in asset classes (primary residences, other buildings owned directly, undeveloped land, indirectly owned real estate assets), meaning that no asset-by-asset comparison can be conducted for the time being.

Household surveys conducted by Insee are routinely linked with income tax returns in order to complement the data with information on income. Thanks to this post-processing, linking respondents with housing wealth tax returns is straightforward because tax unit identifiers are the same in income tax and wealth tax returns. Linking survey respondents with their 2018 housing wealth tax returns yields a subsample of 859 wealthy households. Based on these returns, I compute an estimate of gross housing wealth as the sum of the gross reported value of the primary residence[24] and the reported value of all other real estate assets owned by the tax unit.

Advantages and limitations of benchmarks I compare reported housing wealth to two different benchmarks because each of them has significant limitations. On the one hand, statistical housing wealth estimates do not rely on any information reported by households, as opposed to tax returns. However, three limitations may call into question their reliability. First, they are computed for 2017 Q1 whereas survey data is collected from September 2017. Second, the statistical model described in 2.2.2 provides unbiased market value estimates conditionally on observable characteristics (see André et Meslin (2025)), but does not account for unobserved asset quality. As a consequence, housing wealth estimated might be systematically biased, if the position in the wealth distribution is correlated with asset quality. Third, the asset share owned by each household

[24] The reported value of the primary residence benefits from a 30% discount, so I compute the gross reported value by dividing the reported value by 0.7.

is not available in cadastral data, and hence in the benchmark database. This point is significant in practice, because wealthy households are more likely than the rest of the population to share the ownership of very expensive assets (luxury properties, or large rental investments), inducing potentially large mistakes in housing wealth estimates. On the other hand, housing wealth reported in wealth tax returns has three key advantages compared to statistical housing wealth estimates. First, it is almost perfectly contemporaneous with the survey data collection period, as housing wealth must be reported as of January 1st, 2018. Second, it is *reported by households themselves*, so that the problems regarding unobserved asset quality and mismeasurement of shares can be assumed away. Third, taxpayers are very unlikely to report overestimated asset values because that would result in a higher tax liability. As a consequence, housing wealth reported in tax returns can be considered as a reliable lower bound of true housing wealth. However, housing wealth tax returns have two limitations: complex partial reporting rules apply for assets held in usufruct or bare ownership, and some indirectly owned real estate assets excluded from the survey definition (such as shares in real estate investment funds and life insurance contracts invested in real estate) are included in tax returns-based wealth estimates. Robustness checks described in appendix 8.4 show that my conclusions are robust to these limitations.

Results For each external benchmark, I compute the average relative difference between reported housing wealth and benchmark estimates, expressed as a share of benchmark values. For instance, if the relative difference between reported housing wealth and statistical estimates amounts to -12%, this means that on average reported housing wealth is 12% lower than statistical estimates. Results are presented in figures 12 and 13, with two findings. First, the two benchmarks suggest a clear pattern: underreporting increases with housing wealth, from roughly 20% for households with a benchmark housing wealth comprised between 1.5M€and 2.5M€, to roughly 40% at the very top of the distribution (benchmark housing wealth above 5M€). Second, underreporting affects almost exclusively assets other than the primary residence, whereas the value reported in the survey for the primary residence is roughly is line with wealth tax returns and roughly 10% lower than statistical estimates (consistent with the findings of section 5.3). All in all, this comparison exercise yields another unambiguous message: housing wealth is systematically underreported in the survey, the magnitude of this underreporting increases with housing wealth and reaches 40% at the very top of the distribution.

6 Potential remedies

In this section, I suggest potential remedies to some of the biases highlighted in this paper. Although most of these remedies rely on leveraging administrative data on wealth, I tried to identify solutions that could be implemented when such data is not available.

FIGURE 12. **Comparing reported housing wealth with external benchmarks**

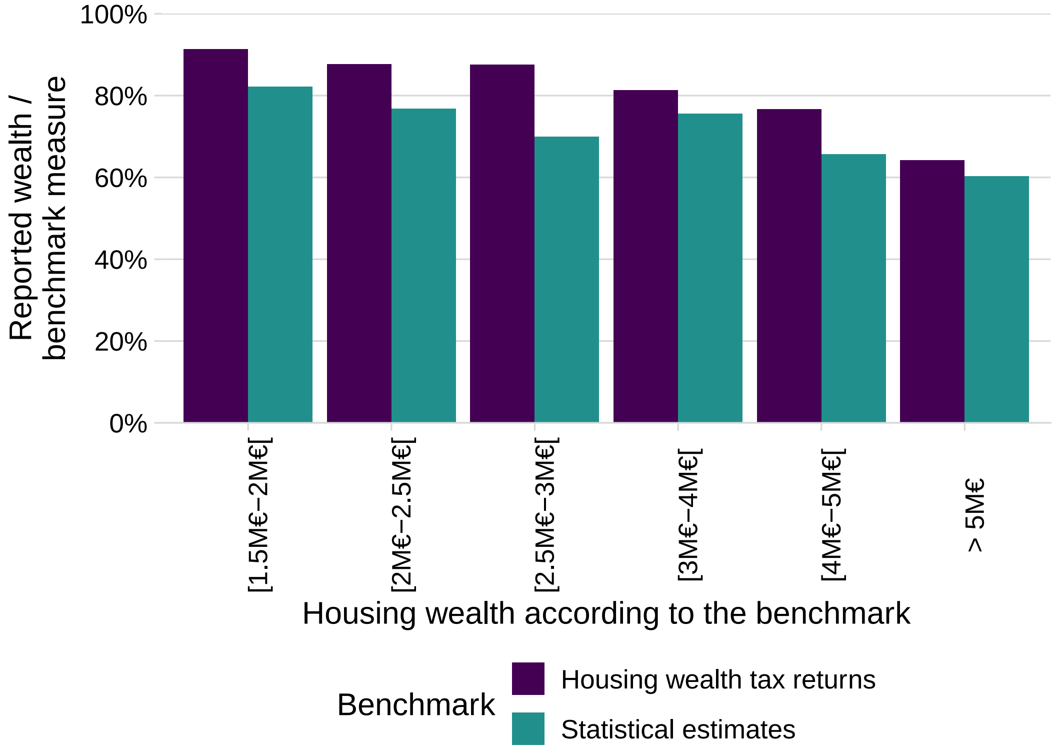
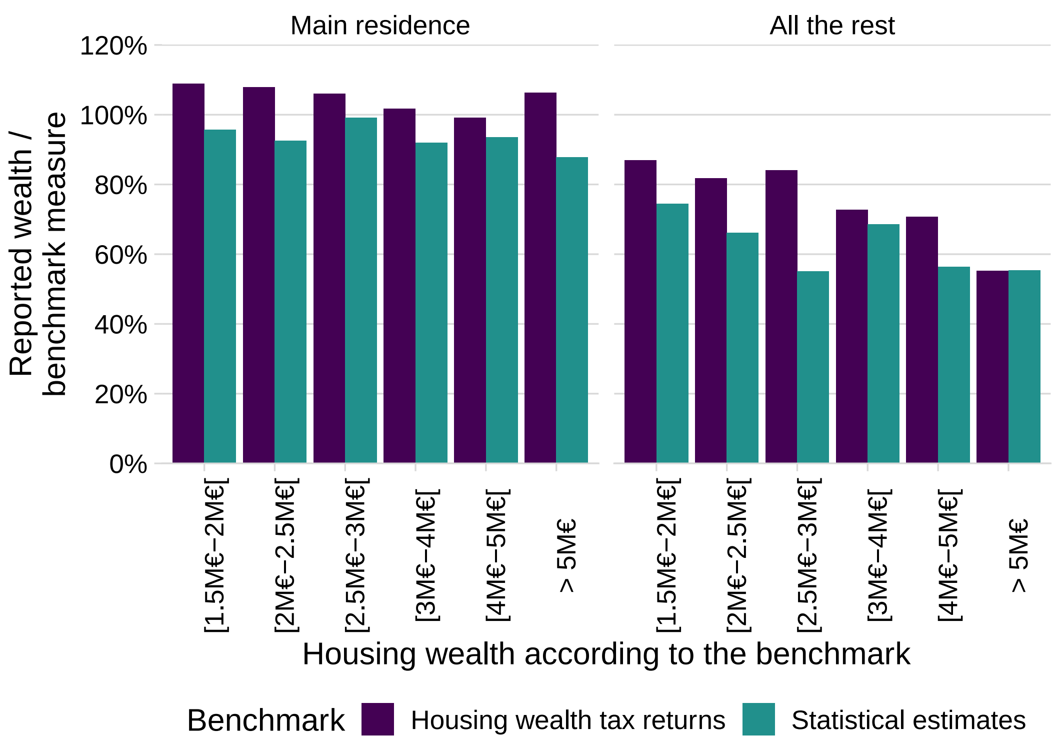


FIGURE 13. **Comparing reported housing wealth with external benchmarks**



6.1 Mitigating the underrepresentation of wealthy households

Section 4 concludes that the underrepresentation of wealthy households in the survey is due to three different problems: selective non-response along with an imperfect unit non-response correction (direct selection problem), a non-representative initial panel subsample because of an imperfect unit non-response correction in the previous wave of the survey (indirect selection problem), and unpredictable distortions in the wealth distribution induced by the calibration step (distortion problem).

Regarding the direct selection problem, one potential avenue could consist in challenging the model used in the homogeneous response groups method: instead of a logit model, the response probability of each household could be estimated using a more flexible, data-driven model allowing arbitrary non-linearities and interactions (such as a random forest). This may help identifying groups with very low response probabilities, resulting in larger weight adjustments.

Two approaches may be considered to housing asset the indirect selection problem and the distortion problem. The most ambitious approach consists in leveraging administrative data to ensure an accurate representation of the wealth distribution: if the position of each respondent in the wealth distribution is known (even approximately) thanks to some external benchmark (such as the benchmark database on housing wealth), then the population totals computed on this benchmark can be used in the calibration step. For instance, the weights of the respondents belonging to the true top 1% could be calibrated so that their total weight in the final sample would amount to exactly 1%. However, this approach is very demanding as such administrative databases are rarely available, particularly regarding financial wealth. A second, more pragmatic approach consists in leveraging information on strata available in the sampling frame, along with standard calibration totals from external benchmarks. For instance, the share of each stratum in the sampling frame could be added to calibration variables: if the top stratum accounts for 0.2% of the sampling frame, then respondents belonging to this stratum would be reweighted so that their total weight in the final sample would amount to exactly 0.2%. This approach might be applied to other variables describing strata (total income, total wealth according to tax returns...). This approach can be thought of as an indirect way to correct for selection on observables.

6.2 Mitigating underreporting of assets and wealth

Improving the data collection process Administrative data could be used to improve data collection in at least two ways. The first, more ambitious approach consists in providing interviewers with a list of assets owned by each household, based on available administrative data, potentially along with a statistical market value estimate. Respondents would then be asked to validate the asset list and to provide additional information such as the share they own in each asset and their own market value estimate. Unfortunately, this approach is likely to raise serious feasibility concerns. First, it is unclear that it would be compatible with legal requirements regarding personal data protection.

Second, the results presented in section 5 showed that information is imperfectly shared within households, and it is not difficult to think of situations where providing such a list to one household member could lead to intrafamilial conflicts, inducing significant reputational risks for statistical institutes. Third, and probably most importantly, providing respondents with this list of assets may result in wealthy households being frightened and thus even more reluctant to participate. This problem is likely to occur in practice, given that wealthy households frequently express the fear that their answers might be used against them, although they are repeatedly told that information reported in the survey would never be used for tax auditing purposes.

A second, less ambitious approach consists in leveraging the results presented in section 5 to improve the questionnaire, in three different ways:

- The question on housing assets could be easily made more consistent with the conceptual framework used by respondents: households would thus be asked how many housing units they own, either in full ownership or in usufruct (rather than in bare ownership). Though this change may seem self-evident, its consequences must be carefully analyzed before implementing it because it would imply a significant methodological break, particularly if some panel households have to answer different questions in the different waves of the survey.
- Some additional questions could be added to the questionnaire to help respondents remember their assets. For instance, given that housing units whose ownership is shared between several other households are more likely to go unreported, respondents could be explicitly asked whether they own a housing units jointly with other family members.
- Respondents could be asked to report more precise information on their assets (such as the detailed address), allowing for a (potentially partial) one-to-one mapping with administrative data. However, asking for such detail may increase the respondents' burden and induce additional mistrust, in case they fear that their answers might be used against them.

Improving data post-processing Administrative data could be leveraged in multiple ways to provide useful data-driven guidance to statisticians when performing survey data post-processing, provided that information collected on housing assets is sufficiently precise to allow for a one-to-one mapping between reported assets and assets actually owned by respondents (at least for a large share of housing assets). If this is the case, at least three avenues could be explored. Regarding asset underreporting at the intensive margin, the number of housing units per housing asset could be corrected based on administrative data when it is inaccurate. Regarding asset underreporting at the extensive margin, statisticians could add unreported assets to the list of assets reported by respondents (although some information would be missing for these additional assets, such as the precise share owned and how the asset was acquired). Regarding asset value imputation, statisticians could use the machine learning algorithm underlying the benchmark database to improve the asset valuation methodology.

7 Conclusion

This paper investigates the biases affecting survey-based estimates of top wealth shares by linking the French wealth survey with a new benchmark database on housing wealth of French households. Compared to this benchmark, only 10% to 20% of the top 10% of the housing wealth distribution is missing in the survey (both in population share and in wealth share), but this proportion increases to 40%-50% for the top 1%. The rest of the paper aims at explaining why.

I link all sampled households to the benchmark database, respondents and non-respondents alike, and use this linked data to measure the effect of each step of the survey process on top wealth shares, based on an innovative decomposition approach. The strong oversampling of wealthy households in the French wealth survey makes it possible to focus on the top of the housing wealth distribution and to isolate the specific behaviors of wealthy households. I conclude that the downward biases in the top 1% population and wealth shares come in equal parts from two causes: wealthy households are strongly underrepresented in the survey mostly because they are somewhat more difficult to contact and much more reluctant to participate than the rest of the population, and asset underreporting is more intense among wealthy households.

I show that the weight adjustment procedure compensates this underrepresentation only partially and that calibration unexpectedly distorts the wealth distribution. I then compare reported housing assets with the assets actually owned by respondents and show that households tend to report assets they have full legal and daily economic control upon and that asset underreporting increases sharply with the number of housing units owned by households and among high net wealth households. I compare market values reported by respondents for their primary residence with statistical estimates and prices observed in real estate transaction data and conclude that households at the bottom of the housing wealth distribution tend to overestimate the value of their home, whereas households at the top tend to underestimate it. I finally compare reported housing wealth with both the benchmark database and wealth tax returns and show that wealth underreporting among the wealthiest households amounts to approximately 40%.

Based on these findings I suggest potential improvements to wealth survey methodology. Future work could explore two directions. First, the approach introduced in this paper should be applied to wealth surveys in other countries so as to assess the external validity of my findings. Second, detailed methodological tests must be carried out to determine whether and how administrative data can be leveraged to improve the accuracy and reliability of wealth measurement in surveys.

References

Alvargonzález, P., C. Barcelo, O. Bover, L. Cobreros, L. Crespo, N. El Amrani, S. García-Uribe, C. Gento, M. Gómez, E. Villanueva, et al. (2024): "The Spanish Survey of Household Finances (EFF): description and methods of the 2020 wave," *Banco de Espana Occasional Paper Forthcoming*, 2405.

André, M., et O. Meslin (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., et O. Meslin (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.

Bach, S., A. Thiemann, et A. Zucco (2019): "Looking for the missing rich: Tracing the top tail of the wealth distribution," *International Tax and Public Finance*, 26(6), 1234–1258.

Blanchet, T., I. Flores, et M. Morgan (2022): "The weight of the rich: improving surveys using tax data," *The Journal of Economic Inequality*, 20(1), 119–150.

Bricker, J., A. Henriques, J. Krimmel, et J. Sabelhaus (2016a): "Estimating top income and wealth shares: Sensitivity to data and methods," *American Economic Review*, 106(5), 641–45.

——— (2016b): "Measuring income and wealth at the top using administrative and survey data," *Brookings Papers on Economic Activity*, 2016(1), 261–331.

Cantarella, M., A. Neri, et M. G. Ranalli (2024): "Estimating the distribution of household wealth: methods for adjusting survey data estimates using national accounts and rich list data," *Review of Income and Wealth*, 70(3), 551–580.

Caron, N. (2005): "La correction de la non-réponse par repondération et par imputation," *Document de travail*, (M0502).

Haziza, D., et J.-F. Beaumont (2017): "Construction of Weights in Surveys: A Review," *Statistical Science*, 32(2), 206 – 226.

Johansson-Tormod, F., et A. Klevmarken (2022): "Comparing Register and Survey Wealth Data," *International Journal of Microsimulation*, 15(1), 43–62.

Juster, F., J. P. Smith, et F. Stafford (1999): "The measurement and structure of household wealth," *Labour Economics*, 6(2), 253–275.

Kennickell, A. (2015): "Dirty and unknown: Statistical editing and imputation in the SCF," *Statistical Journal of the IAOS*, 31(3), 435–445.

Kennickell, A. B. (2017a): "Getting to the top: Reaching wealthy respondents in the SCF," *Statistical Journal of the IAOS*, 33(1), 113–123.

——— (2017b): "Lining up: Survey and administrative data estimates of wealth concentration," *Statistical Journal of the IAOS*, 33(1), 59–79.

——— (2019): "The tail that wags: differences in effective right tail coverage and estimates of wealth inequality," *The Journal of Economic Inequality*, 17(4), 443–459.

Kennickell, A. B., et R. L. Woodburn (1999): "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth*, 45(2), 193–215.

Kopczuk, W., et E. Saez (2004): "Top Wealth Shares in the United States, 1916-2000: Evidence From Estate Tax Returns," *National Tax Journal*, 57(2), 445–487.

Lustig, N., et al. (2020): "The 'missing Rich' in Household Surveys: Causes and Correction Approaches," Discussion paper, ECINEQ, Society for the Study of Economic Inequality.

Lynn, P. (2012): "Longitudinal Survey Methods for the Household Finances and Consumption Survey: a report prepared for the European Central Bank," .

Meriküll, J., et T. Rõõm (2021): "Are survey data underestimating the inequality of wealth?," *Empirical Economics*, 62, 1–36.

Meyer, B. D., W. K. Mok, et J. X. Sullivan (2015): "Household surveys in crisis," *Journal of Economic Perspectives*, 29(4), 199–226.

Network, C. (2023): "Household Finance and Consumption Survey: Results from the 2021 wave," Discussion paper, ECB Statistics Paper.

Osier, G. (2016): "Unit non-response in household wealth surveys," Statistics Paper Series 15, European Central Bank.

Piketty, T. (2014): *Capital in the twenty-first century*. Harvard University Press.

Piketty, T., et G. Zucman (2015): "Wealth and inheritance in the long run," in *Handbook of income distribution*, vol. 2, pp. 1303–1368. Elsevier.

Saez, E., et G. Zucman (2016): "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, 131(2), 519–578.

Vermeulen, P. (2016): "Estimating the top tail of the wealth distribution," *American Economic Review*, 106(5), 646–650.

——— (2018): "How fat is the top tail of the wealth distribution?," *Review of Income and Wealth*, 64(2), 357–387.

8 Appendix

8.1 Supplementary figures for section 4.1

Table A1. Descriptive statistics on participation behavior, by stratum and subsample

Stratum	Overall response rate	Decomposition of the overall response rate		
		In-survey-scope rate	Contact rate	Participation rate
New entrants subsample				
Very HNW households, urban	**48.8%**	90.8%	74.9%	71.7%
Very HNW households, rural	**47.4%**	90.0%	74.2%	71.0%
Other HNW households	**53.6%**	89.5%	77.2%	77.7%
All HNW households	***50.0%***	*90.0%*	*75.4%*	*73.6%*
Older households	**60.4%**	90.4%	80.4%	83.1%
High business income	**59.9%**	90.5%	83.5%	79.3%
High capital income	**62.1%**	91.2%	84.1%	80.9%
High wages	**62.3%**	91.5%	84.7%	80.3%
All other households	**59.7%**	87.9%	82.8%	81.9%
All standard households	***60.5%***	*89.7%*	*82.6%*	*81.6%*
All households	**58.5%**	**89.7%**	**81.3%**	**80.3%**
Panel subsample				
Very HNW households, urban	**71.2%**	96.9%	88.4%	83.2%
Very HNW households, rural	**67.6%**	97.1%	83.5%	83.3%
Other HNW households	**71.6%**	97.4%	84.7%	86.8%
All HNW households	***70.3%***	*97.2%*	*85.1%*	*84.9%*
Older households	**73.8%**	94.8%	86.7%	89.8%
High business income	**79.7%**	98.5%	90.4%	89.6%
High capital income	**82.9%**	99.0%	94.0%	89.1%
High wages	**75.2%**	98.1%	87.7%	87.4%
All other households	**77.6%**	98.6%	88.7%	88.7%
All standard households	***76.6%***	*97.5%*	*88.4%*	*88.9%*
All households	**75.8%**	**97.4%**	**88.0%**	**88.4%**
All households	**63.5%**	**91.9%**	**83.3%**	**82.9%**

Notes: all figures are unweighted.

8.2 Supplementary figures for section 5.2

TABLE A2. Determinants of reporting used in part 5.2

Variable	Values
Characteristics of dwellings	
Type of dwelling	*House*, Flat
Year of purchase	1979 and before, 1980-1989, 1990-1999, *2000-2009*, 2010 and after
Use of dwelling	*Main residence*, secondary residence, rental, vacant
Number of dwellings owned by the household in the building	*1*, 2, 3, 4, 5-9, 10+
Characteristics of ownership rights	
Type of ownership right	*Full ownership*, Bare ownership, Usufruct
Number of households owning jointly the dwelling	*1*, 2, 3, 4+
Is ownership intermediated by an SCI?	Yes, No
Does the recipient of the dwelling's property tax bill belong to the household?	*Yes*, No
Does the household earn rental income?	*No*, Yes, Yes from this property
Wealth of households	
Is the household classified as high net wealth?	*Standard*, High net wealth, Very high net wealth
Number of dwellings owned by the household	0, *1*, 2, 3-4, 5-9, 10+
Position in the housing wealth distribution	P0-50, P50-P75, *P75-P90*, P90-P95, P95-P99, P99-P100
Characteristics of households	
What subsample does the household belong to?	*New entrant*, Panel
Standard of living	D1 to D9, P90-P95, P95-P100 (ref: *P51-P60*)
Number of individuals in the household	*1*, 2, 3, 4+
Is one of the household members working or looking for work?	*Yes*, No
Is there a couple in the household?	Yes, *No*
Does any of the household members have a non-French citizenship?	Yes, *No*
In which region does the household live?	26 administrative regions (ref: *Paris region*)
Characteristics of respondents	
How many respondents are there?	*1*, 2+
Gender of respondents	*Men only*, Women only, Men and women
How many owners of the dwellings are among the respondents?	0, *1*, 2+
Age of oldest respondent	00-39, 40-49, 50-59, 60-69, 70-79, 80+
Highest degree of respondents	Graduate degree, Undergraduate degree, Higher vocational education, *Upper secondary*, Lower secondary, Primary school or less
Relation to the household's reference person	*Reference person*, spouse, other person
Interview quality, reported by the interviewer	
Respondent's ability to express him/herself	*Excellent*, Good, Average, Poor
How mistrustful was the respondent before the interview?	*Not at all*, Somewhat, Very
Did the respondent use documents?	*Frequently*, Sometimes, Rarely, Never
How reliable were the answers on income and wealth?	*High*, Average, Low
Did the respondent hid some information?	Yes, *No*

Notes: the reference category is indicated in italics.

8.3 Supplementary figures for section 5.3

FIGURE A1. Comparing reported asset values and actual market values

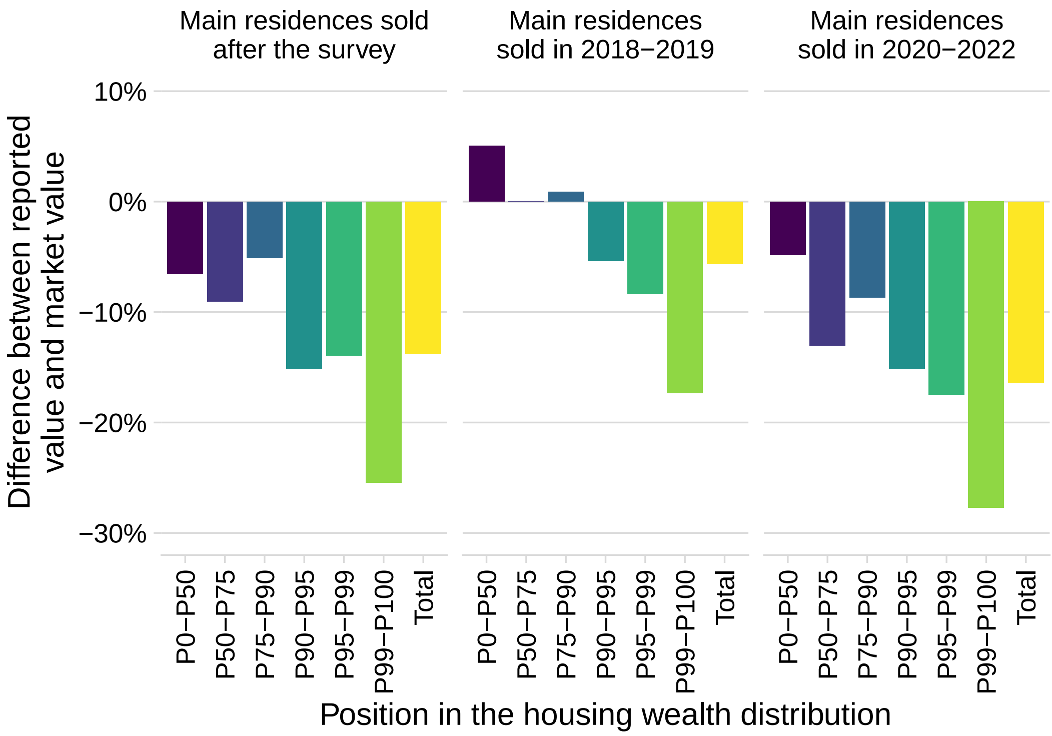


TABLE A3. Determinants of asset valuation used in part 5.3

Variable	Values
Characteristics of dwellings	
Type of dwelling	*House*, Flat
Year of purchase	1979 and before, 1980-1989, 1990-1999, *2000-2009*, 2010 and after
Use of dwelling	*Main residence*, secondary residence, rental, vacant
Number of dwellings owned by the household in the building	*1*, 2, 3, 4, 5-9, 10+
Characteristics of ownership rights	
Type of ownership right	*Full ownership*, Bare ownership, Usufruct
Number of households owning jointly the dwelling	*1*, 2, 3, 4+
Is ownership intermediated by an SCI?	Yes, No
Does the recipient of the dwelling's property tax bill belong to the household?	*Yes*, No
Does the household earn rental income?	*No*, Yes, Yes from this property
Wealth of households	
Is the household classified as high net wealth?	*Standard*, High net wealth, Very high net wealth
Number of dwellings owned by the household	0, *1*, 2, 3-4, 5-9, 10+
Position in the housing wealth distribution	P0-50, P50-P75, *P75-P90*, P90-P95, P95-P99, P99-P100
Characteristics of households	
What subsample does the household belong to?	*New entrant*, Panel
Standard of living	D1 to D9, P90-P95, P95-P100 (ref: *P51-P60*)
Number of individuals in the household	*1*, 2, 3, 4+
Is one of the household members working or looking for work?	*Yes*, No
Is there a couple in the household?	Yes, *No*
Does any of the household members have a non-French citizenship?	Yes, *No*
In which region does the household live?	26 administrative regions (ref: *Paris region*)
Characteristics of respondents	
How many respondents are there?	*1*, 2+
Gender of respondents	*Men only*, Women only, Men and women
How many owners of the dwellings are among the respondents?	0, *1*, 2+
Age of oldest respondent	00-39, 40-49, 50-59, 60-69, 70-79, 80+
Highest degree of respondents	Graduate degree, Undergraduate degree, Higher vocational education, *Upper secondary*, Lower secondary, Primary school or less
Relation to the household's reference person	*Reference person*, spouse, other person
Interview quality, reported by the interviewer	
Respondent's ability to express him/herself	*Excellent*, Good, Average, Poor
How mistrustful was the respondent before the interview?	*Not at all*, Somewhat, Very
Did the respondent use documents?	*Frequently*, Sometimes, Rarely, Never
How reliable were the answers on income and wealth?	*High*, Average, Low
Did the respondent hid some information?	Yes, *No*

Notes: the reference category is indicated in italics.

FIGURE A2. Determinants of asset valuation behavior (1)

FIGURE A3. Determinants of asset valuation behavior (2)

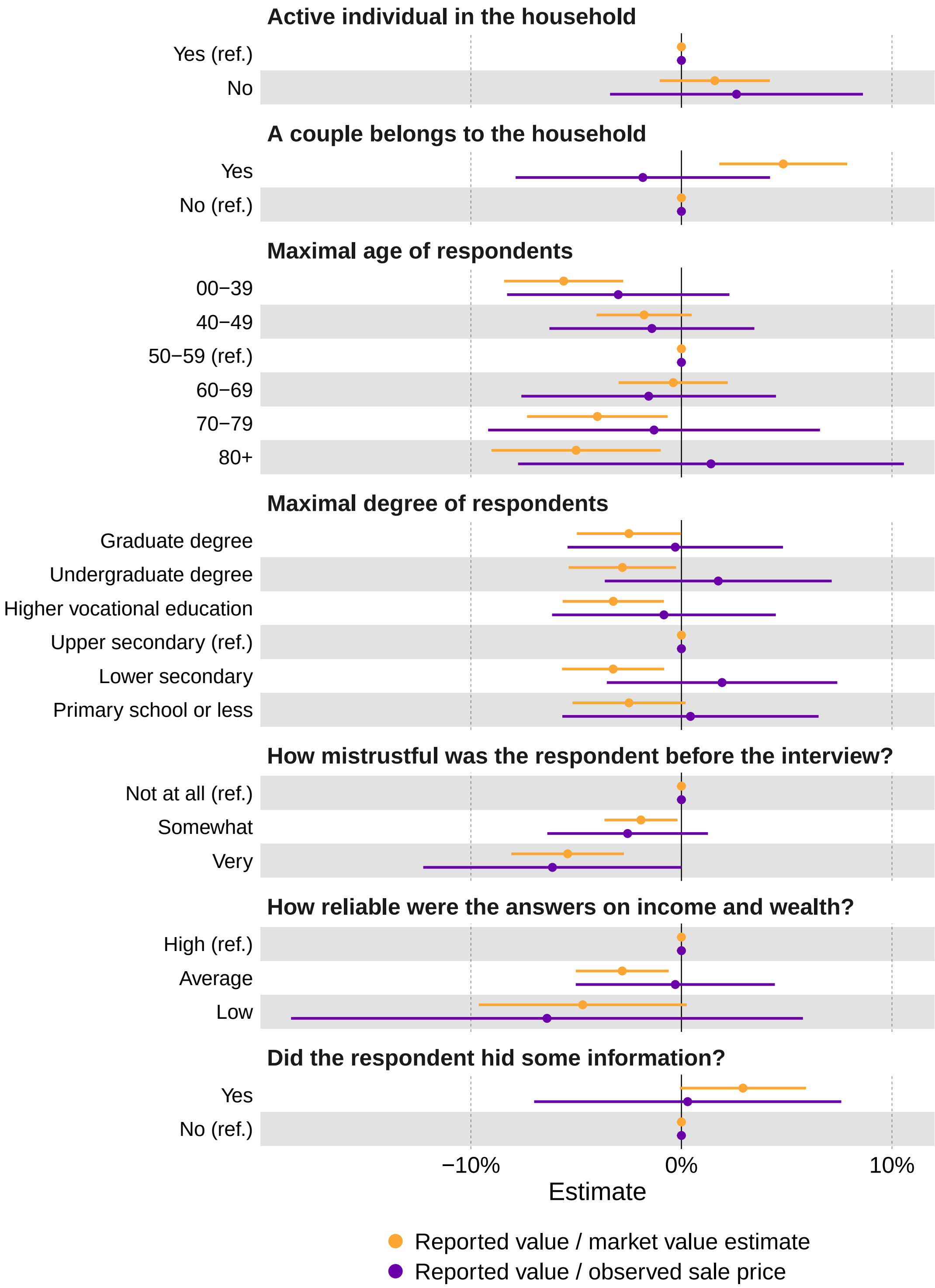
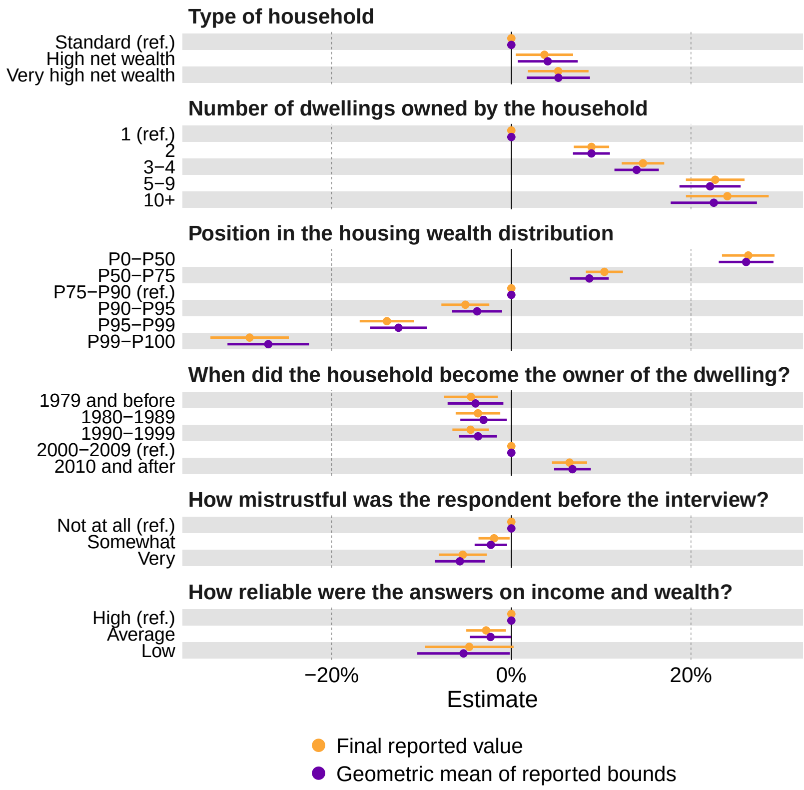
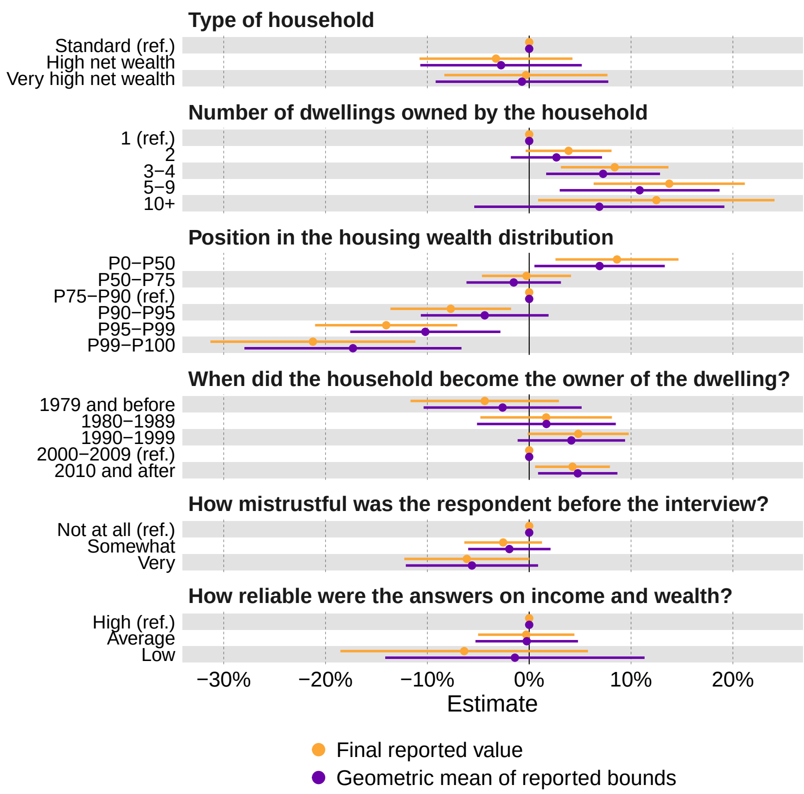


Figure A4. Determinants of asset valuation behavior (robustness check)

(a) Market value estimate



(b) Observed sale price



8.4 Supplementary figures for section 5.4

The housing wealth estimates based on wealth tax returns used in section 5.4 include some indirectly owned real estate assets excluded from the survey definition (such as shares in real estate investment funds and life insurance contracts invested in real estate), so that these estimates may be biased upward compared to wealth reported in the survey. I compute alternative housing wealth estimates based on wealth tax returns, by excluding all indirectly owned real estate assets. These alternative estimates do not include assets owned through SCIs, implying that they may be biased downward compared to wealth reported in the survey. Figure A5 shows that using these alternative estimates reduces the *level* of underreporting, but not the slope of underreporting with respect to housing wealth. I conclude that the key finding that underreporting increases with housing wealth is not driven by comparability problems between survey data and wealth tax returns.

Complex reporting rules apply for assets held in usufruct or bare ownership: in some cases only a fraction of the value of the asset must be reported in tax wealth returns, and the precise fraction depends on the age of the owner. As a consequence, housing wealth estimates based on wealth tax returns used in section 5.4 may be biased downward compared to wealth reported in the survey. Figure A6 shows that restricting the comparison to wealthy households owning only housing units in full ownership does not modify significantly the results. I conclude that the key finding that underreporting increases with housing wealth is not driven by reporting rules regarding assets held in usufruct or bare ownership.

FIGURE A5. Comparing reported housing wealth with tax returns: robustness check

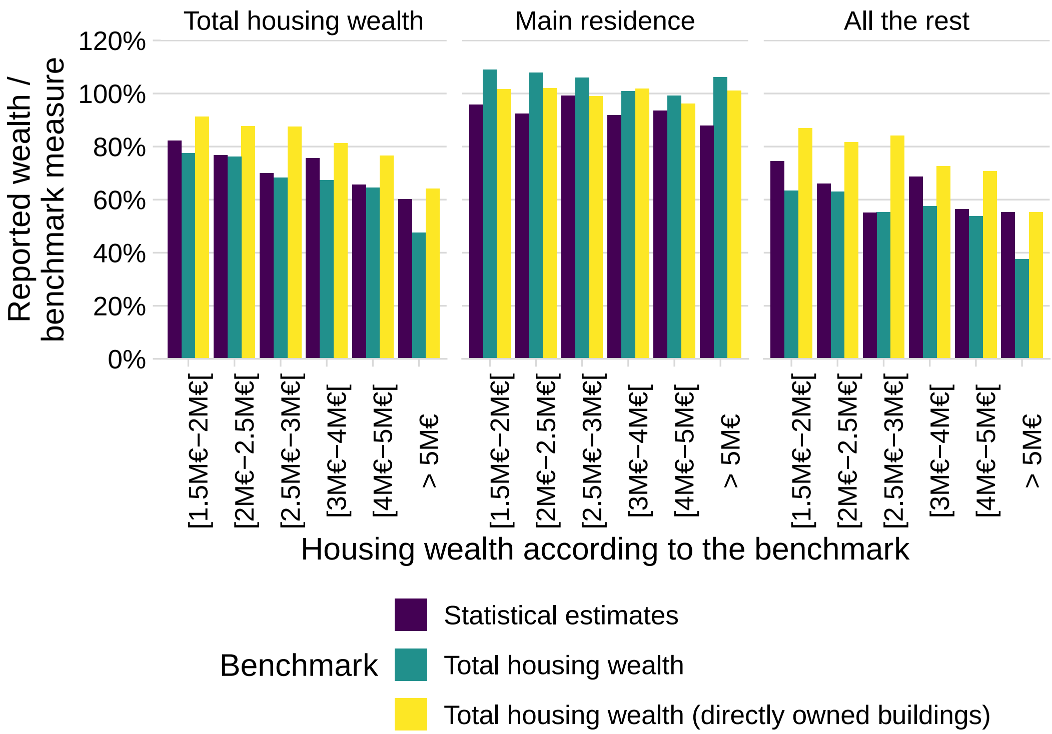
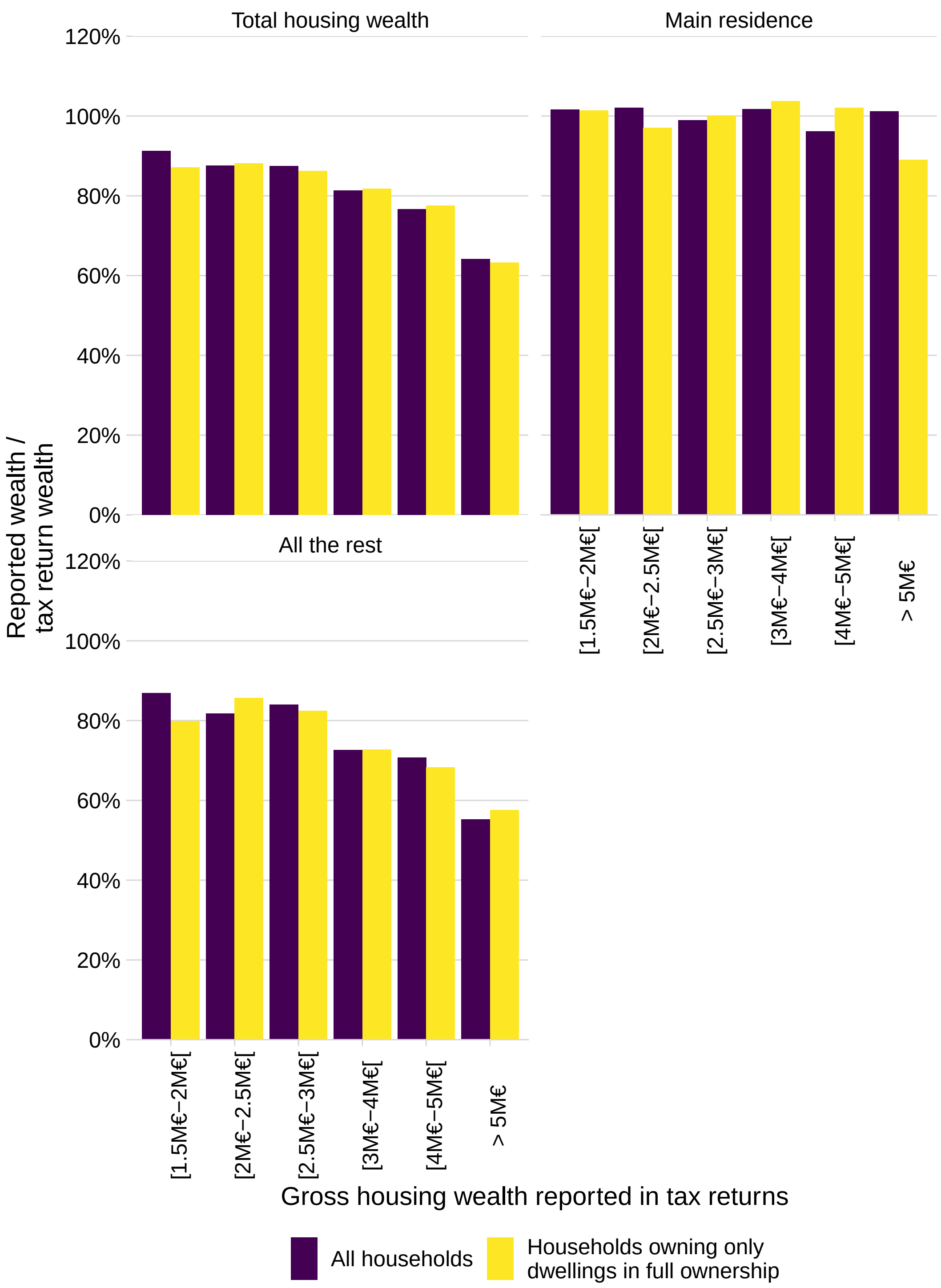


Figure A6. Comparing reported housing wealth with tax returns: robustness check



Série des Documents de Travail « Méthodologie Statistique »

9601 : Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population.
G. DECAUDIN, J.-C. LABAT

9602 : Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises.
N. CARON, P. RAVALET, O. SAUTORY

9603 : La procédure FREQ de SAS - Tests d'indépendance et mesures d'association dans un tableau de contingence.
J. CONFAIS, Y. GRELET, M. LE GUEN

9604 : Les principales techniques de correction de la non-réponse et les modèles associés.
N. CARON

9605 : L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration.
P. RAVALET

9606 : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT).
S. LOLLIVIER, M. MARPSAT, D. VERGER

9607 : Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes.
N. CARON, D. LE BLANC

9701 : Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?
J.-C. DEVILLE

9702 : Modèles univariés et modèles de durée sur données individuelles.
S. LOLLIVIER

9703 : Comparaison de deux estimateurs par le ratio stratifiés et application aux enquêtes auprès des entreprises.
N. CARON, J.-C. DEVILLE

9704 : La faisabilité d'une enquête auprès des ménages.
1. au mois d'août.
2. à un rythme hebdomadaire

C. LAGARENNE, C. THIESSET

9705 : Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine.
P. GIRARD.

9801 : Les logiciels de désaisonnalisation TRAMO & SEATS : philosophie, principes et mise en œuvre sous SAS.
K. ATTAL-TOUBERT, D. LADIRAY

9802 : Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation.
J.-C. DEVILLE

9803 : Pour essayer d'en finir avec l'individu Kish.
J.-C. DEVILLE

9804 : Une nouvelle (encore une !) méthode de tirage à probabilités inégales.
J.-C. DEVILLE

9805 : Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish.
J.-C. DEVILLE

9806 : Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel POULPE.
N. CARON, J.-C. DEVILLE, O. SAUTORY

9807 : Estimation de données régionales à l'aide de techniques d'analyse multidimentionnelle.
K. ATTAL-TOUBERT, O. SAUTORY

9808 : Matrices de mobilité et calcul de la précision associée.
N. CARON, C. CHAMBAZ

9809 : Échantillonnage et stratification : une étude empirique des gains de précision.
J. LE GUENNEC

9810 : Le Kish : les problèmes de réalisation du tirage et de son extrapolation.
C. BERTHIER, N. CARON, B. NEROS

9901 : Perte de précision liée au tirage d'un ou plusieurs individus Kish.
N. CARON

9902 : Estimation de variance en présence de données imputées : un exemple à partir de l'enquête Panel Européen.
N. CARON

0001 : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT) (version actualisée).
S. LOLLIVIER, M. MARPSAT, D. VERGER

0002 : Modèles structurels et variables explicatives endogènes.
J.-M. ROBIN

0003 : L'enquête 1997-1998 sur le devenir des personnes sorties du RMI - Une présentation de son déroulement.
D. ENEAU, D. GUILLEMOT

0004 : Plus d'amis, plus proches ? Essai de comparaison de deux enquêtes peu comparables.
O. GODECHOT

0005 : Estimation dans les enquêtes répétées : application à l'Enquête Emploi en Continu.
N. CARON, P. RAVALET

0006 : Non-parametric approach to the cost-of-living index.
F. MAGNIEN, J. POUGNARD

0101 : Diverses macros SAS : Analyse exploratoire des données, Analyse des séries temporelles.
D. LADIRAY

0102 : Économétrie linéaire des panels : une introduction.
T. MAGNAC

0201 : Application des méthodes de calages à l'enquête EAE-Commerce.
N. CARON

C 0201 : Comportement face au risque et à l'avenir et accumulation patrimoniale - Bilan d'une expérimentation.
L. ARRONDEL, A. MASSON, D. VERGER

C 0202 : Enquête Méthodologique Information et Vie Quotidienne - Tome 1 : bilan du test 1, novembre 2002.
J.-A. VALLET, G. BONNET, J.-C. EMIN, J. LEVASSEUR, T. ROCHER, P. VRIGNAUD, X. D'HAULTFOEUILLE, F. MURAT, D. VERGER, P. ZAMORA

0203 : General principles for data editing in business surveys and how to optimise it.
P. RIVIERE

0301 : Les modèles logit polytomiques non ordonnés : théories et applications.
C. AFSA ESSAFI

0401 : Enquête sur le patrimoine des ménages - Synthèse des entretiens monographiques.
V. COHEN, C. DEMMER

0402 : La macro SAS CUBE d'échantillonnage équilibré
S. ROUSSEAU, F. TARDIEU

0501 : Correction de la non-réponse et calage de l'enquêtes Santé 2002
N. CARON, S. ROUSSEAU

0502 : Correction de la non-réponse par répondération et par imputation
N. CARON

0503 : Introduction à la pratique des indices statistiques - notes de cours
J-P BERTHIER

0601 : La difficile mesure des pratiques dans le domaine du sport et de la culture - bilan d'une opération méthodologique
C. LANDRE, D. VERGER

0801 : Rapport du groupe de réflexion sur la qualité des enquêtes auprès des ménages
D. VERGER

M2013/01 : La régression quantile en pratique
P. GIVORD,
X. D'HAULTFOEUILLE

M2014/01 : La microsimulation dynamique : principes généraux et exemples en langage R
D. BLANCHET

M2015/01 : la collecte multimode et le paradigme de l'erreur d'enquête totale
T. RAZAFINDROVONA

M2015/02 : Les méthodes de Pseudo-Panel
M. GUILLERM

M2015/03 : Les méthodes d'estimation de la précision pour les enquêtes ménages de l'Insee tirées dans Octopusse
E. GROS
K. MOUSSALAM

M2016/01 : Le modèle Logit Théorie et application.
C. AFSA

M2016/02 : Les méthodes d'estimation de la précision de l'Enquête Emploi en Continu
E. GROS
K.MOUSSALAM

M2016/03 : Exploitation de l'enquête expérimentale Vols, violence et sécurité.

T. RAZAFINDROVONA

M2016/04 : Savoir compter, savoir coder. Bonnes pratiques du statisticien en programmation.
E. L'HOUR
R. LE SAOUT
B. ROUPPERT

M2016/05 : Les modèles multiniveaux
P. GIVORD
M. GUILLERM

M2016/06 : Econométrie spatiale : une introduction pratique
P. GIVORD
R. LE SAOUT

M2016/07 : La gestion de la confidentialité pour les données individuelles
M. BERGEAT

M2016/08 : Exploitation de l'enquête expérimentale Logement internet-papier
T. RAZAFINDROVONA

M2017/01: Exploitation de l'enquête expérimentale Qualité de vie au travail
T. RAZAFINDROVONA

M2018/01: Estimation avec le score de propension sous
S. QUANTIN

M2018/02: Modèles semi-paramétriques de survie en temps continu sous
S. QUANTIN

M2019/01 : Les méthodes de décomposition appliquées à l'analyse des inégalités
B. BOUTCHENIK
E. COUDIN
S. MAILLARD

M2020/01 : L'économétrie en grande dimension
J. L'HOUR

M2021/01 : R Tools for JDemetra+ - Seasonal adjustment made easier
A. SMYK
A. TCHANG

M2021/02 : Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman
L. CASTELL
P. SILLARD

M2021/03 : Conception de questionnaires auto-administrés
H. KOUMARIANOS
A. SCHREIBER

M2022/01 : Introduction à la géomatique pour le statisticien : quelques concepts et outils innovants de gestion, traitement et diffusion de l'information spatiale
F. SEMECURBE
E. COUDIN

M2022/02 : Le zonage en unites urbaines 2020
V. COSTEMALLE
S. OUJIA
C. GUILLO
A. CHAUVET

M2023/01 : Les réseaux de neurones appliqués à la statistique publique : méthodes et cas d'usages
D. BABET
Q. DELTOUR
T. FARIA
S. HIMPENS

M2023/02 : Redressements de la première vague de l'enquête epicov : un exemple de correction des effets de sélection dans les enquêtes multimodes
L.CASTELL
C. FAVRE-MARTINOZ
N. PALIOD
P. SILLARD

M2023/03 : Appariements de données individuelles : concepts, méthodes, conseils
L.MALHERBE

M2023/04 : Victimations déclarées et effets de mode : enseignements de l'expérimentation panel multimode de l'enquête cadre de vie et sécurité
L. CASTELL
M. CLERC
D. CROZE
S. LEGLEYE
A. NOUGARET

M2024/01 : Estimation en temps réel de la tendance-cycle : apport de l'utilisation des filtres asymétriques dans la détection des points de retournement
A.QUARTIER-LA-TENTE

M2024/02 : La disponibilité des coordonnées de contact dans fidéli-nautile - quels enseignements pour les protocoles de collecte ?
G. CHARRANCE (INED)

M2024/03 : Discuter l'existence d'un effet de sélection dans un cadre multimode grâce à une analyse de sensibilité - Application aux enquêtes annuelles de recensement
L. COURT
S. QUANTIN

M2024/04 : Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+
A. SMYK
K. WEBEL

M2025/01 : Les estimations par capture-recapture ou par système multiple : quelques éléments théoriques
P. ARDILLY
H. KOUMARIANOS

M2025/02 : Tests cognitifs pour les enquêtes auto-administrées : quelques éléments de méthode
D. GUILLEMOT
J. DIRAND
C. FLUXA

M2025/03 : Statistiques fondées sur des données administratives - esquisse d'un cadre général
H. KOUMARIANOS
P. RIVIÈRE

M2025/04 : Peut-on estimer un effet de mesure sur une enquête à partir d'un essai croisé ab/ba : la question de la non-réponse non ignorable dans l'enquête test emploi du temps
Loreline COURT
Simon QUANTIN

M2025/05 : L'apport des technologies cloud pour industrialiser le processus d'innovation statistique
R. AVOUAC
T. FARIA
F. COMTE

M2025/06 : Le data editing : Définition et principes généraux
N.CARON

M2025/07 : Estimation de la variance pour l'enquête longitudinale Histoire de vie et Patrimoine : une comparaison des approches analytique et par bootstrap
K. LARBI
J. RUBIN

M2025/08 : What's Wrong with Survey-based Top Wealth Shares? Evidence from Housing Wealth of French Households
O. MESLIN