

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel study at DIW Berlin

978-2018

A Novel Sampling Strategy for Surveying High-Worth Individuals – An Application Using the Socio-Economic Panel

Carsten Schröder, Charlotte Bartels, Markus M. Grabka, Martin Kroh, Rainer Siegers

SOEPPapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at
<http://www.diw.de/soeppapers>

Editors:

Jan **Goebel** (Spatial Economics)
Stefan **Liebig** (Sociology)
David **Richter** (Psychology)
Carsten **Schröder** (Public Economics)
Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)
Denis **Gerstorff** (Psychology, DIW Research Fellow)
Elke **Holst** (Gender Studies, DIW Research Director)
Martin **Kroh** (Political Science, Survey Methodology)
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)
Thomas **Siedler** (Empirical Economics, DIW Research Fellow)
C. Katharina **Spieß** (Education and Family Economics)
Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: soeppapers@diw.de



A Novel Sampling Strategy for Surveying High-Worth Individuals – An Application Using the Socio-Economic Panel

Authors: Carsten Schröder, Charlotte Bartels, Markus M. Grabka,
Martin Kroh, Rainer Siegers

DIW Berlin / SOEP, July 2018

Acknowledgments. 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; and Deborah Anne Bowen for editing.

Abstract

High-worth individuals are typically underrepresented or completely missing in population surveys. The lack of a register-based sampling frame on high-worth individuals in many countries challenged previous attempts to sample high-worth individuals in voluntary scientific surveys. In a novel research design, we draw on register data on the shareholding structures of companies as a sampling frame. Our design builds on the empirical regularity that high-worth 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 identified individuals who are both German residents and registered shareholders of companies. In a feasibility study, we interviewed 124 households from a gross sample of 2,000 anchor persons. Our analysis shows that register data on shareholding structures correctly identifies the individuals' rank 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-worth individuals and the top-end of the wealth distribution.

1 Introduction

There are major deficits in the currently available data on high-worth individuals and households in Germany, but also in many other countries. The main reason is that in many countries wealth taxes have been abolished, so that register data on high-worth individuals or households do not exist. Lack of reliable data impedes to draw substantive conclusions in empirical studies on the entire wealth distribution. The introduction of dual income taxation in many European countries since the 1990s exacerbated this problem, because income tax data no longer systematically record dividends and interest income. Germany introduced a withholding tax on dividends and interest income in 2009, which cannot be linked to individual income tax declarations. Before, income tax statistics could be used to at least approximate assets at the top end of the distribution. Correspondingly, the German federal government's Fifth Poverty and Wealth Report (ARB5) emphasizes the urgent need to improve the data infrastructure on high-worth individuals (see of Labour and Affairs [2017]).

In the three large representative population surveys in Germany that provide data on private wealth,¹ there is no household with net worth (the value of total financial and material assets net of the value of total liabilities) higher than 100 million euros. The surveys do not even provide a sufficient number of cases for reliable statistical analysis of individuals/households with net worth above 10 million euros (Westermeier and Grabka [2015]),

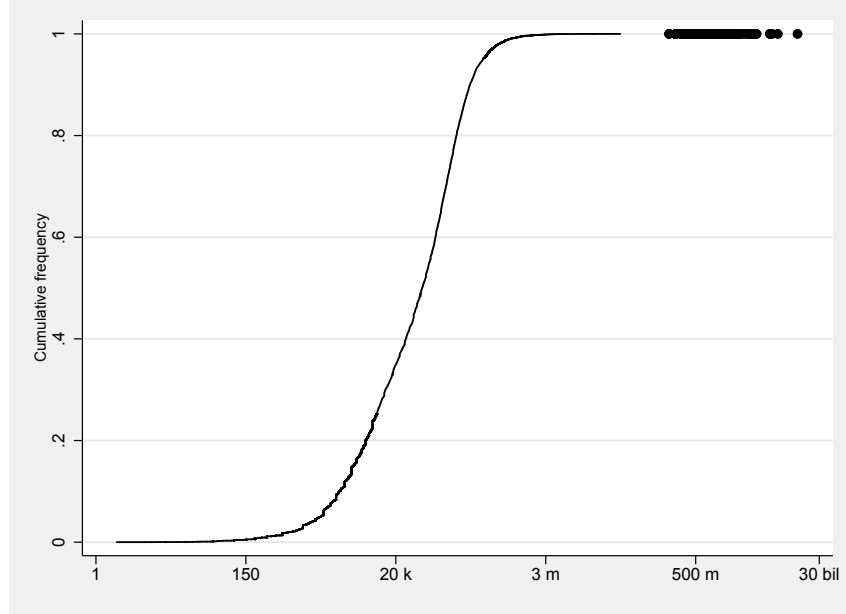
¹These are the German Panel of Household Finances (PHF), the Socio-Economic Panel (SOEP), and the Income and Expenditure Survey (EVS).

although rich lists report a significant number of billionaires and multimillionaires in Germany.

Thus, in all of the existing scientific surveys in Germany, there is a major gap in representation of the high-worth group. *We define high-worth individuals as those whose assets place them at the top end of the wealth distribution, e.g., the top 1%, based on existing population surveys or above that level.* Figure 1 depicts the data gap, taking the SOEP as an example. It shows the cumulative density of individual net worth according to the SOEP² as well as the richest individuals according to *Manager Magazin*. The vertical lines delineate the data gap. The figure shows that the number of cases in the SOEP with assets in the double-digit millions is already very small. In fact, the highest net-worth recorded in the SOEP are “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 among individuals with net worth in the low millions.*³ There are a number of reasons for the limited information available on individuals in the high-worth category. First, it is simply unlikely *per definitionem* that high-worth individuals will end up in a random sample. Accordingly, their numbers are small in random sample surveys of a few thousand individuals. Second, the individual willingness to participate in surveys declines systematically with

²The SOEP is one of the very few population surveys which collects information about assets at the individual level for all adult household members.

³In the past, this data gap has been filled using estimates based on rich lists and the assumption of a Pareto distribution, that is, the assumption that there is a log-log linear relationship between assets and the empirical distribution. This approach rests on a number of assumptions that entail major uncertainties. Westermeier and Grabka [2015] show that such an estimation approach is virtually incapable of providing reliable estimates of trends.



Note: Net worth in euros. Continuous line: Individual net worth in SOEP (positive net worth only); Dots: individuals from *Manager Magazin*'s Top 500 rich list for 2014. Net worth was plotted on a log scale. 2012 SOEP data.

Figure 1: Cumulative density of net worth in the SOEP and *Manager Magazin*

assets (Westermeyer and Grabka [2015]), reducing the share of high-worth individuals in surveys to levels even below their actual share in the population. Third, in Germany – in contrast to many other countries – there are no register data available, e.g., from a wealth tax, that would allow for sufficiently precise identification of high-worth individuals for a survey. Oversampling individuals with an address in regions with above-average tax revenues, as done in the PHF Survey of the German Bundesbank, has not substantially improved this data situation. A non-representative convenience sample, as used in the sample of high-worth individuals in Germany (HViD), does not

provide the basis for generalizable statements.⁴ *The aim of the present feasibility study was therefore to test whether Germany’s research data infrastructure on high-worth individuals can be improved through use of an innovative sampling and surveying concept.* Our concept draws on two considerations: First, individuals with high worth are most likely to hold at least part of their assets in the form of shares.⁵ Second, companies are required to publish information not only on their financial and profit situation but also on their shareholder structures – in the form of names, addresses, and shareholdings. We use this publicly available information to identify shareholders living in Germany. From this population, we draw a random sample of precisely identified high-worth individuals, 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 challenges tackled by the following steps:

1. **Identifying individuals with substantial shareholdings.** A key condition for the successful implementation is to construct a database as complete as possible including the names, addresses, and shareholdings of German residents with substantial shareholdings⁶ 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

⁴In the case of HViD, another sampling criterion was used: financial assets of at least 1 million euros (Ströing et al. [2016]).

⁵Not only is entrepreneurship a source of wealth; many wealthy people also invest in companies.

⁶Bureau van Dijk defines a threshold of at least 0.1% of all shares.

turnover. 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 worldwide and their ownership structures. In some cases, these structures are highly complex, for instance, if companies are intertwined in parent-subsidary relationships. This created the technical challenge of converting a global company database with complex corporate ownership structures into an individual database.

2. Identifying high-worth individuals. Our concept rests on the assumption that individuals with high shareholdings also have high net worth. However, *Orbis* does not contain direct information on the monetary value of shareholdings for individuals, but only the percentage share they hold of the company and financial figures of the company (revenues, profits, etc.). In order to monetarize individual shareholdings, we therefore had to first estimate the company value. Using that, we monetarized each ownership percentage. This allowed us to calculate total shareholdings for each individual.

3. Shareholder Contactability and Willingness to be Surveyed.

The shareholders from the individual database have to be contactable,⁷ and willing to complete the SOEP questionnaire, which includes questions about private assets. We therefore tested the quality of the addresses in the data base.

⁷To make contact, we need current, accurate address information. The present study faced the challenge, that *Orbis* sometimes listed the company address as private address, which complicated direct personal contact with the target person.

The key findings of this feasibility study can be summarized as follows:

1. Our Orbis-based database contains around 1.5 million German residents owning substantial shares in German or foreign companies. *This number of cases enables the concept to be implemented comprehensively in the framework of a full-scale study.*
2. The high-worth individuals listed in *Manager Magazin* are also included in our Orbis database. *This suggests that our concept successfully captures high-worth individuals at the very top of the wealth distribution.*
3. In total, we passed on 2,000 names and addresses from our individual database to Kantar Public, the survey institute responsible for the SOEP. The list is a stratified random sample limited to five regions.⁸ After 124 individuals were interviewed successfully, the fieldwork phase was ended as this number already substantially exceeded the planned pretest sample size of 100 respondents. Willingness to participate was thus 34% percent for the total of 534 contactable cases with valid address information. Combined with the size of the individual database (1.5 million individuals), this response rate would allow for a full-scale survey. *High-worth individuals can therefore not only be identified precisely but also interviewed.*
4. The average net worth of the 124 individuals surveyed in the pretest was over 10 million euros. This exceeds the SOEP average by a factor

⁸Differentiated by urban / rural and suburban areas.

of 100. The highest individual net worth in the pretest was over 200 million euros. Thus, we succeeded in interviewing at least one person who could be on the *Manager Magazin* list. *The pretest results therefore suggest that our concept is capable of effectively closing the data gap in scientific surveys in the high-worth area.*

5. The estimated shareholdings of respondents from our database are strongly positively correlated with the individual net worth stated in the interview. *This would allow for effective and relatively economical oversampling of individual shareholdings at the top end of the wealth distribution in a full-scale study.*

Overall, our concept promises to overcome the problem of systematic under-coverage of high-worth individuals in surveys in Germany as well as in other countries. A full-scale survey is likely to provide valuable insights into the concentration of wealth in Germany, 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-worth individuals; social engagement; and so on. Furthermore, a full-scale study will improve the data infrastructure on the top-end of the *high* income distribution, since income and wealth are strongly positively correlated. The following Sections of this report are structured as follows: Section 2 explains how the base population is identified by describing the Orbis company database, the transformation of the data into a database of individuals, and the method used to evaluate com-

pany shareholdings. Section 3 discusses the suitability of the Orbis database for representation of companies and high-worth individuals and the methods used to evaluate the companies. Sampling and weighting methods are explained in Section 4. Section 5 presents and discusses the results of the pretest, and Section 6 summarizes our conclusions from the feasibility study. The Outlook and Recommendations section at the end provides suggestions for full-scale implementations of our proposed sampling approach.

2 Population identification

Our approach builds on an individual database containing as complete as possible information on the names, addresses and shareholdings of persons residing in Germany with “substantial” shareholdings in companies internationally. Relevant persons are individuals with high shareholdings, who are expected to have large private asset holdings. Neither owners of (small) corporations, for instance, in the trade or service sectors, nor persons with small shares in large companies (small shareholders) are of concern here, but rather persons with high-value shareholdings in large companies. We have created 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 updated data until August 2017. Orbis is based primarily on the published balances of the companies. Various monetary financial figures are listed, such as turnover, profits, equity, and liabilities, as well

as number of employees, headquarters or shareholding structure.⁹ Balance sheets and company accounts also contain the names and addresses of individuals with substantial shareholdings. Our approach targets this group. Orbis is designed as a company database, where we can directly retrieve, for instance, a company’s sales information, number of employees, or ownership structure. Our approach, however, requires individual shareholder data. We therefore transform Orbis’s company database into a shareholder database. Schematically, we carry out this transformation in three steps:

1. Orbis includes several sub-databases, and in the first step we use the Orbis contact database (`Contacts`) in order to extract all persons residing in Germany who currently have shareholdings in at least one active company internationally. Here we filter (`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.

⁹See Orbis Homepage.

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

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

2.1 Monetization of percentage shareholdings

For an accurate survey of high-worth individuals, all percentage shareholdings in our database have to be monetized for each individual and then cumulated across the firms in which the individual holds shares. 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 in descending order by their cumulative shareholdings in order to draw a stratified random sample to be 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 have to differentiate between the value of direct corporate shareholdings (Shareholder - Direct) and their total value including indirect shareholdings (Shareholder - Total). This distinction is important

because financial statements are either consolidated or non-consolidated (Consolidation code). Consolidated means that the values of the subsidiary company are listed on the balance sheet of the parent company.

If the financial statement is consolidated, the percentage of direct shareholdings plus the company value from the consolidated financial statement are sufficient to determine the value of the shareholding. If the financial statement is not consolidated, 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.¹⁰

The value of shareholdings in a listed company is equivalent to the percentage shareholding multiplied by the market capitalization of the company. The vast majority of companies in Germany is not listed. Thus, we need a standardized evaluation method for all companies. We draw on companies' turnover as these figures for a great many companies compared to other financial figures such as profit. This method facilitates a comprehensive and uniform way of evaluating companies and *cumulative shareholdings* in Germany and other countries. Of course, the revenue of a company is not equivalent to the company value; however, a strong positive correlation between the two can be expected. The frequently used capitalised earnings method in accounting relies on the same reasoning. Further, an exact evaluation in euros for the sampling procedure is not necessary, as the shareholders were only sorted by cumulative shareholdings. Their actual assets are collected when surveying the sampled high-worth individuals with SOEP instruments.

¹⁰When generating the data sets for the pretest sampling, we only took two levels of shareholding into account to speed-up the data queries.

Because companies disclose the figures from their balance sheets at different times, we use the most recent turnover according to Orbis with a maximum delay of five years (that is, the most recent turnover between 2012 and 2018). In the absence of turnover information,¹¹ we applied a two-step imputation:

1. For companies with turnover data, the econometrical imputation relies on various explanatory variables from the balance sheets, including: the latest available turnover in the period prior to 2012, the latest available number of employees, the latest balance sheet total, and the number of subsidiary companies and shareholders.
2. For companies whose turnover is missing but explanatory variables are available, we predicted the turnover based on the above model.
3. If both the turnover and explanatory variables were missing, we implemented a randomization procedure in which these companies were assigned an evenly distributed discrete random variable from 1 to 100. On this basis, we assigned a turnover of up to 700,000 euros¹² from the 100 percentiles of the observed turnover distribution.

We extracted more than 1.5 million shareholders from the Orbis database who have shareholdings in more than 1.6 million companies worldwide. About 93 percent of these companies have a German identification number. This is the well-known “home country bias” from the literature (see Coval and Moskowitz [1999]).

¹¹Small corporations with headquarters in Germany (Art. 264 German Commercial Code (HGB)) must only fulfill their duty of disclosure in accordance with § 326(2) HGB by *filing* with the German Federal Gazette.

¹²700,000 euros is the turnover threshold for small corporations (Art. 267a HGB), which allows them to fulfill the otherwise applicable disclosure requirements to a lesser extent.

3 Suitability of the Orbis data for selecting a sample of high-worth individuals

The data extracted from Orbis are suited for accurately drawing a sample of high-worth individuals if the following aspects apply:

1. *Representation of high-worth individuals.* If it is true that high-worth individuals hold corporate shareholdings, they should also appear as shareholders in Orbis. To verify this, we researched the 100 richest people, or more specifically, families, according to *Manager Magazin*, in the individual database we created.
2. *Suitability of the approach for company valuation.* As described above, we determined cumulative shareholdings from company turnovers and shareholdings in percent. We validated this valuation method. First, we sorted all shareholders with their cumulative shareholdings in descending order. Wealthy individuals, according to *Manager Magazin*, should position themselves at the top of the ranking. On the other hand, 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 conditions and, consequently, allows for an accurate representative sample of high-worth individuals.

3.1 Representation of high-worth individuals

Manager Magazin publishes an annual overview of the presumably 100-500 wealthiest German nationals. The assets specified therein are *estimates* based on freely accessible information. Here we use the data from the 2014 edition of *Manager Magazin* as our company figures in Orbis show the largest coverage for the year 2014.¹³ If our working hypothesis according to which wealthy individuals also have shareholdings is true, and further, if our individual database is complete, we should be able to find the wealthiest Germans according *Manager Magazin* in our shareholder database. The search for the wealthiest Germans in our shareholder database did not prove to be a trivial task, as *Manager Magazin* lists individuals in addition to family associations. For individuals, the estimated net worth is owned by the person in question. With family associations, it is distributed across the family members. 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 worth is estimated at 31 billion euros. Dieter Schwarz is in the fifth place: He owns shares, for instance, in Lidl and Kaufland, and his *individual* total net worth is estimated at 14.5 billion euros. The Oetker family holds tenth place, and 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

¹³The list contains 500 individuals/family associations.

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*. Because of several identical values for estimated net worth, this includes a total of 103 individuals and family associations. 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 comprise the family association. The identification of the “head of the family” was generally straightforward. There was some uncertainty regarding (distant) family members.
2. Next, 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 households listed in *Manager Magazin*. For 101 of 103 individuals or family associations, we were able to identify at least one family member in the shareholder database.¹⁴ 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.¹⁵ Overall, these

¹⁴Of whom, according to Orbis, 88 individuals live in Germany.

¹⁵Of these 404 individuals, we were able to identify 331 in our shareholder database. The difference results from individuals who are deceased, from shares having been sold over time, etc. Thirty-nine of these 331 live abroad, 292 in Germany, and 281 are listed as shareholders in at least one company.

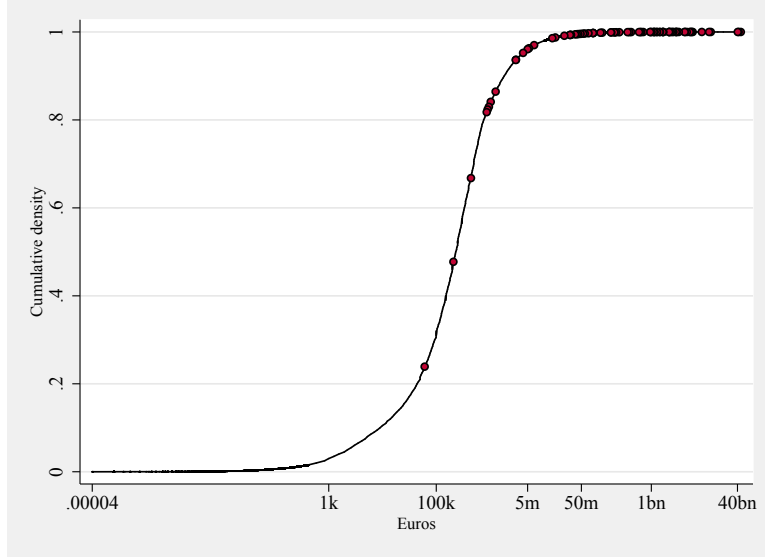
results show that our shareholder database covers virtually the complete list of the wealthiest Germans according to *Manager Magazin*.

3.2 Validity of the estimated cumulative shareholdings

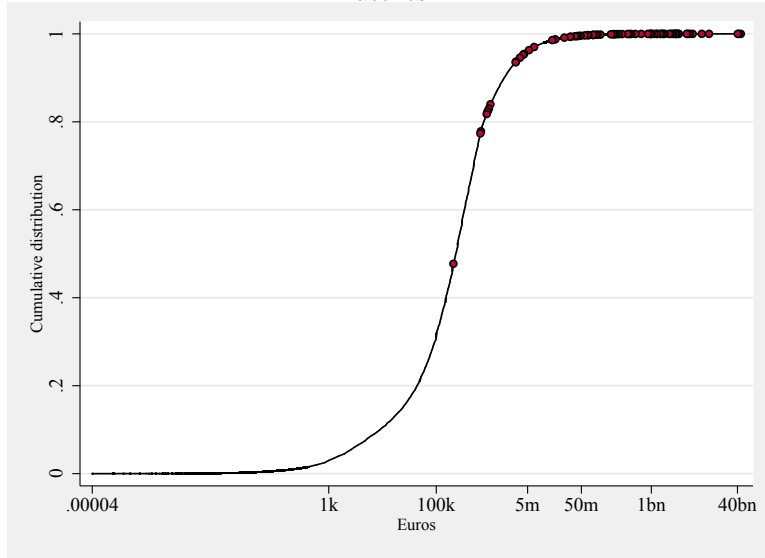
The Figures 2 and 3 show the empirical distributions of the cumulative shareholdings of all 1.5 million German shareholders. Red circles show where the 100 wealthiest Germans according to *Manager Magazin* are positioned within this distribution. Further, we calculated cumulative shareholdings for each member identified in the shareholder database and assigned them to *one person representing the family*. While Figure 3 only shows individuals in the form of red circles, in Figure 2 these also show the representative of the family association. In both figures there are two graphs: The graph at the top takes all cases into consideration in which a link (secure or unsecure) was possible between the individual database and the wealthiest Germans according to *Manager Magazin*. The graph at the bottom takes into account only definitely linkable cases.¹⁶

A highly robust and consistent finding emerges: The wealthy 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 with family associations (due to the distribution of wealth within the family association; delimitation of the family association; precision of estimated net worth in Manager Magazin).

¹⁶Of the 103 wealthiest Germans in *Manager Magazin*, 81 could be identified in Orbis. When considering only individual persons, 25 could be assigned with only two unsecure assignments (see Column “Single persons*”).



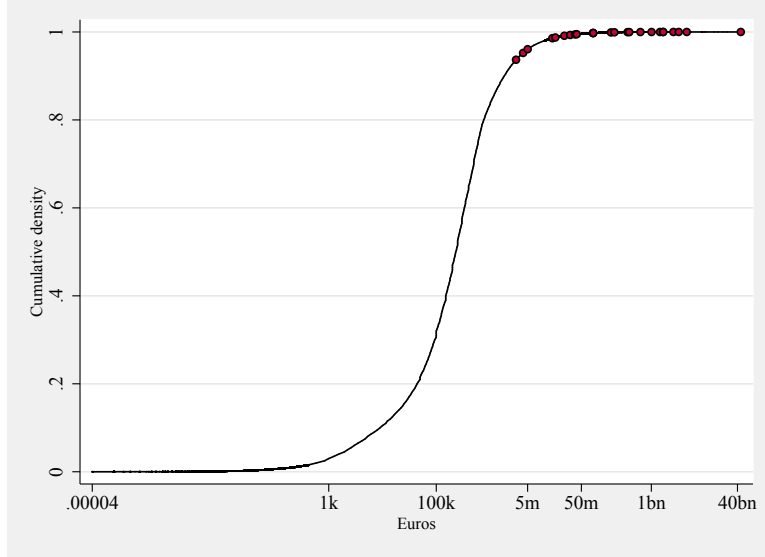
(a): Cumulative shareholdings in Euro including uncertain matches.



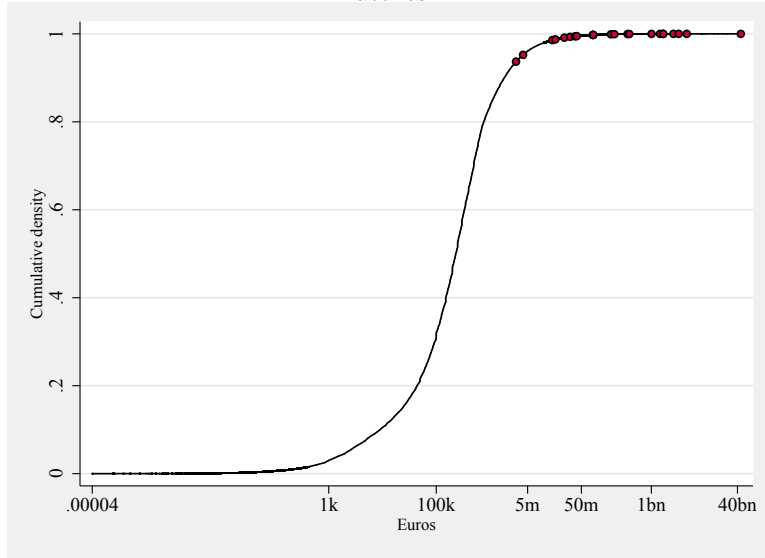
(b): Cumulative shareholdings in Euro. Uncertain matches excluded.

Note: Shareholder data from Orbis and Manager-Magazin (2014). Solid line: Empirical density of cumulative shares; red dots: matched cases from Manager-Magazin; (a): uncertain matches included; (b): only certain matches.

Figure 2: Positions of the richest Germans according to Manager-Magazin in the distribution of cumulative shares – single individuals and families.



(a): Cumulative shareholdings in Euro including uncertain matches.



(b): Cumulative shareholdings in Euro. Uncertain matches excluded.

Note. Shareholder data from Orbis and Manager-Magazin (2014). Solid line: Empirical density of cumulative shares; red dots: matched cases from Manager-Magazin; (a): uncertain matches included; (b): only certain matches.

Figure 3: Positions of the richest Germans according to Manager-Magazin in the distribution of cumulative shares – single individuals only.

Table 1 summarizes these graphical findings descriptively. We again sorted the cumulative shareholdings in ascending order across 100 percentiles. Each percentile includes exactly 1% of the shareholders. The first percentile includes the shareholders with the lowest and the 100th percentile those with the highest cumulative shareholdings.

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 shareholder database. The higher the value, the stronger their concentration in the top percentiles. In fact, the average is between 96.5 and 99.2. *The 100 wealthiest Germans are positioned almost without exception in the highest percentiles on the distribution of cumulative shareholdings.*

Table 1: Richest 100 Germans in the distribution of cumulative shares

	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
Percentil				
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. * including uncertain matches.

4 Sampling procedure and pretest fieldwork

4.1 Sampling procedure

The sampling procedure for the pretest survey is based on the shareholder database we created. The *sampling frame* consists of approx. 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 clear indication of a business rather than a residential address¹⁷. As a result, we eliminated 1.4% of the total cases.

The target population of the sample results from those individuals who belong to the top percentage of the adult population in Germany (see Section 2.1) according to the distribution of cumulative shareholdings (approx. 660,000 individuals). Nearly one million individuals with the lowest cumulative shareholdings were therefore excluded.

As 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 areas¹⁸

¹⁷For closer examination, we passed on addresses with a high level of duplication to the survey institute before the actual pretest. Certain addresses, especially in downtown urban locations, are residential buildings, yet in many cases they are commercial buildings.

¹⁸For approx. 0.8% of persons in the target population, there is no (correct) postal code in the database. They are therefore not included in the sampling frame.

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 below).

In total, 1,111 PSUs were created. Figure 4 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,¹⁹ which should exhibit the following characteristics:

1. *Heterogeneity of regional structure.* Address quality, response behavior, and socio-economic composition may vary between rural, suburban, and urban areas. Thus, we selected sampling points that comprise downtown, peripheral and suburban, 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 showed that contactability and willingness to participate decrease as wealth increase, a stratified random 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 distribution of cumulative shareholdings. Between the selected regions the distribution across strata differs such that

¹⁹To 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.

the selection probability of the gross sample varies across regions and strata between 5.3% and 34.3%.

Out of the 2,000 addresses, approx. 19% were incomplete or outdated, but an additional register search of the sampled persons lead to the accurate address. In only 13% of all cases, a postal address of a sampled person was unavailable. This indicated an overall good address quality for a random sample that is not based on residential registry data (see Kroh et al. [2015]). 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 2/3rd of the sample. This extremely high value indicates that future surveys in high-worth individuals need to invest into 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%. However, fieldwork ended prematurely because the agreed upon number of successful interviews (100) had been exceeded. The adjusted response rate specified can be considered the lower limit of a maximum attainable rate. Referring the 124 shareholdings to those 532 persons who could indeed be contacted during fieldwork, a rate of approx. 23% results.²⁰

Initial analyses demonstrate that the proportion of interviews realized decreases as the value of cumulative shareholdings (according to the strata)

²⁰The initial response rate for pure random samples in SOEP is about 35%, suggesting a surprisingly good response rate for the pretest sample.

increases. It rises with age and is higher among men than among women. Among the individual regions (and interviewees) the differences are small.

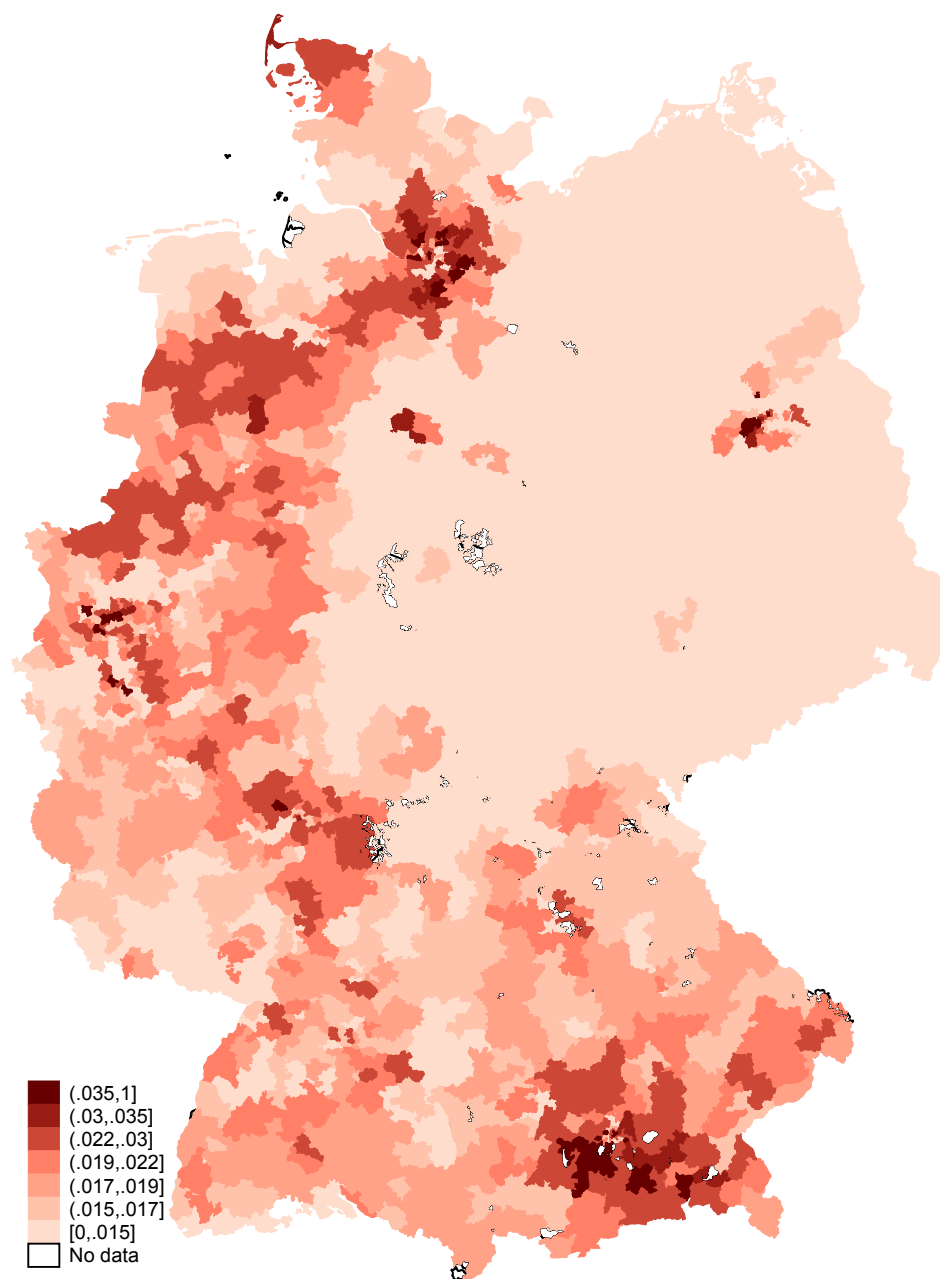
Address quality as well as willingness to participate are sufficient for a full-scale survey.

4.2 Fieldwork

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.

5 Pretest interview results

The analysis described below had two objectives: first, to compare net worth from the pretest and the SOEP; and second, to compare net worth across the three strata in the pretest. To do so, we combined the pretest and SOEP data and evaluated them comparatively. We refrained from further analysis of the content, such as the origin of assets or portfolio composition, because the number of cases from the pretest does not allow for generalizable statements. The following comparison includes the pretest-data for high worth individuals and microdata from the population representative German Socio Economic Panel Study (SOEP). The SOEP is an ongoing panel study



Note: We refrain from providing the boundaries of primary sampling units (PSUs). Thus connecting PSUs with same color cannot be distinguished.

Figure 4: Share of target persons in PSU-specific adult population.

of persons living in private households in Germany. The first wave was conducted in 1984 and since then is replicated each year. In year 1990, the SOEP was extended to the former German democratic republic before reunification. In year 2012, more than 20,000 adult respondents took part in the survey.

5.1 Asset module

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 facilitated a direct comparison of information on net worth and other characteristics of the SOEP respondents.²¹ The latest available information on the personal balance sheet was used for the comparative evaluation of the pretest and SOEP.²²

²¹Individual questions from the SOEP household questionnaire and the HVID study questionnaire were used.

²²The “Your personal balance sheet” module has been surveyed in the SOEP at five-year intervals since 2002. In 2012, the “vehicles” and “educational loans” categories were not included. However, these are quantitatively negligible.

The SOEP wealth module encompasses a total of 12 asset and debt positions:²³

1. Value of owner-occupied property
2. Outstanding debt for owner-occupied property
3. Value of other real property
4. Outstanding debt for other real property
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

5.2 Statistical imputation of the pretest data

All voluntary population surveys face the issue of non-response. Non-response may be total (unit non-response) or pertain to individual questions (item non-response). Item non-response is prevalent with sensitive questions. These include questions about income and wealth (see R. Frick et al. [2010]). Unit non-response was not corrected with weight adjustment because the

²³For 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.

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 illustrating 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 (assets).

In the pretest implementation of MICE, the explained proportion of total variance (R^2) was between 12% and 57% depending on the wealth aggregate for the filter questions and between 9% and 90% for the wealth levels. Thus, the explanatory power of the model is more than satisfactory, in particular for a cross-sectional dataset.

5.3 Wealth in SOEP and Pretest

Individual net worth was used for asset comparison between the SOEP and pretest and for comparing the three strata from the pretest.²⁴

Table 2 allows for a descriptive comparison of net worth in the SOEP and the pretest. Columns 2 and 3 describe the net worth 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

²⁴The results for other wealth aggregates are qualitatively comparable.

Table 2: Individual net worth in Mio. Euro – SOEP and Pretest

	SOEP	Strata 1-3	Pretest		
			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
Beobachtungen	25.803	124	23	41	60

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

magnitude, at around -4 million euros in the SOEP and -2.5 million euros in the pretest.²⁵

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

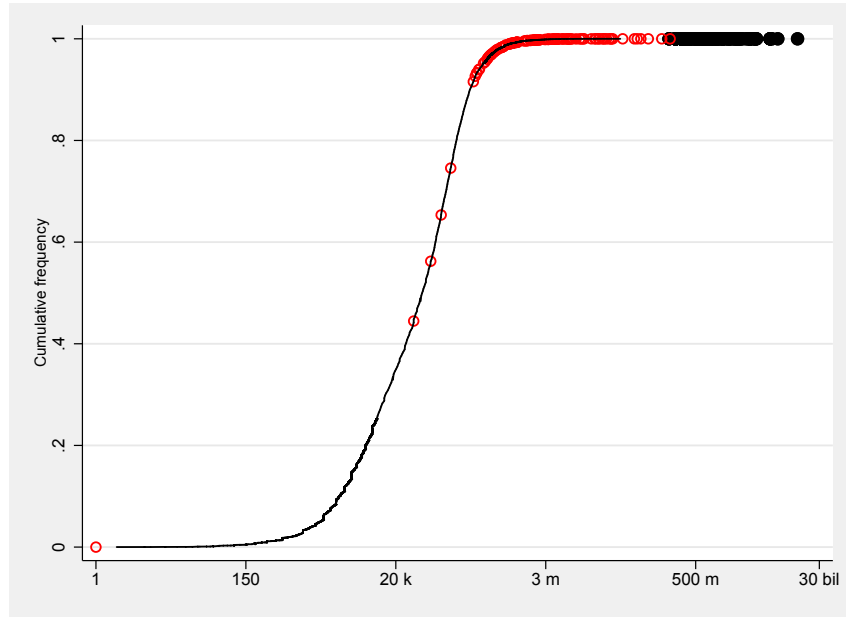
Thus our approach is successful in surveying high-worth individuals. Indeed, net worth in the 25th percentile in the pretest is at the same level as the 99th percentile in the SOEP.

Columns 4 and 6 of Table 2 illustrate the extent to which stratification

²⁵A total of five interviewees in the pretest also have negative net worth, which may be surprising at first glance. However, shareholders are not excluded from indebtedness, for example, in order to finance entrepreneurial activities. Another explanation is that investment values deteriorate if companies become insolvent in the meantime or get into financial stress.

across the cumulative shareholdings in the pretest successfully differentiates persons in the upper sectors of the net worth distribution. In total, we divided the shareholder population into three strata along the distribution of cumulative shareholdings. While average net worth for strata 1 and 2 do not systematically differ, considerably greater net worth is found in stratum 3: Across percentiles, the net worth in stratum 3 is significantly (factor 3 to 10) greater than that in the two lower strata. The mean value of stratum 3 is 200 times greater than in the SOEP.

Stratification seems to facilitate a structured sampling procedure along the data gap in the existing surveys.



Note. Solid line: Individual net worth from SOEP (positive net worth only) in 2012; black dots: single individuals from top 500 according to Manager-Magazin (2014); red circles: Pretest. Wealth plotted on a log scale

Figure 5: Cumulative density of net worth from SOEP, Manager-Magazin and Pretest

Figure 5 again demonstrates that successfully the pretest data close the existing data gap. It supplements Figure 1 by the pretest net worth shown by red circles.

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

The pretest data are distributed over the entire data gap, which suggests that this approach is capable of closing the gap if used in a large, full-scale survey.

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

To estimate the difference between the pretest and SOEP, we used the model,

$$NW_i = \alpha + \beta \times \mathbf{X}_i + \beta^{Pre} \times D_i^{Pretest} + \epsilon_i \quad (1)$$

with the dependent variables, NW_i , the logarithmic net worth of person, i . As covariates we include:

1. One 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, β^{Pre} , of the $D_i^{Pretest}$, the larger the difference in average assets between the pretest and SOEP interviewees, and the more successful the sampling procedure with respect to targeting survey of high-worth 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,

$$NW_i = \alpha + \beta \times \mathbf{X}_i + \beta^{Strat} \times \mathbf{K}_i^{strat} + \epsilon_i. \quad (2)$$

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

Table 3 summarizes the regression results. Here we only considered observations with non-negative net worth. Columns *I* and *III* show the results for both model specifications in the baseline variant without socio-demographic control variables. Column *I* only takes into account the dummy for the pretest. Here the constants are positive and significant. Their value is equivalent to $\exp(10, 58) = 39,340$ euros and describes the average individual net worth in the SOEP. The coefficient for the pretest dummy is $\beta^{Pre} = 4,121$, which equals average assets in the pretest of $\exp(10, 58 + 4, 121) = 2,384,830$ euros. Column *II* uses dummies for the

Table 3: Regression results on the level of net worth and stratum

	I	II	III	IV
Pretest	4,121*** (0.194)		1.652*** (0.187)	
<i>Stratum</i>				
1		4.037*** (0.339)		1.869*** (0.332)
2		3.291*** (0.401)		1.322*** (0.379)
3		4.723*** (0.224)		1.828*** (0.224)
Age			0.173*** (0.004)	0.173*** (0.004)
(Age) ²			-0.001*** (0.000)	-0.001*** (0.000)
Female			-0.209*** (0.021)	-0.209*** (0.021)
<i>Employment</i>				
Employee			0.033 (0.030)	0.033 (0.030)
Self-employed without employees			0.484*** (0.057)	0.486*** (0.057)
Self-employed with 1-9 employees			1.096*** (0.065)	1.096*** (0.064)
Self-employed with 10 and more employees			2.003*** (0.127)	1.966*** (0.125)
Log Household net income			1.029*** (0.022)	1.029*** (0.022)
Constant	10.58*** (0.013)	10.58*** (0.013)	-2.569*** (0.192)	-2.568*** (0.192)
<i>N</i>	19,102	19,102	19,101	19,101
Adj. <i>R</i> 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 worth.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 the SOEP average. Further, the value of the coefficient for the third stratum is considerably higher than the other two strata. This reinforces the descriptive finding that the highest assets are indeed to be found in stratum 3.

In columns *III* and *IV*, socio-demographic variables supplement the explanatory variables of the baseline variant. The results show that after controlling for observed characteristics, the net assets in the pretest are also greater than in the SOEP. A simple screening of the control variables considered here would therefore under no circumstances have attained the accuracy of the concept suggested here. *The regression results therefore demonstrate that the net worth in the pretest is not only significant and quantitatively much higher than in the SOEP, but also that the stratification enabled a targeted sampling along the entire asset spectrum of the data gap.*

6 Conclusions of the Study

There is currently a broad discussion in society and in the research community on issues of being wealthy or poor. It is often criticized that a very small percentage of the population holds a large and growing share of the society's wealth and that they pass 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 in society based on their individual performance. Despite the extensive discussion of 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. Here, all major population surveys show glaring data gaps on individuals with assets beyond the low single-digit millions. In this report, we presented the results of a feasibility study in which we tested an innovative concept for closing this data gap. All of the results clearly show that our concept for collecting data on high-worth individuals and their shareholdings would provide the basis for the first comprehensive survey of high-worth individuals in Germany if implemented in the framework of a larger-scale study:

1. An important advantage of the proposed concept is that the population of high-worth individuals is surveyed using the same survey instruments as the main survey population. This enables researchers to directly compare high-worth individuals with existing SOEP data on the German population.
2. Based on the 2,000 addresses, 124 were surveyed successfully. This is equivalent to an unadjusted response participation of 6%, 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%.
3. The respondents are rich: 56% of the 124 pretest respondents have individual net worth in the millions; approximately 20% have net worth of at least 10 million euros and 5% over 50 million euros. In addition, at least one respondent had net worth 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 turnover (as done in the German Panel of Household Finances): First, it is substantially more precise in defining the population of high-worth individuals. This reduces the costs per surveyed individual. Second, the results of the pretest indicate that if the concept were implemented in the framework of a large-scale survey, it could successfully close the data gap at the top of the distribution in Germany for the first time. A key advantage of the concept in comparison to convenience samples like HViD is that randomized sampling enables inference statistics and thus provides the basis for the first generalized statements about the target group of high-worth individuals in Germany. 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” (including social engagement, charitable donations, and contractual vs. actual working hours). Income tax statistics are silent about the working hours invested by high-worth individuals to gain business income from large partnerships of the German Mittelstand. However, to understand the degree of meritocracy prevalent in a society, the amount of effort invested for a given return is an essential information. The data could also provide valuable insights into the validity of the assumption shared by many studies both in Germany and worldwide that wealth at the upper tail is Pareto-distributed.²⁶ A full-scale study conducted as a panel survey could,

²⁶Recent work by Brzezinski [2014] rejects the log-log-linear relationship implied by the Pareto law.

in the long term, offer important insights into the intra- and intergenerational transmission of wealth, the character of the country's economic elite and the impact of taxes and external shocks (for instance, in the form of capital market volatility) on the various parts of the wealth distribution.

7 Outlook and Recommendations

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 revenues in the Orbis database, which was the basis for evaluating cumulative shareholdings in the individual database and thus for the stratified sample. Although the results of the pretest show that even this simple model can be used successfully to stratify the total population by the distribution of wealth, we see substantial potential to improve the imputation model further. This should be done prior to a full-scale study.
2. The pretest showed that the address information is of sufficiently good 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. The response rate was between 6% and 23% depending on how the sample was defined. This may be regarded as an interval for a possible full-scale study. The rate will probably be below 23%, 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 SOEP wealth module can also be optimized for high-worth individuals, for instance, by facilitating to enter high nominal assets (long series of numbers) and by asking detailed questions about diverse forms of shareholdings.
5. In some cases, high assets are put into (family) foundations. This applies to the Albrecht family in Germany, whose assets in the individual database are not as high as one would expect at first glance. Prior to a full-scale study, efforts should therefore be made to determine whether registries of foundations could be used to more accurately estimate the actual value of companies owned by individuals in this group.
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 our targeted wealth segment. Further, address data are sometimes blocked for privacy issues.

References

- Federal Ministry of Labour and Social Affairs. Lebenslagen in Deutschland - Der Fünfte Armuts- und Reichtumsbericht der Bundesregierung. 2017.
- Christian Westermeier and Markus M. Grabka. Große statistische Unsicherheit beim Anteil der Top-Vermögenden in Deutschland. *DIW Wochenbericht*, 82(7):123–133, 2015. URL <https://ideas.repec.org/a/diw/diwwob/82-7-3.html>.
- M. Ströing, M. M. Grabka, and W. Lauterbach. Hochvermögende in Deutschland. *DIW-Wochenbericht Nr.42/2016*, 2016.
- Joshua D. Coval and Tobias J. Moskowitz. Home Bias at Home: Local Equity Preference in Domestic Portfolios. *Journal of Finance*, 54(6):2045–2073, December 1999. URL <https://ideas.repec.org/a/bla/jfinan/v54y1999i6p2045-2073.html>.
- Martin Kroh, Simon Kuehne, Jan Goebel, and Friedrike Preu. The 2013 iab-soep migration sample (m1): Sampling design and weighting adjustment. *SOEP Survey Papers, Series C, Data Documentation*, (271), 2015.
- Joachim R. Frick, Markus Grabka, and Jan Marcus. Editing und multiple imputation der vermögensinformation 2002 und 2007 im soep. *SOEP Survey Paper*, (146), 12 2010.
- Patrick Royston and Ian White. Multiple imputation by chained equations (mice): Implementation in stata. *Journal of Statistical Software, Arti-*

cles, 45(4):1–20, 2011. ISSN 1548-7660. doi: 10.18637/jss.v045.i04. URL <https://www.jstatsoft.org/v045/i04>.

Michal Brzezinski. Do wealth distributions follow power laws? evidence from “rich lists”. *Physica A: Statistical Mechanics and its Applications*, 406(C):155–162, 2014. doi: 10.1016/j.physa.2014.03.0. URL <https://ideas.repec.org/a/eee/phsmap/v406y2014icp155-162.html>.