the review of income and wealth

Review of Income and Wealth
Series 0, Number 0, Month 2019
DOI: 10.1111/roiw.12452

A NOVEL SAMPLING STRATEGY FOR SURVEYING HIGH NET-WORTH INDIVIDUALS—A PRETEST APPLICATION USING THE SOCIO-ECONOMIC PANEL*

Carsten Schröder

SOEP at DIW Berlin and Free University Berlin

Charlotte Bartels

SOEP at DIW Berlin

Markus M. Grabka

SOEP at DIW Berlin

Johannes König

SOEP at DIW Berlin

Martin Kroh

SOEP at DIW Berlin and University of Bielefeld

AND

Rainer Siegers

SOEP at DIW Berlin

High-wealth individuals are typically underrepresented or completely missing in population surveys. The lack of comprehensive national registers on high-wealth individuals in many countries challenged previous attempts to remedy this under-representation. In a novel research design, we draw on public data on the shareholding structures of companies as a sampling frame. Our design builds on the empirical regularity that high-wealth individuals are likely to hold at least part of their assets in the form of shareholdings. Based on data from over 270 million companies worldwide, we select all individuals who are both German residents and registered shareholders of companies. In a pretest, we interviewed 124 households from a gross sample of 2,000 anchor persons. Our analysis shows that values of shareholdings from register data highly correlate with individual ranks in the wealth distribution, that the quality of personal information, particularly the residential address, is sufficiently high for subsequent interviewing, and that the approach can fill a major data and research gap in the study of high-wealth individuals.

JEL Codes: C83, D31

Keywords: sampling method, wealth, top wealth

Note: We would like to thank the German Federal Ministry of Labour and Social Affairs for financial support; Bureau van Dijk for technical support; Paul Brockmann, Konstantin Göbler, Christoph Halbmeier, Fabian Nemeczek, Christopher Prömel, Axel Ramstein, and Thomas Rieger for research assistance; Deborah Anne Bowen and Adam Lederer for editing.

*Correspondence to: Carsten Schröder, SOEP at DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany (cschroeder@diw.de).

1. Introduction

Wealth concentration is on the rise in many developed and developing countries (Alvaredo *et al*., 2017). In light of the crucial importance of wealth in modern economies and its high concentration, the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz-Sen-Fitoussi report) and the G20 Data Gaps Initiative emphasize the need to improve information on the distributions of wealth, income, and consumption.

Indeed, there are major deficits in the available data on high-wealth individuals and households, here defined as the top 1 percent, in many countries. Many countries do not impose a common general tax on wealth holdings and, thus, lack population registers with detailed wealth information. There is also evidence that a great deal of the wealth of the wealthiest is hidden in tax havens and, as such, unobserved in administrative data (Alstadsaeter *et al*., 2018). The resulting limitations of register data not only impedes drawing substantive conclusions in empirical studies on the entire wealth distribution, but also consistently estimating inequality or concentration indices like the Gini coefficient or quantile ratios, which are sensitive to the inclusion of the richest households in the sample. The introduction of dual taxation of labor and capital income in many European countries since the 1990s exacerbated the data problem.

Household surveys are another potential data source. Such surveys allow researchers to link wealth holdings with the broad set of surveyed background variables in domains like employment, family composition, psychology, health, education, etc. Yet, it is unlikely, by definition, that high-wealth individuals end up in a probabilistic sample of a few thousand individuals. Furthermore, the individual willingness to participate in surveys declines systematically with increasing assets (see D'Alessio and Faiella, 2002; Sànchez Muñoz, 2011; Westermeier and Grabka, 2015), reducing the share of high-wealth individuals in surveys to levels even below their actual share in the population. While the low sampling probability of the wealthiest reduces the precision of statistical estimates, the non-random nonresponse introduces bias.

Administrative register data can help in oversampling wealthy households in scientific surveys. Successful examples that rely on tax data include the Survey of Consumer Finances (conducted by the US Federal Reserve Board), Encuesta Financiera de las Familias (Bank of Spain), and Enquête Patrimoine (Bank of France/INSEE) (Vermeulen, 2018). Following these positive examples, the Eurosystem's Household Finance and Consumption Network (HFCN) made attempts to oversample relatively wealthy households. Yet, administrative data for an appropriate sampling frame was unavailable for all Euro-area countries except France and Spain (see Finances and Network, 2013). As an alternative, oversampling individuals by regions with above-average tax returns, as exemplified in the PHF Survey of the German Bundesbank, was implemented, but did not substantially improve the representation of high-wealth individuals in the survey.

In sum, there is a major data gap regarding the representation of the high-wealth group in many countries. Figure 1 depicts this data gap, taking the German Socio-Economic Panel (SOEP) as an example. The SOEP is an ongoing panel

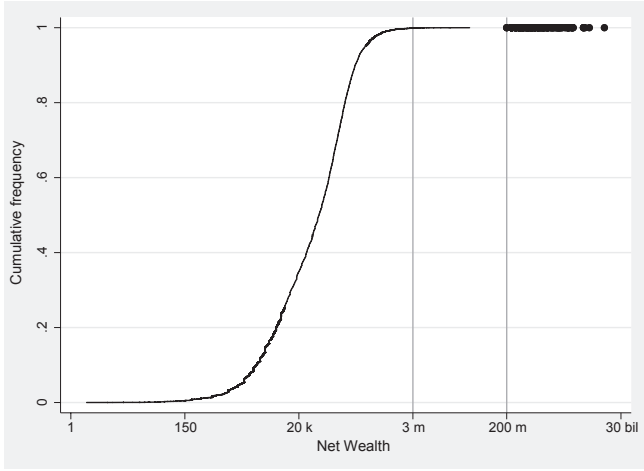


Figure 1. Cumulative Density of Net Wealth in the SOEP and *Manager Magazin*
Note: Net wealth in euros. Solid line: Individual net wealth in SOEP (positive net wealth only); Dots: individuals from Manager Magazin's Top 500 rich list for 2014. Net wealth was plotted on a log scale. 2012 SOEP data.

study of persons living in private households in Germany. The first wave was conducted in 1984 and since then is repeated each year. In 2012, more than 20,000 adult respondents took part in the survey. See Goebel *et al*. (2019) for documentation on the SOEP. It shows the cumulative density of individual net wealth according to the SOEP[1] as well as the richest individuals according to *Manager Magazin*. In the spirit of Forbes 500 the Manager Magazin provides a rich list for Germany. The vertical lines delineate the data gap. The figure shows that the number of cases with assets in the double-digit millions in the SOEP is already very small. In fact, the highest net wealth recorded in the SOEP is "only" around 40 million euros. According to *Manager Magazin* there are at least 100 billionaires in Germany. This means that there is a glaring data gap concerning the top tail of the wealth distribution. Several studies fill this gap using estimates based on rich lists and the assumption of a Pareto distribution; that is, the assumption of a log-log linear relationship between assets and the empirical distribution (e.g. Chakraborty and Waltl, 2018; Vermeulen, 2018).

We propose a novel sampling strategy for oversampling the group of high-wealth individuals in scientific surveys in countries where appropriate administrative tax data are lacking, and then show our results from a pilot study in Germany. Improving existing research data infrastructures with our novel sampling strategy will increase the precision of and reduce the bias in wealth measurements. For example, large uncertainty exists regarding the share of wealth held by the top percentile of the distribution. Furthermore, the strategy can improve comparability of wealth inequality measures between countries that cannot draw on administrative data for oversampling (and adopt our sampling strategy) and those that can.

[1]The SOEP is one of the very few population surveys which collects information about assets at the individual level for all adult household members.

Our concept draws on two considerations: First, individuals with high wealth are likely to hold at least part of their assets in the form of company shares. Entrepreneurship is an important way to accumulate large fortunes. At the same time, individuals with large fortunes frequently invest in companies. Second, companies are not only required to publish information on their financials but also on their shareholding structures—in the form of names, private addresses, and shareholdings. We use this publicly available information to determine shareholders living in Germany, who are invested in at least one company around the globe. From this population, we draw a probabilistic sample of individuals with substantial shareholdings, stratified according to the value of their shareholdings, and survey them using standard SOEP instruments including questions on their asset holdings. The implementation of this concept results in a number of specific challenges that go beyond 'standard' data quality issues in surveys. We tackle these specific challenges as follows:

1. **Individuals with substantial shareholdings.** A key condition for the successful implementation is to construct a database as complete as possible including the names, private addresses, and shareholdings of German residents with substantial shareholdings. The data provider (Bureau van Dijk) defines the threshold of at least 0.1 percent of all shares of a company to be substantial. in companies worldwide. The focus is not on free-floating shares or small business owners (in particular small partnerships), but on individuals with substantial business shareholdings in companies with high revenue. We used the global company database *Orbis* of the business information publisher Bureau van Dijk (BvD), which documents information on more than 270 million companies around the globe and their ownership structures. In some cases, these structures are highly complex, for instance, if companies are intertwined in parent-subsidiary relationships. This creates the technical challenge of converting a global company database with complex corporate ownership structures into an individual database.
2. **High-wealth individuals.** Our concept rests on the assumption that high net wealth individuals have substantial shareholdings. However, Orbis does not contain direct information on the monetary value of shareholdings for individuals, rather only the percentage share of the company they hold and financial figures of the company (revenues, profits, etc.). In order to monetized individual shareholdings, we must first estimate the company value. Using that, we monetized each ownership percentage. This allowed us to calculate the value of total shareholdings for each individual.
3. **Shareholder contactability and willingness to be surveyed.** The shareholders from the individual database have to be contactable to conduct the survey. To make contact, we need current, accurate address information. The present study faced the challenge, that Orbis sometimes lists the company address as the private address, which complicates direct personal contact with the target person. Therefore, we tested the quality of the addresses in the data base. Further, they have to be willing to participate in the survey.

Our feasibility study provides five important insights. First, the Orbis-based database is large, enabling the concept to be implemented comprehensively in the framework of a full-scale study. In Germany, for example, we find 1.5 million residents owning substantial shares in German or foreign companies. Second, high-wealth individuals listed in *Manager Magazin* are included in the Orbis database, suggesting that our concept successfully captures high-wealth individuals at the very top of the wealth distribution. Third, the willingness of the Orbis population to participate in the SOEP survey was larger than expected, with an unadjusted response rate of 6 percent and a cooperation rate of 23 percent, which is an important prerequisite for a full-scale implementation. Fourth, the pretest results suggest that our concept is capable of effectively closing the data gap in scientific surveys in the high-wealth area: The average net wealth of the 124 pretest individuals was over 10 million euros. This exceeds the SOEP average by a factor of 100. The highest individual net wealth in the pretest was over 200 million euros. Fifth, the estimated shareholdings of respondents from our database are strongly positively correlated with the individual net wealth stated in the SOEP interview. Hence, our strategy allows for effective oversampling of individuals from the top end of the wealth distribution.

Overall, our concept promises to overcome the problem of systematic under-coverage of high-wealth individuals in surveys, and not just in Germany. A full-scale survey is likely to provide valuable insights into the concentration of wealth, particularly if designed as a panel study. It could also generate important new insights on the following topics: inter- and intra-generational transmission of wealth; wealth accumulation over the life course and the determinants thereof; the composition of wealth; personality traits of high-wealth individuals; social engagement; and so on. Furthermore, a full-scale study will improve the data infrastructure at the top-end of the income distribution, since income and wealth are strongly positively correlated.

The note is structured as follows: Section 2 explains how the base population is determined by describing the Orbis company database, the transformation of the data into a database of individuals, and the method used to monetize company shareholdings. Section 3 discusses the suitability of the Orbis database for representation of companies and high-wealth individuals. Sampling and weighting methods are explained in Section 4. Section 5 presents and discusses the results of the pretest. Section 6 illustrates a weighting scheme for a full-scale boost sample. Section 7 summarizes our conclusions from the feasibility study and provides suggestions for full-scale implementations of our proposed sampling approach.

2. Proxying the Reference Population

Our approach builds on an individual database containing as complete as possible information on the names, private addresses and shareholdings of persons residing in Germany with substantial shareholdings in companies internationally. Relevant persons for our sample are individuals with large shareholdings, because

we expect them to have large private asset holdings.[2] Neither owners of (small) corporations, for instance, in the trade or service sectors, nor persons with small shares in large companies (less than 0.1 percent) are of concern here, but rather persons with high-value shareholdings. We create a database for this group of individuals using the Orbis company database from the business information provider Bureau van Dijk (BvD).

Orbis records comprehensive information on more than 270 million companies,—including banks and insurers—all over the world, and is regularly updated. We use data as recent as August 2017. Orbis is based primarily on the published balance sheets of the companies. Various financial figures are listed, such as revenue, profits, equity, and liabilities, as well as the number of employees, headquarters, and shareholding structure. Balance sheets and company accounts also contain the names and addresses of individuals with substantial shareholdings. Our approach targets this group. While Orbis is designed as a company database, our approach requires individual shareholder data. Therefore, we transform Orbis' company database into a shareholder database. We carry out this transformation in three steps:

1. Orbis includes several sub-databases: in the first step we use the Orbis contact database in order to extract all persons residing in Germany who currently have shareholdings in at least one active company internationally. Here we filter (in Orbis: Search strategy) by the following traits:

- Contact gender: Male, Female, Unspecified
- Contact type: Individual
- Country of residence: Germany
- Current position: Shareholder not being a director in the same company, Shareholder also being a director in the same company
- Legal status: Active companies (including those with unknown status)

The result of this search is a database comprising the target group of shareholders but also persons without shareholdings who hold an official position in a company (e.g. executive director). The group with zero shareholdings is excluded.[3]

2. Every company in Orbis has a unique identification number (in Orbis: BvD ID number). This enables us to merge information on revenue, profits, number of employees, or the exact percentage of the shareholding from the company to the shareholder database created in 1.
3. The shareholder database created in 1. is merged with the company information extracted in 2.

As a result, our individual database contains the percentage shareholding and address data for about 1.5 million shareholders as well as various financial figures of the respective companies.

[2]If shares are owned by a couple or a family, all family members are listed separately in Orbis.
[3]In principle, the specification of the trait "current position" should exclude all cases with zero shareholdings. However, this did not turn out to be the case. We communicated these as well as other detected data problems and inconsistencies, to the data provider, hoping that this communication leads to data fixes in future releases.

2.1. *Monetization of Shareholdings*

To determine the population of high-wealth individuals, all percentage shareholdings in our database must first be monetized for each individual, and then accumulated across the firms the individual invests in. The result of this step is a single value per shareholder (*cumulative shareholdings*), which proxies the value of the individual's international shareholdings. Following this, the shareholders are sorted by their cumulative shareholdings to draw a stratified probabilistic sample that is surveyed at a later stage.

In the monetization process, we not only evaluated direct shareholdings but also all shares in subsidiary companies. In the Orbis company database, we must differentiate between the value of direct corporate shareholdings and their total value including indirect shareholdings. This distinction is important because financial statements are either consolidated or non-consolidated. Only consolidated statements include the values of the subsidiary companies on the balance sheet of the parent company.

If the financial statement is consolidated, the percentage of direct shareholdings together with the company value from the consolidated financial statement are sufficient to determine the value of the shareholding. If the financial statement is not consolidated, then the value of shares can be determined from the direct investment in the parent company (multiplied by the non-consolidated value of the same) *plus* the percentage share values in the subsidiaries.[4]

The value of shareholdings in a listed company is equivalent to the percentage shareholding multiplied by the market capitalization of the company. Not only in Germany, most companies are not listed on the stock exchange. Thus, we need a standardized valuation method for all companies. We use companies' revenue because it is more frequently available than other financial figures such as profit.[5] This method facilitates a comprehensive and uniform way of valuing companies and *cumulative shareholdings*. Of course, the revenue of a company is not equivalent to company value; however, a strong positive correlation between the two is expected. Bi-variate correlations between operating revenue, total assets, and enterprise value are high. We observe these three variables for a subset of the 1,504,611 companies in Germany. For these companies, the correlation between operating revenue and total assets is 0.875, for operating revenue and enterprise value it is 0.881, and for total assets and enterprise value it is 0.885. The frequently used capitalized earnings method in accounting relies on this high positive correlation between revenue and market capitalization. Further, an approximate valuation of cumulative shareholdings in euros that allows us to assign shareholders to broadly defined wealth strata, is sufficient for our sampling strategy. Their actual wealth portfolios are assessed when we survey the sampled individuals with SOEP's wealth questionnaire. Because companies disclose figures from their balance sheets with different time lags, we use the most recent revenue according to

[4]When generating the data sets for the pretest sampling, we only took two levels of shareholding into account to speed-up the data queries.

[5]Our method assumes a monotonic increasing relationship between revenues and company value. However, revenue relative to value may vary widely, depending on the business model, sector, size, ownership structure, period, etc.

Orbis with a maximum delay of five years (that is, the most recent revenue between 2012 and 2017). In the absence of revenue information,[6] we apply a three-step imputation:

1. For companies with revenue data between 2012 and 2017, we estimate revenues based on the following list of explanatory variables from the balance sheets: the latest available revenue, the latest available number of employees, the latest balance sheet total, the number of subsidiary companies, and the number of shareholders.
2. For companies with missing most recent revenues but available explanatory variables, we predict revenue based on the above model.
3. For companies with missing revenues and missing explanatory variables, we implement a randomization procedure in which these companies are assigned an evenly distributed discrete random variable from 1 to 100. With this, we assign a revenue of up to 700,000 euros from the 100 percentiles of the observed revenue distribution below 700,000. 700,000 euros is the revenue threshold for small corporations (Art. 267a HGB), which allows them to fulfill the otherwise applicable disclosure requirements to a lesser extent. This applies to 380,341 of the total of 1,504,611 companies.

We extract more than 1.5 million shareholders from the Orbis database who have shareholdings in more than 1.6 million companies worldwide. This evidences the well-known "home country bias" from the literature (see Coval and Moskowitz, 1999).

3. Suitability of the Orbis Data for Selecting a Sample of High-Wealth Individuals

The data extracted from Orbis are suited for drawing a sample of high-wealth individuals if it meets the following requirements:

1. *Representation of high-wealth individuals*. Income-tax data show that top income earners are business owners in Germany, while income from renting and leasing is negligible for this subpopulation (Bartels, 2019). This implies that large real estate wealth is frequently held by these people through a business, which, in turn, is included in our shareholder database. To verify that high-wealth individuals are included in our Orbis database, we research the 100 richest people, or more specifically, families, according to *Manager Magazin*, in our shareholder database.
2. *Suitability of the approach for company valuation*. As described above, we determine cumulative shareholdings from company revenues and shareholdings in percent. We validate this valuation method by, first, sorting all shareholders according to their cumulative shareholdings. Wealthy

[6]Very small corporations with headquarters in Germany (Art. 264 German Commercial Code (HGB)) must fulfill their duty of disclosure in accordance with § 326(2) HGB. Yet, there is no publishing requirement.

individuals, according to *Manager Magazin*, should position themselves at the top of the ranking. Second, the private assets of those persons surveyed with SOEP instruments should correlate strongly and positively with their cumulative shareholdings.

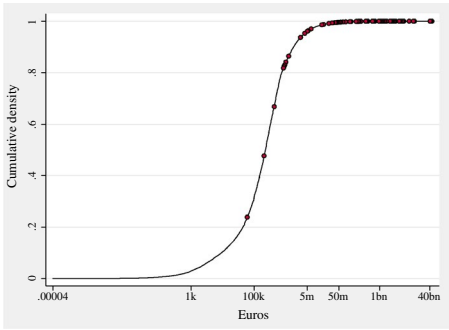
Below we demonstrate that the approach fulfills the above requirements and, consequently, allows us to effectively sample high-wealth individuals.

3.1. *Representation of High-Wealth Individuals*

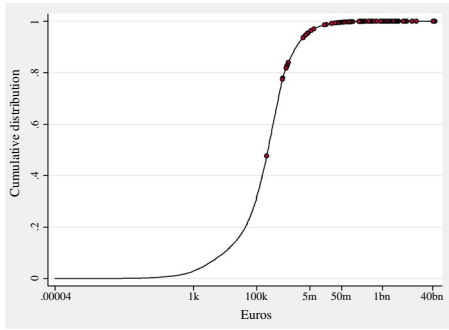
Manager Magazin publishes an annual overview of the presumably 100(500) wealthiest Germans. The assets specified therein are *estimates* based on freely accessible information and expert guesswork. Here we use the data from the 2014 edition of *Manager Magazin* as our company figures in Orbis show the greatest coverage for 2014. The list contains the 500 wealthiest individuals/family associations. If our working hypothesis that wealthy individuals also have substantial shareholdings is true, and further, if our individual database is complete, we should be able to find the wealthiest Germans according to *Manager Magazin* in our shareholder database. The search for the wealthiest Germans in our shareholder database is not trivial, as the observational unit of *Manager Magazin*'s wealth lists is both family associations and individuals. Of the 100 wealthiest Germans according to *Manager Magazin*, 38 are individuals and 62 are family associations. In 2014, Stefan Quandt, Johanna Quandt, and Susanne Klatten hold first place; they have shareholdings in BMW and Altana, among others, and their *shared* net wealth is estimated at 31 billion euros. Dieter Schwarz is in the fifth place: He owns shares, for instance, of Lidl and Kaufland, with his *individual* total net wealth estimated to be 14.5 billion euros. The Oetker family is in tenth place, but its members are not further specified. This means that before searching the shareholder database, we had to investigate family relationships for each family association; and in the case of common names, the companies in which the individuals are assumed to have shareholdings as well. Due to the complexity of the task, we only conducted this research on the 100 wealthiest Germans listed in *Manager Magazin*. To verify the representation of the wealthiest Germans in our shareholder database, we proceeded in two steps:

1. Using Wikipedia articles as well as newspaper and magazine reports available online, we investigated which individuals belong to a family association. The identification of the "head of the family" was generally straightforward. There was some uncertainty regarding (distant) family members.
2. We researched all individuals and family members in the shareholder database. In addition, we compared various traits: last name, first name, birth year, and shareholdings.

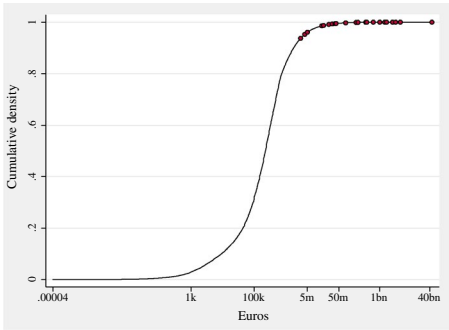
In total, we were able to match 404 individuals from the shareholder database to the 100 wealthiest German individuals or family associations listed in *Manager Magazin*. Effectively, the list entails 103 entries because some units share the same



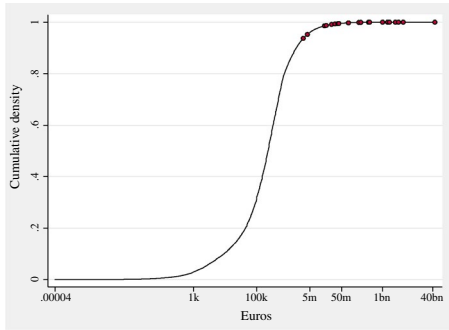
(a): Cumulative shareholdings in euro including uncertain matches. (families and individuals)



(b): Cumulative shareholdings in euro. Uncertain matches excluded. (families and individuals)



(c): Cumulative shareholdings in euro including uncertain matches. (single individuals)



(d): Cumulative shareholdings in euro. Uncertain matches excluded. (single individuals)

Figure 2. Positions of the Richest Germans According to *Manager Magazin* in the Distribution of Cumulative Shares

Note: Shareholder data from Orbis and Manager Magazin (2014). Solid line: Empirical density of cumulative shares; red dots: matched cases from Manager Magazin.

ranks. For 101 of 103 individuals or family associations, we were able to find at least one family member in the shareholder database. According to Orbis, 88 of these individuals reside in Germany. The remaining two unidentified cases are explained by one death and one family association (the Engelhorn family), who hold several places in the *Manager Magazin* list but whose members cannot be clearly matched. Overall, these results show that our shareholder database covers virtually the complete list of the wealthiest Germans according to *Manager Magazin*.

3.2. *Validity of the Estimated Cumulative Shareholdings*

Figure 2 shows the distribution of cumulative shareholdings for all 1.5 million German shareholders in the ORBIS data. Red dots indicate the position of one of the top 100 individuals or families on the *Manager Magazin* list. Further, we calculated cumulative shareholdings for each member in the shareholder database and assigned them to one person representing the family. Red dots indicate

TABLE 1
PERCENTILE POSITIONS OF THE RICHEST 100 GERMANS ACCORDING TO CUMULATIVE SHAREHOLDINGS

	Single Persons*	Single Persons	Single Persons & Families*	Singles Person & Families
Mean	99.15	99.24	96.52	97.24
SD	1.65	1.58	11.27	7.77
Percentile				
10	96.08	98.58	93.63	93.53
50	99.90	99.90	99.97	99.95
N. obs.	25	23	81	76

Note: Own computations using *Manager Magazin* (2014) and Orbis.
*Indicates the inclusion of uncertain matches.

either heads of families and individuals in panels (a) and (b) or they indicate individuals only in panels (c) and (d). Panels (a) and (c) include certain and uncertain matches in the Orbis database and the *Manager Magazin*. Panels (b) and (d) only use matches that are certain.[7] *A highly robust and consistent finding emerges: The wealthiest Germans according to Manager Magazin are concentrated at the top of the distribution of cumulative shareholdings in the Orbis database*. This is particularly true for individuals: here, the uncertainties are smaller than for family associations (due to the distribution of wealth within the family association; delineation of the family association; precision of estimated net wealth in *Manager Magazin*).

Table 1 summarizes these graphical findings. We sort the cumulative shareholdings across 100 percentiles. The row "Mean" indicates the average percentile position of the 100 wealthiest Germans according to *Manager Magazin* in the distribution of cumulative shareholdings in the Orbis shareholder database. The higher the reported "Mean", the stronger the concentration of the wealthiest in the top percentiles of cumulative shareholdings. In fact, depending on the sample definition, "Mean" varies between 96.5 and 99.2. *The 100 wealthiest Germans according to Manager Magazin highly concentrate in the highest percentiles of the distribution of cumulative shareholdings*. This demonstrates that cumulative shareholdings correctly identify the very rich, satisfying one of the above-mentioned crucial requirements of our sampling strategy.

4. SAMPLING PROCEDURE AND PRETEST FIELDWORK

The sampling procedure employs the shareholder database we created. The *sampling frame* consists of approximately 1.5 million shareholders (see Section 2) residing in Germany with international shareholdings. Individuals who share an address with at least five other shareholders are excluded, as this is a strong indication of a business rather than a residential address. For closer examination, we passed on addresses shared by a high number of individuals to the survey institute

[7]Of the 103 wealthiest Germans in *Manager Magazin*, 81 could be found in Orbis. When considering only individual persons, 25 could be assigned with only two uncertain assignments (see Column "Single persons∗" in Table 1).

before the actual pretest. Some addresses, especially in downtown urban locations, are residential buildings, yet in many cases they are commercial buildings. Address sharing is quite equally distributed across our shareholder database, i.e. it is not particularly frequent among top shareholders. As a result, we eliminated 1.4 percent of the total cases.

The target population of the sample are individuals from the top percentage of the adult population in Germany, about 600,000 individuals. For this reason, we dropped about 900,000 individuals from the total Orbis shareholder population.

As is common in face-to-face representative population surveys, we used a two-step sample design. In the first step, a number of regions were selected (*Primary Sample Units, PSU*), and in the second step, addresses within these regions (*Secondary Sample Units, SSU*). This two-step method facilitates surveying in face-to-face mode, because the addresses are spatially clustered.

PSUs are based on postal code areas. Neighboring postal code areas[8] were clustered into PSUs such that a similar number of persons from the target population lives in each PSU. Consequently, the geographical reach of the PSUs varies (see Figure 3).

A total of 1,111 PSUs were created. Figure 3 illustrates the regional differences with regard to the proportion of the target population in the total population. For the pretest, 400 addresses were selected in five regions,[9] conditional on having the following characteristics:

1. *Heterogeneity of regional structure*. Address quality, response behavior, and socio-economic composition may vary between rural, sub-urban, and urban areas. Thus, we selected sampling points that comprise downtown, peripheral and suburban areas, and rural areas.
2. *Low geographical dispersion*. To minimize travel for the interviewer in each region, regions with a high number of individuals in the target population in a smaller area were selected for the pretest.

As past experience shows, contactability and willingness to participate decrease as wealth increases. A stratified probabilistic sample of addresses was selected within the regions according to the value of cumulative shareholdings. This divided the target population into three strata: 1/7 of the addresses drawn came from the lower third, 2/7 from the middle third and 4/7 from the upper third of the top 600,000 in the distribution of cumulative shareholdings. Between the selected regions, the distribution across strata differs such that the selection probability of the gross sample varies across regions and strata between 5.3 percent and 34.3 percent.

Out of the 2,000 addresses, approximately 19 percent were incomplete or outdated, but an additional register search of the sampled persons lead to an accurate

[8]For approximately 0.8 percent of persons in the target population, there is no (correct) postal code in the database. Therefore, they are not included in the sampling frame.

[9]To guarantee a sufficient number of non-surveyed persons for a possible full-scale survey, four adjacent PSUs were clustered within these regions for the pretest, from which the survey addresses were ultimately selected.

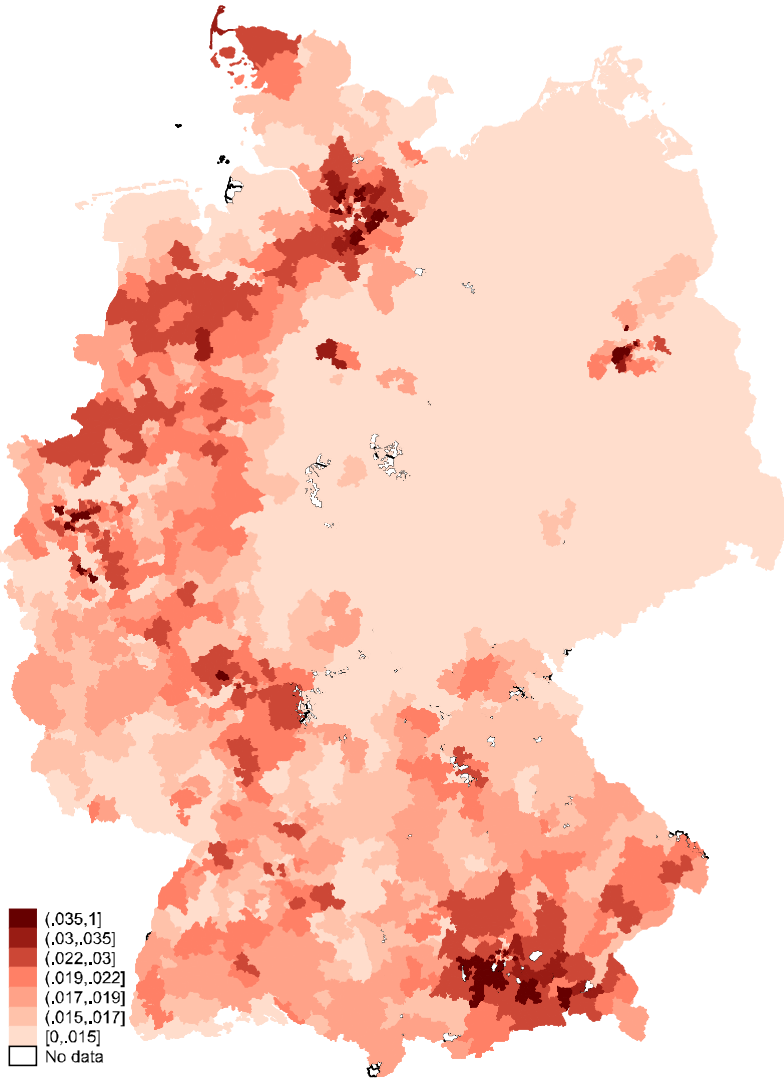


Figure 3. Share of Target Persons in PSU-Specific Adult Population
Note: We refrain from providing the boundaries of primary sampling units (PSUs). Thus connecting PSUs with same color cannot be distinguished.

address. In only 13 percent of all cases, a postal address of a sampled person was unavailable. This indicates an overall good address quality for a probabilistic sample that is not based on residential registry data (see Kroh *et al*., 2015). The likelihood of an invalid address does not systematically vary with a person's position in the shareholding distribution. Of the adjusted gross sample of 1,652 households, the fieldwork organization was unable to establish personal contact within the planned fieldwork period of about three months with the sampled person in 1,120 cases, that is 2/3rds of the sample. This high value indicates that future surveys in

TABLE 2
Response Rates in the Pretest

	Stratum 1	Stratum 2	Stratum 3
Share of Gross Sample	14.3	28.5	57.3
Share of Net Sample	18.5	33.1	48.4
Response rate	8.1	7.2	5.2

Note: Compiled from Kantar Fieldwork Report "Hochvermögende" Table 12.

high-wealth individuals need to invest in additional measures to guarantee personal contact with respondents and should extend the fieldwork period, as this target population is highly mobile and difficult to contact.

A total of 124 interviews were carried out successfully. This results in an unadjusted response rate of 6 percent. However, fieldwork ended prematurely because the agreed upon number of successful interviews (100) had been exceeded. The unadjusted response rate can be considered to be the lower limit of a maximum attainable response rate. Relating the 124 successful interviewees to the 532 persons who could indeed be contacted during fieldwork (difference between 1,652 and 1,120 individuals), results in a cooperation rate of approximately 23 percent. The initial response rate for probability samples of the general population in SOEP is about 35 percent, suggesting a surprisingly good response rate for the pretest sample.

Initial analyses demonstrate that the proportion of interviews realized decreases as the value of cumulative shareholdings (according to the strata) increases. It rises with age and is higher among men than among women. Among the individual regions (and interviewees) the differences are small.[10]

In sum, address quality as well as willingness to participate are sufficient for a full-scale survey.

5. Pretest Interview Results

The analysis described below had two objectives: first, to compare net wealth from the pretest and the SOEP; then, second, to compare net wealth across the three strata in the pretest. To do so, we combined the pretest and SOEP data for comparison. We refrained from further in-depth analysis because the number of cases from the pretest does not allow for generalizable statements.

5.1. *Pretest Response Rate*

Table 2 shows the response rates along the wealth strata as well as the shares of each stratum in the gross and net samples. Firstly, the adjusted overall response

[10]Fieldwork for the pretest began on November 6, 2017, and ended on February 2, 2018. The fieldwork institute Kantar Public conducted computer-assisted personal interviews (CAPI). The average interview length was 58 minutes. Through the use of laptops, interviewees had the option of answering sensitive questions about income and wealth without being seen directly by the interviewer. Due to the interview duration, personal verbal contact was important to avoid respondents breaking off the interview.

rate is 7.5 percent. Each of the strata contributes a response rate value roughly grouped around that mean with the lowest response rate in the third stratum. This is unsurprising as the top-wealth holders are predominantly in the last stratum and responses are bound to drop off for this group.

However, the share from stratum 3 in both the gross and net sample is very encouraging. Over half of the gross sample and close to half of all realized observations come from the wealthiest stratum.

5.2. *Self-Reported Asset Values*

The questionnaire used in the pretest was based on the regular SOEP individual questionnaire in 2017, including the SOEP questionnaire module, "Your personal balance sheet." This facilitates a direct comparison of information on net wealth and other characteristics of the SOEP respondents. The latest available information on the personal balance sheet was used for the comparative evaluation of the pretest and SOEP. The "Your personal balance sheet" module is surveyed in the SOEP at five-year intervals since 2002.

The "Your personal balance sheet" module of the SOEP questionnaire asks for individual wealth in order to increase the probability of fully recording all wealth components of all household members. Further, documenting wealth on the individual level opens up a whole set of research questions, that can be studied on intra-household wealth distribution, financial decision making, bargaining power, etc. However, it comes at the expense of potentially inconsistent responses and double counting (Sierminska *et al*., 2010). The order of the SOEP questionnaire directly addresses these concerns: First, it is asked if the individual holds a particular asset. Second, the market value of the asset is recorded. Finally, it is documented if the individual is the only holder, or, if ownership is shared, the share that is held by the individual.

The "Your personal balance sheet" module encompasses a total of 12 asset and debt positions:

1. Value of owner-occupied real estate assets
2. Outstanding debt for owner-occupied real estate assets
3. Value of other real estate assets
4. Outstanding debt for other real estate assets
5. Value of building loan contracts
6. Value of financial assets
7. Surrender value of life insurance and private pension insurance
8. Value of company or shareholdings in companies
9. Value of tangible assets
10. Outstanding debt in consumer loans
11. Value of vehicles
12. Outstanding debt in educational loans

For each position, a question is asked to filter whether the person possesses assets or debt according to the position specified. If this is true, a question about the value follows. In 2012, the "vehicles" and "educational loans" categories were not included. However, these are quantitatively negligible.

5.3. *Statistical Imputation of Missing Self-Reports*

All voluntary population surveys face the issue of non-response. Non-response may be total (unit non-response) or pertain to specific, individual, questions (item non-response). Item non-response is prevalent with sensitive questions. This is particularly true for questions about income and wealth (Frick *et al.*, 2010). Unit non-response is not corrected for with a weight adjustment because the pretest results are not generalizable. As item non-response can lead to systematic distortion—especially with information on income and assets—we statistically imputed missing values.

As with the wealth variables in the SOEP, we use the "multiple imputation by chained equations" (MICE) procedure for imputation (see Royston and White, 2011). This allows for the consideration of covariates of various scale levels and reflects uncertainty regarding the imputed values. MICE is based on a series of regression models, which simultaneously determine each variable that lacks a response with other variables. This preserves the variability of the data structure.

To group individuals into rough categories in each component of wealth an auxiliary variable was surveyed that indicates whether a respondent's value for that category falls within a specified boundary. We anticipated that there would be high rates of non-response in each component of wealth and additionally surveyed these boundary indicators to alleviate this issue. The non-response rates for the boundary indicators is significantly lower than the non-response on the wealth items themselves. Therefore, we use these boundaries in imputation as the boundary indicator constrains the range of true but missing values.

Thus, in the pretest implementation of the imputation procedure we use the boundary indicators in two steps: 1. If there is any non-response on a boundary indicator, we impute it using socio-demographics and the Orbis stratum variable. 2. We multiply impute each of the wealth components grouped by each category of the corresponding boundary indicator. Thus we ensure that none of the imputed values lie outside the range implied by the indicator.

5.4. *Wealth in SOEP and Pretest*

Individual net wealth was used for asset comparison between the SOEP and pretest as well as for comparing the three strata from the pretest. The results for other wealth aggregates are qualitatively comparable.

Table 3 allows for a descriptive comparison of net wealth in the SOEP and the pretest. Columns 2 and 3 describe the net wealth distributions using mean values, minimums, maximums, and percentiles. The comparison shows that only the minimum in the SOEP and the pretest are of about the same magnitude, at around -4 million euros in the SOEP and −2.5 million euros in the pretest.

The average net wealth in the pretest is higher than 10 million euros; thus more than one hundred times greater than in the SOEP (approximately 90,000 euros). In fact, net wealth across the entire pretest wealth distribution far exceeds that in the SOEP. The median of the pretest is around 2.3 million euros, while less than 20,000 euros in the SOEP. The top percentile value of the pretest is around 157 million euros. In the SOEP it is around 838,000 euros.

TABLE 3
Individual net wealth in million euro—SOEP and Pretest

		Pretest			
	SOEP	Strata 1–3	Stratum 1	Stratum 2	Stratum 3
Minimum	−4	−2.5	−2.5	−1.2	−2.5
Mean	0.09	10.3	4.2	3.3	17.4
Maximum	39.3	207	19.4	28.9	207
Percentile					
25	0	0.7	.,5	0,5	1.5
50	0.02	2.3	2	1.1	3.6
75	0.1	6.6	6	3.3	16.2
90	0.2	23.4	10.8	7.6	52.7
95	0.3	42.6	16.8	11.8	89
99	0.8	156.6	19.4	28.9	207
Observations	25.803	124	23	41	60

Note: Own computations using SOEP 2012 (unweighted) and Pretest (unweighted).

Thus our approach is successful in surveying high-wealth individuals. Indeed, net wealth in the 25th percentile of the pretest is at about the same level as the 99th percentile in the SOEP.

Columns 4 and 6 of Table 3 illustrate the extent to which stratification across the cumulative shareholdings in the pretest successfully differentiates persons in the upper tail of the net wealth distribution. In total, we divided the shareholder population into three strata along the distribution of cumulative shareholdings. While average net wealth for strata 1 and 2 do not systematically differ, considerably greater net wealth is found in stratum 3: Across percentiles, the net wealth in stratum 3 is significantly (a factor of 3 to 10 times) greater than that in the other two. The mean value of stratum 3 is 200 times greater than in the SOEP.

In sum, stratification based on shareholdings is an effective method of oversampling high-wealth individuals.

Figure 4 again demonstrates that the pretest data successfully close the existing data gap. It supplements Figure 1 with the pretest net wealth shown by red circles.

While less than 1 percent of SOEP respondents are in the group of millionaires, around 70 percent of the pretest respondents are millionaires. Twenty-four pretest respondents have a personal net wealth of at least 10 million euros, six are wealth over 50 million euros. In fact, at least one person with a net wealth of about 207 million euros was interviewed, a value in the range of the 500 wealthiest Germans according to *Manager Magazin*.

Thus, we find that the pretest data are distributed over the entire data gap, suggesting that our approach is capable of closing the gap.

The descriptive analysis does not test whether the differences between the SOEP and pretest (as well as among the pretest strata) are significant, nor does it control for covariates. Respondents in the pretest could have higher net wealth, because they are on average older and better educated than the SOEP average, or are more frequently male.

To estimate whether the wealth difference between the pretest and SOEP is significant, we used the model,

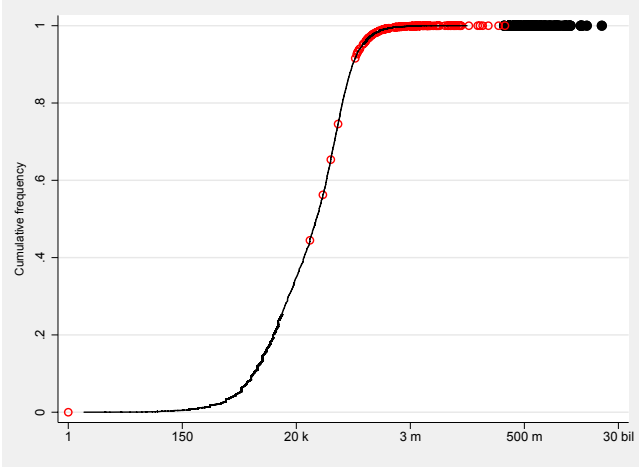


Figure 4. Cumulative Density of Net Wealth From SOEP, *Manager Magazin* and Pretest
Note: Solid line: Individual net wealth from SOEP (positive net wealth only) in 2012; black dots: single individuals from the top 500 according to Manager Magazin (2014); red circles: Pretest. Wealth plotted on a log scale.

$$NW_i = \alpha + \boldsymbol{\beta} \times \mathbf{X}_i + \gamma \times D_i^{Pretest} + \epsilon_i \tag{1}$$

with the dependent variable, NW_i, the logarithmic net wealth of respondent, i, from the pilot study. As covariates we include:

1. A vector of socio-demographic traits of each surveyed person i, $\mathbf{X}_i$. Included here are: age, occupational status, and gender.
2. A 0-1-coded dummy variable, $D_i^{Pretest}$, which specifies whether the person participated in the pretest or the SOEP. If the person participated in the pretest, the dummy assumes the value 1. The larger the regression coefficient, γ, of the $D_i^{Pretest}$, the larger the conditional difference in average assets between the pretest and SOEP interviewees, as well as the more successful the sampling procedure with respect to targeting of high-wealth individuals.

A further specification allows for determining the selectivity of the strata. K_i^{Strat} is a system of categorical dummy variables with *Strat* = (1,2,3), where SOEP cases were coded as 0. The estimation equation is,

$$NW_i = \alpha + \boldsymbol{\beta} \times \mathbf{X}_i + \boldsymbol{\gamma}^{Strat} \times \mathbf{K}_i^{strat} + \epsilon_i. \tag{2}$$

The more strongly the coefficients γ^{Strat} increase with the stratum, the more clearly the strata discriminate between wealth levels.

Table 4 summarizes the regression results. Here we only considered observations with non-negative net wealth. Columns I and III show the results for both model specifications in the baseline variant without socio-demographic

TABLE 4
Regression results on the level of net wealth and stratum

	I	II	III	IV
Pretest	4.121***		1.652***	
	(0.194)		(0.187)	
Stratum				
1		4.037***		1.869***
		(0.339)		(0.332)
2		3.291***		1.322***
		(0.401)		(0.379)
3		4.723***		1.828***
		(0.224)		(0.224)
Age			0.173***	0.173***
			(0.004)	(0.004)
$(Age)^2$			−0.001***	−0.001***
			(0.000)	(0.000)
Female			−0.209***	−0.209***
			(0.021)	(0.021)
Employment				
Employee			0.033	0.033
			(0.030)	(0.030)
Self-employed without			0.484***	0.486***
Employees			(0.057)	(0.057)
Self-employed with 1–9			1.096***	1.096***
Employees			(0.065)	(0.064)
Self-employed with			2.003***	1.966***
10 and more employees			(0.127)	(0.125)
Log Household			1.029***	1.029***
Net income			(0.022)	(0.022)
Constant	10.58***	10.58***	−2.569***	−2.568***
	(0.013)	(0.013)	(0.192)	(0.192)
N	19,102	19,102	19,101	19,101
Adj. R^2	0.032	0.033	0.374	0,374

Note: Data from SOEP v.33.1 and Pretest; own calculations. Robust standard errors in parentheses. Dependent variable: Log of net wealth.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

control variables. Column I only takes into account the dummy for the pretest. The regression constant is 10.58 and reflects the average of logged values of individual net wealth in the SOEP. The coefficient for the pretest dummy is $\gamma = 4.121$, which reflects the average difference in logged values of assets between pretest and SOEP. Remember that the mean of logged values is not equal to the logged mean of the original non-logged variable, i.e. $E(log(y)) \neq log(E(y))$. Column *II* uses dummies for the three strata of the pretest. All three strata coefficients are significant and positive. This means that the average net assets in each pretest level are higher than in the SOEP. Further, the value of the coefficient for the third stratum is considerably higher than the other two strata's coefficients. This reinforces the descriptive finding that the highest assets are indeed found in stratum 3.

In columns III and IV, socio-demographic variables supplement the explanatory variables of the baseline regression. The results show that after controlling for explanatory variables, the net assets in the pretest are still greater than in the

SOEP. Thus, a simple screening of the control variables considered here would under no circumstances have attained the efficacy of the sampling strategy suggested here.

We conclude from the regression results that the net wealth in the pretest is significantly and quantitatively much higher than in the SOEP, and that the stratification enabled a targeted sampling of respondents with wealth holdings located in the data gap.

6. Weighting Scheme for a Boost Sample

This section describes a weighting approach for a probability boost sample of 5,000 households drawn from the Orbis database. As the sample size of the pretest ($n = 124$) is too small to draw any inferences for the underlying population, we outline the general idea of the weighting approach instead of implementing it empirically on the basis of pretest data.

A design weight approach builds on the selection probabilities assigned during sampling (Horvitz and Thompson, 1952). Section 3 details the two-stage sampling procedure of, first, sampling regional sample points (*Primary Sampling Units, PSU*) and, second, persons residing in these areas (*Secondary Sampling Units, SSU*). The sampling frame consists of $N = 600{,}000$ SSU allocated to $M = 1,111$ PSU, i.e. $\sum_{m=1}^{M} N_m = N$. The number of SSU per PSU, N_m, varies between 400 and 600. In probability proportional to size (PPS) sampling at the first stage, sampling probability of, for instance, 100 PSU is $\pi_m = 100 \times \frac{N_m}{N}$. Sampling of, for instance, 50 SSU from each of the 100 sampled PSU gives the sampling probability at stage 2, which is $\pi_n = 50 \times \frac{1}{N_m}$. Thus, the joint selection probability is $\pi_{n|m} = \pi_m \times \pi_n = 100 \times \frac{N_m}{N} \times 50 \times \frac{1}{N_m} = \frac{1}{120}$. The sampling design can be augmented, of course, by stratification in the first stage (e.g. by federal state) as well as the second stage (e.g. by level of shareholding).

SOEP surveys all members of a household. Thus, the sampling probability of a person π has to be corrected for the selection probabilities of all persons of the same household, h, eligible for sampling, which is $\pi = \sum_{1}^{N_h} \pi_{n|m}$. The number of eligible persons per household can be established before sampling, for instance, by assuming that persons with the same postal address in the Orbis database form a household. Alternatively, the number of eligible persons per household can be established after sampling, by determining all members of a participating household in the Orbis database.[11]

The inverse probability weighting, known as the Horvitz-Thompson-Estimator (Horvitz and Thompson, 1952), represents the design weight of the gross sample, which is $w = \pi^{-1}$. This design-weight approach can be extended by non-response weighting to consider selective response probabilities in the survey. In fact, the wealth of individual information from the Orbis database as well as publicly available contextual information on residential addresses permits us to

[11]The first approach allows a preferable self-weighting sampling design (i.e. PPS), however, it builds on the assumption of the same postal address indicating a shared household. Both approaches permit design weights correcting the realized sample of households selected from a person register.

predict response probabilities drawing on a large number of explanatory variables. In turn, the inferred cross-sectional weights adjust the sample to a large set of margins of the underlying sampling population (Kroh *et al*., 2018).

Thus, the weighting approach described so far permits inferences regarding the underlying population of persons listed in the Orbis database, i.e. the top percentile of *shareholders* in Germany. The target population, however, is the top percentile of the *wealth holders* in Germany, raising the issue of coverage error. More specifically, the sample does not cover persons of the top percentile of the wealthy without significant shareholdings. Thus, these cases represent false-negative cases, or type-II-errors of the sample. The sample, in turn, contains false-positive cases that represent type-I-error, i.e. sampled persons of the top percentile of shareholders who are not among the top percentile of the wealthy.

Section 5.4 details that 70 percent of the respondents from the pretest report to be millionaires, the threshold for being in the top 1 percent in SOEP. Hence, 30 percent of the pretest respondents are not among the top-wealth population (type-I-error). We expect to reduce this false-positive rate in the planned boost sample by half by optimizing the monetization of shareholdings and intensifying stratification. To avoid the strong assumption that this coverage error is uninformative for inferences drawn from the data, we propose integrating the Orbis sample in the existing SOEP samples of the general population. The reason for this iss that members of the top-wealth population excluded from sampling of the new sample are included in the existing SOEP samples, albeit in small numbers.

Table 5 provides a stylized description of the integrated weighting approach and estimated incidences under the assumption of a 15 percent type-I-error rate. Column "Base" defines our base population, approximately 60 million adults residing in Germany. This population can be broken up into the top 1 percent ("Wealthy") and the ("Non-wealthy"). Ideally, the top 1 percent should coincide with the 600,000 shareholders with the highest cumulative shareholdings.

For the purpose of this illustration, we assume a type-II-error rate of 15 percent, where roughly 10 percent wealthy persons (60,000) do not have any shareholdings and about 5 percent (30,000) have insignificant shareholdings. The non-wealthy population, about 59.4 million individuals can also be divided into three groups: those being in the top-percentage of shareholders and representing the type-I-error (90,000 persons), those with insignificant shareholdings, and those without any shareholdings. Column "Orbis" defines the Orbis population. By definition, the Orbis population does not include persons without any shareholdings. Hence, the wealthy group in Orbis comprises 540,000 persons (top and insignificant shareholders), while the non-wealthy group in Orbis comprises 960,000 persons.

Column "SOEP" details the most recent SOEP wave, comprising roughly 25,000 adult respondents, in terms of the aforementioned sub-populations. Column"Boost" contains the group composition of the intended boost sample under the assumption of a 15 percent type-I-error rate, while "Integrated" contains the composition of the integrated dataset. The final column reports expected mean weights for the integrated sample that permit inferences about both the

TABLE 5
PLANNED WEIGHTING SCHEME ON THE BASIS OF POPULATION ESTIMATES

		Population		Sample			Weights
		Base	Orbis	SOEP	Boost	Integrated	Mean
Wealthy	Top shareh	510,000	510,000	210	4,250	4,460	114
(Top-percentile)	All other shareh	30,000	30,000	15	–	15	2,000
	Without shareh	60,000	–	25	–	25	2,400
Subtotal		600,000	540,000	250	4,250	4,500	133
Non-wealthy	Top shareh	90,000	90,000	50	750	800	113
(All other percentiles)	All other shareh	870,000	870,000	330	–	330	2,636
	Without shareh	58,440,000	–	24,370	–	24,370	2,398
Subtotal		59,400,000	960,000	24,750	750	25,500	2,329
Total		60,000,000	1,500,000	25,000	5,000	30,000	2,000

shareholding and the wealthy population in Germany. Inference on the latter group comes with a loss of efficiency of the estimation as weights vary considerably.

7. Conclusion and Outlook

In the public debate and in the research community, it is often criticized that a very small percentage of the population holds a large and growing share of society's wealth and that they transfer their wealth to the next generation at low tax rates. This does not fit the idea of a meritocratic society, where people ideally "earn their place" based on their individual effort.

Despite extensive discussion regarding this topic, there is still a lack of empirical data on the actual concentration of wealth in many countries. This is certainly true for Germany, where all major population surveys show glaring data gaps on individuals with assets beyond the low single-digit millions.

In this note, we present the results of a feasibility study in which we test a novel concept for closing this data gap. All of the results clearly show that our concept for collecting data on high-wealth individuals and their shareholdings would provide the basis for the first comprehensive survey of high-wealth individuals in Germany:

1. An important advantage of the proposed sampling strategy is that the population of high-wealth individuals is surveyed with the same survey instruments as the previously surveyed population, which allows direct comparisons of the two groups.
2. Based on the 2,000 addresses, 124 individuals were surveyed successfully. This is equivalent to an unadjusted response rate of 6 percent, although it should be kept in mind that the fieldwork phase was ended ahead of schedule. Willingness to participate among those contacted was around 23 percent.
3. The respondents are rich: 56 percent of the 124 pretest respondents have individual net wealth in the millions; approximately 20 percent have net wealth of at least 10 million euros and 5 percent over 50 million euros. In addition, at least one respondent had net wealth in the range of the 500 richest Germans according to *Manager Magazin*.

The concept presented and tested here has two major advantages over a sample that is stratified, for instance, according to the level of regional tax revenue: First, it is substantially more precise in defining the population of high-wealth individuals. This reduces the costs per surveyed high-wealth individual. Second, the results of the pretest indicate that if the sampling strategy were implemented in the framework of a full-scale survey, it could successfully close the data gap at the top of the distribution in Germany and other countries for the first time. A key advantage of the strategy is that probabilistic sampling enables inference and thus provides the basis for valid statements about the target group of high-wealth individuals.

Conducting a survey of this kind in a full-scale study could provide valuable data on the generation, concentration, and transmission of wealth as well

as socio-demographics, personality traits, and activities of the "rich." Income tax statistics are silent about the working hours invested by high-wealth individuals to gain business income from large partnerships. However, to understand the degree of meritocracy prevalent in a society, the amount of effort invested for a given return is essential information. The data could also provide valuable insights into the validity of the assumption shared by many studies worldwide that wealth at the upper tail is Pareto-distributed. A full-scale study conducted as a panel survey could, in the long term, offer important insights into the intra- and intergenerational transmission of wealth, the character of the country's economic elite, as well as the impact of taxes and external shocks (for instance, in the form of capital market volatility) on the various parts of the wealth distribution.

One should note, however, that our approach is not free of limitations. First, wealth measures for smaller top groups than the top 1 percent, such as the top 0.1 percent or top 0.01 percent, are unlikely to be improved as the number of sampled individuals will be too small to draw statistical inference. Studying the very affluent remains an endeavor that requires a full survey, possibly in the form of administrative wealth tax data. Second, standard survey problems such as partial unit non-response, mismeasurement, and item non-response also apply to our approach. Third, inter-temporal mobility into and out of the group of the affluent is only captured if the sample is refreshed on a regular basis, which is costly.

The following aspects should be taken into consideration when preparing for a possible large-scale survey:

1. Due to the temporal restrictions and enormous effort required to create an individual shareholder database, we used a relatively simple model for imputing missing information on company revenues in the shareholder database. While the results of the pretest show that even this simple model suffices to stratify the total population by wealth, at least between the second and third stratum, a more sophisticated model could further improve the efficacy of stratification via shareholdings. Furthermore, since the distribution of cumulative shareholdings is rather flat over a wide range, stratification might become more effective in future implementations by increasing the upper bound of the first stratum and a higher sampling probability in the third stratum.
2. The pretest showed that the address information is of sufficient quality. Nevertheless, some portions of the address data were not usable, for instance because addresses were out of date or incomplete, or because they were company rather than private addresses. In a full-scale study, further efforts should be undertaken to improve the quality of the address data in the individual database, for instance, by checking address registries.
3. An unadjusted response rate of 6 percent and a cooperation rate of 23 percent affirm the feasibility of a full-scale study. Our previous experience with field work suggests that these rates will probably be lower, since (a) the fieldwork organization used very experienced interviewers in the pretest; (b) regions with short travel distances and a high density of shareholders were selected; (c) the effort required to check address registries by hand in a

full-scale study would be much greater than in a pretest with a small number of cases.

4. The "Your Personal Balance Sheet" module can be optimized for high-wealth individuals. For instance, we can facilitate the entry of high nominal assets (long series of numbers) and ask more detailed questions about shareholdings.
5. High assets are sometimes put into (family) foundations. Prior to a full-scale study, efforts should therefore be made to estimate the wealth invested in foundations—if appropriate data is available.
6. By definition, the proposed strategy will not sample wealthy individuals who are not invested in at least one company with at least 0.1 percent. This should, however, be more of an issue for persons with assets in the low million euro range, but a minor limitation in the targeted top-wealth segment.

References

Alstadsaeter, A., N. Johannesen, and G. Zucman, "Who Owns the Wealth in Tax Havens? Macro Evidence and Implications for Global Inequality," *Journal of Public Economics*, 162, 89–100, 2018.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman, "Global Inequality Dynamics: New Findings from WID.world," *American Economic Review*, 107, 404–09, 2017.

Bartels, C., "Top Incomes in Germany, 1871–2014," *Journal of Economic History*, 79, 669–707, 2019.

Chakraborty, R. and S. R. Waltl, "Missing the Wealthy in the HFCS: Micro Problems with Macro Implications," European Central Bank Working Paper Series, 2018.

Coval, J. D., and T. J. Moskowitz, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *Journal of Finance*, 54, 2045–73, 1999.

D'Alessio, G. and I. Faiella, "Nonresponse Behaviour in the Bank of Italy's Survey of Household Income and Wealth," *Banca d'Italia Temi di Discussione*, 2002.

Finances, H. and C. Network, "The Eurosystem Household Finance and Consumption Survey. Methodological Report of the First Wave," ECB Statistics Paper Series, 2013.

Frick, J., M. Grabka, and J. Marcus, "Editing und Multiple Imputation der Vermögensinformation 2002 und 2007 im SOEP," SOEP Survey Paper, 2010.

Goebel, J., M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schroeder, and J. Schupp, "The German Socio-Economic Panel (SOEP)," *Journal of Economics and Statistics*, 239, 345–60, 2019.

Horvitz, D. G. and D. J. Thompson, "A Generalization of Sampling Without Replacement from A Finite Universe," *Journal of the American Statistical Association*, 663–85, 1952.

Kroh, M., S. Kuehne, J. Goebel, and F. Preu, "The 2013 IAB-SOEP Migration Sample (M1): Sampling Design and Weighting Adjustment," SOEP Survey Papers, Series C, Data Documentation, 2015.

Kroh, M., S. Kuehne, R. Siegers, and V. Belcheva, "Documentation of Sample Sizes and Panel Attrition of SOEP-Core (1984 until 2016)," SOEP Survey Papers, 480, 2018.

Royston, P., and I. White, "Multiple Imputation by Chained Equations (MICE): Implementation in Stata," *Journal of Statistical Software, Articles*, 45, 1–20, 2011.

Sànchez Muñoz, C., "The Euro-area Household Finance and Consumption Survey: Survey Mode, Oversampling Wealthy Households and Other Methods to Reduce Non-Response Bias," presented at the UNECE Conference of European Statisticians, 2011.

Sierminska, E., J. Frick, and M. Grabka, "Examining the Gender Wealth Gap," *Oxford Economic Papers*, 62, 669–90, 2010.

Vermeulen, P., "How Fat is the Top Tail of the Wealth Distribution?," *The Review of Income and Wealth*, 64, 357–87, 2018.

Westermeier, C., and M. Grabka, "Große statistische Unsicherheit beim Anteil der Top-Vermögenden in Deutschland," *DIW Wochenbericht*, 82, 123–33, 2015.