

Schröder, Carsten et al.

Working Paper

Improving the coverage of the top-wealth population in the Socio-Economic Panel (SOEP)

SOEPpapers on Multidisciplinary Panel Data Research, No. 1114

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Schröder, Carsten et al. (2020) : Improving the coverage of the top-wealth population in the Socio-Economic Panel (SOEP), SOEPpapers on Multidisciplinary Panel Data Research, No. 1114, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/229846>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1114²⁰²⁰

SOEP papers
on Multidisciplinary Panel Data Research

Improving the Coverage of the Top-Wealth Population in the Socio-Economic Panel (SOEP)

Carsten Schröder, Charlotte Bartels, Konstantin Göbler, Markus M. Grabka, Johannes König, Rainer Siegers,
Sabine Zinn

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 (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)

Sabine **Zinn** (Statistics)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)

Denis **Gerstorff** (Psychology, DIW Research Fellow)

Katharina **Wrohlich** (Gender Economics)

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



Improving the Coverage of the Top-Wealth Population in the Socio-Economic Panel (SOEP)

Carsten Schröder, Charlotte Bartels, Konstantin Göbler, Markus M. Grabka, Johannes König, Rainer Siegers, Sabine Zinn

Abstract

We have developed and implemented a new sampling strategy to better represent very wealthy individuals in the German Socio-Economic Panel (SOEP). Our strategy is based on the empirical regularity that the very wealthy have at least part of their assets invested in businesses, and that businesses document shares of relevant shareholders in their books. Our results show that combined analysis of the SOEP and the new sample (SOEP-P) provides meaningful insights into individuals at the top end of the wealth distribution in Germany. The SOEP-P population has about 21 times higher net wealth on average than the regular SOEP population. Comparing millionaires (net worth \geq 1M Euros) to the rest of the population, a number of noticeable differences emerge. Millionaires differ from the non-rich population with regard to portfolio allocation, demographics, careers, personality, satisfaction with life, and satisfaction with other important life domains.

Keywords: wealth distribution, oversampling, top wealth, wealth portfolios

JEL-Codes: D31, D14, C83

Acknowledgement: We would like to thank the German Federal Ministry for Labour and Social Affairs (BMAS) and the German Federal Ministry for Education and Research (BMBF) for the financial funding of the project "Improvement of the survey data infrastructure regarding the top wealth population of the Socio-Economic Panel (SOEP)". Charlotte Bartels, Markus M. Grabka, Johannes König, and Carsten Schröder acknowledge the financial support by Deutsche Forschungsgemeinschaft (project "Wealthholders at the Top" (WATT), project number: 430972113). We thank Marcelo Avila, Paul Brockmann, Hannah Eger, and Natascha Hainbach for outstanding research assistance.

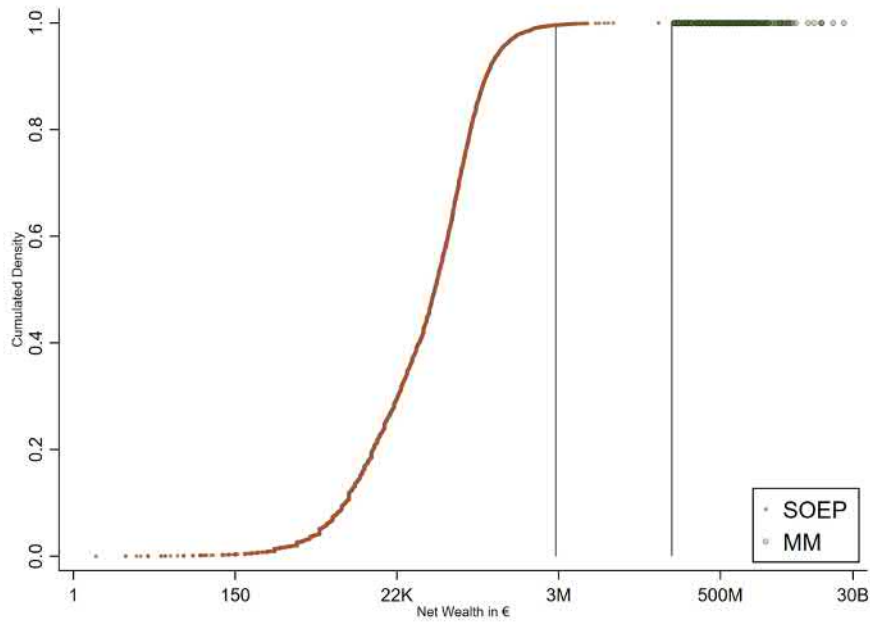
1 Introduction

There is high public interest in the “rich,” the population at the upper end of the wealth distribution. This interest has only increased in the wake of the financial crisis, with its diverse negative economic outcomes, and with rising income inequality, the increasing concentration of wealth (see, e.g., [Atkinson, Piketty, and Saez, 2011](#)), and the growing influence of wealthy individuals on public life, for example, through political connections ([Corneo, 2006](#)) and philanthropic donations (e.g., in education and health; [Andreoni and Payne, 2013](#)).

Despite the interest in and importance of this group, there is a severe lack of data on high net worth individuals in Germany and many other countries. This severely limits the possibility for empirical research on the wealth distribution, on the origins of wealth, and the social and philanthropic activities of wealthy people. This problem is exacerbated in Germany, where only limited information is reported on dividends and interest since the introduction of the final withholding tax in 2009, especially for individuals with high taxable income. As a result, income tax data can no longer be used to approximate wealth at the upper end of the distribution.

Figure 1 gives an idea as to how large the data gap is in Germany based on Socio-Economic Panel (SOEP) data. It shows the empirical distribution of individual net wealth compared to the net wealth of the richest people in Germany in 2017 according to *Manager Magazin*. The vertical lines mark the data gap. The figure shows the very small number of wealthy respondents (with assets in the double-digit millions) in the SOEP: In 2017, the SOEP had only 23 respondents with net assets of more than 5 million euros. Only eight of them provided all the information needed to determine their net worth, such that 15 figures are statistical imputations. According to *Manager Magazin*, there are 100 billionaires in Germany alone (plus an unspecified number of millionaires or multimillionaires).

There are several reasons for this data gap. First, the probability of randomly selecting individuals with high net worth (at the top of the current distribution) into a survey is, by definition, low. As a result, in random surveys with a few thousand respondents, the case numbers are correspondingly low. Second, willingness to participate falls systematically with increasing wealth ([Westermeier and Grabka, 2015](#)). This means that the percentage



Note: Empirical distribution function of net assets in euro. Brown dots (SOEP): Individual SOEP net assets in 2017 (positive net assets only); Green dots (MM): Individuals in the top 500 of the 2014 *Manager Magazin* “rich list”. Assets were mapped on a log scale.

Figure 1: Data gap in the area of high wealth in the SOEP

of high net worth individuals in surveys is actually lower than their percentage in the overall population. Third, in contrast to many other countries, Germany does not have any official data available on private wealth that would provide the basis for a targeted survey of high wealth individuals. Attempts at oversampling individuals with addresses in regions with above-average tax revenue, for instance, in the Deutsche Bundesbank survey Panel on Household Finances (PHF), have not substantially improved the data situation. Non-representative convenience samples, as used in the High Net Worth Individuals in Germany (HVID) survey, do not provide the basis for conclusions that can be generalized.¹

To improve the data situation, we have developed a new strategy for oversampling high net worth individuals and implemented it in the field using the regular SOEP survey instruments. The survey was funded by the German Federal Ministry of Labour and Social Affairs.

The basic idea for the new sampling strategy comes from the empirical observation that almost all high net worth individuals hold at least part of their wealth in shares. Companies publish information not only on their financial situation and corporate profits

¹See [Lauterbach et al. \(2015\)](#).

but also on their ownership structures, including the names, addresses, and holdings of their shareholders. Our strategy uses this information to identify individuals in Germany with significant company shareholdings (top shareholders). From this population, we selected a random sample, stratified by the value of the shareholdings, and surveyed them using standard SOEP survey instruments with a special focus on wealth. We used the module “Your personal financial statement” to capture the size and composition of individual wealth holdings. The steps involved in implementing this strategy are described in [Schröder, Bartels, Grabka, König, et al. \(2019\)](#).

Comparing SOEP-P and SOEP, key results are as follows:

1. SOEP-P makes a decisive contribution to closing the data gap described above: The SOEP-P population has, on average, about 21 times higher net worth than the SOEP population. SOEP-P includes 881 individuals with net worth of at least one million euros. This lays the foundation for the first representative analyses of this population group to date, and in particular, integrated analyses with SOEP data.
2. Looking at the individual net worth of adult SOEP respondents, the Gini coefficient is about 0.78. If one integrates SOEP-P target respondents from corporate registers into this population, the Gini for the individual net wealth distribution increases to 0.81, and the share of the richest 1% increases from 22% to 29% of total wealth.
3. Whereas in the SOEP population, wealth is held mainly in owner-occupied housing and traditional conservative financial products, such as fixed deposit or savings accounts, in SOEP-P, assets are held mainly in company shares and non-owner-occupied real estate.
4. The two populations differ systematically in their sociodemographic characteristics. In the SOEP-P population, older people, married people, people living in western Germany, and men are overrepresented.
5. SOEP-P respondents are also more often active on the labor market (78% in SOEP-P vs. 60% in SOEP). There are similar proportions of retired people in SOEP-P and SOEP (22% vs. 20%).

6. The adjusted monthly net household income of SOEP-P respondents is, at 7,838 euros, significantly higher than in SOEP. All percentiles of the income distribution are higher in SOEP-P than in SOEP, which shows that SOEP-P is also well suited to analyzing high-income earners.
7. SOEP-P respondents differ from SOEP respondents in aspects of personality. They are less agreeable (Big Five personality trait) and more willing to take risks than the average SOEP respondent.

Comparing millionaires with individuals in the three other segments of the wealth distribution (lower half, middle quartile, and upper quartile) based on the integrated data from SOEP and SOEP-P (SOEP + SOEP-P) illustrates the unique living situation and characteristics of this population group:

1. Millionaires are predominantly male, have higher than average education, and at 56 years of age are older than the population average. An above-average percentage of millionaires live in West Germany and a below-average percentage are first or second-generation immigrants.
2. Three-quarters of millionaires are either self-employed or entrepreneurs. Many hold managerial positions and/or work as managing directors or partners in a company.
3. While more than 80% of self-employed people in the lower half of the wealth distribution are “solo self-employed” (that is, they operate their own one-person business), this is true of fewer than 20% of millionaires. Forty-six percent of self-employed millionaires work in businesses with ten or more employees, which are therefore classified as belonging to the classic SME sector.
4. Millionaires also rate their life satisfaction higher on average, both overall and in specific areas of life (e.g., income, living environment, etc.). Satisfaction with leisure time is an exception: Here, millionaires do not differ from rest of the population. This may be due at least in part to working hours: Millionaires work 47 hours a week on average, about 10 hours more than the rest of the working population.
5. Millionaires have an adjusted monthly net household income of over 7,600 euros on average, more than three times that of non-millionaires. A higher percentage of

millionaires save part of their earnings, which is also on average a larger amount in both absolute and relative terms compared to the rest of the population. Their savings rate is far above-average at around 21%. This means that millionaires accumulate wealth more quickly than the rest of the population, which is likely to lead to a further widening of the wealth distribution.

The following parts of this report are structured as follows: Chapter 2 presents the state of the research. Chapter 3 describes the data used in the report: SOEP and SOEP-P, and discusses the sampling and weighting strategy as well as the statistical imputation of the data in detail. Chapter 4 presents the results in four sections: Section 4.1 provides a comparative overview of the sociodemographic composition of the SOEP and SOEP-P samples. In the following two sections, 4.2 and 4.3, we show how SOEP and SOEP-P can be used to construct an integrated distribution of net wealth. On this basis, we describe wealth inequality and the concentration of wealth and trace their development since 2002. In the concluding section of this chapter, Section 4.4, we provide a comparative analysis of millionaires and non-millionaires with regard to their sociodemographics, personality traits, satisfaction, and asset portfolios based on the integrated SOEP + SOEP-P sample. Chapter 5 discusses the results and their implications. The Appendix contains additional information, including details on the description of the SOEP-P sampling procedure.

2 State of the Art

Due to limited available data, studies on the distribution of wealth in Germany come to very different conclusions, especially those focusing on the top 1% share of total wealth. The majority of studies on the wealth distribution in Germany are based on one of three large representative population surveys, which we describe briefly in the following.

The Income and Consumption Survey (EVS) began in 1963 and has been fielded at five-year intervals since 1978. It has included Germany's new (former East) federal states since 1993, and the most recent available wave of data is from 2018. Since 1978, EVS has also included questions on household wealth. The SOEP survey included questions on wealth in the 2002, 2007, 2012, and 2017 survey waves. The Deutsche Bundesbank survey Panel on Household Finances (PHF) was initially conducted in 2010 and 2011 and

continued in 2014 and 2017.

In all three of these surveys, EVS, SOEP, and PHF, there are very few households with net wealth of 10 million euros or more. Yet according to the 2017 *Forbes* list, there were 100 billionaires in Germany in that year alone, and there are also an unknown number of (multi-)millionaires. All available scientific household surveys in Germany point to a substantial data gap in the top wealth sector.

Some studies have attempted to overcome this limitation by supplementing survey data with observations from “rich lists”, which are based on expert estimates. [Bach, Thiemann, and Zucco \(2019\)](#), for example, used *Manager Magazin*, whereas [Westermeier and Grabka \(2015\)](#) and [Vermeulen \(2018\)](#) used the *Forbes* list. These authors based their analyses on the assumption that top wealth follows a Pareto distribution. Their studies show that adding observations from rich lists leads to significantly higher figures for the wealth concentration than without these observations.

Along with survey data, some studies use administrative data based on wealth tax data. However, Germany has not had a wealth tax since 1996. [Baron \(1988\)](#) used the available wealth tax data to analyze the concentration of wealth in West Germany up to 1980, [Dell \(2008\)](#) up to 1995.

Table 1 provides an overview of the estimated wealth shares of the top 10%, 5%, and 1% and the Gini coefficient. Based on the survey data, the estimated share of total net wealth of the top 1% of households is 13% (EVS), 18% (SOEP), and 24% (Household Finance and Consumption Survey, HFCS). The lower percentage reported in the EVS is mainly due to the fact that company shares were not included.

If top wealth holdings are imputed in HFCS, their share increases from 24% (raw) according to [Bach, Thiemann, and Zucco \(2019\)](#) to 31% (34% in [Vermeulen, 2018](#)). The assumption that the top of the wealth distribution follows a Pareto distribution is just as controversial as the validity of rich lists. [Raub, Johnson, and Newcomb \(2010\)](#) compared deceased individuals’ tax data with their wealth ranking according to the *Forbes* list and found that the latter overestimated their net wealth by about 50%, since, among other things, it did not take billionaires’ debts sufficiently into account. [Blanchet, Fournier, and Piketty \(2017\)](#) showed that more flexible functional forms approximate the upper end of the wealth distribution better than Pareto.

An international comparison helps to put the data that are available for Germany into context. [Wolff \(2017\)](#) estimates, based on the Survey of Consumer Finances (SCF), that the top 1% in the United States hold approximately 40% of total wealth. Wealthy households are well represented in the SCF data—in contrast to the currently available German data—because in the SCF, tax returns from the Internal Revenue Service are used in sample selection. [Saez and Zucman \(2016\)](#) also find that the top 1% hold 40% of total wealth in the United States, but based on capitalized income tax data. Studies from other countries, such as the United Kingdom, France, and Sweden, find significantly lower shares of wealth among top wealth holders. Also based on administrative tax data, they are more comparable to the order of magnitude of the uncorrected survey data in Germany. In all three countries, the top 1% hold an estimated 20% of total wealth ([Alvaredo, Atkinson, and Morelli, 2018](#); [Garbinti, Goupille-Lebret, and Piketty, 2020](#); [Lundberg and Waldenström, 2018](#); [Roine and Waldenström, 2009](#)).

Table 1: Studies on wealth inequality in Germany

Year	Data	Gini	Wealth shares in percent of the						Source	
			Top 10%		Top 5%		Top 1%			
			raw	imp.	raw	imp.	raw	imp.		
1995	Wealth tax							12	Dell (2008)	
2003	EVS	0.64			46				Frick, Grabka, and Hauser (2010)	
2013	EVS	0.72	52		35			13	BMAS (2017)	
2010	HFCS				46	54		24	34	Vermeulen (2018)
2010	HFCS	0.75	34	51				24	31	Bach, Thiemann, and Zucco (2019)
2012	SOEP							39	31-34	Westermeier and Grabka (2015)

Note: The lower limit of the range of estimates of [Bach, Thiemann, and Zucco \(2019\)](#) and [Vermeulen \(2018\)](#) indicates the estimate based on the original survey data; the upper limit indicates the estimate after imputation of top wealth.

3 Data

3.1 SOEP

The data we used in our analyses on individuals and households in Germany come from the Socio-Economic Panel (SOEP). The SOEP is a representative longitudinal survey based at

DIW Berlin that has been carried out annually since 1984, providing a reliable long-term picture of the economic and social circumstances of households in Germany. Over the years, various new subsamples have been added, either to compensate for the panel mortality that is typical for voluntary household surveys or to increase the number of cases in specific groups such as immigrant and high-income populations. [Goebel et al. \(2019\)](#) provides detailed information on the SOEP survey, including the questionnaires, fieldwork, data preparation, weighting, and quality management.

The SOEP wealth module: In addition to questions that are included on an annual basis, the SOEP survey also includes rotating modules that are included at longer intervals. One of these is “Your personal financial statement”, which contains questions on the amount and composition of individual wealth and has been fielded in 2002, 2007, 2012, and 2017. We used the SOEP data collected in 2017 in our SOEP-based analyses. The SOEP-P data are from 2019.

Since 2017, the wealth module has covered twelve different wealth and debt components:

1. Owner-occupied housing
2. Other real estate (including undeveloped property, holiday and weekend homes)
3. Financial wealth (savings accounts, savings and mortgage bonds, shares and investment certificates)
4. Assets from private insurance policies (life insurance and private pensions, including so-called Riester pension plans)
5. Building loan balance
6. Business assets (in solely-owned businesses or co-owned business partnerships or corporations; after deduction of operating liabilities)
7. Tangible property, in the form of valuable collections of gold, jewelry, coins, or art
8. Value of motor vehicles
9. Mortgages on owner-occupied housing

10. Mortgages on other real estate
11. Consumer debt
12. Remaining student loan debt

When categorizing the eight wealth components, it should be noted that wealth invested in companies may be included in two components: financial investments (in the form of shares and investment certificates) and business assets.

The difference between gross wealth (components 1-8) and liabilities (components 9-12) results in total net assets, which is normally used in analyses of the wealth distribution at the individual level, and which we use here as well. We note that, in order to facilitate intertemporal comparison, we have elected not to include the value of vehicles and education debt into the net wealth aggregate of the SOEP population.

3.2 SOEP-P

The SOEP Top Shareholder (SOEP-P) sample is a stratified random sample surveyed in 2019. The target population consisted of persons residing in Germany who fell into the top percentile of company shareholdings.

The sampling concept is based on an empirical regularity: Almost all high net worth individuals hold at least part of their wealth in company shares. Our calculations based on data from the Panel on Household Finances (PHF) show that in the lower percentiles, around 20% of households own shares in companies in Germany, and that this figure is as high as 80% in the top percentile. A similar pattern appears in other countries as well. [Wolff \(2017\)](#), for example, shows that 94% of the top 10% of wealthy people in the United States own company shares.

Companies are obligated by law to publish information on their ownership structures. The service provider Bureau van Dijk (BvD) has made this globally published data available in a company database called ORBIS. ORBIS contains data on more than 270 million companies worldwide, including balance sheet figures as well as shareholder structures—including the names and addresses of individuals who own a significant percentage of company shares.²

²A significant percentage of company shares is defined as starting at 0.1% of total company shares.

The base population for SOEP-P consists of shareholders residing in Germany (ORBIS data from Fall 2018) who own shares in at least one company worldwide. From the approximately 1.7 million shareholders residing in Germany, we selected the 600,000 with the highest cumulative monetized value of shareholdings.

Companies' market values can only be taken directly from the ORBIS database for companies that are listed on the stock exchange. However, ORBIS contains a large number of other key company figures, including company turnover. To obtain uniform and comparable market values for all companies, we used company turnover.³

We excluded foundations, non-commercial partnerships under civil law (GbRs) (which come into existence, for example, when a group of self-employed people form a partnership), and clubs or associations that are engaged in commercial activities, which are not subject to the general reporting obligations, meaning that no comparable market values could be obtained.

A random sample was selected from the remaining population of shareholders with significant shareholdings, stratified across the size of the shareholdings, and regionally clustered. This sample was contacted by the survey institute Kantar and interviewed using the SOEP questionnaires.⁴ The SOEP wealth module mentioned above, which was also included in the 2017 SOEP survey, was also part of the questionnaires. This makes it possible to directly compare the wealth positions of the regular SOEP population and the SOEP-P population without making any additional assumptions.

The annex to this report contains more detailed information on the ORBIS database, the valuation of shareholdings, and the quality of the data.

3.2.1 Sample selection and weighting

The ORBIS database forms the framework for SOEP-P. Data derived from this database, dated January 2018, were used to select the sample. ORBIS contains company ownership data on about 1.7 million shareholders who reside in Germany. The target population for the sample is the top 1% of the adult population (approx. 600,000 individuals) with the

³See Appendix 5 and [Schröder, Bartels, Grabka, Kroh, et al. \(2018\)](#) for a detailed description of how these cumulative share values were constructed.

⁴A report on the survey methodology and fieldwork for SOEP-P can be found in [Glemser, Huber, and Rathje \(2020\)](#).

highest cumulative value of shareholdings estimated from the available company data.

As is typically done in representative population surveys using face-to-face interviews, we used a two-stage sampling plan: In the first stage, we selected a number of regions (primary sample units, PSU), and in the second, addresses within these regions (secondary sample units, SSU).

The PSUs are based on the zip code areas within Germany's federal states. Neighboring zip code areas⁵ were combined into PSUs in such a way that each PSU had similar numbers of people from the target population (targeted respondents). First, ORBIS was used to determine how many target respondents live in the individual zip code areas. Subsequently, neighboring zip code areas were merged in an iterative process until each resulting PSU contained approximately 500 target respondents (at least 360 and at most 800).

A total of 1,275 PSUs were created, 250 of which were then selected at random. The probability of selecting a PSU varied with the percentage of target respondents in it. This type of selection is known as probability proportional to size (PPS) sampling. For the eastern German federal states (with the exception of Berlin), we doubled the probability of selection due to the lower incidence of target respondents⁶ in order to increase the analytical potential for eastern Germany.

Based on the results of the pretest (Schröder, Bartels, Grabka, König, et al., 2019), we assumed that the probability of success in contacting target respondents and their willingness to participate would vary according to the gender of the target respondent. Furthermore, we expected these probabilities to increase with age and to decrease with the monetary value of the cumulative company shareholdings. Accordingly, a random sample of addresses was selected, stratified by gender, age, and value of shareholdings. This resulted in 12 sampling strata, each of which was assigned a different sampling probability.⁷ The sampling probabilities are between 7% and 62% depending on the sampling stratum.

A total of 31,000 target respondents were selected at the SSU level in two equal-sized tranches, which were used as the gross sample. These individuals are referred to in the

⁵There was no (correct) zip code for about 0.8% of all target respondents listed in the database. These individuals were not taken into account in the sample selection.

⁶The percentage of target respondents in the adult population is 1.4 times higher in the western federal states (plus Berlin) than in the eastern states, where only 12% of target respondents reside.

⁷A total of 131,867 target respondents live in the selected PSUs. 31,000 of these individuals were selected.

following as anchors.

For this report, we created preliminary weights that allow for analysis of the survey data on the anchor respondents from 2019 together with 2017 survey data from the other SOEP samples. It should be noted that these are preliminary weights that were designed specifically for the analyses presented in this report and under the associated time restrictions. Final weights that are in line with the usual SOEP standards are expected to be available in the first half of 2021. These will differ from the preliminary weights used here in the following respects:

- Analysis of non-response mechanisms and consideration of these in the weighting.
- Consideration of the household context, whereas in SOEP-P, only anchor respondents are weighted.
- Improved consideration of the overlap between the SOEP-P and SOEP populations.
- Integration of SOEP-P and 2019 SOEP survey data.

We began by calculating design weights for the anchor respondents, taking into account the different sampling probabilities depending on age, gender, place of residence (East/West) and the level of shareholdings. We then adjusted the marginal distributions (gender, age, region of origin (East/West, city/country), and estimated value of company shares from ORBIS). The resulting weights are thus suitable for a separate analysis of SOEP-P and extrapolation of results to the target population.

Appropriate weights are also necessary for an integrated analysis of SOEP-P and SOEP data. To this end, we adjusted the weights for respondents in the existing SOEP samples who belong to the same base population as SOEP-P respondents, so that the addition of new respondents to SOEP samples does not lead to an erroneous increase in population totals. This meant that we had to identify all cases from the existing SOEP samples that are also in the SOEP-P target population. We checked whether any respondents in existing SOEP samples are listed in ORBIS and also had sufficiently large company shareholdings to be included in SOEP-P. We were able to identify 70 overlapping individuals.

The integrated weights result from a joint adjustment along the margins for SOEP-P respondents and the SOEP respondents in the target population to the aforementioned

marginal distributions. The weights for respondents in existing SOEP samples who do not belong to the target population of SOEP-P were not changed.

3.2.2 Imputation

SOEP-P exhibits not only a problem of general (unit) non-response, that is, the failure of some respondents to complete the survey, but also a problem of partial (item) non-response, the failure of respondents to complete some questions. To be able to use respondents' data despite these issues, and to avoid bias in subsequent data analyses, we imputed the missing data. Imputation, the estimation of a substitute value or values (to reflect statistical uncertainty) for missing items, must take the underlying mechanism of non-response into account (see [Rubin, 1987](#)). If one does not impute missing data, this means that one implicitly assumes the data are missing completely at random. [Rubin \(1987\)](#), who introduced the term “missing completely at random” (MCAR), pointed out that it is rarely the case in survey data that non-response depends neither on observable nor on non-observable factors. If non-response does depend on such factors, relevant observations should not be ignored or removed from the analysis population.

[Rubin \(1987\)](#) distinguished two mechanisms of non-MCAR missingness: 1) data are missing at random (MAR) if observed factors are solely responsible for the missing data, and 2) data are missing not at random (MNAR) if non-observed or unobservable factors also play a role in the missing data. Common imputation procedures provide consistent and unbiased results under MAR, while MNAR requires more accurate specifications. What we mean here by accurate is that the mechanism of missingness must be modelled directly in order to be able to take corrective action in subsequent analyses. However, since this mechanism includes both observable and non-observable factors, any MNAR models are generally context-dependent and potentially susceptible to model misspecification.

To identify whether missing values are MCAR, we used a procedure proposed by [Little \(1988\)](#). The basic idea underlying the procedure is to answer the question of whether a given variable A can explain the item non-response in another variable B (and vice versa). If B explains the missingness in A, MCAR is rejected. However, the reverse conclusion does not apply because there is no knowledge of the non-observed components or their influence on each other. For this reason, there has been no procedure available for testing

the MAR hypothesis until recently. [Breunig \(2019\)](#) provided a solution: that of using instrument variables and thus exogenous variation for the MAR test.

Against this backdrop, it is advantageous to use an imputation procedure that produces consistent imputation results independent of the non-response mechanisms. Here, consistent means that the joint distribution can be reconstructed completely. Non-response among top wealth holders is precisely the type of missingness that could vary systematically with observable or latent variables. [Lillard, Smith, and Welch \(1986\)](#), for example, pointed out that item non-response is highest at the margins of the income distribution, while [Bollinger et al. \(2019\)](#) demonstrated that particularly in the case of income, response behavior gives evidence of MNAR.

There are several challenges associated with the imputation of SOEP-P:

1. To date, SOEP-P is only a cross-sectional study. Thus, no “historical” information from previous years can be taken into account in imputation processes. Thus, unlike in SOEP, it is not possible to use wealth reported in a previous year to estimate wealth in the current year even though this is a good predictor.
2. The SOEP questionnaire offers a wealth of data that can be used as auxiliary variables for latent variables and thus potentially to counteract the MNAR problem.
3. The SOEP also offers the possibility to logically restrict missing values. For the wealth components, for example, there are filter questions that determine whether the respondent owns the respective wealth component. If respondents report that they do not own a given wealth component, the corresponding euro value is missing for that component. This is not a classic case of item non-response: Rather, the use of the filter for logical restriction indicates that a given value should not be imputed. The chosen imputation method should therefore take this logical structure into account.

One of the most common approaches to imputing multivariate data is multivariate imputation by chained equations (MICE) ([Rubin, 1976](#); [Rubin, 1987](#); [Rubin, 1996](#)). In MICE, statistical models in the form of conditional distributions are specified for each variable with missing information. Modern Markov chain Monte Carlo (MCMC) methods

are used to determine *a posteriori* distributions, on which the imputations are ultimately based. Central to this is the chained structure of the approach, since both complete variables as well as variables with missing values serve as explanatory factors to make efficient use of all available information. Rubin has pointed out that a single imputed value nevertheless leads to an underestimation of the variance in the imputed dataset, independent of the imputation method used. The solution is to construct several imputed datasets in which the imputed values are allowed to differ across imputed datasets. Estimations are then carried out in parallel across all imputed datasets and then aggregated. Estimators for the within- and between-imputation variance are then determined to represent the entire variance without bias.

MICE, like most imputation procedures, is suitable if the MAR assumption applies to the individual univariate models. Problems arise if the *a posteriori* distribution does not converge, or if relevant variables are missing when the models are specified. The reconstruction of the joint distribution of all variables can also fail. The greatest challenge lies in considering the aforementioned logical data structure in the chained procedure. Without this, the imputation procedure will be biased. For example: When imputing missing values for the variable “owner-occupied property”, euro amounts should only be imputed for those respondents who own such property. The same logic should be applied to missing euro amounts for company shares. However, since the percentages of respondents who either own no owner-occupied housing and/or no company shares may diverge, logically missing values are first imputed for the chained model and then logically restricted again. Due to the use of logically restricted values in the chained imputation, MICE leads to biased results.

For these reasons, we have developed a novel imputation procedure based on Bayesian Networks (BN). Bayesian networks provide a graphic representation of the probabilistic relationships in multivariate data (Pearl, 1988). Based on this structure, it is possible to reproduce the complexity of the joint distribution using conditional and unconditional independence between variables to determine local distributions. Both the network structure and the resulting parameters can be determined through the use of learning algorithms. This can be done in a supervised, semi-supervised, or unsupervised manner. Supervised means the entire network is specified by an expert; semi-supervised means that information

available *a priori* is entered into the structure and parameter algorithms; and unsupervised means completely data-driven, that is, all relationships in the network can be learned automatically (Scutari, Vitolo, and Tucker, 2019).

Although imputation on the basis of BN has been discussed in the literature, none of the proposed methods are capable of dealing with complex databases in which both continuous and discrete information is available (see Di Zio et al., 2004; Hruschka, Hruschka, and Ebecken, 2007; Niloofar and Ganjali, 2014; Niloofar, Ganjali, and Rohani, 2013; Rancoita et al., 2016; Romero and Salmerón, 2004). Here we use the approach proposed by Goebler (2019). It can be used with both discrete and continuous data and can be enhanced with any amount of *a priori* information. An advantage of this approach is that, in contrast to MICE, the structural algorithm takes over the specification of the individual models. In concrete terms, this means that for a variable to be imputed, the network structure is searched for those nodes (variables) that provide non-redundant information. This selection of nodes is called Markov Blanket and forms the basis of the imputation procedure (Pearl, 1988). Subsequently, the variables to be imputed are sorted according to their percentage of missing values and iteratively replaced using MCMC methods. This results in a random value being drawn from the *a posteriori* distribution based on the Markov Blanket for each piece of missing data.

This approach ensures that all necessary information is always used efficiently for missing values. Due to the probabilistic structure of the imputation approach, survey filters can be used as nodes in the network structure. This allows logical restrictions to be modelled directly, thus ensuring that they are neither directly nor indirectly used in imputation. In this way, BN imputation avoids the types of bias that can occur in MICE (see Goebler, 2019).

4 Empirical analyses

The empirical analyses are made up of four parts. Section 4.1 compares selected sociodemographic characteristics of adult respondents in SOEP and SOEP-P to obtain an overview of the composition of the two samples with their similarities and differences. In addition, the determinants of capital formation as well as the amount, distribution, and composition

of wealth are described comparatively.

An integrated overview of the SOEP and SOEP-P samples in 2017 is presented in Section 4.2, while a time series of selected indicators of the net wealth distribution in Germany, based on the SOEP sample with SOEP-P integrated into it in 2017, is presented in Section 4.3. Finally, Section 4.4 compares the characteristics of the (multi-)millionaire and non-millionaire populations using the integrated sample. In all analyses, only data from adult respondents were used. Furthermore, all subsequent results are extrapolated using cross-sectional weights.

In all of the analyses presented below, it should be noted that only the anchors in the SOEP-P sample are currently used and only a preliminary framework for population weighting is available.

4.1 Comparing SOEP and SOEP-P sample composition

The SOEP population is representative of the population of private households in Germany. SOEP-P represents a subpopulation: households in which at least one member has significant shareholdings in at least one company worldwide. This leads us to expect that, in addition to assets, the two samples also differ in other characteristics. We test this expectation below based on a number of individual and household characteristics.

When interpreting the results of the comparison, it should be noted that the populations of SOEP-P and SOEP are not completely disjoint: The SOEP contains respondents with high net worth and respondents with corporate investments, but a much lower percentage than in SOEP-P. The comparisons thus reflect differences in sample design. They do not serve in direct comparison of characteristics of people with high and low wealth. This is described in Section 4.4.

4.1.1 Sociodemographic composition

Table 2 gives an overview of the sociodemographic composition of the two samples. There are systematic differences, for example, in the proportion of female respondents: While there are approximately equally sized groups of women and men in the SOEP sample, women make up only about 22% of the SOEP-P sample. This lower proportion can be explained by the anchor concept used in SOEP-P and by findings from the 2019 DIW Berlin

Women Executives Barometer: Even on the supervisory boards of the 200 highest-earning companies in Germany in 2018, the percentage of women was just 27% (Holst and Wrohlich, 2019).

Table 2: Comparison of sociodemographic characteristics in SOEP and SOEP-P

Sociodemographic characteristics	Percentage shares	
	SOEP	SOEP-P
<i>Attributes</i>		
Women	51.06	21.66
With migrant background	23.20	12.35
New federal states	17.23	12.23
<i>Age groups</i>		
18-24	8.60	0.10
25-49	38.13	28.69
50-64	27.08	48.00
65+	26.20	23.20
<i>Labor market status</i>		
Active	60.32	78.44
Inactive	17.25	1.68
Retired	22.43	19.88
<i>Occupational status</i>		
Self-employed	8.62	74.23
Worker	19.69	1.03
Civil servants	5.74	0.64
Trainees	9.28	0.05
Employee	56.67	24.04
Employee with managerial funct.	2.17	22.29
Manager and shareholder	2.14	39.63
<i>Size of self-employed enterprise</i>		
No employees	61.61	12.50
1-9 employees	32.19	44.73
10 and more employees	6.20	42.77
<i>School education</i>		
No Degree	3.16	0.00
Secondary Level I	11.98	43.83
Secondary Level II	56.81	26.23
Upper Secondary Degree	28.05	29.94

Source: SOEP v35 and SOEP-P; own calculations with weighted values.

There are also systematic differences in the age structure between the two samples. Respondents are significantly older in SOEP-P, where the age group of 50 to 64-year-olds is 21 percentage points larger than in the SOEP sample. The relatively high average age

in SOEP-P can also be explained by the sampling strategy: Shareholdings and wealth accumulate over the life course, although inheritances received at an early age and successful business start-ups can also explain high wealth in younger age cohorts (Bönke, Grabka, Schröder, and Wolff, 2020; Cagetti, 2003; De Nardi and Fella, 2017; Wolff, 1981).

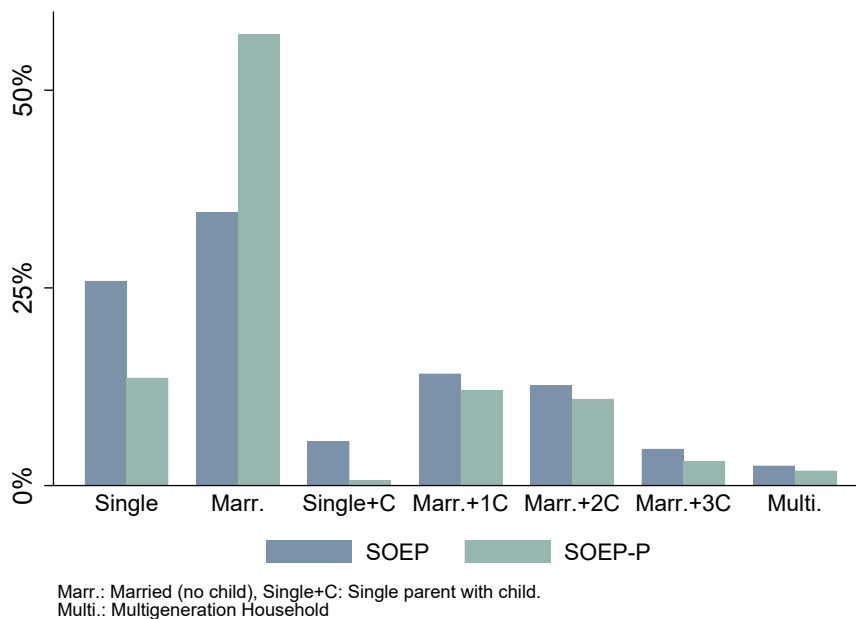
The two samples also differ systematically in the percentage of respondents living in the federal states of the former East and West Germany (including the state of Berlin). SOEP-P has a higher proportion of respondents in the former West, at about 88%. This may be relatively unsurprising, given the lower average wealth and income in eastern Germany (Bönke, Grabka, Schröder, Wolff, and Zyska, 2019; Bönke, Schröder, and Schulte, 2010; Grabka, Halbmeier, et al., 2019). In contrast, SOEP-P has a significantly lower share of respondents with a migrant background, at about 12%, compared to around 23% in SOEP. This finding is consistent with international studies that have found lower wealth among migrants than among autochthon populations (Carroll, Rhee, and Rhee, 1994; Dustmann, 1997; Merkle and Zimmermann, 1992; Porpiglia, Sierminska, and Mathä, 2011).

At over 78%, the share of employed adults in SOEP-P is about 18 percentage points higher than in SOEP. The two populations also differ systematically by the type of work: As expected, almost 75% of SOEP-P respondents are self-employed (SOEP: approx. 9%). In addition, a much higher percentage of SOEP-P respondents hold management positions. The proportions of retired people are similar in the two samples at just over 20% in SOEP and just below 20% in SOEP-P.

Individuals who are not active on the labor market make up a small minority of the SOEP-P sample (approx. 2%) but approximately 17% of the adult population in SOEP. The lower percentage of economically inactive individuals in SOEP-P is mainly due to the higher employment rate among those aged 50 and older. With regard to the number of people employed in the companies of self-employed persons in the two samples, SOEP-P respondents are less often “solo self-employed” (that is, operators of one-person businesses) and therefore more often have employees in their businesses.

The two groups also differ in educational attainment, especially with regard to Secondary Level I, which is more frequently reported by the SOEP-P population. This can be attributed to a cohort effect, since the SOEP-P respondents are above average in age and therefore in many cases only completed compulsory education.

The two samples also show clear differences in household type. As Figure 2 shows, married people without children make up the majority of the population in both data sets, but the proportion in SOEP-P is about 59%, which is about 24 percentage points higher. Conversely, the share of single households and single parents in SOEP-P is significantly lower (13% and 0.5%, respectively) than in SOEP (26% and 5.6%). The share of married couples with children is somewhat higher in SOEP, at about 25%. The share of multigenerational households is below 5% in both data sets.



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 2: Distribution of household types in SOEP and SOEP-P

4.1.2 Personality traits

In addition to sociodemographic characteristics, the SOEP also collects data on the personality traits of respondents. These include the “Big Five” from the five-factor model of personality psychology. According to this model, people’s personality can be classified along the higher-order dimensions of openness to experience (open-mindedness), conscientiousness (perfectionism), extraversion (sociability), agreeableness (consideration, cooperativeness, empathy), and neuroticism (emotional instability and vulnerability).

The SOEP measures the Big Five personality traits through self-assessments on a scale

from 1 (disagree completely) to 7 (agree completely). In addition, respondents rate their willingness to take risks on a scale from 0 to 10: The higher the value on the scale, the higher the willingness to take risks.

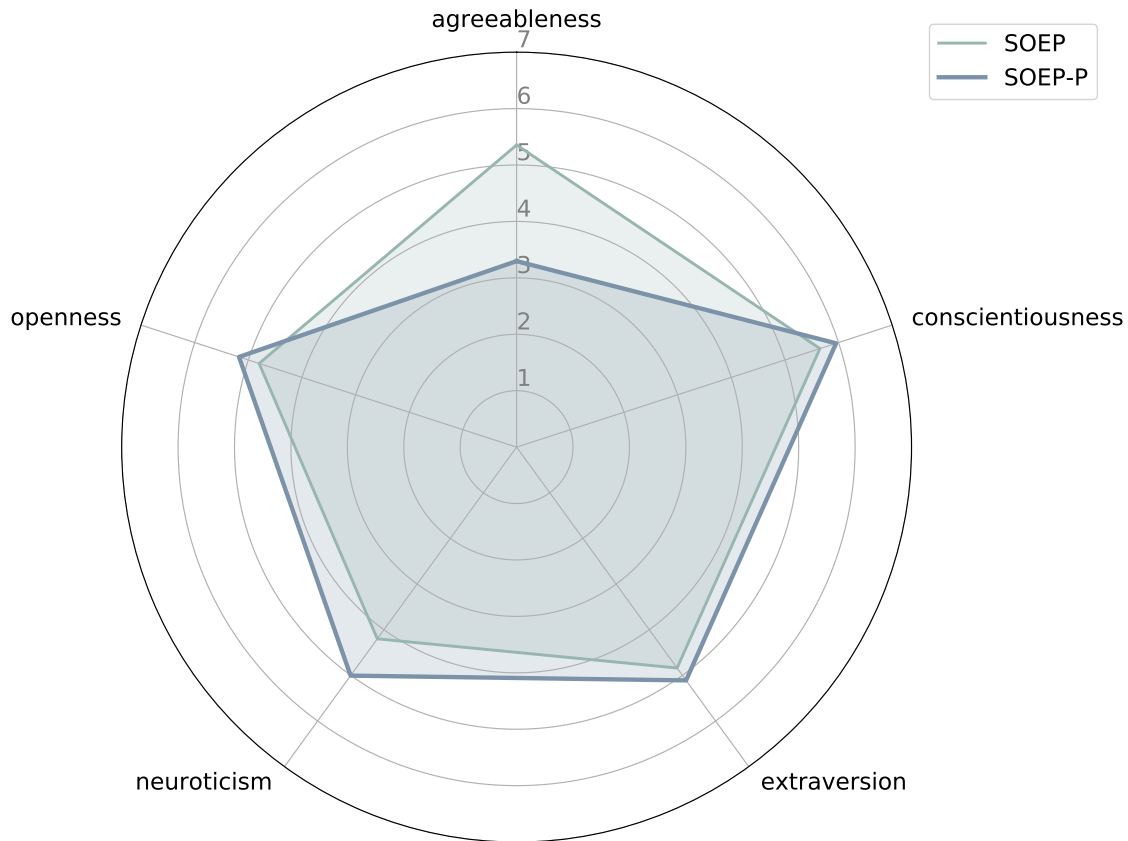
The results for the Big Five are presented here as a network diagram. There is an axis for each higher-order dimension. The same orientation applies to all axes; the value for the corresponding dimension increases with the distance from the center. For each sample, the means are then connected with a line. Thus, in both samples, many people consider themselves to be very conscientious (mean of about 6) but not very neurotic (mean of about 3). For only one higher-order dimension there is a notable difference between the SOEP-P and SOEP population (see Figure 3): answers from SOEP-P respondents result in a lower value for agreeableness.

With regard to the willingness to take risks, SOEP-P respondents have a much higher willingness than SOEP respondents (Figure 4). The shares of the SOEP-P population with high scores of 8, 9, and 10 are over 20%, just above, and just below 10%, respectively. In SOEP, on the other hand, the shares for these three scores are less than half as high. This pattern confirms the previous empirical evidence that independence correlates positively with risk tolerance (Caliendo, Fossen, and Kritikos, 2014).

4.1.3 Wealth formation and its determinants

Savings rates and income are the key determinants of wealth formation, along with inheritances, gifts, and returns on portfolio positions (König, Schröder, and Wolff, 2020). The first two components, savings and income, are the subject of this chapter. Savings behavior is mapped using information from direct questions about savings in the SOEP survey. In the SOEP, the first step is to determine whether households are saving for major purchases and emergencies (precautionary saving) or to accumulate wealth, and, if so, to determine how much money they are putting away in savings per month.

In the following, as our income concept, we use equivalent monthly net household income. Equivalent or means-tested income takes into account differences in the material needs of different types of households. For example, a couple with two children needs a higher income than a single person to achieve a certain standard of living, but their income does not have to be four times as high. Equivalence scales are used to determine equivalent



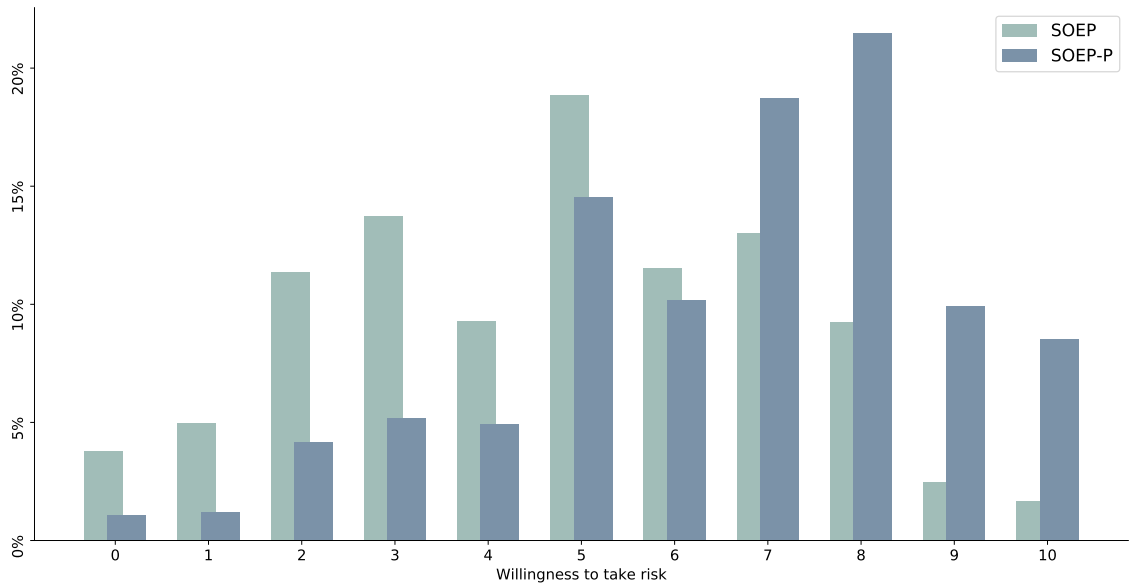
Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 3: The Big Five in SOEP and SOEP-P

income. Here, we use the modified OECD equivalence scale.⁸

Savings: Figure 5 shows the proportions of the SOEP and SOEP-P populations that are putting away money in savings, either to accumulate wealth or as a precaution against unexpected expenses and emergencies. About 50% of the SOEP-P population—almost twice the percentage of the SOEP population—put away money in savings with the goal of building wealth. For larger purchases or emergencies, over 60% of the SOEP-P population and nearly 55% of the SOEP population are putting away money in savings. Not only is there a higher percentage of SOEP-P respondents who actively save: Those who are saving save higher amounts. This is shown by the conditional averages in euros in Table 3. SOEP respondents who save based on the wealth motive put away an average of about 500 euros per month, and those who save based on precautionary motives put away an average

⁸In the OECD-modified scale, the first adult receives a weighting of 1, every other adult 0.5, and every child (up to 14 years of age) a weighting of 0.3, according to which a couple with two children under 14 years of age needs 2.1 times the income of a single person to achieve the same standard of living.

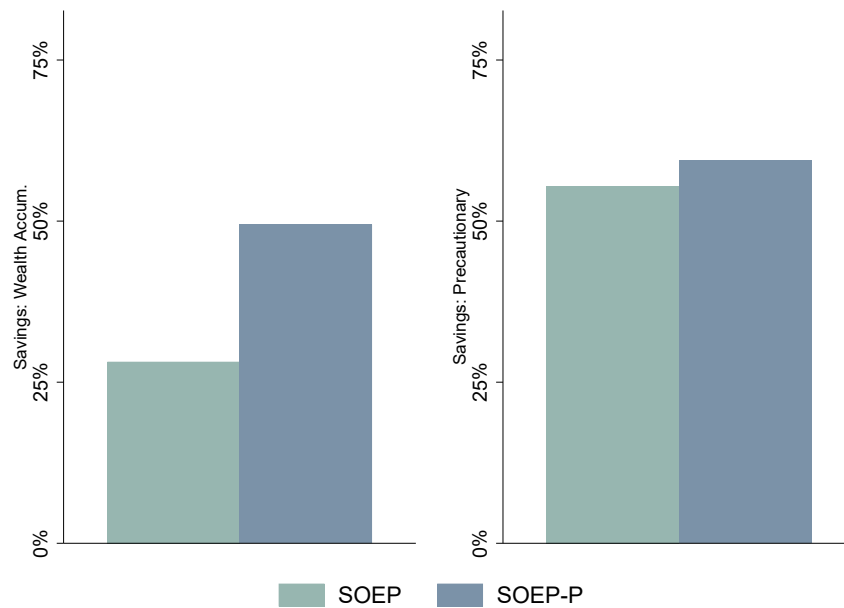


Source: SOEP v35 and SOEP-P; own calculations with weighted values. Risk preferences on a scale from very risk-averse (0) to very risk-tolerant (10).

Figure 4: Risk preferences in SOEP and SOEP-P

of around 410 euros per month. In SOEP-P, these conditional averages are many times higher, at about 2,200 euros and 1,200 euros per month, respectively. In both populations, there are large differences between respondents in terms of savings activity. This is shown by the high values for the Gini coefficient and major differences across the conditional distribution: The Gini coefficients for savings in the SOEP for the wealth motive and the retirement motive are 0.58 and 0.51, respectively. The corresponding values for SOEP-P are higher, at around 0.68 and 0.59, respectively. Among respondents who are saving, the value of savings driven, for example, by the wealth motive in SOEP is 20 euros per month in the lowest, 250 euros per month in the middle, and 4,000 euros in the 99th percentile. In SOEP-P, the corresponding values are 40 euros, 1,000 euros, and 20,000 euros.

Income: Table 3 also shows the distribution of the monthly equivalent net household income in both data sets. On average, it is about 7,800 euros per month in SOEP-P, almost four times as high as in SOEP (2,000 euros). The large difference in income is, of course, a key explanation for the large differences in savings behavior between SOEP and SOEP-P. The difference also shows that SOEP-P contributes to improving the data situation at the



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 5: Proportion of people in SOEP and SOEP-P who are saving for precautionary or wealth accumulation motives

upper end of the income distribution.⁹

Not only is the average income far higher in SOEP-P than in SOEP: Every percentile value of income is higher in SOEP-P than in SOEP (stochastic dominance). This is shown in Figure 6 in the form of the so-called Pen’s parade (Pen, 1971). To construct this parade, households are sorted in ascending order according to their equivalent net household income and, for each income level, the percentage of households that have no more than this income is subtracted. According to this, the lowest quintile (20%) of SOEP households have an equivalent net household income of at most 1,160 euros per month, while the corresponding value in SOEP-P is 2,000 euros. In the top quintile, the value is 2,530 euros in the SOEP and more than twice of that, at around 5,450 euros, in SOEP-P.

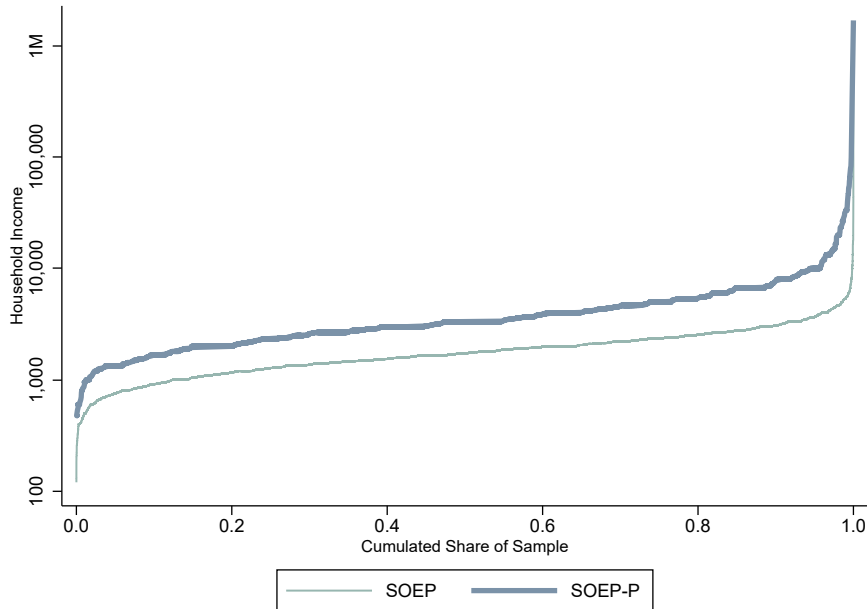
Savings and income: Figure 7 shows how the amount of monthly savings for wealth accumulation or precautionary savings develop along sample-specific deciles of equivalent

⁹It might have been expected that there would also be an even clearer difference in net household income between SOEP and SOEP-P. One possible explanation for the fact that this expectation is not fully met is that, for tax reasons, it is attractive in Germany to retain corporate profits rather than distribute them. Thus, the value of the companies gradually increases, but this is not reflected in higher current income for the shareholders.

Table 3: Savings for precautionary or wealth accumulation motives and net household income

	SOEP			SOEP-P		
	Wealth Acc. Savings	Precaution. Savings	HH net Income	Wealth Acc. Savings	Precaution. Savings	HH net Income
p1	20	25	495	40	50	952
p5	50	50	750	110	100	1,333
p10	70	100	905	200	170	1,667
p25	130	120	1,278	480	300	2,333
p50	250	250	1,733	1,000	500	3,333
p75	500	500	2,333	2,000	1,000	5,000
p90	1,000	1,000	3,067	3,999	2,500	7,778
p95	1,500	1,200	3,667	5,000	4,000	10,000
p99	4,000	2,500	5,385	20,000	10,000	33,333
p99.9	12,000	5,000	12,000	150,000	35,000	1,700,000
p75/p50	2.00	2.00	1.35	2.00	2.00	1.50
p90/p50	4.00	4.00	1.77	4.00	5.00	2.33
p99.9/p50	48.00	20.00	6.92	150.00	70.00	510.00
Gini	0.580	0.513	0.289	0.676	0.589	0.647
Mean	502	407	1,995	2,236	1,185	7,838

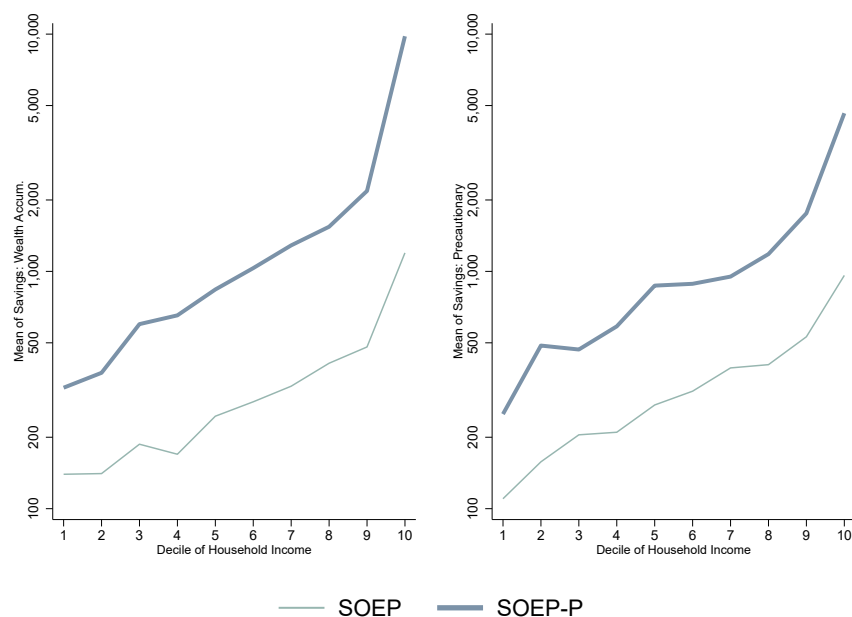
Source: SOEP v35 and SOEP-P; own calculations with weighted values. Only those individuals who are putting away money with the motive of building wealth or precautionary savings are considered, i.e., those who answered with a value greater than zero.



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 6: Pen's parades of equivalent net household income

net household income. Only individuals with a strictly positive savings amount for the particular form of savings are taken into account, and therefore, *conditional mean values* are shown. In both data sets, these conditional values increase with income: Savings driven by the wealth accumulation motive increase in SOEP-P from about 300 euros in the first decile to about 900 euros in the fifth to just under 10,000 euros in the tenth decile. The corresponding values in the SOEP are significantly lower at about 140, 240, and 1,200 euros, respectively. The top income decile in SOEP saves less than the seventh decile in SOEP-P, at around 1,200 euros and 1,300 euros, respectively. The patterns are similar for precautionary savings: savings rise across income deciles and are consistently higher in SOEP-P. Specifically, the value in the first decile of the equivalent net household income in SOEP-P is about 250 euros (SOEP: 110 euros), about 900 Euro in the fifth decile (SOEP: 275 Euro), and about 4,600 Euro in the tenth decile (SOEP: 1,000 Euro).



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

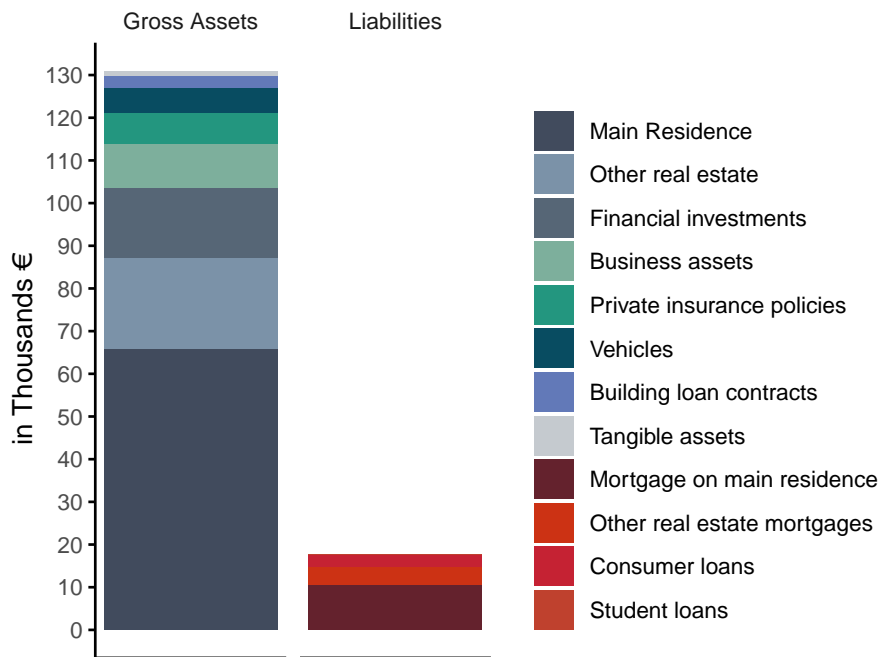
Figure 7: Savings for precautionary and wealth accumulation motives along the deciles of equivalent net household income

4.1.4 Wealth holdings and portfolio composition

As described above, the SOEP and SOEP-P wealth modules distinguish between a total of eight wealth and four debt components. The eight wealth components add up to gross

wealth, the accumulated debt components add up to debt; the difference between gross wealth and debt is called net wealth.

Figure 8 depicts the average level of individual gross and net wealth in the SOEP as well as the portfolio composition. Gross wealth averages approximately 130,000 euros. About half of this is accounted for by owner-occupied housing. Other important types of wealth are other real estate and financial investments. Business assets, private insurance, and vehicles play a relatively minor role. Tangible assets and building loan contracts are of lesser importance. Gross assets are offset by debts of around 18,000 euros on average. These consist mainly of debts from the purchase of real estate (including other real estate debts in addition to owner-occupied housing). Consumer and student loans are of minor importance.



Source: SOEP v35; own calculations with weighted values.

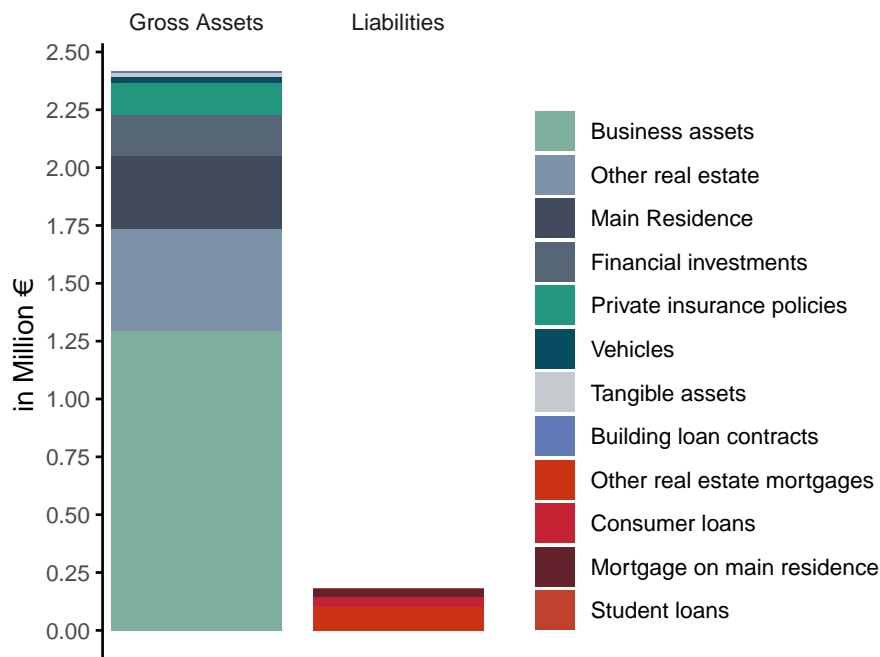
Figure 8: Wealth portfolio in SOEP

Figure 9 depicts the portfolios of SOEP-P respondents in analogous fashion. At around 2.4 million euros, average gross wealth in SOEP-P is about 18 times higher than in SOEP. There are also systematic differences in the composition of gross wealth between the two samples: Whereas owner-occupied housing makes up about half of respondents' wealth in

SOEP, it plays a smaller role in SOEP-P at just 14%.¹⁰

As expected, business assets are the most important component of wealth in SOEP-P, accounting for more than 50%, whereas in SOEP, business assets make up just 8% of wealth. SOEP-P respondents also have slightly higher investments in other forms of real estate than SOEP respondents, at around 21% and 16%, respectively. The other wealth positions such as financial investments, private insurance policies, etc. play a relatively minor role in both samples.

The SOEP-P population also has debt averaging about 189,000 euros, or about 8% of gross wealth. This means that the ratio of debts to gross wealth is significantly smaller in SOEP-P than in the SOEP, where it is around 14%. Real estate debt is dominant in both samples, although mortgages on other real estate in SOEP-P are significantly higher in absolute terms at more than 100,000 euros.



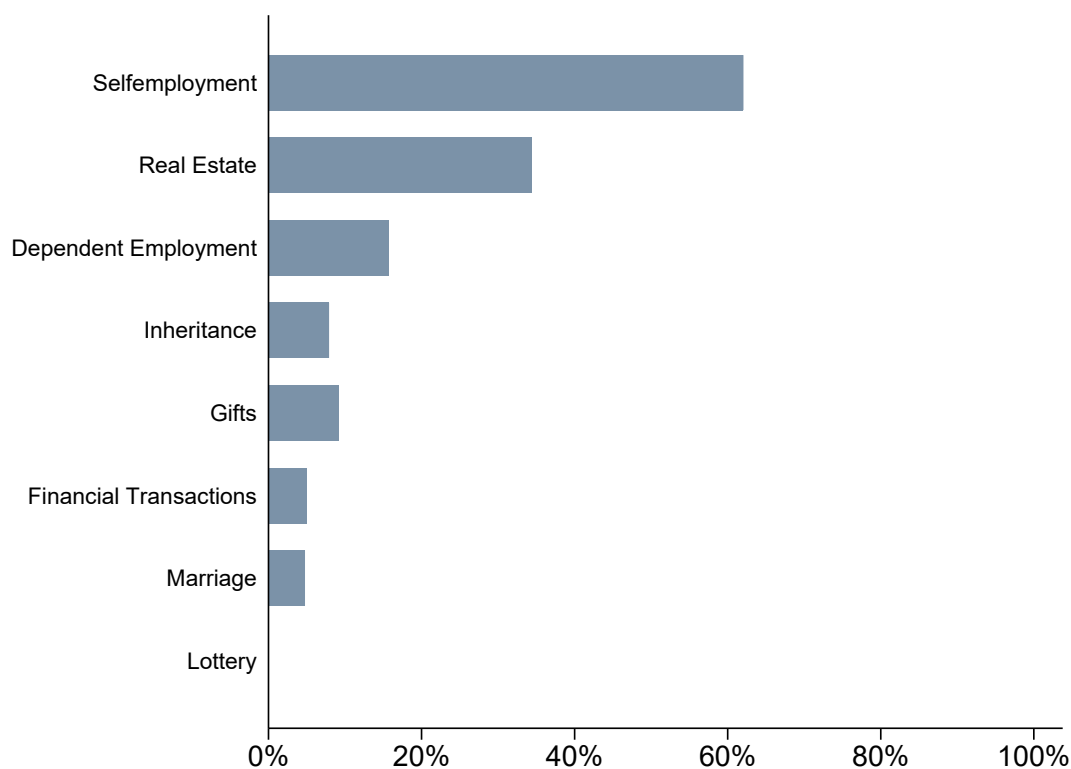
Source: SOEP-P; own calculations with weighted values.

Figure 9: Wealth portfolios in SOEP-P

We can perform a more in-depth analysis of the factors that influence wealth accumulation for SOEP-P. Specifically, participants are asked about the main determinants of

¹⁰The proportion of respondents living in owner-occupied housing is around 50% in SOEP and close to 80% in SOEP-P.

their wealth. Figure 10 shows that respondents consider self-employment to be a crucial factor: More than 60% of SOEP-P respondents report that their wealth comes from self-employment; about 20% report wealth from dependent employment. Real estate ownership is seen as another important factor with more than 30%. Here, it should be kept in mind that real estate ownership may also be associated with self-employment, for example, if the real estate was entered into a property management company. This finding is also consistent with earlier studies from the ViD and HViD studies (Ströing, Grabka, and Lauterbach, 2016). SOEP-P respondents report that marriage, financial market transactions, and also intergenerational transfers from inheritances and gifts have played a subordinate role in wealth accumulation.



Source: SOEP-P; own calculations with weighted values.

Figure 10: Self-report data on factors influencing wealth formation in SOEP-P

4.1.5 Distribution of net wealth

To describe the distributions of individual net wealth in both samples, Table 4 shows different percentile values, percentile ratios, and Gini and Theil coefficients. Besides the

point estimates, lower and upper bounds define the 95% confidence intervals based on 500 bootstrap replications.¹¹

The Gini coefficient corresponds to twice the area between the uniform distribution line and the Lorenz curve and assumes values in the interval [0; 1] if the underlying variable of interest is not negative. Since there is also negative net wealth, the Gini here is not limited to this value range and can also assume values greater than 1. However, we have also calculated the Gini for a censored distribution (Gini*) in which all assets are positive.

Table 4: Distribution of net wealth in SOEP and SOEP-P

	SOEP			SOEP-P		
	lower bound	estimator	upper bound	lower bound	estimator	upper bound
p1	-25,000	-22,260	-20,000	-305,869	-13,208	0
p5	-3,900	-3,000	-2,060	31,616	46,100	65,000
p10	0	0	0	86,000	111,000	134,600
p25	0	0	0	280,000	300,000	335,000
p50	20,000	22,000	24,000	690,000	752,005	800,000
p75	120,000	123,620	127,100	1,609,000	1,800,000	1,987,595
p90	256,000	265,000	276,031	3,770,821	4,315,000	4,825,098
p95	395,000	407,700	426,000	5,884,999	6,810,000	7,484,231
p99	920,000	1,050,000	1,200,000	17,608,876	39,724,100	54,710,000
p99.9	3,108,000	4,019,000	4,418,795	56,536,952	104,050,000	130,800,000
p75/p50	5.24	5.62	6.17	2.17	2.39	2.60
p90/p50	11.19	12.05	13.35	5.03	5.74	6.49
p99.9/p50	136.30	182.68	212.93	71.43	138.36	183.93
Gini	0.769	0.783	0.796	0.690	0.734	0.769
Gini*	0.750	0.761	0.773	0.679	0.724	0.758
Theil*	1.198	1.316	1.463	1.068	1.299	1.470
Mean	101,903	107,649	114,727	1,855,974	2,235,626	2,674,914

Source: SOEP v35 and SOEP-P; own calculations with weighted values. Bootstrap confidence intervals with 500 replications specified as lower and upper bounds

* All negative and null values are recoded to 0.01 since the Theil index is defined only for strictly positive values. The Gini is also adjusted accordingly for comparison purposes.

In the SOEP, the lowest percentile has negative net wealth of about 25,000 euros. The value for the 25th percentile is zero euros, the median is about 20,000 euros, and for the 75th percentile about 120,000 euros. Beyond this point, wealth rises sharply to around 260,000 euros for the 90th percentile, around 400,000 euros for the 95th percentile, and around 1 million euros for the 99th percentile. The variance of wealth in SOEP-P

¹¹Bootstrapping is a method of resampling that uses replacement. Here, 500 random permutations of the original data set are created. For each permutation, the statistic of interest is then calculated. The 500 statistics determine the confidence band.

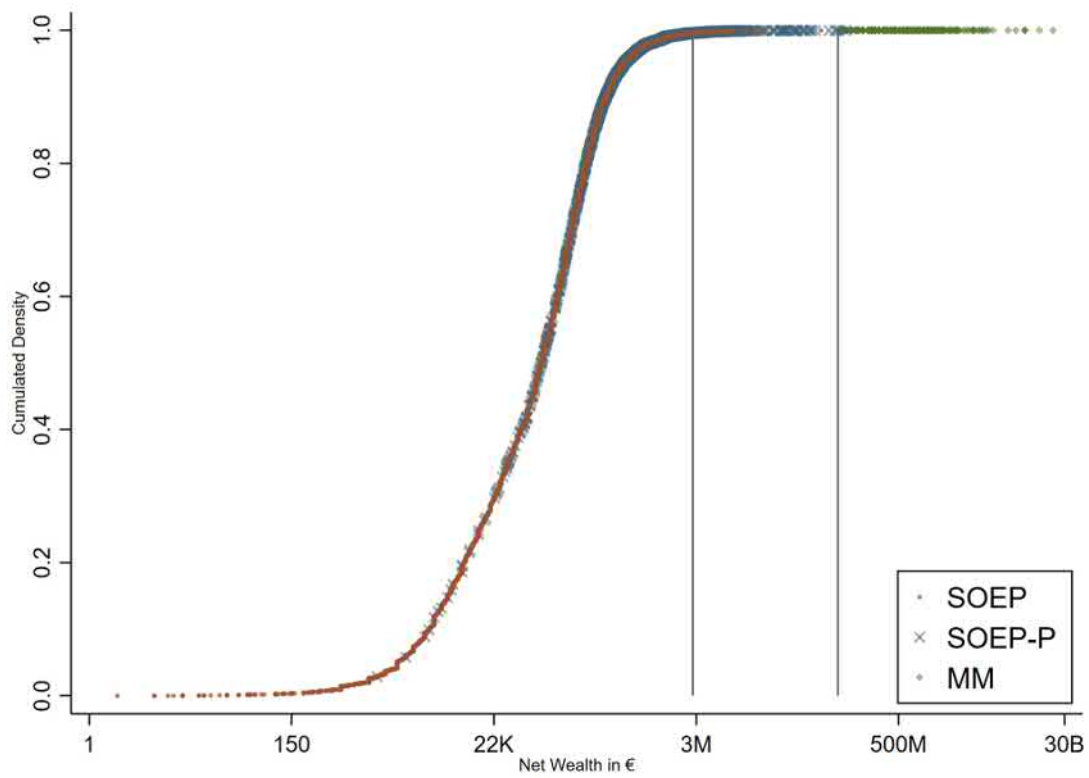
is significantly greater. The lowest percentile has debt of around 13,000 euros. Beyond the lowest percentile, wealth increases rapidly and is far above the SOEP values in all subsequent percentiles. For example, the values for the 25th percentile in SOEP-P are already 300,000 euros and over 750,000 euros for the median. The 75th percentile has wealth of 1.8 million euros and the 99th percentile has 40 million euros. The differences in the average wealth are striking: Average wealth in SOEP-P is 21 times higher than in SOEP, at over 2.2 million euros and around 100,000 euros, respectively.

Since people already possess significant wealth in the lower percentiles of SOEP-P, in contrast to SOEP, the wealth inequality in SOEP-P is significantly lower: The 90-50 and 75-50 percentile ratios are around 5.7 and 2.4, respectively, in SOEP-P, and 12.8 and 6.0, respectively, in SOEP. The Gini coefficient is also lower in SOEP-P than in SOEP at around 0.73 and 0.79, respectively. In summary, these findings show that SOEP-P respondents are strongly concentrated at the upper end of the net wealth distribution. As a result, many cases in SOEP-P fall within the data gap described above. This is evident in Figure 11, which is familiar from Chapter 1 but is supplemented here by SOEP-P. As noted at the beginning, the continuous function describes the empirical distribution of positive net wealth according to SOEP. The green dots are cases from *Manager Magazin*. The blue crosses are new in this figure and indicate individual data points from SOEP-P. In total, there are 881 cases in SOEP-P with wealth of at least 1 million euros, 353 cases with wealth of at least 3 million euros, and 185 cases with wealth of at least 5 million euros.

4.2 Integrated wealth distribution

As shown in the previous chapter, 4.1.5, many SOEP-P respondents have high wealth and many fall into the area of the data gap. SOEP and SOEP-P were used separately in the descriptions. However, the comparative advantage of the present project is that SOEP and SOEP-P used the same survey instruments (questionnaires) and use an integrated set of weights (see Chapter 3). This makes it possible to integrate the data into a single data set and thus to conduct joint analyses based on a larger number of cases than in the two individual data sets, especially in the area of high wealth. One limitation is that the two data sets were collected in two different years.

Accordingly, the SOEP wealth data refer to the survey year 2017, and the SOEP-P



Source: SOEP v35 and SOEP-P and MM; own calculations with weighted values. Empirical distribution function of net wealth in euros. Orange line: Individual SOEP net wealth in 2017 (positive net wealth only); blue crosses: SOEP-P population in 2019 (positive net wealth only); green diamonds: Individuals in the top 500 of the *Manager Magazin* “rich list” from 2014. Wealth was mapped on log scale.

Figure 11: Data gap in the area of high wealth in SOEP+SOEP-P

data refer to 2019, and it can be assumed that wealth in Germany has risen in the last two years. Thus, the SOEP figures for 2017 probably underestimate the situation in 2019 to some degree, but wealth in certain segments of the distribution may also have fallen.

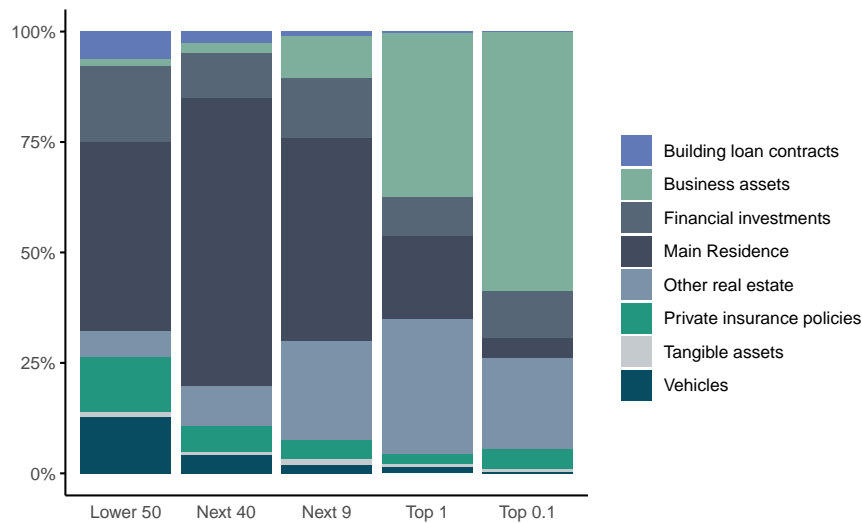
4.2.1 Portfolios

Figure 12 shows how the portfolios are composed along the integrated asset distribution. For the bottom 50% of respondents in SOEP+SOEP-P, the market values of vehicles represent the central wealth component with a portfolio share of around 25%. Financial investments in the form of savings deposits, savings and mortgage bonds, stocks, or investment shares account for just under 16%. Building loan contracts and private insurance policies account for just under 14%. In the range between the 50th and 90th percentile, around 68% of wealth comes from owner-occupied residential property or other real estate. The portfolio share invested in real estate remains stable in the following nine percentiles. In the highest percentile, it falls to just over 50% and in the highest 0.1th percentile to only about 25%. At the same time, the share of business assets increases: while these play only a very minor role for the lower 99%, they rise to 30% in the top percentile and to over 50% in the top 0.1th percentile. The high importance of business assets in the top percentile is consistent with empirical findings based on PHF for Germany (see Schröder, Bartels, Grabka, König, et al., 2019) and the Survey of Consumer Finance for the United States (Wolff, 2017).

It is noteworthy that, contrary to expectations, financial assets do not play a significant role for top wealth. One possible explanation is that corporate profits are not distributed but retained. It is also evident that a comparatively high proportion of private insurance companies are in the top 0.1% of the distribution. Here, however, there is a clustering of self-employed people, who make private provisions for old age, for instance, through private life insurance policies.

4.2.2 Wealth inequality and wealth concentration

It is often argued that survey data underestimate the true inequality and concentration of wealth. In order to quantify this underestimation, the following section compares corresponding measures for two distributions: one based solely on individual net wealth from SOEP, and the other based on SOEP+SOEP-P.



Source: SOEP v35 and SOEP-P; own calculations with weighted values. Top 1% includes the Top 0.1%.

Figure 12: Portfolios in SOEP+SOEP-P by wealth groups

Table 5 describes the inequality of both net wealth distributions using selected percentile values, as well as percentile ratios and the Gini coefficient. Through the integration of SOEP-P, wealth increases significantly from the 95% percentile onwards from approximately 410,000 euros (SOEP) to 440,000 euros (SOEP+SOEP-P). In the 99% percentile, wealth increases from about 1.05 million euros to 1.33 million euros with the integration of SOEP-P. These results confirm that many SOEP-P respondents rank in the top 1% of the wealth distribution.

Integration also means an increase in the measured inequality using the 99.9-50 percentile ratio and the Gini coefficient: The percentile ratio increases from about 183 to 241; the Gini coefficient from 0.78 to 0.81. When interpreting the Gini coefficient, it should be noted that it is not limited to the 0-1 interval due to negative wealth. The Theil index, which is also given, is only defined for strictly positive wealth. Although the Theil index is generally not defined for the 0-1 interval, it has the advantage that, like other indices in the entropy family, it can be decomposed additively by subgroup. If all assets that are not strictly positive are recoded to 0.01, the Gini coefficient decreases slightly. Like the Gini coefficient, the Theil coefficient increases noticeably with the integration of SOEP-P.

A key indicator of wealth concentration is the proportion of total assets held by a particular quantile at the top end of the individual net wealth distribution, e.g., the top

Table 5: SOEP net wealth distributions after integration of SOEP-P

	SOEP			SOEP+SOEP-P		
	lower bound	estimator	upper bound	lower bound	estimator	upper bound
p1	-25,000	-22,260	-20,000	-25,000	-22,260	-20,000
p5	-3,900	-3,000	-2,060	-3,600	-3,000	-2,000
p10	0	0	0	0	0	0
p25	0	0	0	0	0	0
p50	20,000	22,000	24,000	20,700	22,800	25,000
p75	120,000	123,620	127,100	122,500	126,000	130,200
p90	256,000	265,000	276,031	266,000	279,236	287,500
p95	395,000	407,700	426,000	417,400	438,000	457,500
p99	920,000	1,050,000	1,200,000	1,171,367	1,330,000	1,430,000
p99.9	3,108,000	4,019,000	4,418,795	4,664,803	5,490,000	8,011,008
p75/p50	5.24	5.62	6.17	5.13	5.53	5.96
p90/p50	11.19	12.05	13.35	11.22	12.25	13.24
p99.9/p50	136.30	182.68	212.93	201.84	240.79	349.68
Gini	0.769	0.783	0.796	0.798	0.809	0.820
Gini*	0.750	0.761	0.773	0.778	0.789	0.800
Theil*	1.198	1.316	1.463	1.490	1.627	1.798
Mean	101,903	107,649	114,727	120,190	126,694	135,015

Source: SOEP v35 and SOEP-P; own calculations with weighted values. Bootstrap confidence intervals with 500 replications given as lower and upper bounds.

* All negative and zero assets are recoded to 0.01 as the partial index is only defined for strictly positive values. The Gini is also adjusted accordingly for comparison purposes.

1%. Table 6 gives these percentages for the three samples defined above. In addition to the top 10%, the top 5%, 1%, and 0.1% of these three individual net wealth distributions are also shown.

In the SOEP, the top 10% hold just under 60% of total wealth, the top 5% about 44%, the top 1% about 22%, and the top 0.1% about 8%. For the integrated distribution SOEP+SOEP-P, these percentages rise to just under 64%, 50%, 28%, and just under 12%, respectively. The concentration of wealth at the top or upper bound of the distribution is therefore clearly underestimated by the available survey data.

4.3 Development of the wealth distribution over time

The SOEP wealth module is often used to describe the intertemporal development of private wealth in Germany. As shown above, the SOEP alone underestimates the concentration of wealth. The integration of SOEP-P means that, in contrast to previous years, it is now

Table 6: Total wealth shares in SOEP with SOEP-P

	SOEP	SOEP+SOEP-P
Top 10%	58.9 (57.1; 61.2)	64.1 (62.3; 66.1)
Top 5%	43.9 (41.4; 46.7)	50.6 (48.0; 53.2)
Top 1%	21.6 (18.5; 25.1)	29.0 (25.8; 32.4)
Top 0.1%	7.3 (4.6; 10.7)	12.8 (9.8; 15.8)

Source: SOEP v35 and SOEP-P; Own calculations with weighted values. Bootstrap confidence intervals with 500 replications given in parentheses.

possible to guarantee better representation of people with high wealth. However, this also has an effect on the comparability of indicators of inequality over time. This section shows the development of wealth in Germany from 2002 to 2017, with the results for last year based first on SOEP and second on SOEP+SOEP-P. This allows a quantitative assessment of the strength of the effect.

Table 7 shows the statistics for percentile values and ratios, Gini coefficient, and mean values at a 5-year interval from 2002 onwards. Below the 25th percentile of the wealth distribution, the distribution is relatively stable over time: the lowest percentile is negative 20,000 euros, and the 25th percentile is zero. The median rises by almost 30% from around 17,500 euros in 2002 to 22,000 euros in 2017. In the upper half of the distribution, a clear dynamic can be observed: in the 75th percentile, wealth increases from approximately 100,000 euros in 2002 to approximately 125,000 euros in 2017. In the 99th percentile, wealth rises from just under 800,000 euros to over 1,000,000 euros over the same period. These dynamics correspond to a relative increase of approximately 25% and 35%, respectively. The percentile ratios as well as the Gini coefficient are correspondingly stable; e.g. the ratio between the 90th and 50th percentile is 12.26 in 2002 and remains virtually unchanged at 12.05 in 2017.

The integration of SOEP-P in the last year of the study does not change the picture for the lower half of the wealth distribution, but reinforces the dynamics in the upper half, especially in the top percentile. It results in a further 3.6% increase in the median to 22,800 euros, a 1.9% increase in the 75th percentile to 126,000 euros, and a 27% increase

in the 99th percentile to over 1.3 million euros. In addition to a significant increase in average assets from approximately 110,000 to 125,000 euros, the Gini coefficient also rose significantly from 0.783 for SOEP to 0.806 for SOEP+SOEP-P. Comparable figures are found with the alternative use of the Theil coefficient, which is 1.315 for SOEP and 1.575 after integration of SOEP-P. This again shows that the SOEP-P population is particularly well represented at the top end of the wealth distribution.

Table 7: Time series of net wealth with and without integration of SOEP-P

	2002 SOEP	2007 SOEP	2012 SOEP	2017 SOEP	2017 SOEP+SOEP-P
p1	-19,750 (-20,704; -17,000)	-27,000 (-32,000; -21,480)	-21,500 (-25,000; -19,900)	-22,260 (-25,000; -20,000)	-22,260 (-25,287; -20,000)
p5	-1,800 (-3,000; -1,000)	-4,000 (-4,534; -3,000)	-3,200 (-4,000; -2,500)	-3,000 (-3,900; -2,000)	-3,000 (-3,749; -2,000)
p10	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)
p25	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)	0 (0; 0)
p50	17,500 (16,000; 19,382)	16,000 (15,000; 17,500)	20,000 (18,600; 21,500)	22,000 (20,000; 24,001)	22,800 (20,800; 25,000)
p75	100,000 (97,500; 104,000)	95,000 (91,000; 99,800)	104,571 (100,000; 108,800)	123,620 (120,000; 127,000)	126,000 (123,000; 130,000)
p90	214,600 (207,600; 222,000)	215,000 (203,400; 223,000)	220,500 (213,000; 229,000)	265,000 (255,000; 277,000)	278,544 (267,500; 286,750)
p95	326,750 (315,000; 343,641)	325,000 (311,900; 340,000)	330,870 (317,000; 350,000)	407,700 (394,420; 427,500)	437,800 (419,300; 457,000)
p99	765,000 (716,224; 824,500)	790,000 (733,000; 868,076)	841,708 (790,000; 917,500)	1,050,000 (920,000; 1,200,000)	1,329,800 (1,171,367; 1,430,000)
p99.9	2,745,001 (2,425,600; 3,455,000)	3,933,000 (3,169,737; 4,740,000)	3,208,887 (2,525,000; 4,156,396)	4,019,000 (2,860,000; 4,418,795)	5,450,000 (4,530,000; 7,446,000)
p75/p50	5.71 (5.25; 6.18)	5.94 (5.52; 6.33)	5.23 (4.95; 5.48)	5.62 (5.19; 6.10)	5.53 (5.11; 5.99)

Table 7: Continuation

	2002 SOEP	2007 SOEP	2012 SOEP	2017 SOEP	2017 SOEP+SOEP-P
p90/p50	12.26 (11.13; 13.13)	13.44 (12.26; 14.40)	11.03 (10.38; 11.81)	12.05 (11.04; 13.23)	12.22 (11.07; 13.36)
p99.9/p50	156.86 (137.04; 198.58)	245.81 (188.20; 302.18)	160.44 (124.66; 207.82)	182.68 (135.88; 210.42)	239.04 (197.46; 329.41)
Gini	0.775 (0.763; 0.787)	0.797 (0.786; 0.810)	0.774 (0.762; 0.789)	0.783 (0.770; 0.797)	0.806 (0.794; 0.818)
Gini*	0.756 (0.745; 0.768)	0.773 (0.762; 0.784)	0.751 (0.740; 0.764)	0.761 (0.751; 0.775)	0.786 (0.775; 0.796)
Theil*	1.323 (1.197; 1.459)	1.399 (1.286; 1.504)	1.252 (1.159; 1.372)	1.316 (1.197; 1.476)	1.575 (1.443; 1.704)
Mean	86,371 (81,841; 91,101)	86,750 (82,082; 91,692)	89,206 (84,737; 94,579)	107,649 (100,945; 114,474)	124,789 (118,352; 131,282)

Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Bootstrap confidence intervals with 500 replications given in parentheses.

*All negative and null values are recoded to 0.01, since the partial index is defined only for strictly positive values.

*The Gini is also adjusted accordingly for comparison purposes.

4.4 Millionaires in Germany

The goal of this chapter is a comparative analysis of sociodemographic, economic, and qualitative characteristics along the distribution of individual net wealth using the integrated SOEP+SOEP-P dataset. In concrete terms, the adult SOEP+SOEP-P population is assigned to four disjoint wealth segments:

1. *The bottom 50%*: People with individual net wealth up to around 23,000 euros.
2. *The middle 25%*: People from the median to the 75th percentile (around 126,000 euros).
3. *The top 24%*: People from the 75th to around 99th (more precisely: 98.6th) percentile (< 1 million euros).
4. *Millionaires*: People with individual net wealth of at least 1 million euros.

Studying millionaires as a separate group becomes much more meaningful through the integration of SOEP-P than with SOEP alone, because integration increases the observed number of millionaires from 317 (SOEP 2017) to 1,198.

4.4.1 Sociodemographic characteristics

Table 8 allows comparative examination of the sociodemographic characteristics of the integrated SOEP+SOEP-P population in the lower 50%, middle 25%, upper 24%, and millionaires.

For all of the characteristics considered, there are clear differences across the four segments of the wealth distribution—and especially between the group of millionaires and the three groups of non-millionaires.

- *Gender*. The percentage of women decreases slightly from about 53% in the lower 50% to about 48% in the upper 24%. In contrast, the population of millionaires is clearly dominated by men: only 31% of millionaires are women.
- *Age*. The age distribution shifts significantly across the wealth distribution: While people over 49 years of age account for only about 38% in the lowest wealth segment,

Table 8: Sociodemographic characteristics by wealth groups

Corresp. segment	Shares in percent			
	p0 - p50	p50 - p75	p75 - p98.6	Millionaires
Wealth bandwidth (in thousands)	$w \leq 22.8$	$22.8 < w \leq 126$	$126 < w < 1,000$	$w \geq 1,000$
<i>Attributes</i>				
Women	53.02	50.50	47.63	30.69
With migrant background	30.64	18.64	12.41	13.62
New federal states	20.26	20.25	8.06	6.05
<i>Age groups</i>				
18-24	16.28	1.43	0.12	0.00
25-49	45.51	37.80	23.33	23.39
50-64	19.27	31.99	38.65	36.71
65+	18.95	28.78	37.91	39.90
<i>Labor market status</i>				
Active	58.63	65.21	59.32	61.59
Inactive	24.98	9.89	8.80	5.11
Retired	16.39	24.91	31.88	33.31
<i>Occupational status</i>				
Self-employed	3.93	8.41	19.13	72.56
Worker	22.84	19.99	12.25	1.69
Civil servants	3.57	6.55	9.59	3.68
Trainees	17.09	1.81	0.29	0.12
Employee	52.57	63.24	58.73	21.95
Employee with managerial funct.	0.86	1.85	5.24	22.22
Manager and shareholder	1.75	1.55	4.01	25.62
<i>Size of self-employed enterprise</i>				
No employees	83.40	69.11	47.04	19.58
1-9 employees	15.48	28.53	44.34	34.08
10 and more employees	1.12	2.36	8.62	46.34
<i>School education</i>				
No Degree	5.65	0.90	0.54	0.00
Secondary Level I	16.06	8.98	7.88	14.07
Secondary Level II	59.27	59.81	48.98	31.80
Upper Secondary Degree	19.01	30.31	42.60	54.13

Source: SOEP v35 and SOEP-P; own calculations with weighted values.

they make up over 60% in the middle segment and over 75% in the upper segment as well as in the group of millionaires.

- *Migration background.* The percentage of the population with a background of migration is overrepresented in the lower 50% and significantly underrepresented in the group of millionaires, at 31% and 14%, respectively.
- *Place of residence.* The population residing in the new federal states is also overrepresented in the lower 50% and clearly underrepresented in the group of millionaires, at about 20% and 6%, respectively.

- *Labor market status.* The age gradient described above is reflected in the share of retirees: as expected, the share rises from approx. 16% in the lower 50% to approx. 25% in the middle and over 30% in the upper wealth segment, and to about one third in the group of millionaires. In contrast, it is surprising that the working population is quite stable at around 60% across all wealth groups, despite the significant differences in age structure. By contrast, the share of the population that is currently economically inactive (e.g., due to child rearing, child care, etc.) falls sharply from around 25% in the bottom 50% to around 10% in the middle and upper wealth segment, and to just 5% in the group of millionaires. There is almost no one in the group of millionaires who is of working age and economically inactive.
- *Professional status.* White-collar workers are by far the most strongly represented group of working people. They account for between approximately 53% and 63% of the lower three wealth groups, but only 22% of millionaires. Here, self-employed are in the majority, accounting for over 70% of millionaires, whereas they make up only about 4% of the lower wealth segment and only 19% even in the upper segment. Other groups of employed people whose shares systematically decrease along the distribution of wealth are blue-collar workers (from approx. 23% to below 2%) and apprentices or trainees (from approx. 17% to below 1%).
- *Management positions.* As was to be expected, the proportion of employees in management positions rises significantly across the wealth distribution. Employees who state that they hold management responsibilities in their jobs, or are partners or managing directors in companies, remain below 2% in the lower 75% of the wealth distribution. In contrast, among millionaires these individuals are strongly represented at almost 50% of employees.
- *Size of company among self-employed people.* Of the self-employed in the lower wealth segment, a vast majority, around 83%, are “solo self-employed”, operating their own one-person business. A further 16% have a maximum of nine employees. The number of employees in the following three wealth segments rises significantly: In the middle and upper segments, the share of solo self-employed people falls to approx. 69% and approx. 47%, respectively. Approximately 29% and 44%, respectively, report

having 1-9 employees. Among the millionaires, only about 20% are solo self-employed. About 34% have 1-9 and about 46% have 10 or more employees.

- *Educational level.* An appropriate level of education and vocational training is the key to an upward career trajectory, and on the labor market, the reward for higher education is typically higher income. Accordingly, there is a clear positive correlation between educational attainment and wealth: the share of individuals without a school-leaving certificate (completion of Secondary Level I, which extends up to the end of grade 10 or compulsory education; Secondary Level II, which extends beyond grade 10 and includes vocational and university-track schools) in the lower wealth segment is around 6% (16%; 60%) and among millionaires less than 0.01% (14%; 32%). By contrast, only around 19% of the lower 50% have an Abitur, the highest level of secondary education. The share of persons with an Abitur increases across the four wealth groups and reaches 54% in the group of millionaires.

4.4.2 Personality traits

There are various arguments suggesting that personality and socioeconomic status (and thus wealth) are related. [Roberts et al. \(2007\)](#), for instance, argue that there are positive selection effects if a career path taken is compatible with the individual's personality. [Cook, Vance, and Spector \(2000\)](#) argue that selection effects could also be at work in the sense that people with certain personality traits are selected for certain high-status positions. An example would be the preference for more extraverted, less neurotic, and less conscientious people for positions that require a high level of communicative ability. [Judge and Ilies \(2002\)](#) also argue that personality has a more direct influence on performance. This correlation would be triggered, for example, if more conscientious and less neurotic individuals formulated clearer personal goals, which in turn had an effect on their work performance.

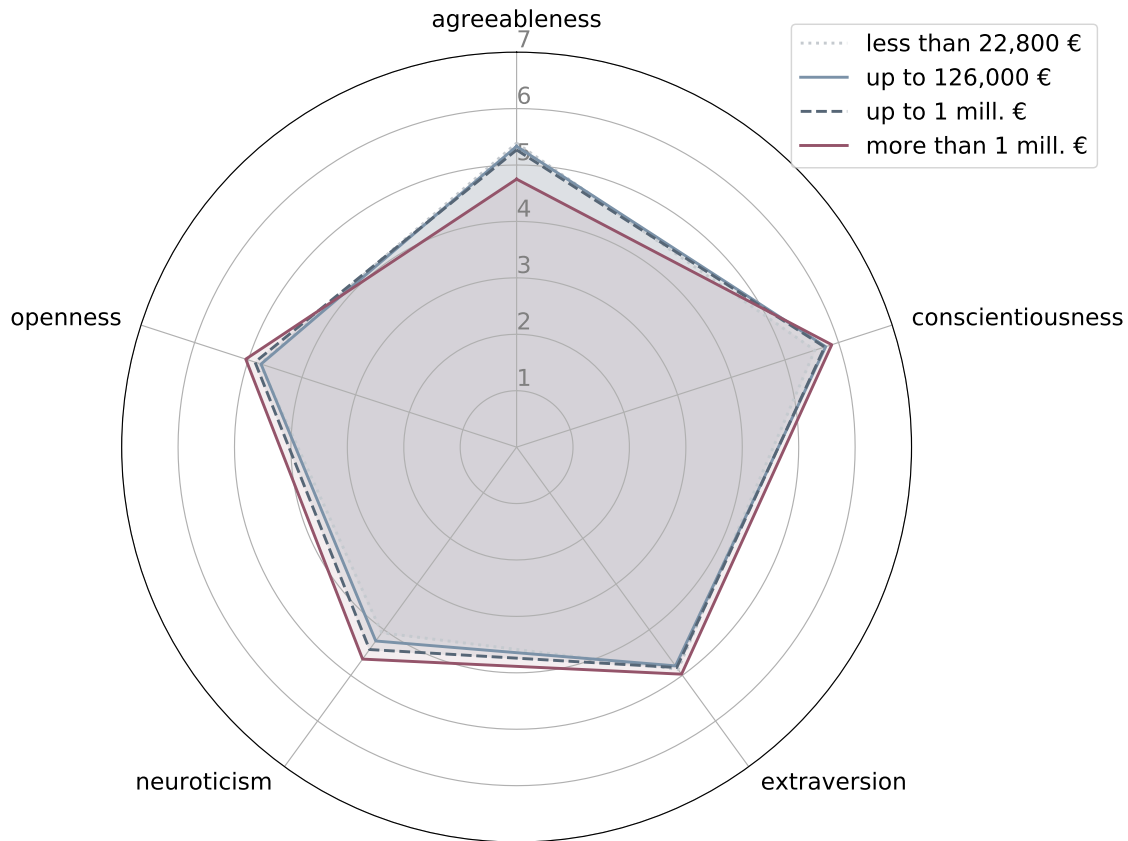
The literature review by [Leckelt et al. \(2019\)](#) cites a number of empirical studies showing that the personality traits of people with high economic resources and in high professional positions differ systematically from the rest of the population (e.g., [Ozer and Benet-Martinez, 2006](#); [Roberts et al., 2007](#)). [Ackerman and Heggstad \(1997\)](#), for example,

report that extraverted individuals are more likely to become entrepreneurs and take on leadership roles (Judge, Bono, et al., 2002). This personality-career relationship is also associated with higher work performance and success (Judge, Higgins, et al., 1999). Other studies show that high socioeconomic status is positively correlated with extraversion, conscientiousness, and narcissism (Grijalva et al., 2015) and negatively correlated with neuroticism (e.g., Anderson and Cowan, 2014). Ng et al. (2005) describe similar patterns for employment careers: In their meta-analysis, they report that salary and promotion are positively correlated with extraversion, conscientiousness, openness to experience (for salary only), and locus of control (for salary only) and negatively correlated with neuroticism. Leckelt et al. (2019) show that the Big Five personality traits of people with high wealth are different from the population average.

However, all of the aforementioned studies base their empirical analyses either on comparatively small samples and/or convenience samples, which limits their explanatory power. In the following, we describe the personality traits and willingness to take risks of millionaires in comparison to the lower 50%, middle 25%, and upper 24% of the integrated SOEP+SOEP-P population. When interpreting the results, please note that only bivariate statistical correlations are presented. Thus, no statistical control is carried out for larger differences in the composition of the four subpopulations. For example, if a characteristic differs between the group of millionaires and the lower 50%, this may be due to differences in wealth as well as in the higher average age of the former subpopulation.

Figure 13 shows the average values for the personality traits of the respondents in the four wealth segments. The mean for each characteristic is given for each of the four subpopulations. For the three dimensions of conscientiousness, extraversion, and openness, there are no systematic differences between subpopulations. In contrast, values for neuroticism rise across the wealth segments and fall for tolerance: the group of millionaires in particular reports being less tolerant and emotionally stable (neuroticism).

Figure 14 shows how the average willingness of the four subpopulations to take risks differs, showing the proportion of the respective subpopulation with each self-rated value from 0 to 10. For the lower three subpopulations, there are no systematic differences in the willingness to take risks. In contrast, millionaires are significantly more willing to take risks: more than 12% of millionaires fall into the top two categories, while only around 5%



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

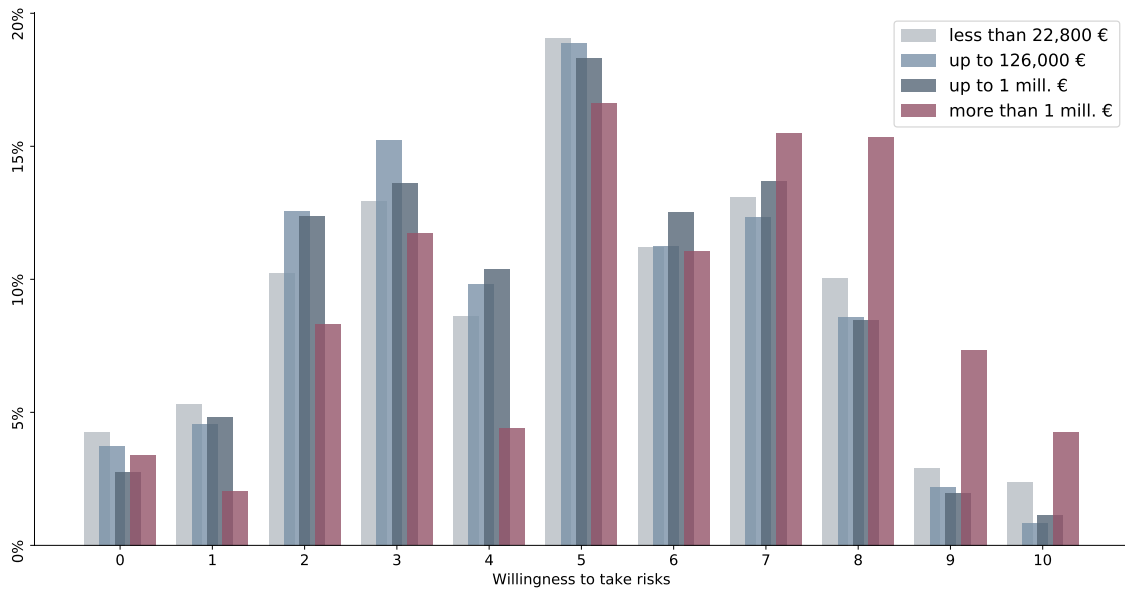
Figure 13: The Big Five in SOEP+SOEP-P by wealth groups

of the other subpopulations do. This is in line with studies showing that the self-employed are generally more willing to take risks (Caliendo, Fossen, and Kritikos, 2014).

4.4.3 Satisfaction with different areas of life

Gross national product is an important indicator of social prosperity at the macro level, while individual income or wealth are important micro-level indicator. However, these indicators do not take into account other important dimensions of well-being, such as health or access to cultural or educational institutions.

Various institutions and expert commissions have therefore called for more comprehensive measures of prosperity, taking into account not only economic but also social and ecological factors (European Commission: Beyond GDP, OECD: Best Life Initiative; Bundesregierung, 2019). Such measures often also make use of survey respondents' self-assessments in different areas, from job satisfaction to satisfaction with life as a whole.



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 14: Willingness to take risks in SOEP+SOEP-P by wealth groups

Results show that there is not necessarily a linear relationship between subjective satisfaction in one area and the objective factors that underlie it. This applies, for example, to the relationship between working hours and job satisfaction: people who are not allowed to work are just as dissatisfied as those who work much more than they would like to. This also applies to the relationship between income and satisfaction: According to the so-called Easterlin paradox, income and satisfaction are positively correlated, but there is a threshold value above which well-being does not increase further (Easterlin, 2001).

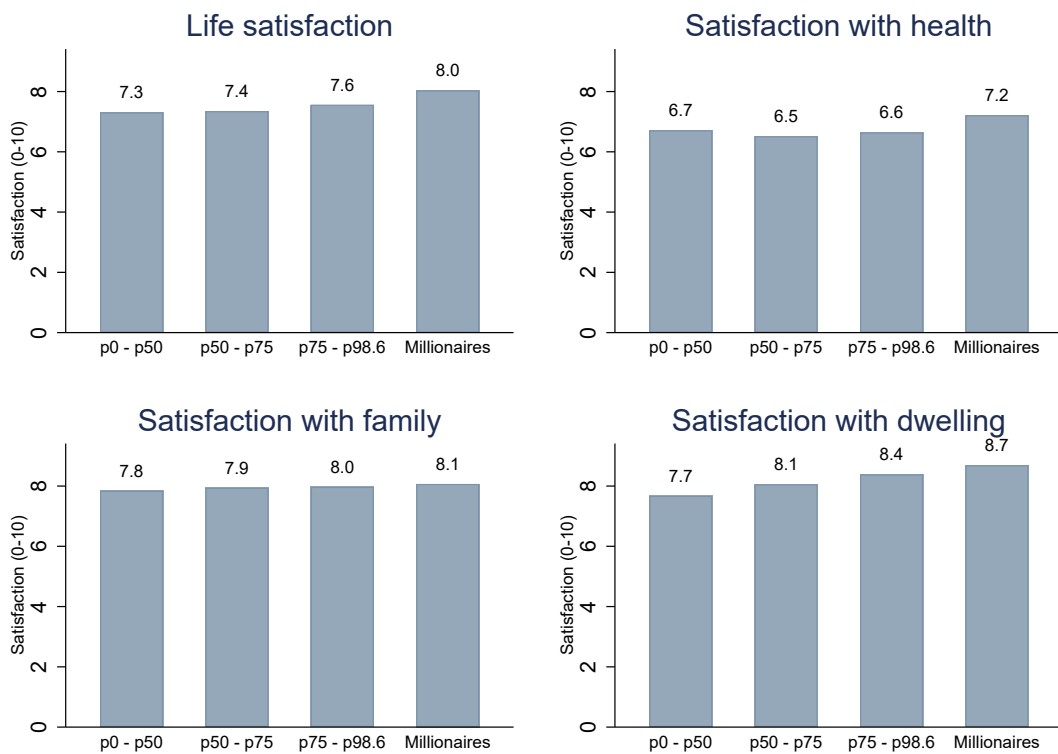
In the following, we compare subjective satisfaction in the four subgroups—from the bottom 50% to the millionaires—and then, as far as the data allow, attempt to provide an explanation for the results using objective indicators. We expect, for example, that satisfaction with work will depend on the nature of the work or the amount of overtime.

In SOEP, satisfaction with “life as a whole” is surveyed as an aggregate indicator, as is satisfaction in various areas of life: health, sleep, work, household activities, household income, personal income, housing, leisure time, child care, family life, residential area, and social security. Satisfaction is measured on scale with 11 values (0 to 10), where 0 means “completely dissatisfied” and 10 means “completely satisfied.”

Figure 15 shows that, on average, respondents are satisfied with their lives and that

satisfaction increases across the four wealth segments. Average life satisfaction is 7.3 in the lower 50%, 7.4 in the middle 25%, 7.6 in the upper 24%, and as high as 8.0 among millionaires. Even if this finding does not refute the saying that “money alone doesn’t make you happy,” it does show that wealth correlates with life satisfaction. Contrary to the Easterlin hypothesis, there are also no clear satiation effects in the upper range of the wealth distribution.

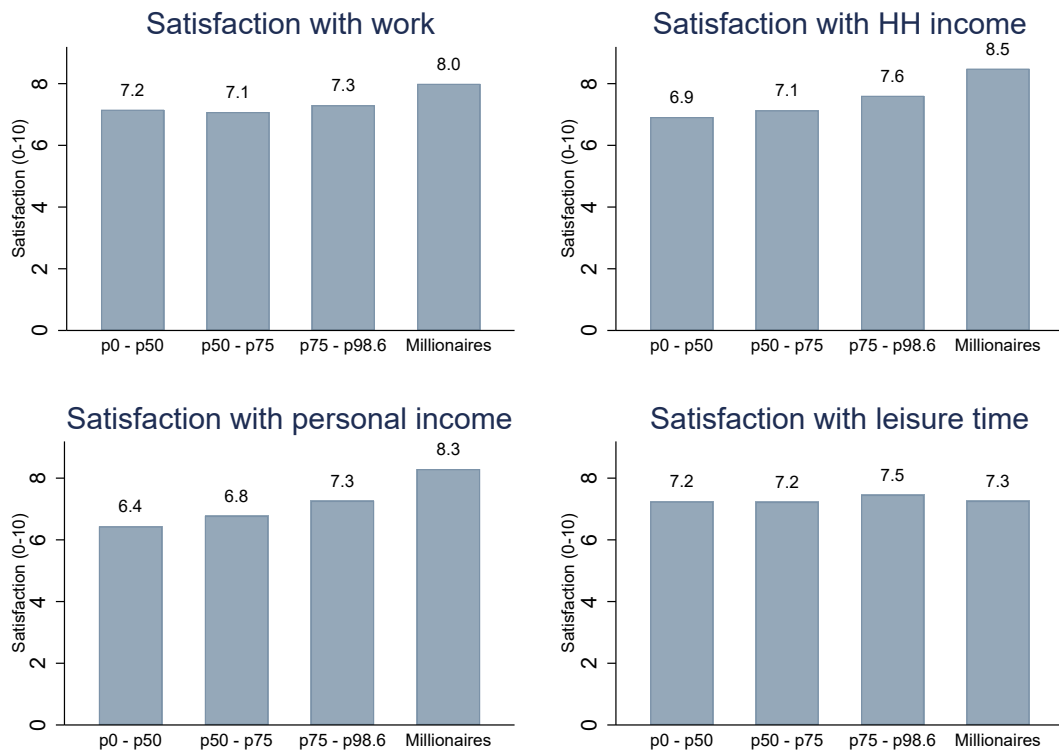
The positive bivariate correlation between wealth and satisfaction is also confirmed for the dimensions of income, work, housing, and health and family. In this context, a clear increase in area-specific satisfaction is evident especially among millionaires compared to the three lower segments (for comparison, see, e.g., the dimensions of work and income in Figures 15 and 16).



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 15: Satisfaction in selected areas of life by wealth groups

A plausible explanation for the positive correlation between personal or household income and wealth is the high positive correlation between income and wealth (see Chapter 4.4.4). A plausible explanation for the increase in job satisfaction with wealth is the more

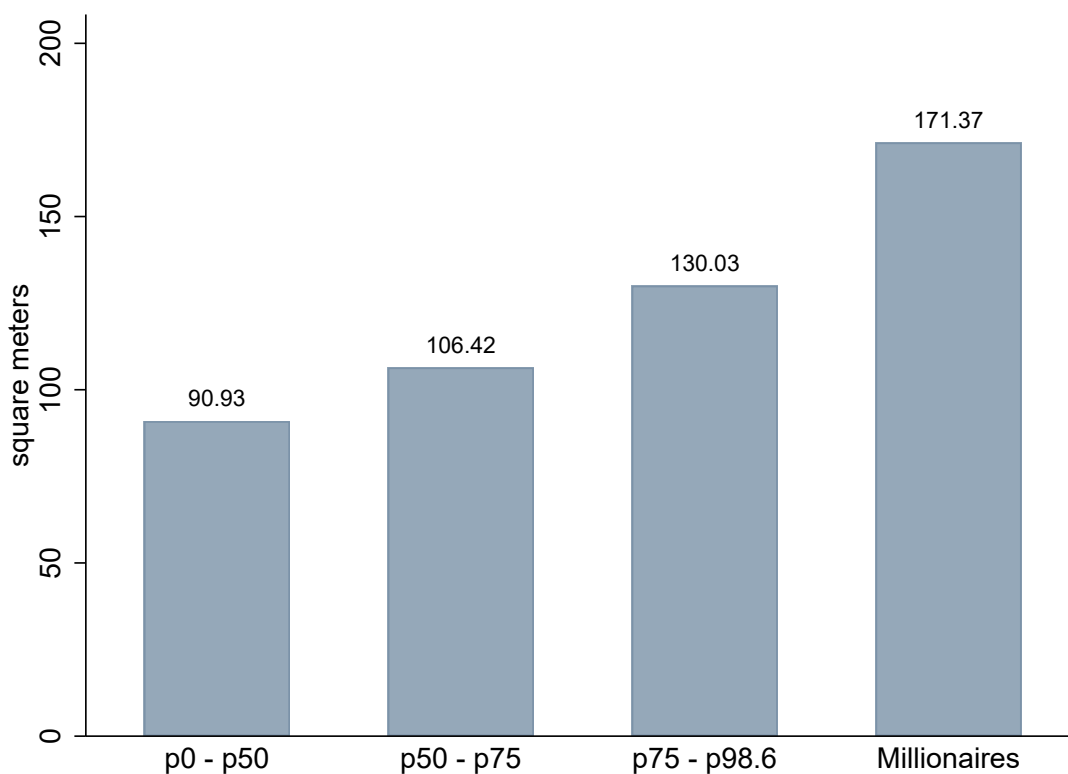


Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 16: Satisfaction in further selected areas by wealth groups

interesting content or nature of the work, and the higher degree of autonomy at work associated with a higher occupational status. One indication of this is that people with high wealth are more likely to hold managerial positions or be self-employed (see Table 8). The fact that greater material resources also increase satisfaction with one's dwelling can be explained by an overall better living environment, better building features, or more living space. Figure 17 shows, for example, that the average amount of living space increases from about 90 square meters in the lower 50% to over 171 square meters among millionaires. At the same time, however, the average household size does not increase significantly, meaning that millionaire households have far more per capita living space than people in the lower wealth segments.

Only in the dimension of leisure time do the results fail to show a positive correlation between satisfaction and wealth. Here, the average level of satisfaction rises from around 7.2 to 7.5 in the lower three wealth segments and then falls slightly to 7.3 among millionaires. A plausible explanation is the workload, measured by actual weekly working hours: employed



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

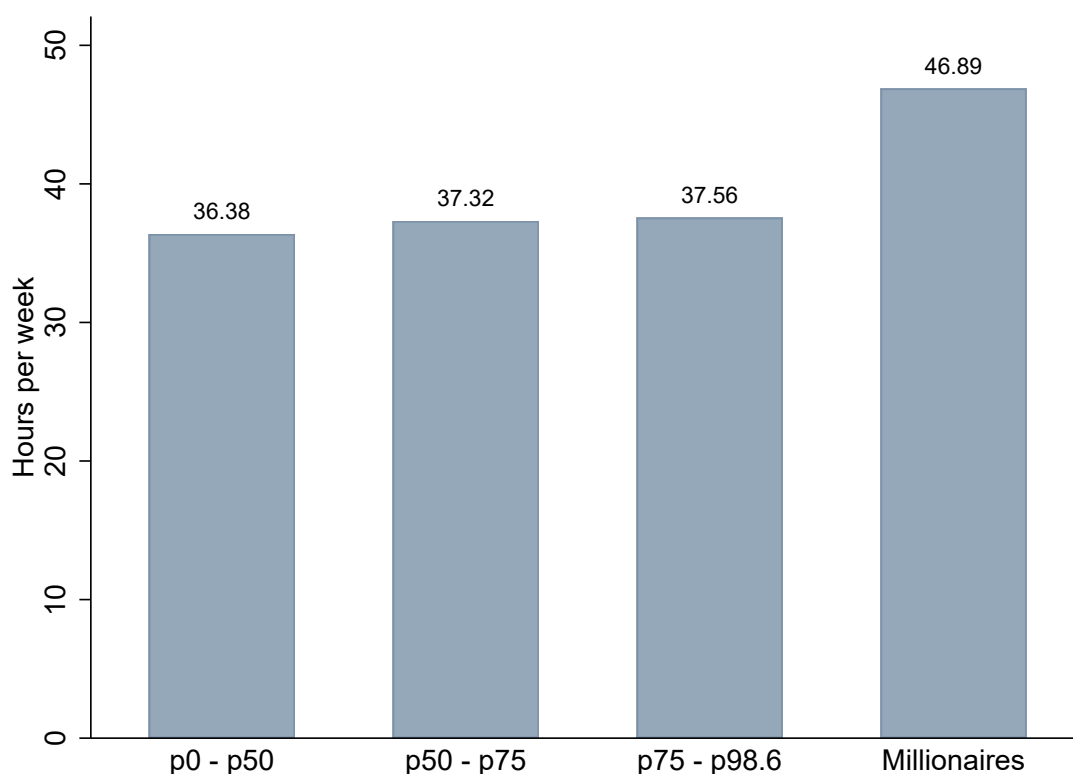
Figure 17: Living area in square meters by wealth groups

millionaires state that, at around 47 hours per week, they work almost 10 hours more than working people in the three lower wealth segments (see Figure 18). Accordingly, although this population does not lack the financial means to spend their free time as they like, their high workload is likely to decrease the amount of leisure time available to them.

4.4.4 Determinants of wealth formation

In the following, as in our comparative analysis of SOEP and SOEP-P in Section 4.1.3, we consider the key determinants of wealth formation, i.e. income and savings behavior, but now differentiated between the four wealth segments.

Figure 19 shows how equivalent net household income changes across the integrated SOEP+SOEP-P distribution of individual net wealth. As can be expected, the annual equivalent net household income rises as wealth increases. However, the increase across the first three segments of the wealth distribution is relatively moderate. For the bottom 50%, net household income is around 1,900 euros per month, while it is around 2,200

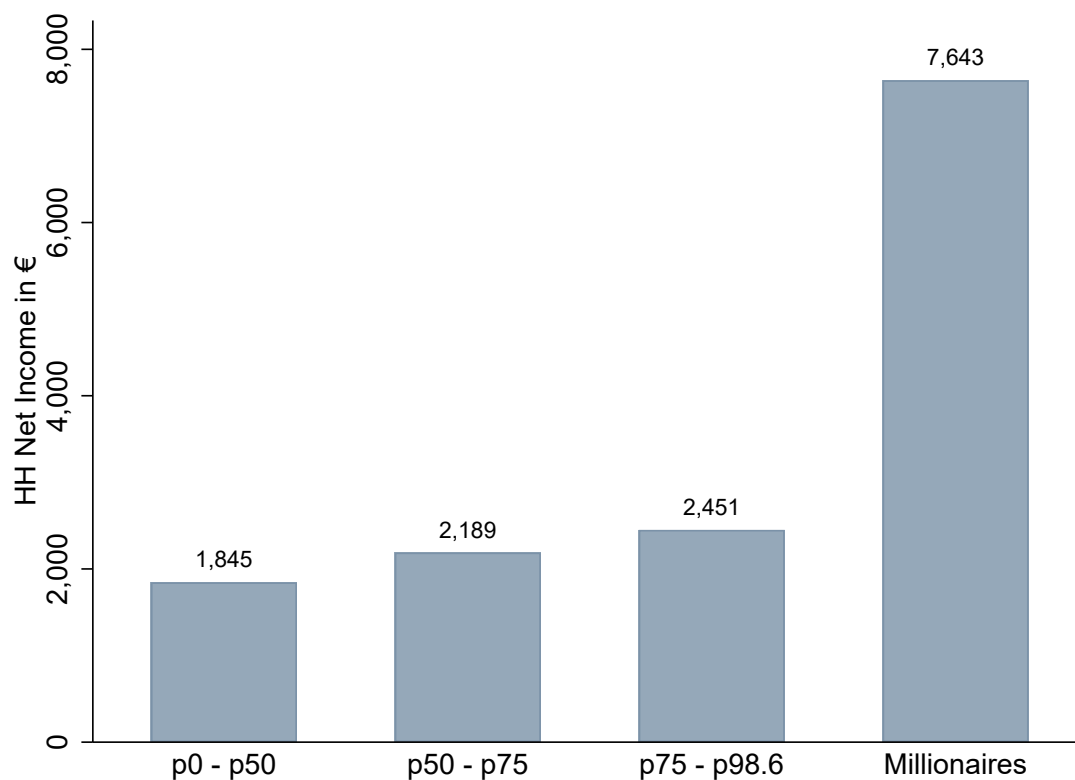


Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 18: Weekly working hours among the working population by wealth groups

euros in the middle 25%, and just under 2,500 euros in the top 24%. Comparing the lower and upper segments of the wealth distribution, we see a rise in average income of around 33%. For millionaires, on the other hand, the equivalent net household income rises again significantly to around 7,600 euros per month. This means that millionaires have considerably more financial resources at their disposal for further wealth accumulation than the rest of the population. Whether they use or consume these resources for further wealth accumulation is shown by the percentage of savers and the amount of money they save.

Figure 20 shows how the percentage of people who save with the motives of accumulating wealth and making provisions for old age change across the four wealth segments. The percentage of people saving to accumulate wealth increases from approx. 33% in the lower segment to approx. 37% in the middle and upper segments, and is around 49% in the millionaire segment. Measured by the proportion of people who save by wealth segment, however, the most important motive is that of making provisions for old age: approx. 62%



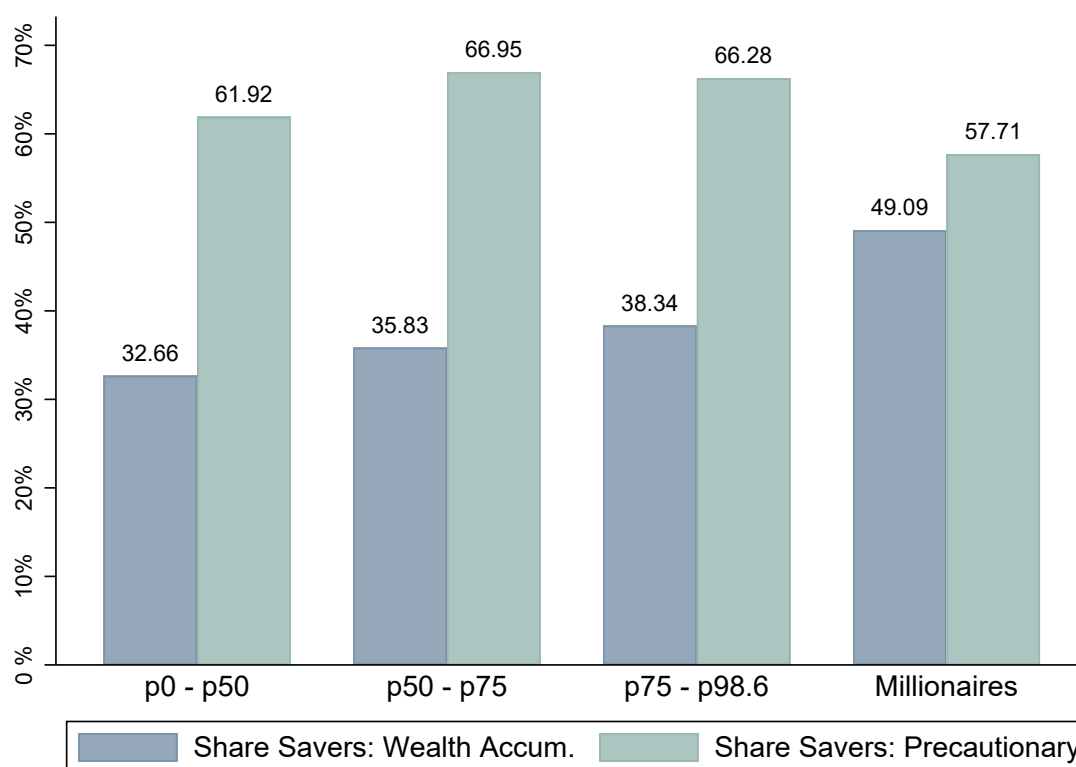
Source: SOEP v35 and SOEP-P; own calculations with weighted values

Figure 19: Equivalent monthly household net household income in SOEP+SOEP-P by wealth groups

of the lower 50% and approx. 67% of the middle 25 and upper 24% and only 58% in the millionaire segment save with this motive.

Not only do equivalent net household income and the proportion of savers with the motive of accumulating wealth increase with rising wealth: as Figure 21 shows, the amount of conditional annual savings also increases among those who save.

For savers with the motive of accumulating wealth, the lower 50% save 345 euros per year, the middle 25% just over 400 euros, and the upper 24% almost 680 euros. This means that the savings with the motive of wealth accumulation are about twice as high in the upper segment as in the lower segment. Savings also increase with wealth for savers with the motive of making provisions for old age. Here, the amounts saved per year in the three lower wealth segments are around 320 euros, 380 euros, and 560 euros. Thus, the increase in saving with the motive of making provisions for old age is, at 75%, somewhat lower than for the motive of wealth accumulation, with over 95%. But no matter what the motive is,



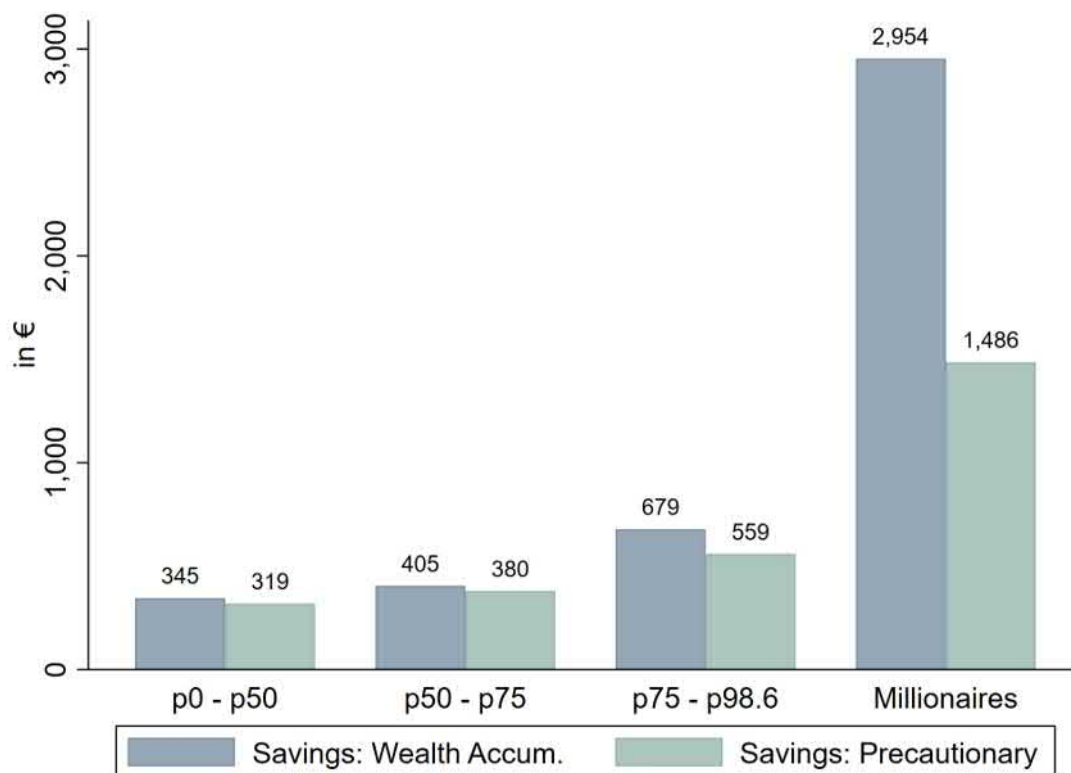
Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 20: Percentage of savers by wealth groups in SOEP+SOEP-P

the conditional savings across the three lower wealth segments increase almost twice as fast as net equivalent income. The propensity to save, measured by the savings ratio, is therefore increasing.

The savings behavior of millionaires is on a completely different and far higher level. On average, millionaires who save out of the motive of accumulating wealth put aside almost 3,000 euros per month. If they are saving for retirement, they save about half as much, at around 1,500 euros. This means that millionaires not only save considerably more than people in the lower wealth segments, but also that the relative importance of the two savings motives is different: while the amount saved out of the motives of saving for retirement and saving to accumulate wealth is similar in the lower three wealth segments. The motive of accumulating wealth clearly takes priority for millionaires, measured by the amount saved.

In sum, it can be said that the percentage of savers increases across the four wealth segments—irrespective of their savings motive—and that millionaires use considerably



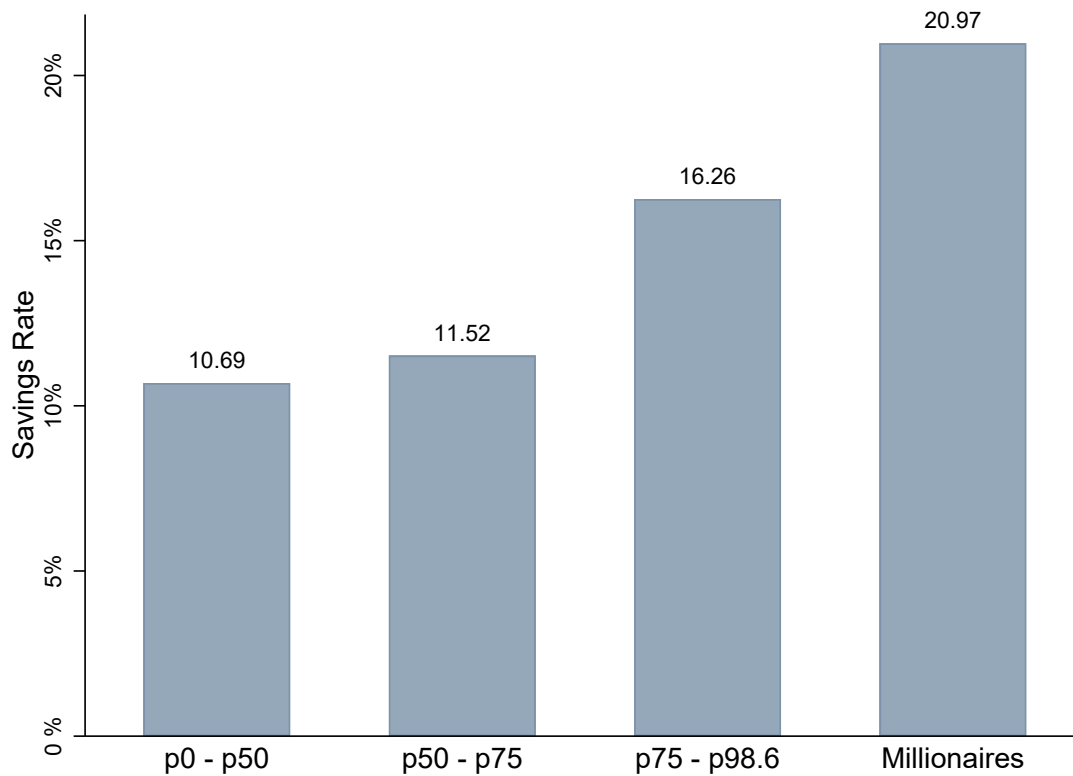
Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 21: Savings in SOEP+SOEP-P by wealth groups

more of their income for wealth accumulation, both in relative and absolute terms, than people in the lower three segments. In fact, the savings rate¹² as shown in Figure 22 rises from around 11% in the lower wealth segment to 12% and 16% in the middle and upper segments, respectively, and to almost 21% among millionaires. However, this also means that the higher the position in the wealth distribution, the higher the absolute amount saved, which means that wealth inequality is likely to increase further, *ceteris paribus*, in the medium term.

Looking at the relative importance of the two savings motives, the picture is mixed. A larger percentage of people save for retirement than to accumulate wealth. However, the amounts saved for retirement are lower than the amounts saved with the motive of wealth accumulation, which is particularly true for the group of millionaires.

¹²The savings ratio in an asset segment is defined as the weighted sum of the savings amounts from the pension and capital formation motive divided by the weighted sum of the net income in this segment.



Source: SOEP v35 and SOEP-P; own calculations with weighted values.

Figure 22: Savings rate in SOEP+SOEP-P by wealth groups

4.4.5 Portfolios

The chapter on millionaires in Germany concludes with a description of the average portfolio composition in the four wealth segments (Table 9).

The upper part of the table shows how individual gross wealth is distributed among the eight wealth components. The lower 50% have gross wealth averaging around 11,000 euros. More than 3,000 euros of this is in vehicles, approx. 2,700 euros in owner-occupied housing, approx. 1,600 euros in financial investments, and approx. 1,500 euros in private insurance policies. The gross wealth of the middle 25% is almost nine times as high at approx. 96,000 euros. The three most important wealth components here are owner-occupied housing at approx. 57,000 euros, followed by financial investments at approx. 13,000 euros, and private insurance policies at approx. 8,300 euros. The top 24% hold gross wealth of approx. 330,000 euros on average. Here, owner-occupied housing is again the most important wealth component at almost 190,000 euros. Other real estate is rising in importance and

Table 9: Wealth portfolio in SOEP+SOEP-P by wealth groups

	p0-p50	p50-p75	p75-p98.6	Millionaires
<i>Wealth</i>				
Owner-occupied housing	2,733	56,738	188,680	574,637
Other real estate	887	5,880	50,055	791,784
Financial investments	1,614	12,657	38,519	333,480
Private insurance	1,474	8,320	17,661	103,255
Building loan assets	789	3,994	5,946	10,337
Business assets	283	1,455	14,269	1,255,887
Tangible fixed assets	114	780	2,779	24,265
Vehicles	3,244	6,571	9,781	33,352
<i>Gross wealth</i>	11,139	96,395	327,690	3,126,997
<i>Debt</i>				
Mortgages on owner-occupied housing	-2,219	-18,598	-19,996	-29,641
Mortgages on other real estate	-897	-2,064	-8,642	-140,907
Consumer loans	-3,964	-1,731	-2,000	-14,567
Student loans	-378	-104	-47	-0.00031
Liabilities	-7,457	-22,496	-30,686	-185,115
<i>Net wealth</i>	3,682	73,899	297,004	2,941,882

Source: SOEP v35 and SOEP-P; own calculations with weighted values.

represents the second-largest component at approx. 50,000 euros. This is followed by financial investments at approx. 39,000 euros. Among millionaires, average gross wealth amounts to approx. 3 million euros. Business assets dominate here at approx. 1.3 million euros, followed by other real estate at approx. 790,000 euros and owner-occupied housing at approx. 580,000 euros. Among millionaires, therefore, investment forms aimed at generating income (income from renting and leasing and commercial operations) dominate.

The lower part of the table shows total debt and the four different liability components, while the last row shows the amount of net wealth. As debt (mainly consumer loans) is, on average, almost as high as the wealth in the lower 50%, the average net wealth is very low, at less than 4,000 euros. In the middle 25%, net wealth is around 74,000 euros. The average debt in this wealth segment adds up to around 22,000 euros, with mortgage loans accounting for the lion's share (around 19,000 euros for owner-occupied housing and a further 2,000 euros or so for other real estate). The top 24% have average net wealth of almost 300,000 euros and debt of 31,000 euros. Here, too, debt is mainly explained by mortgage loans (around 20,000 euros for owner-occupied housing and another approx.

9,000 euros for other real estate). The millionaires have net assets of about 3 million euros on average. They also have debt (approx. 180,000 euros), which are mainly due to mortgages on other real estate (approx. 140,000 euros).

5 Summary

The present report has two main focuses: First, it deals with the methodological question of whether a new sampling strategy - which was used here for the first time to draw a large sample (SOEP-P) - is capable of oversampling individuals with high private wealth and evaluating this new sample integrated with SOEP core data. Second, it deals with substantive questions: How high is the wealth concentration in Germany, and how is the wealth concentration developing? How does the inclusion of the SOEP-P data affect the measurement of the wealth concentration? How do millionaires differ from lower-wealth segments of the population in terms of sociodemographics, personality, life satisfaction, etc.?

As far as the methodological issues are concerned, this report explains the new sampling strategy and demonstrates that it works: Individuals who reside in Germany and have large shareholdings in companies are heavily concentrated in the upper quantiles of wealth distribution: 45% of the anchors in SOEP-P have individual net wealth of at least 1 million euros, compared to only 1% in SOEP. SOEP-P therefore makes a significant contribution to closing the data gap in SOEP in the high wealth range. The report also describes how a weighting framework for integrated analysis of the SOEP-P anchors and the SOEP population can be created, and how the wealth information in SOEP-P can be imputed in the case of item non-response.

As far as the substantive questions are concerned, the intertemporal development of the lower end of the wealth distribution since 2002 shows a high degree of stability over time. Wealth has remained constant at zero in the 25th percentile. In the middle of the distribution, wealth has shown a positive development: the median rose from 17,500 euros in 2002 to 22,000 euros in 2017. In the upper half, the changes are even greater. In the 75th (99th) percentile, wealth increased from about 100,000 euros (765,000 euros) in 2002 to about 124,000 euros (1.05 million euros) in 2017. The integration of SOEP-P in

the last year of the study leaves the picture unchanged in the lower half of the wealth distribution, but the wealth concentration increases: In the 75th (99th) percentile, it rises to around 126,000 euros (1.3 million euros). Overall, after the integration of SOEP-P, the top 1% share of total net wealth is around 29%, and thus at a similar level as reported in [Garbinti, Goupille-Lebret, and Piketty \(2020\)](#) for France or in the Pareto imputation study by [Vermeulen \(2018\)](#) for Germany. The Gini coefficient of net wealth also increases from 0.783 for SOEP to 0.81 for SOEP+SOEP-P.

The comparison of millionaires with the non-millionaire population shows that the two groups differ in other ways beyond wealth: millionaires tend to be male, have an above-average level of education, and are older than the rest of the population. They also have a far above-average equivalent household income, at over 7,600 euros, and save more than the average, which is why they accumulate additional wealth more quickly than the population of non-millionaires. At the same time, an above average percentage of millionaires live in West Germany and a below average percentage have a migration background. Millionaires who are employed are often self-employed or entrepreneurs, often hold management positions and/or work as managing directors or partners in a company. In terms of weekly working hours, millionaires who are employed work well above average at around 47 hours per week. The general life satisfaction of millionaires is also above average.

References

- Ackerman, P. L. and E. D. Heggstad (1997). “Intelligence, Personality, and Interests: Evidence for Overlapping Traits”. In: *Psychological Bulletin* 121.2, pp. 219–245.
- Alvaredo, A., A. B. Atkinson, and S. Morelli (2018). “Top Wealth Shares in the UK Over More Than a Century”. In: *Journal of Public Economics* 162, pp. 26–47.
- Anderson, C. and J. Cowan (2014). “Personality and Status Attainment: A Micropolitics Perspective”. In: *The Psychology of Social Status*. Springer, pp. 99–117.
- Andreoni, J. and A. A. Payne (2013). “Charitable Giving”. In: *Handbook of Public Economics*. Vol. 5. Elsevier, pp. 1–50.
- Atkinson, A. B., T. Piketty, and E. Saez (2011). “Top Incomes in the Long Run of History”. In: *Journal of Economic Literature* 49.1, pp. 3–71.
- Bach, S., A. Thiemann, and A. Zucco (2019). “Looking for the missing rich: Tracing the top tail of the wealth distribution”. In: *International Tax and Public Finance* 26.6, pp. 1234–1258.
- Baron, D. (1988). *Die personelle Vermögensverteilung in der Bundesrepublik Deutschland und ihre Bestimmungsgründe*. Peter Lang. Frankfurt.
- Blanchet, T., J. Fournier, and T. Piketty (Mar. 2017). *Generalized Pareto Curves : Theory and Applications*. Working Papers 201703. World Inequality Lab.
- BMAS (2017). “Analyse der Verteilung von Einkommen und Vermögen in Deutschland”. In: *Fünfter Armuts- und Reichtumsbericht der Bundesregierung*. Ed. by B. für Arbeit und Soziales.
- Bollinger, C. R. et al. (2019). “Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith, and Welch”. In: *Journal of Political Economy* 127.5, pp. 2143–2185.
- Bönke, T., M. M. Grabka, C. Schröder, and E. N. Wolff (2020). “A Head-to-Head Comparison of Augmented Wealth in Germany and the United States”. In: *The Scandinavian Journal of Economics* 122.3, pp. 1140–1180.
- Bönke, T., M. M. Grabka, C. Schröder, E. N. Wolff, and L. Zyska (2019). “The Joint Distribution of Net Worth and Pension Wealth in Germany”. In: *Review of Income and Wealth* 65.4, pp. 834–871.

- Bönke, T., C. Schröder, and K. Schulte (2010). “Incomes and Inequality in the Long Run: The Case of German Elderly”. In: *German Economic Review* 11.4, pp. 487–510.
- Breunig, C. (2019). “Testing Missing at Random Using Instrumental Variables”. In: *Journal of Business & Economic Statistics* 37.2, pp. 223–234.
- Bundesregierung (2019). *Monitoringbericht 2018 – Nachhaltigkeit konkret im Verwaltungshandeln umsetzen – Maßnahmenprogramm Nachhaltigkeit*.
- Cagetti, M. (2003). “Wealth Accumulation over the Life Cycle and Precautionary Savings”. In: *Journal of Business & Economic Statistics* 21.3, pp. 339–353.
- Caliendo, M., F. Fossen, and A. S. Kritikos (2014). “Personality Characteristics and the Decisions to Become and Stay Self-employed”. In: *Small Business Economics* 42.4, pp. 787–814.
- Carroll, C. D., B.-K. Rhee, and C. Rhee (1994). “Are there Cultural Effects on Saving? Some Cross-sectional Evidence”. In: *The Quarterly Journal of Economics* 109.3, pp. 685–699.
- Cook, K. W., C. A. Vance, and P. E. Spector (2000). “The Relation of Candidate Personality with Selection-Interview Outcomes”. In: *Journal of Applied Social Psychology* 30.4, pp. 867–885.
- Corneo, G. (2006). “Media Capture in a Democracy: The Role of Wealth Concentration”. In: *Journal of Public Economics* 90.1-2, pp. 37–58.
- De Nardi, M. and G. Fella (2017). “Saving and Wealth Inequality”. In: *Review of Economic Dynamics* 26, pp. 280–300.
- Dell, F. (2008). “L’Allemagne inégale: inégalités de revenus et de patrimoine en Allemagne, dynamique d’accumulation du capital et taxation de Bismarck à Schröder 1870-2005”. PhD thesis. Paris, EHESS.
- Di Zio, M. et al. (2004). “Bayesian Networks for Imputation”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 167.2, pp. 309–322.
- Dustmann, C. (1997). “Return Migration, Uncertainty and Precautionary Savings”. In: *Journal of Development Economics* 52.2, pp. 295–316.
- Easterlin, R. A. (2001). “Income and Happiness: Towards an Unified Theory”. In: *Economic Journal* 111.473, pp. 465–484.
- Frick, J., M. Grabka, and R. Hauser (2010). *Die Verteilung der Vermögen in Deutschland - Empirische Analysen für Personen und Haushalte*. Edition Sigma.

- Garbinti, B., J. Goupille-Lebret, and T. Piketty (2020). “Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France”. In: *Journal of the European Economic Association* 0.0, pp. 1–45.
- Glemser, A., S. Huber, and M. Rathje (2020). “SOEP-Core 2019: Report of Survey Methodology and Fieldwork”. In: *SOEP Survey Papers* 900.
- Goebel, J. et al. (2019). “The German Socio-Economic Panel (SOEP)”. In: *Jahrbücher für Nationalökonomie und Statistik* 239, pp. 345–360.
- Goebler, K. (2019). “A Novel Imputation Algorithm with Bayesian Networks”. In: *Mimeo*.
- Grabka, M. M., C. Halbmeier, et al. (2019). “Vermögensungleichheit in Deutschland bleibt trotz deutlich steigender Nettovermögen anhaltend hoch”. In: *DIW Wochenbericht* 86.40, pp. 735–745.
- Grijalva, E. et al. (2015). “Narcissism and leadership: A meta-analytic review of linear and nonlinear relationships”. In: *Personnel Psychology* 68.1, pp. 1–47.
- Holst, E. and K. Wrohlich (2019). “Frauenanteile in Aufsichtsräten großer Unternehmen in Deutschland auf gutem Weg: Vorstände bleiben Männerdomänen”. In: *DIW-Wochenbericht* 86.3, pp. 19–34.
- Hruschka, E. R., E. R. Hruschka, and N. F. Ebecken (2007). “Bayesian Networks for Imputation in Classification Problems”. In: *Journal of Intelligent Information Systems* 29.3, pp. 231–252.
- Judge, T. A., J. E. Bono, et al. (2002). “Personality and Leadership: A Qualitative and Quantitative Review”. In: *Journal of Applied Psychology* 87.4, pp. 765–780.
- Judge, T. A., C. A. Higgins, et al. (1999). “The Big Five Personality Traits, General Mental Ability, and Career Success across the Life Span”. In: *Personnel Psychology* 52.3, pp. 621–652.
- Judge, T. A. and R. Ilies (2002). “Relationship of Personality to Performance Motivation: A Meta-analytic Review”. In: *Journal of Applied Psychology* 87.4, pp. 797–807.
- König, J., C. Schröder, and E. N. Wolff (2020). “Wealth Inequalities”. In: *Handbook of Labor, Human Resources and Population Economics*. Ed. by K. F. Zimmermann. Springer International Publishing, pp. 1–38.

- Lauterbach, W. et al. (2015). “HViD – Hochvermögende in Deutschland Abschlussbericht zu den Ergebnissen der Befragung (2014).” In: *Expertise für das Bundesministerium für Arbeit und Soziales (BMAS)*.
- Leckelt, M. et al. (2019). “The Rich are Different: Unravelling the Perceived and Self-reported Personality Profiles of High-net-worth Individuals”. In: *British Journal of Psychology* 110.4, pp. 769–789.
- Lillard, L., J. P. Smith, and F. Welch (1986). “What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation”. In: *Journal of Political Economy* 94.3, pp. 489–506.
- Little, R. J. (1988). “A Test of Missing Completely at Random for Multivariate Data with Missing Values”. In: *Journal of the American Statistical Association* 83.404, pp. 1198–1202.
- Lundberg, J. and D. Waldenström (2018). “Wealth Inequality in Sweden: What can we Learn from Capitalized Income Tax Data?” In: *Review of Income and Wealth* 64.3, pp. 517–541.
- Merkle, L. and K. F. Zimmermann (1992). “Savings, Remittances, and Return Migration”. In: *Economics Letters* 38.1, pp. 77–81.
- Ng, T. W. et al. (2005). “Predictors of Objective and Subjective Career Success: A Meta-analysis”. In: *Personnel Psychology* 58.2, pp. 367–408.
- Niloofer, P. and M. Ganjali (2014). “A New Multivariate Imputation Method Based on Bayesian Networks”. In: *Journal of Applied Statistics* 41.3, pp. 501–518.
- Niloofer, P., M. Ganjali, and M. F. Rohani (2013). “Performance Evaluation of Imputation Based on Bayesian Networks”. In: *Sankhya B* 75.1, pp. 90–111.
- Ozer, D. J. and V. Benet-Martinez (2006). “Personality and the Prediction of Consequential Outcomes”. In: *Annual Review of Psychology* 57, pp. 401–421.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Pen, J. (1971). *Income Distribution: Facts, Theories, Policies*. Praeger.
- Porpiglia, A., E. Sierminska, and T. Y. Mathä (2011). “The Immigrant/Native Wealth Gap in Germany, Italy and Luxembourg”. In: *ECB Working Paper* 1302.

- Rancoita, P. M. et al. (2016). “Bayesian Network Data Imputation with Application to Survival Tree Analysis”. In: *Computational Statistics & Data Analysis* 93, pp. 373–387.
- Raub, B., B. Johnson, and J. Newcomb (2010). “A Comparison of Wealth Estimates for America’s Wealthiest descendants using Tax data and Data from the Forbes 400”. In: *National Tax Association Proceedings, 103rd Annual Conference on Taxation*, pp. 128–135.
- Roberts, B. W. et al. (2007). “The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes”. In: *Perspectives on Psychological Science* 2.4, pp. 313–345.
- Roine, J. and D. Waldenström (2009). “Wealth Concentration over the Path of Development: Sweden, 1873-2006”. In: *The Scandinavian Journal of Economics* 111.1, pp. 151–187.
- Romero, V. and A. Salmerón (2004). “Multivariate Imputation of Qualitative Missing Data using Bayesian Networks”. In: *Soft Methodology and Random Information Systems*. Springer, pp. 605–612.
- Rubin, D. B. (1976). “Inference and Missing Data”. In: *Biometrika* 63.3, pp. 581–592.
- (1987). *Multiple Imputation for Nonresponse in Surveys*. New York [u. a.]: John Wiley & Sons, Inc.
- (1996). “Multiple imputation after 18+ years”. In: *Journal of the American statistical Association* 91.434, pp. 473–489.
- Saez, E. and G. Zucman (2016). “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”. In: *The Quarterly Journal of Economics* 131.2, pp. 519–578.
- Schröder, C., C. Bartels, M. M. Grabka, J. König, et al. (2019). “A Novel Sampling Strategy for Surveying High-Worth Individuals - An Application Using the Socio-Economic Panel”. In: *Review of Income and Wealth* forthcoming.
- Schröder, C., C. Bartels, M. M. Grabka, M. Kroh, et al. (2018). “Machbarkeitsstudie zur Verbesserung der Forschungsdateninfrastruktur im Bereich Hochvermögender mit dem Sozio-oekonomischen Panel (SOEP)”. In: *SOEP Working Paper Series* 981.
- Scutari, M., C. Vitolo, and A. Tucker (Sept. 2019). “Learning Bayesian Networks from Big Data with Greedy Search: Computational Complexity and Efficient Implementation”. In: *Statistics and Computing* 29.5, pp. 1095–1108.

- Ströing, M., M. M. Grabka, and W. Lauterbach (2016). “Hochvermögende in Deutschland”. In: *DIW-Wochenbericht* 42.
- Vermeulen, P. (2018). “How Fat is the Top Tail of the Wealth Distribution?” In: *Review of Income and Wealth* 64.2, pp. 357–387.
- Westermeier, C. and M. Grabka (2015). “Große statistische Unsicherheit beim Anteil der Top-Vermögenden in Deutschland”. In: *DIW Wochenbericht* 82.7, pp. 123–133.
- Wolff, E. N. (1981). “The Accumulation of Household Wealth over the Life-cycle: a Microdata Analysis”. In: *Review of Income and Wealth* 27.1, pp. 75–96.
- (2017). “Household Wealth Trends in the United States, 1962 to 2016: Has Middle Class Wealth Recovered?” In: *NBER WP Series*.

List of Figures

1	Data gap in the area of high wealth in the SOEP	2
2	Distribution of household types in SOEP and SOEP-P	20
3	The Big Five in SOEP and SOEP-P	22
4	Risk preferences in SOEP and SOEP-P	23
5	Proportion of people in SOEP and SOEP-P who are saving for precautionary or wealth accumulation motives	24
6	Pen's parades of equivalent net household income	25
7	Savings for precautionary and wealth accumulation motives along the deciles of equivalent net household income	26
8	Wealth portfolio in SOEP	27
9	Wealth portfolios in SOEP-P	28
10	Self-report data on factors influencing wealth formation in SOEP-P	29
11	Data gap in the area of high wealth in SOEP+SOEP-P	32
12	Portfolios in SOEP+SOEP-P by wealth groups	34
13	The Big Five in SOEP+SOEP-P by wealth groups	45
14	Willingness to take risks in SOEP+SOEP-P by wealth groups	46
15	Satisfaction in selected areas of life by wealth groups	47
16	Satisfaction in further selected areas by wealth groups	48
17	Living area in square meters by wealth groups	49
18	Weekly working hours among the working population by wealth groups	50
19	Equivalent monthly household net household income in SOEP+SOEP-P by wealth groups	51
20	Percentage of savers by wealth groups in SOEP+SOEP-P	52
21	Savings in SOEP+SOEP-P by wealth groups	53
22	Savings rate in SOEP+SOEP-P by wealth groups	54
23	Number of registered enterprises by legal form	68
24	Shares of enterprises without revenues in the last five years	69
25	Number of registered companies by revenues	70
26	Percentage share of companies contained in each database by federal state	71

List of Tables

1	Studies on wealth inequality in Germany	7
2	Comparison of sociodemographic characteristics in SOEP and SOEP-P	18
3	Savings for precautionary or wealth accumulation motives and net household income	25
4	Distribution of net wealth in SOEP and SOEP-P	30
5	SOEP net wealth distributions after integration of SOEP-P	35
6	Total wealth shares in SOEP with SOEP-P	36
7	Time series of net wealth with and without integration of SOEP-P	38
8	Sociodemographic characteristics by wealth groups	41
9	Wealth portfolio in SOEP+SOEP-P by wealth groups	55
10	SOEP net wealth distributions after integration of SOEP-P and <i>Manager Magazin</i>	66

A Distribution of net wealth with *Manager Magazin*

Table 10: SOEP net wealth distributions after integration of SOEP-P and *Manager Magazin*

	SOEP			SOEP+SOEP-P			SOEP+SOEP-P+MM		
	lower bound	estimate	upper bound	lower bound	estimate	upper bound	lower bound	estimate	upper bound
p1	-25,000	-22,260	-20,000	-25,000	-22,260	-20,000	-25,300	-22,260	-20,000
p5	-3,900	-3,000	-2,060	-3,600	-3,000	-2,000	-3,759	-3,000	-2,000
p10	0	0	0	0	0	0	0	0	0
p25	0	0	0	0	0	0	0	0	0
p50	20,000	22,000	24,000	20,700	22,800	25,000	20,700	22,800	25,000
p75	120,000	123,620	127,100	122,500	126,000	130,200	122,500	126,000	130,260
p90	256,000	265,000	276,031	266,000	279,236	287,500	267,789	279,236	287,700
p95	395,000	407,700	426,000	417,400	438,000	457,500	419,500	438,000	455,000
p99	920,000	1,050,000	1,200,000	1,171,367	1,330,000	1,430,000	1,196,668	1,333,000	1,431,845
p99.9	3,108,000	4,019,000	4,418,795	4,664,803	5,490,000	8,011,008	4,760,000	5,650,000	8,011,008
p75/p50	5.24	5.62	6.17	5.13	5.53	5.96	5.15	5.53	5.98
p90/p50	11.19	12.05	13.35	11.22	12.25	13.24	11.24	12.25	13.25
p99.9/p50	136.30	182.68	212.93	201.84	240.79	349.68	205.42	247.81	364.14
Gini	0.769	0.783	0.796	0.798	0.809	0.820	0.816	0.826	0.836
Gini*	0.750	0.761	0.773	0.778	0.789	0.800	0.798	0.807	0.816
Theil*	1.198	1.316	1.463	1.490	1.627	1.798	2.011	2.178	2.325
Mean	101,903	107,649	114,727	120,190	126,694	135,015	132,009	139,007	146,666

Source: SOEP v35, SOEP-P and MM; own calculations with weighted values. Bootstrap confidence intervals with 500 replications given as lower and upper bounds.

* All negative and zero assets are recoded to 0.01 as the partial index is only defined for strictly positive values. The Gini is also adjusted accordingly for comparison purposes.

B Additional information on sampling

B.1 Monetization of percentage shareholdings

Two pieces of information are needed to estimate the value of the shares held by a shareholder in a company: the percentage amount of the shareholding, and the value of the company. While the percentage shareholding is available from the database, the value of the company can only be determined directly for listed companies, namely by way of market capitalization. However, the majority of companies are not listed on the stock exchange. A uniform valuation procedure is therefore required for all companies.

The proxy for the company value is revenue, as revenue is available in the data for a very large number of companies and are highly positively correlated with the value of a company. In addition, an exact valuation in euros is not necessary because the investment values are only used to sort the shareholders and to determine the top 1% of shareholders. As revenues are not available for all companies, missing values were imputed (for details see [Schröder, Bartels, Grabka, König, et al., 2019](#)).

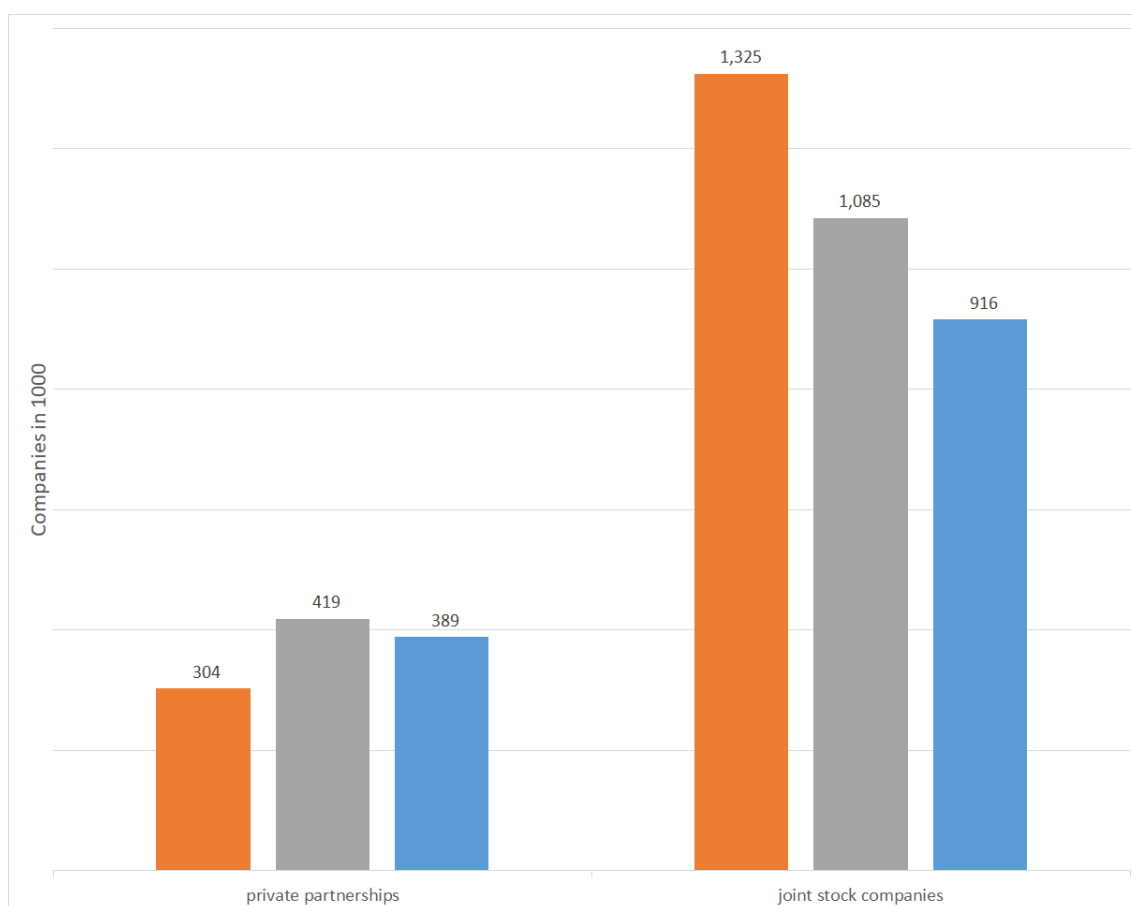
Monetization involves assessing the value not only of direct shareholdings, but also any shares in subsidiaries. In the Orbis database, this means that a distinction must be drawn between the value of direct shareholdings in companies (Shareholder – Direct) and their total value including indirect shareholdings (Shareholder – Total). This distinction is important because financial statements are partly consolidated and partly non-consolidated (Consolidation code). Consolidated means that the balance sheet of the parent company shows the values of the subsidiaries. If the financial statements are consolidated, the percentage of direct shareholding and the enterprise value from the consolidated balance sheet are sufficient to determine the value of the shares. If the financial statements are not consolidated, the value of the shares is the direct shareholding in the parent company (multiplied by the non-consolidated value of the parent company) plus the percentage values of the shares in the subsidiaries.

The share values are then totaled across all companies worldwide in which a shareholder has shares. This is the cumulative shareholding value.

B.2 Representation of companies in Orbis

In order to assess the representation of companies in Germany in Orbis, we first compare the case numbers in Orbis with the trade tax statistics and the business register.

The case numbers for partnerships and joint stock companies according to Orbis is approximately the same as according to the trade tax statistics and the business register. Figure 23 shows that Orbis lists around 300,000 partnerships, whereas the business register lists around 390,000 and the trade tax statistics around 420,000. The lower number of cases in the Orbis database can be explained in particular by the absence of “unlimited partnerships” (*Gesellschaften bürgerlichen Rechts*), which are only listed in the trade register.



Note: Own calculations based on ORBIS (orange), trade tax statistics (grey), 2013 and business register (blue), 2015.

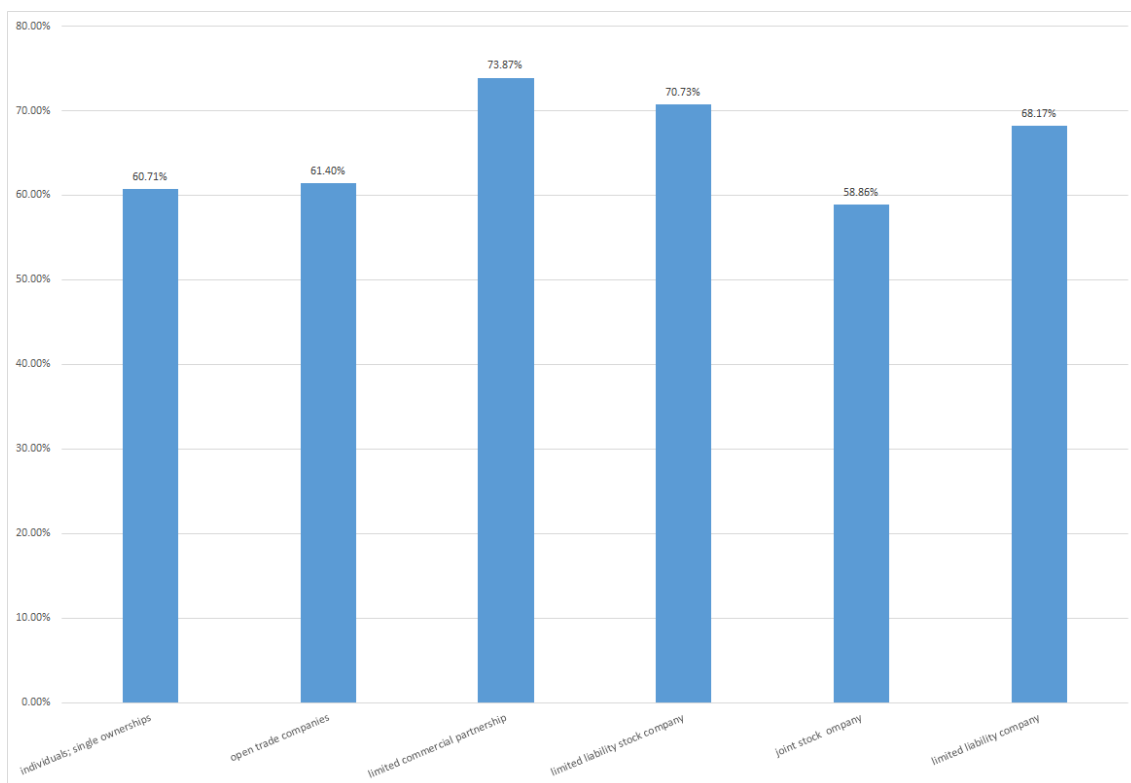
Figure 23: Number of registered enterprises by legal form

Orbis contains around 1.3 million joint stock companies, the business register around 900,000, and the trade tax statistics around 1.1 million. The higher number of joint stock

companies in Orbis compared to the business register is explained by the fact that the register does not include companies with revenues of less than 17,500 euros. Although private limited asset-management companies (*GmbHs*) are joint stock companies, they may be exempt from trade tax, which is another explanation for the different number of cases in Orbis and the trade tax statistics.

The number of small enterprises in the form of sole proprietorships or natural persons is significantly smaller in Orbis than in the other two sources. While Orbis only lists about 140,000 sole proprietorships or natural persons, the business register and trade tax statistics list about 2 million. This undercoverage is not problematic for our project, as it is unlikely that high net worth individuals will hold shares in small businesses only.

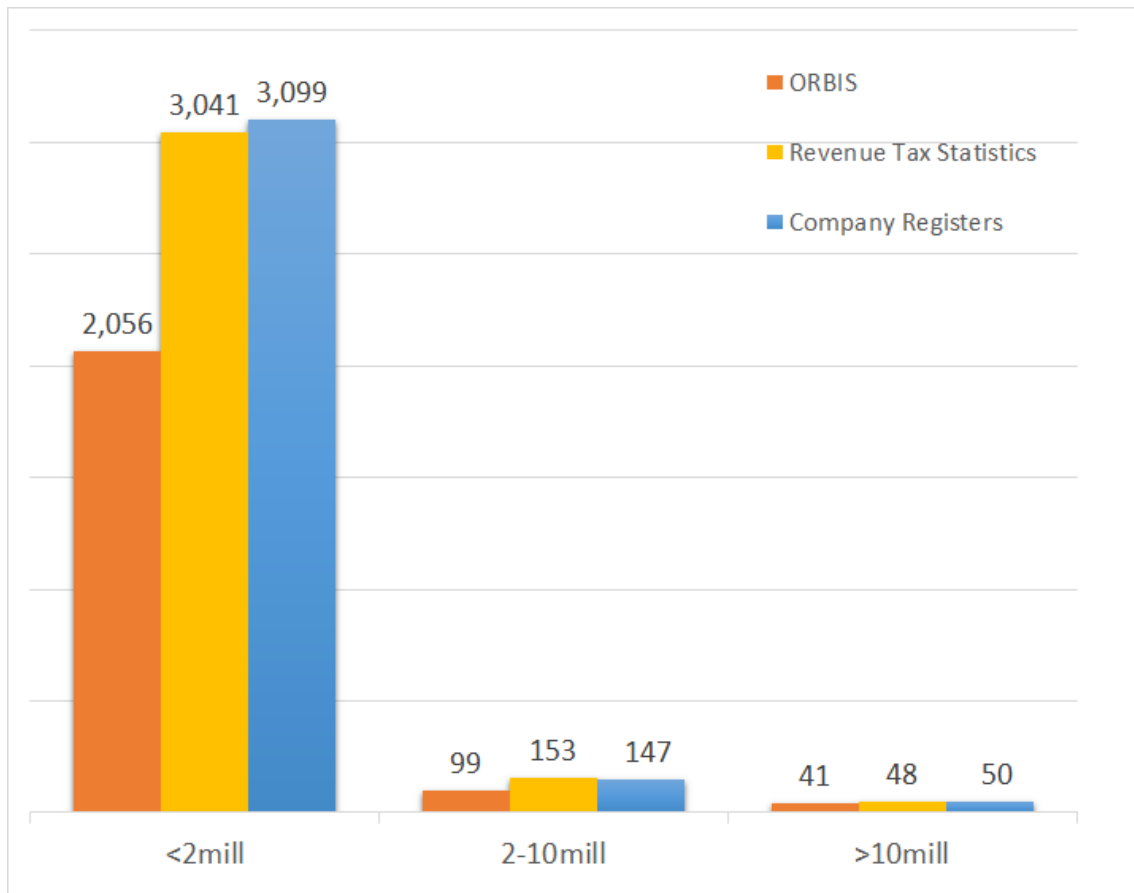
Due to reduced disclosure requirements for certain types of companies, it is not surprising that revenues are only directly available for about 30 to 40 percent of the companies in Orbis. Figure 24 shows the shares of companies active in 2017 without data on revenues. It is also possible that companies did not submit their balance sheets to the Federal Gazette in time.



Note: Proportion of enterprises without revenue information in all enterprises active in 2017. Own calculations based on Orbis.

Figure 24: Shares of enterprises without revenues in the last five years

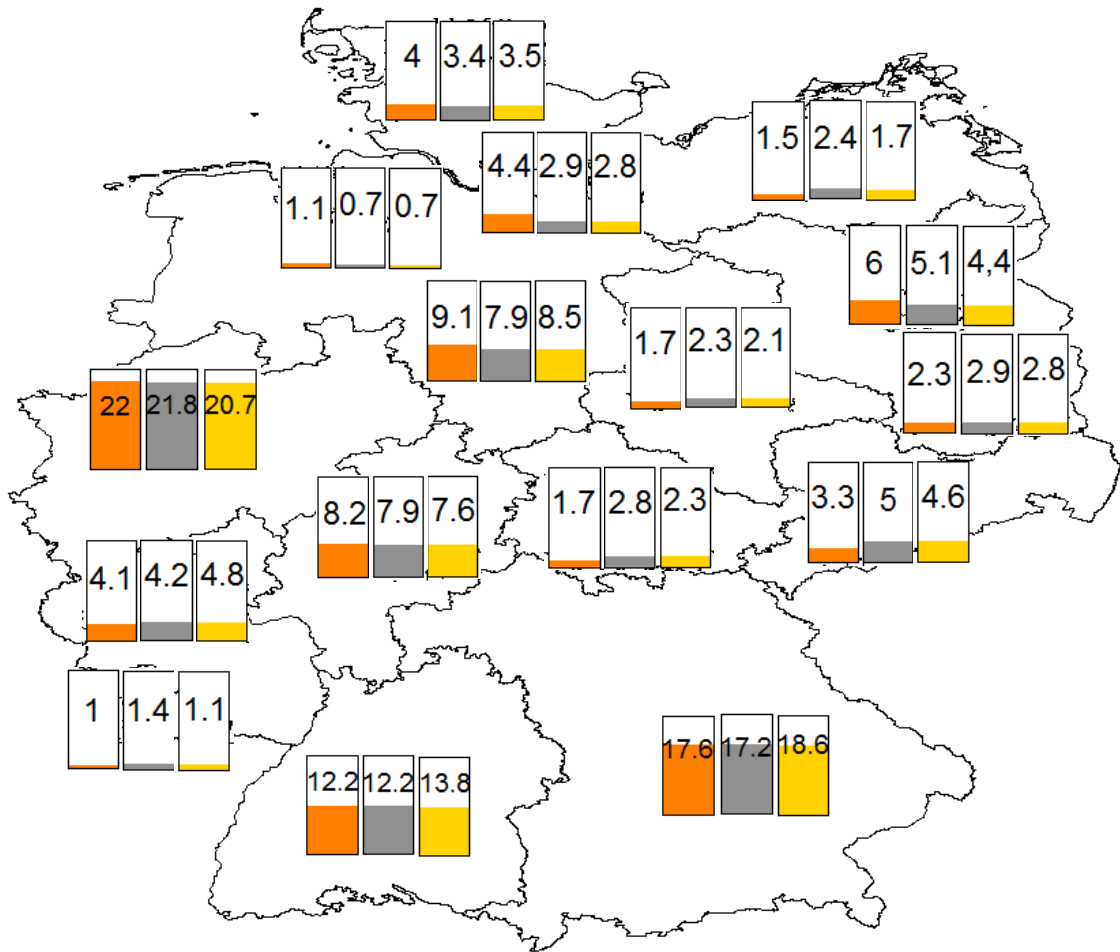
Where revenues were missing, we imputed them statistically for the following evaluations (see Chapter B.1). Based on this augmented database, a comparison with external statistics was possible. Figure 25 shows high levels of agreement. Orbis and the business register contain similar case numbers of enterprises with revenues between 2 and 10 million. The agreement is even higher for revenues of 10 million euros or more. The category below 2 million euros in revenues is systematically under-recorded. Again, the lower number of small businesses in the Orbis database is an explanation.



Note: Own calculations based on Orbis (orange), sales tax statistics (advance notifications, yellow), 2013 and business register (blue), 2015.

Figure 25: Number of registered companies by revenues

The regional distribution of companies according to Orbis and the trade tax statistics is very similar (Figure 26). The majority of the companies (approx. 22%) are located in North Rhine-Westphalia, followed by Bavaria (approx. 17%) and Baden-Württemberg (approx. 12%).



Note: Own calculations based on Orbis (orange), trade tax statistics (grey), 2013 and sales tax statistics (advance returns, yellow), 2015.

Figure 26: Percentage share of companies contained in each database by federal state