Modifying the Sample Design of the Panel on Household Finances in Wave 4: Aims and Challenges

René Bernard[1], Tobias Schmidt[2], Panagiota Tzamourani[3]

[1]Goethe University Frankfurt & Deutsche Bundesbank[1],
Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main
[2]Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main
[3]Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main

Abstract

The sampling design of the Panel on Household Finances (PHF) aims to provide representative statistics of the distribution of household wealth in Germany. In order to capture adequately the whole range of wealth distribution and also deliver information on assets held by a smaller percentage of wealthy households, the sampling design of the first three waves employed stratification based on estimated household wealth and oversampling of wealthy households. In 2020, a modified design has been introduced with the aim of strengthening regional representation while ensuring an overrepresentation of the wealthy in the final sample. The new design features (a) explicit stratification by federal state and (b) oversampling of eastern Germany rather than wealthy areas, while maintaining the existing stratification based on household wealth. Simulations based on samples drawn from the sample of the wave conducted in 2017 verify that overrepresentation of the wealthy in the final sample can most likely be maintained under the new sampling scheme.

Key Words: Panel on Household Finances, wealth distribution, sampling design, oversampling.

1. Introduction

The paper proposes a new sample design for the Panel of Household Finances (PHF), a representative survey of German households.[2] The PHF constitutes the German component of the Eurosystem Household Finance and Consumption Survey (HFCS). It provides a comprehensive picture of German households' balance sheets and covers various aspects relevant to the financial situation of German households, e. g. savings and consumption behaviour, income, employment, as well as expectations and financial attitudes. The PHF was launched in 2010/2011 and takes place every three years. It has a panel component and a refresher component in order to account for panel attrition and changes in the population. Its main mode is face-to-face (CAPI) interviewing.

The survey aims to ensure a good representation of households at the top of the wealth distribution. Thus, the sampling design of the first three waves of the PHF features an oversampling of wealthy regions in Germany. Over time, the repeated oversampling of wealthy households in the refresher components of the PHF together with less attrition at the top of the wealth distribution than in other parts has led to an increasing

[1] The findings reported in this article represent the authors' views and do not necessarily represent the views of the Deutsche Bundesbank or its staff.
[2] See Altmann et al. (2020) and www.bundesbank.de/phf-research for details.

share of the oversampled population from wave to wave. At the same time, there has been growing demand from researchers for more accurate regional statistics and analyses of less wealth households.

The survey team thus decided to adjust the PHF sample design, primarily in order to strengthen the regional representation of the PHF but also to enhance the representation of less wealthy households, while maintaining a substantial number of households at the top of the distribution. The modified design features a regional oversampling, rather than oversampling of wealthy areas. This poses a risk to preserving the large share of the wealthy in the PHF sample in the future. However, the large panel component and the overrepresentation of wealthy households in it ensure that such a risk is mitigated. In addition, the proposed design maintains the existing stratification, which is based on wealth. This ensures that the refresher sample will also be covering wealthy areas[3].

The new sampling design was originally envisaged for the next face-to-face survey wave, scheduled to take place in the spring of 2020. Due to the Covid-19-pandemic, this wave had to be postponed and is now scheduled to take place in 2021. Between September and December 2020, however, a shorter PAPI survey will take place, with the aim of capturing the effects of the Covid-19-pandemic on households' finances. The new sampling design will be applied to this PAPI survey.

The remainder of the paper is laid out as follows: Section 2 describes the sampling design of waves 1 to 3. Section 3 explains how the sampling design will be modified in wave 4. Section 4 shows how the new sampling design will, in theory, impact estimates of wealth indicators. Section 5 provides details on the practical implementation of the new design. Section 6 summarizes our results and concludes the paper.

2. Sampling design of waves 1-3

The sample design of the PHF in waves one, two and three aimed to capture adequately the wealth distribution of households, increase the statistical power for its analysis compared with other available data sources and also deliver information on assets held by a smaller percentage of wealthy households. For this purpose, (a) the stratification was based on data related to household wealth (with only implicit stratification across German federal states ("Bundesländer") and municipality size classes) and (b) the wealthy areas were sampled at a higher rate than their percentage share in the population would indicate. In particular, the strata defined based on micro-geographic indicators (for street sections) and income tax statistics (for small and middle-sized municipalities) are the following[4]:

- wealthy small and middle-sized municipalities (i.e. with fewer than 100,000 adult residents)
- other small and middle-sized municipalities
- wealthy street sections (i.e. street sections in wealthy neighborhoods in cities with 100,000 adult residents or more)
- other street sections (complementary to the above)

Within the above strata, addresses of adults (18 years and older), corresponding to a household, are drawn randomly from population registers of German municipalities ("Einwohnermeldeamtsregister"). In municipalities with fewer than 100,000 residents, individuals are selected by means of a systematic random selection process out of a list of all registered residents ordered by family name. In cities with more than 100,000 residents, addresses from the selected street sections are selected randomly.

[3] Theoretically, the sample design should not affect (asymptotically) the means, but only the precision of the estimates.

[4] See Altmann et al. (2020) for a more detailed discussion of the sampling design in the PHF.

3. Modifying the design of the PHF

For the fourth wave of the PHF, modifications to the sampling design were introduced. The aim of the new design is to strengthen the regional representation of households in the PHF while still maintaining a high degree of representation of wealthy households in the sample. Strengthening the regional representation, rather than repeating the oversampling of wealthy areas, also improves the precision of estimates for the lower end of the wealth distribution.

Better representation of regions and less wealthy households

There are several reasons that led to the decision to introduce a change in the sampling design:
First, there is a need for higher precision regional estimates, in particular for eastern Germany. For studies on inequality within Germany, the inequality between eastern and western Germany is of particular interest (see Kreutzmann et al. 2019 and the discussion therein). In general, distinguishing between eastern and western Germany is typically deemed important when analyzing economic phenomena. For example, research shows that Germans in the eastern part of the country have higher saving rates than Germans in the west (Fuchs-Schündeln, 2008), expect higher inflation (Goldfayn-Frank and Wohlfart, 2019) and exhibit higher levels of conspicuous consumption decades after reunification (Friehe and Mechtel, 2014). Moreover, a more accurate regional representation might allow the use of regional variation for identification purposes in econometric models.

Second, there is a general interest in better measuring the lower end of the wealth distribution. For the analysis of poverty, for example, measuring with precision the bottom part of the distribution is as important as measuring the top part. Other policy-relevant questions also require a solid coverage of less wealthy households and the composition of their balance sheets. For example, recent studies on the transmission of monetary policy show that "hand-to-mouth" households, who own low levels of liquid wealth, exhibit a higher marginal propensity to consume (MPCs). Hence, these households are more important than the top wealthy in determining the response of aggregate demand to monetary policy (Jappelli and Pistaferri, 2014 & 2020). Furthermore, households at the lower end of the wealth distribution are more indebted (relative to their overall net wealth). They therefore pose a higher risk to financial stability. It is thus important to measure precisely the financial positions of these households and gain a better understanding of their economic behaviour.

Preserving the large share of the wealthy.

Despite the increased need for a better coverage of lower parts of the wealth distribution, the PHF also aims to maintain a good representation of households at the top of the distribution. Analysis of the portfolio structure of the wealthy, their reaction to changes in monetary policy or to shocks is still high on the agenda of many researchers and policymakers. Attempts to construct distributional national accounts, which have recently been made, require a good coverage of the wealth aggregates held by the households at the top of the wealth distribution.

How the new design helps to achieve both goals

To reconcile the requirement to achieve a better regional coverage and maintain an overrepresentation of wealth households, the new sampling design features the following:

a) explicit stratification by German federal states ("Bundesländer") in addition to stratification by wealth as in the previous waves
b) oversampling regions in eastern Germany, rather than wealthy regions

Maintaining the stratification based on wealth (as in the previous waves) ensures that the refresher sample also covers the wealthy, despite not oversampling them. Furthermore, the lower attrition rate of the panel component as compared to the refresher component ensures a large share of wealthy households in the wave 4 "cross-sectional" sample, which consists of the refresher and panel households. Table 1 shows the number of households with net wealth above a given threshold as a percentage of the total of sample of each PHF

wave. The percentage of wealthy households in the panel component is higher than the refresher sample in waves 2 and 3. More importantly, the absolute difference in this percentage between the panel and refresher component grew between waves 2 and 3. This is due to the lower attrition rate of wealthy households. Hence, the panel is increasingly important in determining the degree of representation of the very wealthy households in the unweighted total sample. This implies that the presence of the panel component leads to a larger share of wealthy households in the total sample (before weighting) compared to the general population.

Table 1: Households with net wealth above a given threshold in the PHF sample (% of total of sample)

	> €0.8 mn.	> €0.9 mn.	>€1.0 mn.	> €2.0 mn.	> €3.0 mn.	Obs.
Wave 1:						
Total	9.5	8.1	7.0	2.4	1.2	3565
Wave 2:						
Panel	6.4	5.5	4.7	1.5	0.9	2138
Refresher	5.3	4.4	3.7	1.3	0.8	2270
Total	11.8	10.0	8.5	2.8	1.7	4461
Wave 3:						
Panel	11.1	9.4	8.1	2.7	1.2	3243
Refresher	4.1	3.6	3.1	0.9	0.4	1607
Total	15.3	13.1	11.3	3.6	1.7	4942

Source: Panel on Household Finances, Wave 1-3, Deutsche Bundesbank.
Note: Split Households are included only in the total sample. Unweighted figures.

A better regional coverage is obviously achieved directly through explicitly stratifying by German federal states. Additionally, no longer oversampling wealthy regions, which are concentrated in some federal states, also helps in achieving a better regional coverage.

Finally, not oversampling wealthy regions will automatically shift sampling mass to lower parts of the distribution. The oversampling of eastern Germany, where relatively more poor households live than in the west, also plays a part in the better coverage of lower parts of the wealth distribution.

4. Simulations to estimate the impact of the new sample design on wealth estimates

To test the effects of modifying the sample design on net wealth and income estimates, we conducted a simulation study. We construct the sampling base for the simulation by expanding the wave 3 refresher sample using the design weights. We define strata by the intersection of German states and the four strata described in section 2 (wealthy and other small and middle-sized municipalities, wealthy and other street sections in large municipalities). We implemented three different sampling rates: one proportional to the size of the strata, one where eastern Germany is oversampled by 25%, and one where eastern Germany is oversampled at 50%.

Tables 2 and 3 use the wave 3 refresher sample as a benchmark and report how these sampling strategies affect net wealth statistics. Table 2 shows the percentage of households with net wealth above a given threshold. The numbers are similar to those for the refresher households in wave 3, and even slightly higher, except for the top threshold of net wealth above €3 million. The good coverage of the wealthy in the simulations is encouraging. However, they should be interpreted with caution. The fact that we sample from an artificially constructed population, and not from the real population in Germany may influence the simulation results. It is feasible to assume that these percentages would be lower in a sample drawn from the real population.

Table 2: Households with net wealth above a given threshold: wave 3 (refresher) and simulated samples

	> €0.8 mn.	> €0.9 mn.	> €1.0 mn.	> €2.0 mn.	> €3.0 mn.
Wave 3:					
Refresher	4.1	3.6	3.1	0.9	0.4
Simulated Samples:					
Proportional to size	4.7	4.2	3.5	1.0	0.3
Oversampling the East, 25%	4.4	4.0	3.3	1.0	0.3
Oversampling the East, 50%	4.1	3.7	3.1	0.9	0.3

Source: Panel on Household Finances, Wave 1-3, Deutsche Bundesbank.
Notes: Row 1 shows the number of households with net wealth above a given threshold in the refresher sample as a percentage of the total sample of wave 3. Rows 2, 3 and 4 show the percentage of households with net wealth above a given threshold for each simulated sample. All figures are unweighted.

Table 3: Net wealth estimates for wave 3 refresher and simulated samples

Panel A. Mean and median net wealth

	Mean			Median		
	Total	East	West	Total	East	West
Wave 3:						
Refresher	218,126	88,701	255,434	53,500	13,400	69,500
Simulated Samples:						
Proportional to size	209,013	89,410	240,448	56,000	15,800	67,150
Oversampling the east, 25%	210,502	91,661	241,631	54,300	15,700	67,150
Oversampling the east, 50%	211,694	90,786	243,248	54,300	15,120	67,150

Panel B. Standard errors and sample sizes

	Standard Error (of the mean)			Sample Size		
	Total	East	West	Total	East	West
Wave 3:						
Refresher	12,174	11,653	14,731	1,607	270	1,337
Simulated Samples:						
Proportional to size	12,175	7,375	15,154	1,994	415	1,579
Oversampling the east, 25%	12,450	7,328	16,003	1,996	518	1,478
Oversampling the east, 50%	12,743	6,778	16,977	1,989	618	1,371

Source: Panel on Household Finances, Waves 1-3, Deutsche Bundesbank.
Notes: Figures for the wave 3 refresher sample are weighted with the design weights. Figures for the simulation sample with oversampling the east by 25% and 50%, respectively, are weighted with the inverse of the oversampling rate.

Table 3 shows the net wealth estimates from the wave 3 refresher and the simulated samples (mean and its respective standard errors, median as well as the sample size). Using the design-weighted refresher sample of wave 3 as a benchmark, the simulated estimates of the means for the whole sample are close (within 5%) to corresponding estimates of the benchmark (Panel A). The estimate for overall net wealth is 4% lower in the simulated sample. This stems from the drop in the corresponding estimate for western Germany. Across the three simulations, the largest changes in the estimates occur after dropping the oversampling of the wealthy and implementing a proportional-to-size sampling. The changes in the estimates induced by

changing the oversampling rate in the east are smaller than in the west.[5] Most importantly, oversampling the east improves the standard errors of the estimates in the east (Panel B). Thus oversampling the east results in more precise regional estimates, in particular for eastern Germany.

Based on the simulations above, we do expect that the change in the sample design may have some effect on the wealth estimates for the refresher sample. However, because of the presence of the panel component and its importance, both in terms of composition and relative sample size, we expect the effect on the overall net wealth estimates to be relatively small.

5. Practical Implementation

Based on the simulation study, the survey team decided to allocate roughly 32% of the total number of sampling points to eastern Germany for the refresher component of the next PHF wave.[6] A sampling design proportional to population size would have implied allocating about 21% of the sampling points in eastern Germany. Therefore, eastern Germany is oversampled at a rate of 54%.[7]

A sample based on the new design was originally drawn to be used for the fourth wave of the PHF, which was scheduled to take place in spring 2020. Because of the Covid-19-pandemic, the fourth wave, based on CAPI, had to be postponed. A shorter PHF survey, based on PAPI, will be conducted between September and December 2020. As an intermediate survey between waves, it will use the sample drawn following the new design.

Assuming a response rate of 16.1% (AAPOR 2) for the refresher component and 69.9% for the panel component, as in wave 3, we expect the refresher component to constitute 31% of the full sample in wave 4.[8] After completion of the PAPI survey, an evaluation of the sampling design is envisaged. This will inform the sampling for future refresher samples in the PHF.

6. Conclusions

The paper describes the adjustment to the sample design of the Panel on Household Finances (PHF). The new design aims at increasing the precision of the wealth estimates at the lower end of the distribution, while maintaining a substantial number of households at the top of the distribution, and at achieving a better regional coverage than previous waves. A simulation analysis indicates that both goals can be achieved with a design that combines the regional, wealth-based stratification implemented in the first three waves of the PHF survey, with an explicit stratification by German federal states and an oversampling of eastern Germany. The new design will first be implemented for a PAPI survey scheduled to start in September 2020. The plan is to oversample regions in the east at a rate of 54%. The new design has enough flexibility to oversample both specific regions as well as specific wealth strata in future waves.

Acknowledgements

The authors wish to thank seminar participants of the workshop of Panelsurveys in Germany and the Joint Statistical Meeting for their useful comments and suggestions.

[5] We did, however, notice large changes induced by individual, very wealthy households, for example, in the case where they were the only household within a stratum (for example, in one wealthy small municipality in Saxony). This should be avoided in the actual sample.
[6] The total number of sampling points to be drawn is 170.
[7] One cell in eastern Germany was set manually in order to cover almost all cells in eastern Germany. Therefore, the final oversampling rate deviates from the target rate at 50%.
[8] The gross sample of the refresher component comprises 10,000 addresses. The full sample from wave 3 comprises 4,942 households. For the response rates, see also Altmann et al. (2020).

References

Ann-Kristin Kreutzmann, Philipp Marek, Nicola Salvati, and Timo Schmid. 2019. "Estimating regional wealth in Germany: How different are East and West really?" Deutsche Bundesbank. Bundesbank Discussion Paper 35/2019.

Friehe, Tim and Mario Mechtel. 2014. "Conspicuous consumption and political regimes: Evidence from East and West Germany." European Economic Review, 67: 62–81.

Fuchs-Schundeln, Nicola. 2008. "The Response of Household Saving to the Large Shock of German Reunification." American Economic Review, 98(5): 1798–828.

Goldfayn-Frank, Olga and Johannes Wohlfart. 2019. "Expectation formation in a new environment: Evidence from the German reunification." Journal of Monetary Economics.

Jappelli, Tullio and Luigi Pistaferri. "Reported MPC and Unobserved Heterogeneity." American Economic Journal: Economic Policy, 2020(forthcoming).

Jappelli, Tullio and Luigi Pistaferri. 2014. "Fiscal Policy and MPC Heterogeneity." American Economic Journal: Macroeconomics, 6(4): 107–36.

Kristina Altmann, René Bernard, Julia Le Blanc, Enikö Gabor-Toth, Malik Hebbat, Lisa Kothmayr, Tobias Schmidt, Panagiota Tzamourani, Daniel Werner, and Junyi Zhu. 2020. "The Panel on Household Finances (PHF) – Microdata on household wealth in Germany." German Economic Review, 21(3): 373–400.