EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Robin Chakraborty, Sofie R. Waltl

Missing the wealthy in the HFCS: micro problems with macro implications

No 2163 / June 2018

Household Finance and Consumption Network (HFCN)

This paper contains research conducted within the Household Finance and Consumption Network (HFCN). The HFCN consists of survey specialists, statisticians and economists from the ECB, the national central banks of the Eurosystem and a number of national statistical institutes.

The HFCN is chaired by Ioannis Ganoulis (ECB) and Oreste Tristani (ECB). Michael Haliassos (Goethe University Frankfurt), Tullio Jappelli (University of Naples Federico II) and Arthur Kennickell act as external consultants, and Sébastien Pérez-Duarte (ECB) and Jiri Slacalek (ECB) as Secretaries.

The HFCN collects household-level data on households' finances and consumption in the euro area through a harmonised survey. The HFCN aims at studying in depth the micro-level structural information on euro area households' assets and liabilities. The objectives of the network are:

1) understanding economic behaviour of individual households, developments in aggregate variables and the interactions between the two;
2) evaluating the impact of shocks, policies and institutional changes on household portfolios and other variables;
3) understanding the implications of heterogeneity for aggregate variables;
4) estimating choices of different households and their reaction to economic shocks;
5) building and calibrating realistic economic models incorporating heterogeneous agents;
6) gaining insights into issues such as monetary policy transmission and financial stability.

The refereeing process of this paper has been co-ordinated by a team composed of Pirmin Fessler (Oesterreichische Nationalbank), Michael Haliassos (Goethe University Frankfurt), Tullio Jappelli (University of Naples Federico II), Sébastien Pérez-Duarte (ECB), Jiri Slacalek (ECB), Federica Teppa (De Nederlandsche Bank), Oreste Tristani (ECB) and Philip Vermeulen (ECB).

The paper is released in order to make the results of HFCN research generally available, in preliminary form, to encourage comments and suggestions prior to final publication. The views expressed in the paper are the author's own and do not necessarily reflect those of the ESCB.

Abstract

Macroeconomic aggregates on households' wealth have a long tradition and are widely used to analyse and compare economies, yet they do not provide any information about the distribution of assets and liabilities within the population. The *Household Finance and Consumption Survey* (HFCS) constitutes a rich source of micro data that can be used to link macro aggregates with distributional information to compile *Distributional National Accounts* for wealth. Computing aggregates from this survey usually yields much lower amounts than what is reported by macroeconomic statistics. An important source of this gap may be the lack of the wealthiest households in the HFCS. This article combines a semi-parametric Pareto model estimated from survey data and observations from rich lists with a stratification approach making use of HFCS portfolio structures to quantify the impact of the missing wealthy households on instrument-specific gaps between micro and macro data. We analyse data for Austria and Germany, and find that adjusting for the *missing wealthy* pushes up inequality even further, increases instrument-specific aggregates, has large effects on equity, but explains less than ten percentage points of the micro-macro gap for most other instruments. Additionally, we document that some countries' lack of an oversampling strategy for wealthy households limits the cross-country comparability of wealth inequality statistics.

Keywords: Distributional National Accounts (DINA), National Accounts, HFCS, Micro-Macro Comparison, Semi-parametric Pareto Model for Wealth, Wealth distribution

JEL codes: D31, E01

1 Non-technical summary

Analysing, assessing and comparing economies is usually done by means of macroeconomic aggregates or indicators derived therefrom. These indicators, however, do not tell anything about the distribution of income, consumption, saving and wealth within the population although distributional information is needed to thoroughly design and assess the impact of policies (including monetary policy), for research purposes, as a source of information for the greater public and, generally, to assess an economy in a comprehensive way.

The need for distributional information for the household sector was also pointed out in the *Stiglitz-Sen-Fitoussi report* and by the *G20 Data Gaps Initiative*.[1] The joint *OECD-Eurostat Expert Group on Measuring Disparities in a National Accounts Framework* focuses on distributional indicators for income and consumption, and the *ECB Expert Group on Linking Macro and Micro Data for the Household Sector* (EG-LMM) works on linking micro data obtained from the *Household Finance and Consumption Survey* (HFCS) with macro data from the financial/national accounts to derive distributional indicators for wealth. The article at hand emerged from work for the EG-LMM.

All these initiatives should eventually lead to complementing and breaking-down wealth or components of wealth as reported in the national accounts by homogeneous household groups such as income or wealth groups (e.g., from the poorest 20% to the wealthiest 20%), or types of households determined by the number and age of household members. These *Distributional National Accounts* (DINA) should ideally be compiled in a standardised and harmonised way enabling comparability across countries and time. We believe that distributional figures should be integrated within the national accounts framework, which means that eventually distributional split-ups should sum up to national accounts aggregates. This integration is needed to avoid confusion among users, linking distributional data with other macro-economic indicators such as GDP growth, and enable a consistent and comprehensive discussion about distributional phenomena. We believe that simply having aggregate and distributional figures at the same place could increase the utilization of distributional statistics as users and policy makers are “nudged” towards taking them into consideration.

For the purpose of integration, the comparability of the definitions of instruments in national accounts and the micro source used needs to be assessed and should be aligned to the extent possible. The EG-LMM has worked out a bridging table that classifies HFCS instruments as of high, medium and low conceptual comparability with the financial accounts (EG-LMM, 2017).

When calculating aggregates from the HFCS, these aggregates are usually much lower than national accounts figures even for conceptually highly comparable instruments. A current focus therefore lies on investigating the reasons causing this macro-micro gap. There are many reasons why we would expect differences in aggregates and one of them are what is called here the *missing wealthy*.

The HFCS, just like any other wealth-related survey, does not perfectly reflect the very top of the wealth distribution, as the participation of very wealthy households in such surveys is less likely. Due to the high concentration of wealth at the very top, particularly good information from this part of the distribution would, however, be needed.

The aim of this article is twofold: First, we quantify the contribution of the missing wealthy toward the macro-micro gap, i.e., we measure the increase in coverage ratios (defined as HFCS aggregates over national accounts aggregates) after adjusting for the missing wealthy. The methodology can thus be used to test whether the *missing wealthy* constitute a serious problem

[1] See section 2 for further details and references.

for a particular instrument, country or point in time. Second, we show how such quantification could be used in the compilation of distributional national account figures. We perform a case study for Austria and Germany.

As the HFCS lacks observations from the very top, we substitute the top tail of the empirical wealth distribution implied by the HFCS by a theoretical model – namely a Pareto model jointly estimated from top HFCS and rich list data. In a second step, we break down the Pareto-inflated wealth distribution by instruments (various asset classes and liabilities).

Relying on rich list data is only a second-best strategy, as these lists are compiled by newspapers based on opaque methodology. To decrease the dependency on such less trustworthy data sources and increase the quality of distributional national accounts figures, collecting more and making available existing administrative wealth-related data for statistical purposes would be essential.

Against prior believes, the missing wealthy do not explain large parts of the macro-micro gap for highly comparable instruments (liabilities, bonds, deposits and mutual funds). For most instruments, less than 10 percentage points of this gap can be attributed to the missing wealthy still leaving significant parts (ranging between 32 and 87 percentage points) unexplained. This suggests that there is not *that one single source* explaining the observed under-coverage for these instruments but rather hints toward a longer list of small – but in sum important – reasons.

In contrast to highly comparable instruments, we find that micro data based on aggregates for equity (and to a lesser extend also for real estate) increase heavily when adjusting for the missing wealthy. Also without this adjustment, HFCS totals for equity usually exceed what is observed in the financial accounts. Our methodology pushes up these aggregates even further. The large increase in equity is not surprising as the wealthiest of the wealthy hold great parts of their total wealth in the form of company shares or other forms of equity. Other financial instruments such as mutual funds, bonds or deposits are much less important (in relative terms) for this part of the population.

While a large jump in total equity is clearly expected and any other result would decrease confidence in our methodology, this rise however also points toward a fundamental problem of the macro-micro linkage: the low conceptual comparability of equity as defined in the financial accounts and the HFCS definition of business wealth. Both, the definition of what is intended to be measured and the valuation concepts diverge. As equity forms a large and important part of total wealth, better aligning definitions must therefore be a top priority for future work. Improvements on both sides, the HFCS and the financial accounts, are needed to come up with high-quality and reliable distributional statistics in the future.

2 Introduction

While general macroeconomic statistics such as national accounts show the evolution of overall household wealth and income, they are unable to reveal information on the distribution of totals within an economy, and on how the distribution evolves over time.

In recent years, there has been renewed interest in wealth distributions – and in particular in wealth inequality – among the greater public as well as academics, politicians, and policy makers. Thomas Piketty's *Capital in the Twenty-First Century* has contributed largely to these flourishing discussions (Piketty, 2014). Piketty and Zucman (2014) find increasing wealth-to-income ratios for the top eight developed economies in recent decades and conclude that because of this

> [...] the inequality of wealth, and potentially the inequality of inherited wealth, is likely to play a bigger role for the overall structure of inequality in the twenty-first century than it did in the postwar period. (Piketty and Zucman, 2014, page 1261)

The availability of high-quality data on wealth distributions constitues the first step towards comprehensive analyses and policy action. Since the pioneering work by Kuznets (1955) and Atkinson and Harrison (1978), however, the topic of *systematically* collecting distributional information on wealth (and income) as an addition to macro aggregates has been evoked only recently: The *Commission on the Measurement of Economic Performance and Social Progress* emphasises in the "Stiglitz-Sen-Fitoussi report" the need to give more prominence to the distribution of income, consumption and wealth.[2] This need is also expressed in the *G20 Data Gaps Initiative*.[3] Since 2011 the *World Wealth and Income Database* aims to provide comprehensive and long data series on distributional indicators on income and wealth for a large number of countries around the globe.

Collecting distributional data may ultimately lead to *Distributional National Accounts* (DINA). The aim of DINA is to have distributional estimates available that are consistent with macroeconomic statistics reported in the national accounts (Fesseau and Mattonetti, 2013; Alvaredo et al., 2016; Piketty et al., 2016). Piketty et al. (2016) name the gap to national accounts as the most important limitation that needs to be overcome to rigorously measure income inequality and thus call for DINA. Due to better data availability there is already substantial progress with regard to income, whereas wealth-related indicators stand at the very beginning.

While it is well known that wealth is highly concentrated at the top, recent data show that top income and wealth shares – and in particular wealth-income ratios – have even risen in many developed and developing countries in recent decades (Alvaredo et al., 2017). However, extensive administrative data that would enable a sound analysis of these developments are hardly available in many European countries including Austria and Germany: Limited prevelance of wealth-related taxes[4] disincentivises data collection, and data protection and privacy concerns limit the use of existing administrative data – even by public institutions.

As administrative data is often not available, wealth surveys constitute a distinct source of information leading towards the compilation of DINA. Unfortunately, macro and micro data are not, and cannot always be, completely aligned and definitions as well as concepts differ in

[2] Stiglitz-Sen-Fitoussi report, recommendation 4, available at `http://www.stiglitz-sen-fitoussi.fr`.

[3] Recommendation 16. Report available at: `http://www.imf.org/external/np/g20/pdf/102909.pdf`.

[4] Wöhlbier (2014) reports that "[m]ost net-wealth taxes were removed or scaled down by [EU] Member States between 1995 and 2007."

some instances.[5]

While surveys have the advantage of a long list of variables (including a comprehensive list of socio-economic characteristics of the houshold) compared to the scarcely available administrative data, voluntary surveys suffer from several kinds of reporting errors including a lack of observations from the wealthiest households.[6] Due to the high concentration of wealth at the very top (see Davies and Shorrocks, 2000), even more observations from this part of the distribution would be necessary to ensure an acceptable degree of precision. However, surveys and particularly wealth-related surveys usually fail to capture these households appropriately as they are harder to be contacted and to get hold of.

There is evidence from the Spanish (Bover et al., 2014, Table 5) and US (Kennickell and Woodburn, 1999; Kennickell, 1999a) household wealth surveys that response rates decline with increasing wealth leading to a high unit-non-response rate in this part of the distribution. Such systematic non-response is likely to introduce a unit-non-response bias.[7]

Osier (2016) analyses a potential unit-non-response bias for the first wave of the *Household Finance and Consumption Survey* (HFCS). He finds that "[u]nit non-response is a key concern for the HFCS, as non-response rates are important in some countries and response patterns are not completely random" (Osier, 2016, page 3).

In general, for Austria and Germany a quantification of unit-non-response bias is problematic as no external household level data on wealth can be used for such an analysis. However, "[i]n a wealth survey, the sensitivity of the subject and the time cost of being interviewed, for people with complex assets, should be enough to raise *a priori* concerns" (Kennickell, 2008, page 405).

This underrepresentation of the wealhtiest hoseuholds poses a problem when aggregates of wealth surveys are compared to macro statistics. Aggregated micro data are usually lower than macro statistics as shown in Figure 1 and one important reason are "missing wealthy households"[8] (see also Avery and Elliehausen, 1986; Avery et al., 1988) which contribute to the micro-macro gap through two channels: unit-non-response as well as a lack of precision due to too few observations at the very top. While strategically oversampling the wealthiest households in a survey aims to reduce the latter problem, in principal it is not well suited to correct for unit-non-response bias.

There have been recent attempts to analyse the impact of underrepresentation of wealthy households on the wealth distribution (Bach et al., 2014; Eckerstorfer et al., 2016; Vermeulen,

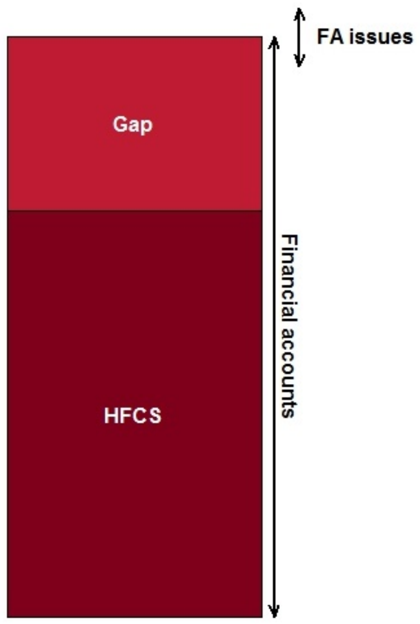
[5] See also Kavonius and Törmälehto (2010); Kavonius and Honkkila (2013); Henriques and Hsu (2014); Andreasch and Lindner (2016); Baranyai-Csirmaz et al. (2017); EG-LMM (2017).

[6] Cowell and van Kerm (2015) survey measurement issues related to wealth distributions and wealth inequality estimated from household surveys.

[7] *Unit-non-response bias* refers to the bias introduced when certain households systematically do not participate or participate less often in a survey. It has to be noted, though, that – when the non-participation is random – low response rates must not necessarily lead to strong bias and vice versa.

[8] Other sources of this gap include conceptual discrepancies (differences in the definition of instruments, different valuation methods), population discrepancies (different scopes of the survey and financial accounts), reporting errors (intentional under- or over-reporting and related behavioural effects, reporting errors due to a lack of knowledge; see D'Aurizio et al. 2006), rounding errors (rounding of amounts by respondents and additional rounding for the purpose of anonymisation), and sampling errors (errors in the sample design). Althoguh financial accounts are supposed to be exhaustive, they are no perfect benchmark due to, for instance, balancing across accounts, difficulties in the valuation of unquoted assets and potential coverage problems of wealth held by residents abroad (see Zucman, 2013). See EG-LMM (2017) for more details regarding population and conceptual discrepancies.

Figure 1: Discrepancies between FA and HFCS aggregates.



Notes: The figure illustrates the most common case for conceptually well comparable instruments: under-coverage, i.e., micro (HFCS) are lower than macro (financial accounts) aggregates.

2016).[9] This is usually done by replacing observed wealth of the richest households in the survey (also referred to as the tail population) by a parametric model – such as a Pareto model – yielding a semi-parametric approach.[10]

When estimating a Pareto distribution, observations from rich lists (such as for instant the Forbes World's billionaires list) may be used to complement the survey. Rich lists add observations at the very top of the distribution, where the survey generally does not provide any information, and thus enhance the reliability of the Pareto model.

The goal of this article is to quantify the contribution of the "missing wealthy" on the gap between macro and micro data for liabilities and a list of asset classes. Thereby, the term "missing wealthy" refers to both issues related to wealthy households: unit-non-response bias and a lack of precision at the top tail.

We propose two complementary approaches to achieve this goal: the analytical approach yields point estimates and a simulation approach complements them with confidence intervals and allows to calculate adjusted Gini coefficients.

We perform a case study for Austria and Germany using the second wave of the HFCS released in December 2016. We chose these countries as they lack additional administrative micro data on wealth to perform strategic oversampling of the wealthy, and correct and cross-check self-

[9]Similar problems arise when estimating the income distribution from survey data. See, for instance, Törmälehto (2017) for details regarding the *European Union Statistics on Income and Living Conditions (EU-SILC)*.

[10]Another way of addressing this issue is to create artificial observations representing the wealthiest households and add them to the survey with corresponding weights. Such an approach is followed by the Hungarian Central Bank (see Baranyai-Csirmaz et al., 2017, page 53ff). These artificial observations are constructed considering information on the very rich obtained from rich lists and other external data sources such as company registers or tax data.

reported wealth components.[11] We thus expect larger effects for these countries. Above that, "[b]oth the composition and the distribution of wealth in Germany and in Austria exhibit considerable similarities" (Fessler et al., 2016, page 29) justifying the comparison of results.

Our findings show that when relying on the HFCS only, one concludes that the richest 1% roughly hold 25% of total wealth in Austria and 24% in Germany. Replacing the top tail by a Pareto distribution estimated from the survey and complemented with data from rich lists increases this share to roughly 43% in Austria and 36% in Germany.[12] We find stronger effects for Austria than for Germany (in terms of wealth shares and changes in the effective oversampling rate) indicating that completely lacking an oversampling strategy for the wealthy as in the case of Austria is problematic.

We find that equity[13] has an increasing importance in the portfolios of the very rich and our adjustments thus have the largest impact on this instrument. In Austria, aggregates for equity increase by roughly 125% and in Germany by roughly 38%. While the survey usually leads to under-coverage for most instruments, aggregates for equity calculated from the HFCS tend to exceed the financial accounts counterparts (see EG-LMM, 2017). HFCS aggregates increase even more after applying our methodology. This is another hint that some of the underlying instruments (in particular the value of self-employed businesses on the HFCS side, and unlisted shares and other equity on the financial accounts side) are not well comparable and differences in valuation concepts are an issue.

Our analysis further shows that for other instruments – namely deposits, bonds and liabilities (loans) – changes are much less pronounced as these instruments seem to be less important for the very rich. For these instruments, coverage ratios[14] increase by roughly 0.5 to 20 percentage points in Austria and by 2 to 9 percentage points in Germany still leaving a significant gap. This shows that the missing wealthy explain only a rather small fraction of the macro-micro gap for financial instruments other than equity. The sources of the remaining gap need further exploration.

We consider our analysis as a first step towards the compilation of wealth-related distributional national accounts. Distributional indicators add important information to aggregates currently found in national accounts and provide deep insights into the structure and distribution of wealth within a society. We therefore demonstrate how the proposed methodology can be used to improve distributional national accounts figures.

While it is not new that Austria and Germany have similar degrees of net worth inequality (Fessler et al., 2016, page 29f), we find also strikingly strong similarities in (the high degree of) inequality on a more disaggregated level. For instance, average financial wealth among the poorest 20% of households equals -11,506 EUR in Austria (-25,127 EUR in Germany) as compared to 325,920 EUR (186,789 EUR) in the top quintile.

This article adjusts HFCS results not only in the dimension of net worth but for a range of

[11] While Austria refrains from oversampling at all, Germany performs oversampling based on geographic but not individual wealth or income data.

[12] Shares for the top 1% after Pareto adjustments are already documented in the literature: Vermeulen (2017) uses the first wave of the HFCS and finds shares of 32%-34% for Germany and 31%-32% for Austria. Bach et al. (2015) find 33% for Germany and Eckerstorfer et al. (2016) 38% for Austria.

[13] See Table 6 for definitions.

[14] We define the *coverage ratio* (CR) for each instrument as the ratio of the HFCS aggregate over the financial accounts aggregate. In case of $CR < 1$ we speak of under-coverage and in case of $CR > 1$ of over-coverage.

different asset classes and liabilities,[15] and is first to quantify the impact of under-represented wealthy households in a household survey on the macro-micro gap for several instruments.

We also add to the literature by providing more detailed information on the wealth distribution and portfolio allocation of the wealthiest households. Such information is so far hardly available for Austria and Germany due to a lack of data sources.

Above that, to our knowledge, our article is first to make use of such a quantification in the compilation of DINA. We calculate instrument-specific distributional indicators, which are useful by themselves, but also aggregate them to an overall indicator for financial wealth. Instrument-specific and aggregated indicators are fully consistent with financial accounts.

The remainder of this article is organised as follows: First, section 3 elaborates on the distribution of net worth and describes how a Pareto model may be used to adjust the distribution at the top. The analytical approach is described in section 4 and the simulation approach in section 5. Data is presented in section 6, section 7 reports empirical results and section 8 demonstrates how our findings can be used to improve the compilation of distributional national accounts. Finally, section 9 concludes. The appendix adds important technical details, performs comprehensive sensitivity analyses, and provides further numerical results.

3 The top of the wealth distribution

In an ideal world, a survey would represent the entire population, i.e., also the most affluent households. Unfortunately, surveys – and particularly wealth related surveys – usually fail to capture these households as they are less willing to participate in a wealth survey, and are harder to be contacted.

Therefore, most countries participating in the HFCS perform some kind of oversampling of wealthy households, i.e., strategically contacting proportionally more wealthy households. Oversampling leads to more observations in a particular part of the distribution than implied by the original sample frame, which makes it necessary to down-weight each observation in this part to guarantee correct population totals.

In general, a survey with a sophisticated sample design yields an unbiased estimate of the aggregate – also without any kind of strategically oversampling wealthy households. Oversampling the wealthy, however, may help to decrease the *variance* of the aggregate and hence increases precision. In the case of a highly skewed distribution or equivalently speaking a high degree of inequality, the upper tail largely contributes to the aggregate and thus oversampling this important part of the distribution may be very effective to get more reliable results for a single survey wave.[16]

It is important to note though, that oversampling does not remove any kind of bias resulting from unit-non-response.

Countries with comprehensive data on wealth or income (e.g., register data maintained for the sake of collecting wealth related taxes or income tax files, which are used to project wealth based on investment returns, see Kennickell, 1999b) can use this extra information to strategically contact larger numbers of wealthy households. However, in Austria and Germany no such data are used (mainly) due to data protection concerns and oversampling is hence largely restricted.

[15]Chakraborty et al. (2016) are to our knowledge the only other authors relying on a semi-parametric Pareto model for a similar instrument-specific analysis. In contrast to our stratification approach, this article only computes average portfolio shares.

[16]At the same time, oversampling the wealthy without increasing the total number of observations negatively affects other parts of the distribution due to larger weights per observation outside the right tail.

Austria refrains from oversampling whereas Germany relies on a regional oversampling strategy (i.e., more observations are drawn from wealthy municipalities or regions within cities; see HFCN, 2016a; Schmidt and Eisele, 2013, and footnote 22).

Ex-post adjustments relying on parametric models are a suitable approach to address *both* issues related to the missing wealthy: precision and bias. There is evidence that the top of the wealth distribution can be well approximated by a Pareto distribution aka the "power law" (see Pareto, 1895).[17] Thus, relying on survey observations only for the less-wealthy and a Pareto model for the wealthy, i.e., a semi-parametric approach, seems to be well suited to get hold of this problem. This idea is not new and has been widely used before.[18]

In general, a Pareto model may be estimated from survey data only. As the very top of the tail is not appropriatly represented by a survey, the estimated distribution is likely to generate a too flat tail. Hence, several articles add observations from rich lists to the estimation procedure to increase confidence in the estimated tail and we follow this strategy (see Bach et al., 2014; Eckerstorfer et al., 2016; Vermeulen, 2016, 2017).[19]

3.1 The Pareto distribution

The standard Pareto distribution is a two-parameter distribution with cumulative distribution function (CDF)

$$F_Y(y) = 1 - \left(\frac{y}{y_0}\right)^{-\vartheta}, \quad y \geq y_0, \tag{1}$$

and density function

$$f_Y(y) = \frac{\vartheta y_0^{\vartheta}}{y^{\vartheta+1}}, \quad y \geq y_0,$$

where $y_0 > 0$ denotes the scale (or threshold) parameter and $\vartheta > 0$ the shape parameter: Decreasing ϑ yields to a reduction of probability mass at the threshold y_0 and at the same time a prolongation of the tail.

There is no universally agreed estimation procedure to determine the threshold y_0. In general, the threshold should be large enough to guarantee that observations follow a Pareto law. The threshold, however, must also be small enough so that there are enough observations above the threshold that can be used to estimate the Pareto shape parameter. For reasons of comparability, we use the same threshold for Austria and Germany, namely one million Euro. Appendix C shows that this is a suitable choice for both countries and performs robustness checks with this regard.

On the contrary, several estimators for the shape parameter are suggested in the literature. When aiming for an estimator based on survey data, a method that accounts for survey weights is needed. Vermeulen (2017) suggests a pseudo maximum likelihood and a weighted regression estimator. Alfons et al. (2013) extend robust estimators for survey data.[20] In this article, we rely on two robust estimators, and Vermeulen's regression method. Additionally, we derive a robust version of the regression method based on quantile regression that is expected to be less sensitive toward extreme observations. Details are provided in Appendix A.

[17] Atkinson (1975), Davies and Shorrocks (2000), Klass et al. (2006), Piketty et al. (2006), Cowell (2011a) and Cowell (2011b) use the Pareto distribution to model top wealth distributions.

[18] See for instance Cowell (2011a); Bach et al. (2015); Eckerstorfer et al. (2016); Vermeulen (2016, 2017).

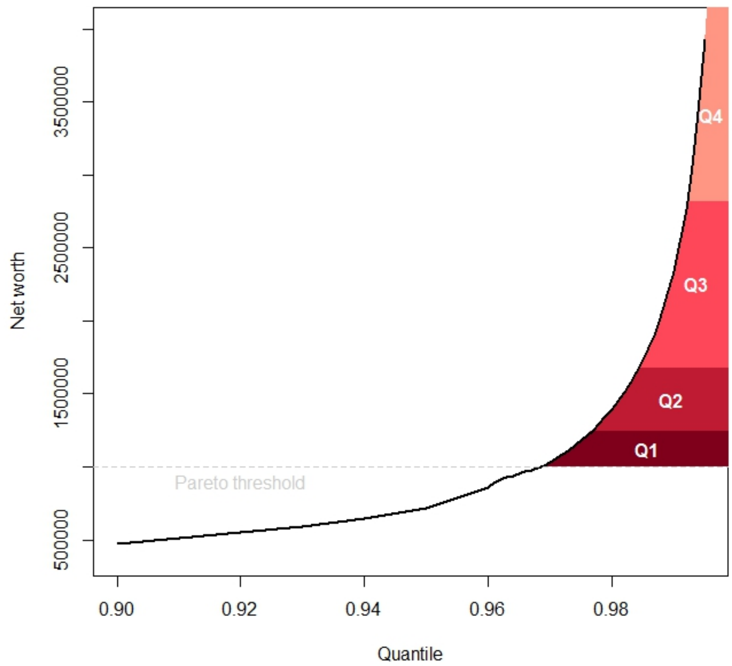
[19] Klass et al. (2006) analyse the wealth distribution implied by the 400 richest people in the United States according to Forbes lists for various years and find that the Pareto distribution is a very well suited model for this top tail.

[20] They implement these estimators (namely the weighted integrated squared error (wISE) and weighted partial density component (wPDC) estimator) in the R package laeken (see Alfons and Templ, 2013).

3.2 Portfolio structure at the very top

A correlation between portfolio structures and the relative position in the wealth distribution is well known. While portfolios of households belonging to the lower deciles of the wealth distribution consist mainly of deposits, the shares of more risky financial assets (mutual funds, bonds, publicly traded shares) increase when moving up the wealth distribution (HFCN, 2016b). The probability to own a private business is significantly higher for households belonging to the top 20% of the net wealth distribution (Arrondel et al., 2014). Appendix B provides detailed distributional information on portfolio structures calculated from second wave HFCS data for several euro area countries.

Figure 2: Illustration of the top tail of the net worth distribution.



Notes: The figure depicts the top of the net worth distribution (for this illustration German HFCS data has been used). Below the Pareto threshold, the empirical distribution is plotted. Above the threshold, the parametric model, i.e., the theoretical Pareto distribution, takes over. $Q1$ to $Q4$ represent the quartiles of the tail distribution. *Source:* HFCS

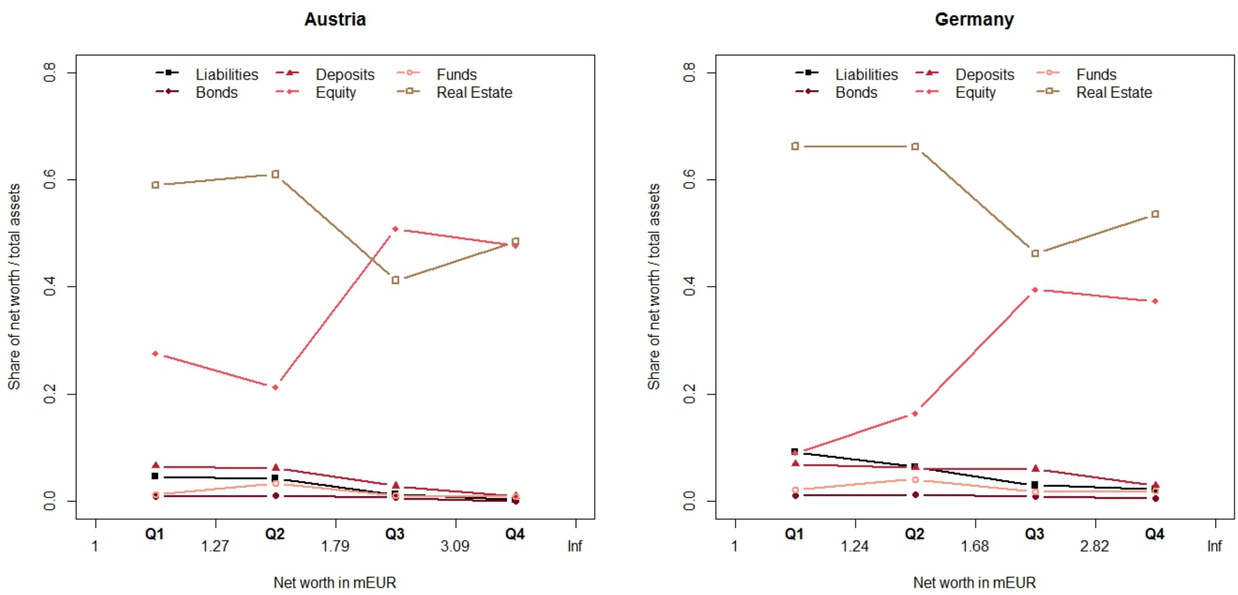
We even find large variation in portfolio structures within the tail: Portfolios of the "wealthy" still differ from portfolios of the "extremely wealthy." To show that, we split up the top tail into four strata $Q1$, $Q2$, $Q3$, and $Q4$ as demonstrated in Figure 2. The number of observations per stratum and average obervation weights are given in Table 1.[21,22]

Table 1: Number of observations and average observation weights per stratum.

	$Q1$	$Q2$	$Q3$	$Q4$	Total tail	Total HFCS
Austria						
No. of observations	24	28	21	12	85	2,997
Average weight	1,493.5	1,452.4	1,526.1	1,484.5	1,487.7	1,288.8
Germany						
No. of observations	124	108	76	71	379	4,461
Average weight	2,407.9	3,350.8	4,104.6	3,727.0	3,264.0	8,893.1

Notes: The thresholds for the strata are determined by quartiles of the adjusted tail distribution. *Source:* HFCS

Figure 3: Change in portfolio structure in the tail.



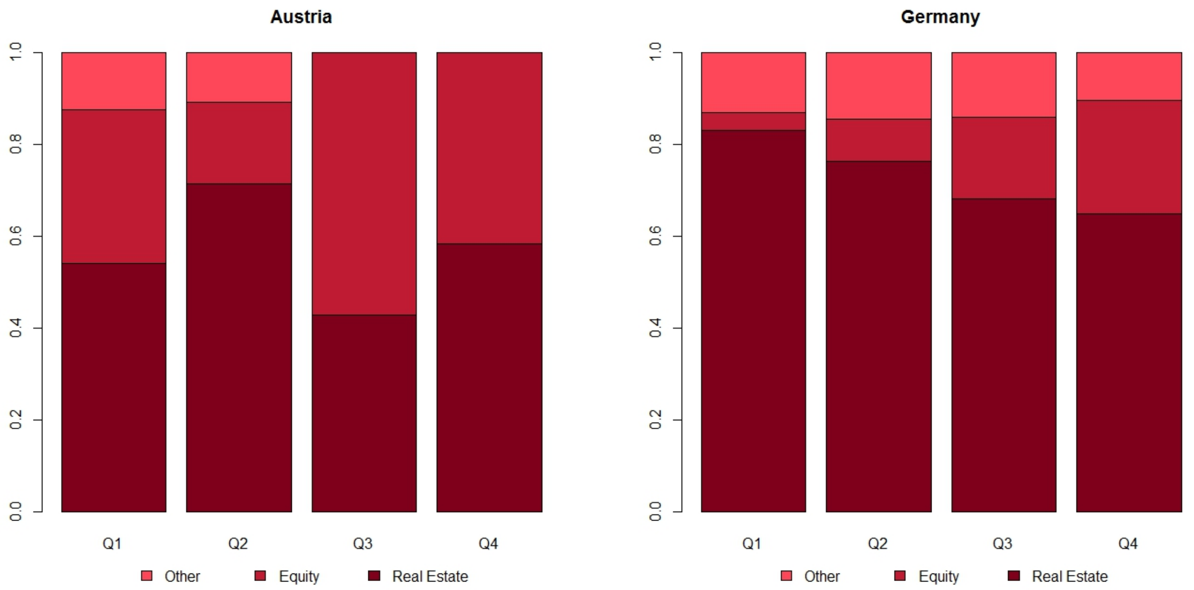
Notes: The figure shows bonds, deposits, equity, mutual funds, and real estate assets as share of total assets and the ratio of liabilities (loans) over net worth (see Table 6 for definitions). Shares are calculated separately for each net worth tail quartile (quartile thresholds are calculated from the semi-parametric Pareto model), e.g., Q4 refers to the wealthiest 25% in the tail. For the figures instrument-specific aggregates are calculated for each quartile and divided by total assets (total net worth) in the respective quartile. *Source:* HFCS

[21]Due to very few tail observations in Austria, we slightly increased $Q4$ by including the top 26% instead of the top 25% and shrinking $Q3$ accordingly. This has the effect that two additional observations enter $Q4$ that stabilize participation rates in bonds since, according to the HFCS, the top ten observations in terms of wealth hold zero assets in the form of bonds. This adaptation has virtually no effect on other instruments.

[22]Table 1 reveals two other important facts worth noting: First, the overall HFCS sample size for Germany is strikingly small implying large average weights per observation. (Note that Germany is roughly ten times the size of Austria in terms of population – see also Table 4.) Second, the geographic oversampling strategy applied by Germany leads to substantial lower weights in the tail than for the rest of the population. In contrast, the average weight of tail observations in Austria is larger indicating that Austria was indeed facing problems of reaching wealthy households.

Figure 3 shows the average portfolio structure for each tail quartile: For the top 25% of the tail population, i.e., $Q4$, equity is much more important than for the rest of the tail population. On the contrary, real estate assets are more important for the lower 50% ($Q1$ and $Q2$) of the tail population. The share of total assets held as deposits, bonds, and mutual funds decreases when moving to the very top of the distribution. Figure 3 also shows the ratio of liabilities over net worth. This ratio also decreases with net worth.

Figure 4: Types of households in the tail.



Notes: The figure shows the share of types of households in the tail split up by quartile. A household is of type "equity" when its largest position in the portfolio is equity and so forth. For Austria, results are less smooth due to fewer tail observations (see Table 4 for totals). *Source:* HFCS

The portfolio allocation for individual observations and for the entire tail can be found in Figure 17 in the appendix. There, each bar represents one single observation in the HFCS. The wider the bar the more households are represented by this observation. It is clearly visible that portfolio structures systematically change when moving from $Q1$ to $Q4$.

Figure 4 shows the share of "types of households" for $Q1$ to $Q4$. A household is classified to be of type "equity" when its largest portfolio position is equity and so forth. The share of "real estate households" decreases when moving up the distribution while the share of "equity households" increases. Portfolios consisting mainly of assets other than real estate and equity are generally more frequent in $Q1$ to $Q3$ than in $Q4$.

Although we observe clear tendencies in changes in portfolio structures, which we exploit in our analysis, the survey (and in particular surveys without oversampling) most probably still lack representative portfolios at the very top and accuracy of our results may thus be limited.[23] For instance, Kennickell (2008) finds that in the 2004 *US Survey of Consumer Finances* out of roughly 400 interviewed households that had direct holdings of government or commercial bonds, approximately 90% of these cases entered the survey through oversampling. Indeed, we find that bond holdings are surprisingly low among households in the top tail in Austria.

[23] As Austria completely refrains from oversampling the very rich, we particularly expect such shortcomings for this country. Additionally, Austria is in terms of population size much smaller than Germany (see also Table 4) and there are thus in general much less tail observations. Therefore, the results for Austria in Figure 4 are less smooth and accuracy – especially for $Q4$ – would most likely increase with introducing oversampling.

Under the assumption that the portfolio structures of the wealthiest of the wealthy do not differ strongly across European countries, we demonstrate in Appendix C how borrowed portfolios from other HFCS countries can be incorporated into our approach to increase reliability. However, we find that there are only marginal changes to general results.

It is debatable whether surveys are in general well suited to capture the detailed portfolio structures at the very top as even countries oversampling the wealthiest households may fail to *perfectly* capture them. If one believes that portfolio structures become even more extreme when, say, moving from the 99th to the 99.5th quantile of the wealth distribution (in terms of for instance relative importance of equity), the results presented in the article would in fact constitute a lower bound. We acknowledge that results could become even more extreme when better reflecting portfolio structures at the very top (by for instance collecting information from asset managers), but we consider it unlikely that the direction of results would change as we can identify clear and plausible trends in portfolio structures in the HFCS.

4 An analytical approach to adjust for the missing wealthy

The analytical method is straight-forward and easy to implement. However, it comes at the cost of loosing micro structure: the dependency among instruments (as each instrument-specific aggregate is calculated independently of all other instruments) as well as the valuable variability resulting from using individual rather than average portfolio shares are lost. Calculating, for instance, the share of total assets (or liabilities) held by the top 1% or distributional measures such as Gini coefficients is not possible.

In the case of $\vartheta > 1$ (which holds true for all our estimations), the Pareto distribution has a finite first moment, i.e., the mean exists and is given by

$$E(Y) = \frac{\vartheta}{\vartheta - 1} \cdot y_0. \tag{2}$$

In this case, the total tail wealth can be estimated by multiplying the number of households belonging to the tail N_{tail} by the mean, i.e.,

$$N_{tail} \cdot \frac{\vartheta}{\vartheta - 1} \cdot y_0.$$

In a survey setting, N_{tail} equals the sum of weights of all households with a net worth greater than y_0.

As input, the analytical method needs a Pareto model for the top tail. The parameter of this model are ideally derived by combining survey and rich list observations as described in section 3. We split the Pareto tail into four strata[24] as shown in Figure 2.

We calculate average wealth for each stratum Q. Therefore, we calculate *expected net worth Y conditional on net worth realising in stratum Q*. Strata are defined via quartiles, which are denoted by $F^{-1}(p)$. Thus, strata are intervals $Q = [F^{-1}(p_1), F^{-1}(p_2))$, with $\{p_1, p_2\} \in \{\{0, 0.25\}, \{0.25, 0.5\}, \{0.5, 0.75\}, \{0.75, 1\}\}$. Stratum-specific expected wealth is given by[25]

$$W(Q) = E(Y|Y \in Q) = \frac{\vartheta y_0}{(1-\vartheta) \cdot (p_2 - p_1)} \left[(1-p_2)^{1-1/\vartheta} - (1-p_1)^{1-1/\vartheta} \right].$$

[24] Four strata seem appropriate for our case study. However, the method can easily be adjusted if a finer split-up is needed. Keep in mind the practical adjustment for Austria as mentionde in footnote 22.

[25] The derivation of this formula is given in Appendix D.

Total stratum-specific net worth is thus given by

$$TW(Q) = (p_2 - p_1) \cdot N_{tail} \cdot W(Q).$$

Note that in case of a split-up by quartiles $p_2 - p_1 = 0.25$.

As a next step, we calculate stratum-specific portfolio shares. Let w_k denote the weight of household k and s_j^k assets held by household k in the form of instrument j over net worth. *Instrument- and stratum-specific tail aggregates* are then given by

$$A_j(Q) = \left(\sum_{k \in Q} s_j^k \cdot \frac{w_k}{\sum_{l \in Q} w_l} \right) \cdot TW(Q).$$

Total instrument-specific aggregates are obtained by summing up stratum-specific tail aggregates and adding them to the non-tail aggregate $A_j(NT)$, which is calculated respecting survey weights:

$$A_j = A_j(NT) + \sum_{Q \in \{Q1,Q2,Q3,Q4\}} A_j(Q).$$

5 A simulation approach to adjust for the missing wealthy

The fact that the top tail of an empirical distribution is replaced by a parametric model can also be exploited for simulation. A simulation creates micro data files that are adjusted for the too flat tail. These micro files allow one to analyse distributional patterns at a disaggregated level. The simulation, which here combines Monte Carlo simulation and bootstrapping, perfectly mirrors the analytical approach, and thus empirical confidence intervals constructed as bootstrap quantiles supplement the analytical results. The proposed algorithm aims to recover observed portfolio structures and thus also enables analyses with this regard.

We combine the semi-parametric model for net worth with a wealth-stratified bootstrap algorithm to conserve the correlation between portfolio structure and net worth. The bootstrap algorithm is a fully non-parametric[26] procedure.

For each household in the tail, total net worth is simulated from the Pareto model. The household is then allocated to its respective net worth stratum and a suitable portfolio is drawn.

Again, we stratify tail observations into four strata[27] ($Q1$ to $Q4$) depending on their position in the net worth distribution. So, a household with simulated net worth of roughly one million Euro is allocated to the lowest quartile and thus a portfolio is drawn from the $Q1$ stratum. In contrast, households with large simulated wealth are assigned a portfolio structure from $Q4$.

The algorithm consists of nine steps and is summarised in Table 2. The stratified bootstrap is repeated B times for each of the M draws from the Pareto distribution each of size N_{tail}. Our choices of $M = 100$ and $B = 100$ yield in total $M \times B = 10,000$ repetitions. In the case of

[26]The correlation between portfolio structures and net worth does not seem to be linear but rather follows distinct functional patterns. Hence, relying on a parametric model such as a Beta regression (The Beta distribution is ideally suited to model shares as it is bounded between 0 and 1.) might (and in fact does) heavily depend on the exact model specification. Semi-parametric approaches with data driven functional forms (such as Generalized Additive Models) would be an obvious candidate to overcome this issue. However, the observed tail is very sparsely populated and it was (at least in our analysis) impossible to estimate stable and trustworthy functional forms.

[27]The algorithm works for any stratification as long as there are sufficient observations per stratum. The choice here is in accordance with the analytical method.

Table 2: Simulation approach.

Step 1.	For a given y_0 estimate the shape parameter of the Pareto distribution ϑ using the combined sample of survey and rich list observations.
Step 2.	Calculate quartiles of the tail implied by the estimated Pareto distribution yielding strata $Q1$, $Q2$, $Q3$, and $Q4$.
Step 3.	Allocate each observation in the survey with a net worth of at least y_0 to its respective stratum $Q1$, $Q2$, $Q3$, or $Q4$.
Step 4.	Draw N_{tail} random numbers from Pareto$(y_0, \hat{\vartheta})$ and allocate them to the strata $Q1$, $Q2$, $Q3$, or $Q4$.
Step 5.	For each simulated net worth draw a random portfolio allocation from observed allocations in the respective stratum $Q1$, $Q2$, $Q3$, or $Q4$ respecting survey weights.
Step 6.	Calculate instrument-specific aggregates for the tail population.
Step 7.	Add the simulated tail aggregates to the non-tail aggregates and calculate instrument-specific coverage ratios.
Step 8.	Repeat steps 5 to 7 B times.
Step 9.	Repeat steps 4 to 8 M times.

B **observations.**

$M \times B$ **observations.**

$N_{tail} \times M \times B$ **observations.**

Germany, the analysis is thus based on $N_{tail} \times M \times B = 12,370.37$ million and in the case of Austria $N_{tail} \times M \times B = 1,264.58$ million simulated portfolios.

Appendix A describes how to calculate quartiles from a Pareto distribution (step 2) and how to sample from a Pareto distribution (step 4).

In step 5, we *draw a random portfolio allocation*. In this context, this means that we draw a full list of shares of assets/net worth but no amounts. Specifically, this list includes the ratio of liabilities over net worth, as well as the share of total assets invested in bonds, deposits, equity, and real estate. The portfolio is then constructed by, first, calculating the amount of liabilities by multiplying its drawn ratio with simulated net worth and, second, using the Euro amount of liabilities to calculate total assets. Third, we multiply total assets with the drawn share of bonds, deposits, equity, and real estate to receive the respective Euro amounts.

Thus, although we replicate observed patterns in terms of portfolio structure, simulated amounts are (on average) larger as net worth is drawn from a Pareto model, which generally increases total tail wealth.

6 Data

The analysis makes use of three different data sources: The first source are data for Austria and Germany collected in the second wave of the HFCS. Second, we use aggregates from financial accounts (FA) as benchmark measures. Third, we use national rich lists that yield information on net worth of the wealthiest individuals/families.

Table 3 reports the codes used in the HFCS and financial accounts as defined in the *European System of National and Regional Accounts* (ESA2010).

Table 3: HFCS and FA codes.

HFCS codes	
DA1140	Value of self-employment businesses
DA1400	Real estate (incl. property used for business activities)
DA2101	Deposits (Value of sight and savings account)
DA2102	Mutual funds, total
DA2103	Bonds
DA2104	Value of non-self-employment businesses
DA2105	Shares, publicly traded
DL1000	Total outstanding balance of household's liabilities
DN3001	Net wealth, excl. public and occupational pensions
FA codes	
F22	Transferable deposits
F29	Other deposits
F3	Debt securities
F4	Loans (long-term and short-term)
F51	Equity
F52	Investment fund shares/units

Source: HFCS, ESA2010

6.1 HFCS data

The HFCS collects household-level data on households' finances and consumption. The second wave, which was released in December 2016, was conducted in 18 euro area countries as well as the non-euro area countries Hungary and Poland in a harmonised way while still leaving room for a country-specific implementation.[28] While some European countries have a long tradition in conducting wealth related household surveys, a harmonised survey for such a large number of countries is a recent development and results of the first wave were only released in 2013. In Austria and Germany[29] the HFCS is the first comprehensive household survey on wealth, which is why there is yet very limited knowledge about the wealth distribution in these countries.

The variables in the HFCS follow common standards and definitions. Therefore, the instruments used in the analysis (assets and liabilities) are in principle comparable between Austria and Germany. In both countries, the survey has been mainly conducted via *Computer Assisted Personal Interviews*.

The HFCS uses probability sampling and the sample size is representative both at the country and at the euro area level. For Austria, the gross sample size was 6,308 which led to 2,997 observations (response rate of 49.8%) representing 3.9 million households while in Germany the gross sample size was 16,221 yielding 4,461 observations (response rate of 19%)[30] representing 39.7 million households.

[28]See HFCN (2016a) for a general documentation and `https://www.hfcs.at/en/` (Austria) and `http://www.bundesbank.de/Navigation/EN/Bundesbank/Research/Panel_on_household_finances/panel_on_household_finances.html` (Germany) for a more detailed country-specific documentation.

[29]In Germany, the HFCS is labelled *Panel on Household Finances* (PHF).

[30]Response rate considering households that are interviewed for the first time (Germany has a panel component). The response rate including panel households is 29% (see HFCN, 2016a).

Table 4: HFCS data.

AT – Austria

	Total Aggregate	Summary Statistics for the tail 25%	Mean	Median	75%
Net worth	998,130	1.220	2.618	1.518	2.040
+Liabilities	66,601	0.000	0.044	0.000	0.042
Total Assets	1,064,730	1.284	2.662	1.609	2.068
Bonds	5,273	0.000	0.012	0.000	0.000
Deposits	98,745	0.028	0.089	0.055	0.100
Equity	194,739	0.000	0.942	0.217	1.000
Mutual Funds	17,070	0.000	0.044	0.000	0.006
Real Estate	686,174	0.458	1.479	0.910	1.300
Number of observations:					2,997
Number of observations in the tail:					85
Number of households:					3,862,526
Number of households above 1 mEUR:					126,458
Total population:					8,543,930
Average number of persons per household:					2.212

DE – Germany

	Total Aggregate	Summary Statistics for the tail 25%	Mean	Median	75%
Net worth	8,500,090	1.202	2.611	1.570	2.472
+Liabilities	1,020,920	0.000	0.118	0.000	0.127
Total Assets	9,521,010	1.267	2.729	1.654	2.574
Bonds	71,916	0.000	0.055	0.000	0.005
Deposits	1,007,795	0.021	0.187	0.070	0.200
Equity	1,326,078	0.001	0.735	0.087	0.500
Mutual Funds	206,863	0.000	0.082	0.000	0.037
Real Estate	5,872,317	0.715	1.487	1.065	1.650
Number of observations:					4,461
Number of observations in the tail:					379
Number of households:					39,672,000
Number of households above 1 mEUR:					1,237,037
Total population:					80,983,000
Average number of persons per household:					2.041

Source: HFCS, Eurostat (population totals for 2014)

Notes: Amounts are in mEUR. 25% and 75% refer to the first and third quartile, i.e., the 25% and 75% percentile. The tail is here defined as households having at least 1 mEUR of net worth.

Table 4 provides HFCS summary statistics. Specifically, details are provided for the wealthiest households here defined as all households with a net worth of at least one million Euro.

Most countries that conduct the HFCS try to oversample the wealthy as it is known that their influence on total wealth is substantial and these households are less likely to participate in surveys. The success of an oversampling strategy can be measured by the *effective oversampling rate* (see also HFCN, 2016a), which is defined for the top $x\%$ as

$$\frac{S(1-x)-x}{x},$$

where $S(1-x)$ is the share of sample households in the wealthiest $x\%$.

For our analysis, the oversampling strategy is of particular relevance as we focus on the top part of the wealth distribution. Austria refrains from oversampling wealthy households which results in an effective oversampling rate of -33% for the top 1% while Germany uses regional indicators for their oversampling which results in an effective oversampling rate of 131% for the top 1%.[31] As we mainly analyse households with a net worth of at least a million Euro we also report the effective oversampling rate for millionaires: For Germany, this rate equals 171% while for Austria -14%.[32] (See also footnote 22.)

The HFCS relies on multiple imputation to fill in gaps due to item-non-response, and erroneous and implausible entries.[33] Imputations are not free of doubt which is why five implicates are provided. For our analysis, for each imputed value we use the average over all five implicates.

Our methodology critically depends on an exact estimate of the number of households owning at least one million Euro. We estimate this number from the HFCS by summing up the weights of the wealthiest households. Results are also provided in Table 4. According to the *World Wealth Report* (Capgemini, 2016) there are 114,000 (121,000) *high net worth individuals* (HNWIs)[34] in 2014 (2015) in Austria and 1,141,000 (1,199,000) in Germany. Although the definitions of HNWIs and millionaires in the HFCS are not perfectly aligned, the very similar results increase our confidence in HFCS numbers.

6.2 Financial accounts data

Aggregates compiled under the ESA2010 framework generally provide a detailed and consistent description of an economy and aim to be exhaustive.[35] Within the European Union, data is well comparable as Member States are legally obliged to follow common accounting rules.

For our comparisons, we use financial accounts data for the household sector, which form an integral part of the ESA2010 accounting framework.

Aggregates from financial accounts are reported in Table 5. The period of measurement in financial accounts are either quarters or years. As fieldwork periods of the HFCS are usually longer than a quarter but shorter than a year, we calculate weighted averages of quarterly financial accounts data using all quarters that overlap with the respective fieldwork period: If, say, the fieldwork period was three months of which one month falls into quarter A and two months fall into quarter B, then quarter A will be weighted with 1/3 and quarter B with 2/3.

6.3 Comparability of HFCS and FA instruments

Although the definitions of variables are similar in both the HFCS and FA, they are not fully comparable. The *ECB Expert Group on Linking Macro and Micro Data for the Household Sector* (EG-LMM) developed a catalogue that assesses the conceptual comparability for each

[31] If e.g. the share of very wealthy (top 1%) households in the net sample is exactly 1%, then the effective oversampling rate of the top 1% would be zero. In general, oversampling is thus successful when the effective oversampling rate is greater than zero. In the case of Austria, due to unit-non-response this rate is even negative (i.e., there are even less wealthy households in the net sample then there would be if all households had the same weight) and thus leads to an up-weighting of observations at the top increasing the influence of every single one. This has potentially a negative effect on the precision of estimates.

[32] We calculate this rate by empirically inverting the empirical cumulative distribution function implied by HFCS data. For Austria this yields in fact the 0.0330 rate and for Germany the 0.0313 rate.

[33] When a respondent generally agrees to participate in the survey but denies to answer some particular questions, these missing values are classified as item-non-responses.

[34] HNWIs are defined as individuals having at least one million USD of investable assets, excluding primary residence, collectibles, consumables, and consumer durables.

[35] While data generally is intended to be exhaustive, for example the value for real estates for the household sector is not yet fully included in practice.

Table 5: Financial accounts data.

AT – Austria					
HFCS fieldwork period:					June 2014 – February 2015
Number of months:					9
	Q2:2014	Q3:2014	Q4:2014	Q1:2015	Weighted
Quarter weight	0.111	0.333	0.333	0.222	Average
Liabilities	165,866	167,271	167,936	171,543	168,286
Bonds	43,058	41,869	40,476	38,860	40,868
Deposits	213,652	212,285	217,867	219,472	215,895
Equity	120,586	120,241	127,252	131,278	125,069
Mutual Funds	45,186	46,549	47,787	51,984	48,018
DE – Germany					
HFCS fieldwork period:					April 2014 – November 2014
Number of months:					8
	Q2:2014	Q3:2014	Q4:2014	Q1:2015	Weighted
Quarter weight	0.375	0.375	0.250	0	Average
Liabilities	1,556,417	1,566,132	1,570,512	1,572,707	1,563,584
Bonds	176,357	168,895	162,198	156,766	170,019
Deposits	1,822,186	1,835,383	1,870,435	1,880,884	1,839,197
Equity	502,825	497,196	508,914	563,395	502,236
Mutual Funds	420,608	431,679	442,504	487,658	430,234

Sources: Oesterreichische Nationalbank, Deutsche Bundesbank

Notes: Numbers are in mEUR. Quarters are weighted according to the fieldwork period. Data is for the household sector only, i.e., S.14 according to ESA2010.

individual instrument (see EG-LMM, 2017). Their findings are summarised in Table 6.

Table 6: Conceptual comparability of HFCS and FA instruments.

	FA instrument	HFCS instrument	EG-LMM classification
Liabilities	F4	DL1000	High
Bonds	F3	DA2103	High
Deposits	F22 + F29	DA2101	High
Equity	F51	DA2104 + DA1140 + DA2105	Medium
Mutual funds	F52	DA2102	High
Real estate	–	DA1400	–
Net worth	–	DN3001	–

Notes: The table reports conceptual comparability as assessed by the EG-LMM (2017). See Table 3 for definitions.

Whereas liabilities, bonds, deposits, and mutual funds are in principal highly comparable, the case is more complicated for equity.[36] While listed shares as part of equity are highly comparable between the HFCS and ESA2010, other parts of equity are not. In particular, there is a valuation discrepancy for unlisted shares and other equity/(non) self-employment businesses: The HFCS relies on self-evaluation of the market value.[37] In financial accounts, "the valuation for unlisted shares and in particular holdings of other equity is less accurate [than for items with quoted market prices] as their valuation requires assumptions and modelling. In the case of unlisted shares, the valuation should mirror that of similar listed shares, the value of own funds, or discounted expected profits. Other equity reflects the value of own funds (conceptually calculated as assets minus liabilities, in practice only book values may be available)" (EG-LMM, 2017, page 8).

We look only at the aggregate of equity since the EG-LMM assessed the comparability of the more detailed break down, i.e., at the level of unlisted shares and other equity, as low.

In general, we refrain from directly comparing net worth between the HFCS and FA. Chakraborty et al. (2016) and EG-LMM (2017) analyse different wealth concepts and conclude that at this early stage a comparison at the level of net wealth suffers from many conceptual difficulties. However, we do provide some results for *financial wealth*, which we define here as the sum of bonds, deposits, equity and mutual funds minus liabilities.

Similarly, we do not link real estate assets yet. This comparison will be possible once data coverage in the national accounts is improved.[38]

Another important issue concerning the comparability between HFCS and FA data (see EG-LMM, 2017, page 17f for more details), emerges from the distinction of producer households (part of the household sector), and quasi-corporations and corporations (part of the non-financial corporations sector) in the system of national accounts. This distinction affects the composition of the household sector balance sheet. In general, if an unincorporated enterprise is considered to be a separate unit, it is recorded in the corporate sector and shares held by households are recorded as other equity on a net basis. In contrast, unincorporated enterprises classified as producer households are part of the household sector, and their assets and liabilities are spread over all instruments. In the HFCS, a concept comparable to producer households does not exist. Assets and liabilities of self-employed businesses are recorded as net values. From this aspect alone, we would thus expect over-coverage for equity and under-coverage for remaining instruments.[39]

6.4 Rich lists

In many countries, journalists create lists of the wealthiest individuals and families, of which the *Forbes World's Billionaires List* (e.g., Forbes, 2015) is probably the most famous one. Forbes lists the name and net worth of individuals and families owning at least one billion

[36] Equity as defined here differs from the HFCS definition of "business wealth," which includes properties used for business purposes. These assets are added to non-business real estate here and show up as part of "real estate." In Austria and Germany, the variable DA1140 also includes land and buildings being part of small farms (excluding the part used as main residence). With this regard the distinction between business real estate and other business wealth is far from perfect.

[37] HFCS question: "What is the net value of your /your household's share of the business? That is, what could you sell it for, taking into account all (remaining) assets associated with the business and deducting the (remaining) liabilities?"

[38] The *ESA Transmission Programme* requires the transmission of annual data on land only by end-2017. The EG-LMM has not yet made a final decision on the conceptual comparability, but expects it to be at least of medium comparability. In general, national accounts separate real estate into AN.111 Dwellings, AN.112 Other buildings/structures and AN.211 Land.

[39] For a more in depth discussion of this issue, see Chakraborty et al. (2016).

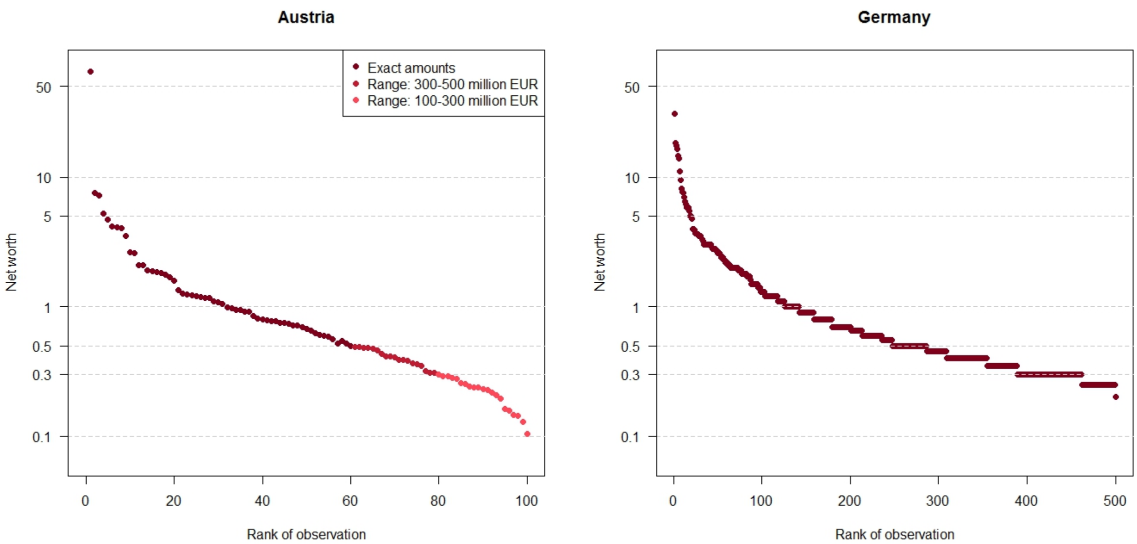
US-dollars every year. It uses a number of different sources and rather opaque methodologies to estimate the wealth of the financial elite around the globe.[40] Such rich lists are in general neither expected to be complete nor completely accurate, but constitute the only public data source reporting an estimate of the individual wealth of the most affluent members of society.

Next to the Forbes list targeting all billionaires in the world, there are several national rich lists. For Austria, the weekly business magazine *Trend* (trend, 2015) publishes a list of the wealthiest 100 Austrians every year. For Germany, the *Manager Magazine* (manager magazine, 2014) publishes the list of the wealthiest 500 Germans.

In this article, we rely on national rich lists. National lists are more comprehensive as they list not only billionaires. For instance in the case of Austria, there are only roughly ten billionaires on the Forbes list per year, whereas the Trend list reports the 100 wealthiest individuals/families. Additionally, the Trend and Manager Magazine lists are measured in Euro, which – compared to the Forbes list measured in USD – does not introduce an additional source of ambiguity due to changes in exchange rates.[41]

Next to various measurement issues there is another major problem when using rich lists to adjust the tail of the wealth distribution: The Forbes, Trend, and Manager Magazine list fail to distinguish between households and family clans, whereas both the HFCS as well as financial accounts use households as unit of recording. This is why we do not directly rely on exact amounts reported in the rich lists but rather use these observations only to increase the quality of our Pareto estimates. We also perform some robustness checks with this regard as reported in Appendix C.

Figure 5: Rich lists.



Notes: The figures plot observations from rich lists. Net worth is in billion Euro. Due to the high degree of skewness, net worth is plotted on a log-scale. For Austria, amounts of less than 500 million Euro are reported as ranges in the Trend list and are here distributed uniformly within the range to avoid clusters. *Sources:* HFCS, Trend, Manager Magazine.

[40] Freund and Oliver (2016) explain that "Forbes uses shareholder information, company financial statements, current exchange rates, and meetings with candidates to estimate individual net worth."

[41] The number of Austrian names on the Forbes list changed from eleven in 2014 to eight in 2015 which is probably at least partly due to a drop in the exchange rate.

Trend list: The magazine lists the names, net worth and major sources of wealth of the richest individuals or families in Austria. Additionally, they classify the "type" of assets: assets hold in the form of a private foundation ("Stiftungsvermögen"), listed and unlisted shares ("Betriebsvermögen"), and inherited assets. This classification shows that in Austria the wealthy store large parts of their assets in the form of private foundations.[42] We perform robustness checks (see Appendix C) on how estimates change when leaving out rich list observations that hold their assets exclusively as foundations.

For individuals with net worth of less than 500 million EUR, Trend does not report the exact amount but only a range. Altogether, there are 60 observations with exact amounts, 19 observations with a net worth between 300 and 500 million, and 21 observations with net worth ranging between 100 and 300 million EUR. To avoid cluster effects in the estimation of the Pareto tail, we assume that the net worth of range observations is uniformly distributed within the respective range and assign them an accordingly drawn random number (the result is shown in Figure 5).

Like Forbes, Trend uses a large number of different sources to compile their rich list. The list is published in June. We use the data from the 2015 list, as this best matches the fieldwork period in Austria.[43]

Manager Magazine list: The Manager Magazine publishes the list of the wealthiest 500 individuals and families in Germany together with their major source of wealth (e.g., real estate assets or the name of a company). Like the Forbes and Trend list, a wide range of sources and methodologies are used when compiling the list. In contrast to the Trend list, "exact" amounts[44] are reported rather than ranges and therefore no adjustment has to be made with this regard. The list does not provide information about the type of assets.

As the field-work period for Germany fully falls into the year 2014, we make use of the 2014 list.[45]

7 Empirical application

7.1 The adjusted wealth distribution

Pareto estimation results depend on the choice of method to estimate the shape parameter. Remember that in this article we rely on four different estimation techniques described in Appendix A. We choose the estimation procedure according to graphical inspection.[46]

[42] We see many private foundations ("Privatstiftungen") in Austria among the rich list as foundations serve as a way to secure the wealth of the family (e.g., the founder of a company can prevent fragmentation of business assets) and a private foundation can benefit from tax-preferred treatments (e.g., to reduce inheritance taxes as well as taxes on capital income). The majority of private foundations in Austia follow private purposes and not charitable or public purposes (Schneider et al., 2010). The claims by households in these private foundations are in the national accounts recorded in the household sector (see Andreasch et al., 2015). The national questionnaire of the Austrian HFCS has a dedicated question asking for the value of private foundations, but in practise this question was hardly ever answered. Therefore, adding information from the rich list in Austria is particular important and complements the HFCS rather well. The analysis by Andreasch et al. (2015) on equity stakes held by private foundations furthermore indicates that wealth in private foundations is rather concentrated even inside the group of private foundations showing similar patterns as the rich list and a Pareto-like behaviour.

[43] For listed shares, they use values as of the beginning of May 2015.

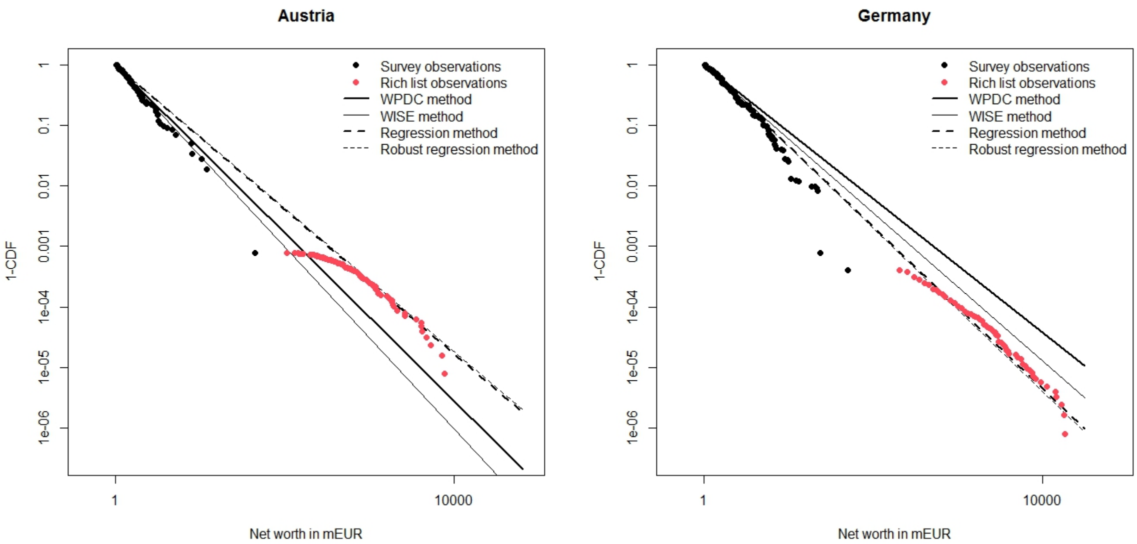
[44] Net worth is estimated and thus cannot be "exact." Above that, amounts are heavily rounded.

[45] For listed shares, they use values as of September 2014.

[46] From theory, we know that the logged random variable to be modelled (here: net worth) and the log of one minus the cumulative distribution function (CDF) have a linear relationship. Thus, we look for a straight line that best fits our observations. This is shown in Figure 6.

Figure 6 shows these choices graphically and Table 7 reports the respective numbers. For both countries, the shape parameter is chosen according to the *regression method* as it seems to be best suited to capture the richest households appropriately.[47]

Figure 6: Estimated Pareto distributions.



Notes: The figures plot the (empirical) complementary cumulative distribution function (which is defined as 1-CDF) on a log-log scale. Survey observations as well as observations from rich lists are included. The graphs are used to decide which estimation method is most appropriate for each country.
Sources: HFCS, Trend, Manager Magazine

Table 7: The choice of shape parameter.

	AT − Austria		**DE − Germany**	
	$\hat{\vartheta}$	Implied tail wealth	$\hat{\vartheta}$	Implied tail wealth
wISE	1.505	376,951	1.225	6,743,922
wPDC	1.389	451,950	1.107	12,853,001
Regression method	1.194	778,285	1.338	4,899,939
Robust regression method	1.186	808,018	1.350	4,767,989

Notes: The implied tail wealth is calculated using formula (2) and is reported in mEUR. Estimation is performed for the combined survey and rich list data.

We draw $M = 100$ samples each of size N_{tail} from the Pareto$(y_0, \hat{\vartheta})$. Table 8 analyses the distribution of net worth when replacing the observed tail by samples from the Pareto distribution.

Both, the wPDC and wISE estimators are desigend to be robust against outliers. As all our rich list observations are somewhat “outliers” these procedures take these observation into account to a lesser degree. The regression and robust regression method are succesful to combine the information from the HFCS as well as the rich lists and lead to almost indistinguishable results. Vermeulen (2017) compares several other methods and also concludes that the regression method including observations from the Forbes list is suited best.

[47]However, we do not find a consistent pattern across methods for the two countries: In Austria, the wISE and wPDC estimators imply the lowest tail wealth whereas for Germany these methods yield the largest tail wealth. Results are quite sensitive towards this choice. In particular for Austria, the gap between the lowest observation on the rich list and the largest observed wealth in the HFCS is substantial (partly as a consequence of lacking an oversampling strategy), and judging which method is the most appropriate is challenging. A more flexible method than what is used here and in the current literature, and which does not impose a constant shape parameter for the entire tail may be more appropriate. We leave this for future research.

Average results of all $M = 100$ replicates as well as coefficients of variation (CoV) are reported. Additionally, the table reports results when exclusively relying on survey data.

Table 8: Pareto estimation results.

AT – Austria	Simulation Mean	Simulation CoV	Survey	Change
Total wealth	1,377,072	0.186	998,130	38.0%
Total tail wealth (simulation)	744,985	0.344	366,043	103.5%
Total tail wealth (analytical)	778,285		366,043	112.6%
Share of tail wealth of total wealth	54.1%		36.7%	17.4 pp
99th Percentile	2.70	0.003	1.94	39.2%
Wealth of top 1%	608,560	0.421	247,125	146.3%
Share of top 1% of total wealth	43.3%		24.8%	19.4 pp
Gini coefficient	79.6%	0.022	72.4%	7.2%
Effective oversampling rate				
of the top 5%	-11.9%		-11.9%	± 0 pp
of the top 1%	-49.9%		-13.2%	-36.7 pp

DE – Germany	Simulation Mean	Simulation CoV	Survey	Change
Total wealth	10,090,679	0.017	8,500,090	18.7%
Total tail wealth (simulation)	4,867,181	0.036	3,276,592	48.5%
Total tail wealth (analytical)	4,899,939		3,276,592	49.5%
Share of tail wealth of total wealth	48.2%		38.5%	9.7 pp
99th Percentile	2.34	0.001	2.34	0.2%
Wealth of top 1%	3,644,236	0.048	2,059,921	76.9%
Share of top 1% of total wealth	36.1%		24.2%	11.9 pp
Gini coefficient	79.7%	0.004	75.9%	3.8 pp
Effective oversampling rate				
of the top 5%	175.3%		175.3%	±0 pp
of the top 1%	130.9%		130.9%	±0 pp

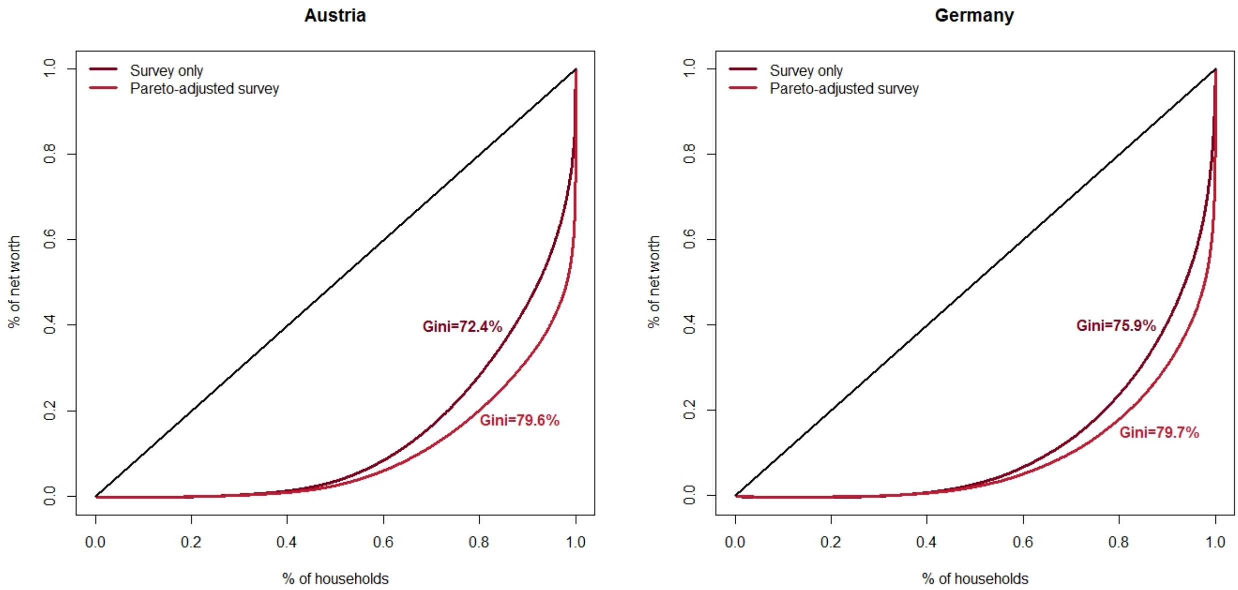
Source: HFCS

Notes: Amounts are in mEUR. Shares (e.g., the top 1%) refer to the share of *households* but not individuals. Likewise, the Gini coefficient is calculated on the household level. CoV refers to the coefficient of variation, i.e., the standard deviation divided by the mean. *Tail* refers to households with a net worth of at least one million EUR. *Analytical* refers to the theoretical mean using the estimated parameters and formula (2). Simulation results are based on $M = 100$ draws from the respective Pareto distribution. "pp" refers to percentage points.

For both countries, the semi-parametric Pareto model leads to higher degrees of wealth inequal-

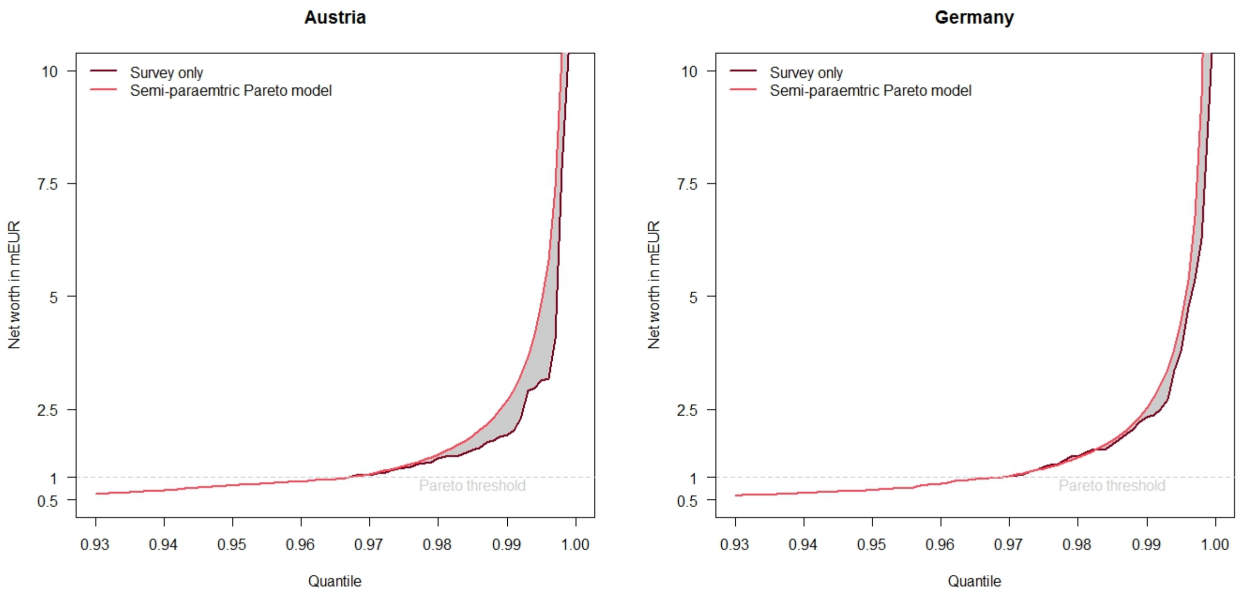
[48] To plot the Lorenz curve, among the 100 simulated Pareto samples the one is chosen which fits best the theoretical Pareto distribution according to a Kolmogorov-Smirnov test. Gini coefficients reported here are the average coefficient over all 100 replicates.

Figure 7: Lorenz curves for net worth.



Notes: The figures show household-level Lorenz curves for net worth before and after adjusting[48] for the missing wealthy. Survey weights are respected. *Source:* HFCS

Figure 8: Top tail of the wealth distribution.



Notes: The figure plots the top tail of the Austrian and German net worth distribution once estimated using the HFCS data only and once using the semi-parametric Pareto model, i.e., the combination of observed survey data and a theoretical Pareto tail estimated from survey and rich list observations. *Source:* HFCS

ity compared to calculations relying on survey data only. In Austria, the top 1% is estimated to own roughly 43% of total wealth whereas when relying on the survey only, this number is estimated to be 25%. In the case of Germany, this kind of inequality increases from 24% to 36%. Together all millionaires in Austria own roughly 54% of total wealth (compared to 37%

when relying on the survey only) and in Germany 48% (compared to 39% when relying on the survey only). Likewise, household-level Gini coefficients increase from 72% to 80% in Austria and from 76% to 80% in Germany. Figure 7 shows household-level Lorenz curves before and after adjusting for the missing wealthy. The Lorenz curve is read as follows: The poorest x% of the households hold y% of total wealth.

These findings are in-line with prior analyses based on HFCS data of the first wave,[49] which all find that adjusting the tail for the missing wealthy pushes up net worth and in particular the net worth at the top. Eckerstorfer et al. (2016) find for Austria that aggregate wealth increases by 28 percentage points. The share of the richest 1% increases by 15 percentage points to 38%. Vermeulen (2016) finds that the share of the top 1% increases by 6-7 percentage points in Germany and by 8-11 percentage points in Austria when adding observations from the Forbes list. Figure 8 plots the shift of the distribution when replacing the top tail by a Pareto model. For both countries, mainly the top 1-2% are affected.

Observations from rich lists can also be used to assess the success to estimate the wealth distribution. According to the rich lists, there are 31 billionaires in Austria (according to the Forbes list, in 2014 there are 11 and in 2015 only eight USD-billionaires) and 142 in Germany. In the survey, however, not a single billionaire is interviewed indicating that the survey really fails to capture this – in terms of aggregate wealth – very important part of the population. When drawing from the Pareto distribution, we create on average eight billionaires in Austria and 121 in Germany, which matches the number on the rich lists reasonably well.

Simulated observations can also be used to assess the effective oversampling rate of the wealthy. The effective oversampling rate gives an indication on how well the wealthy households are represented in the final sample in comparison to their share in the population (see also subsection 6.1). Austria is one of the very few countries that do not oversample at all. For Austria, the effective oversampling rate is even negative, which means that there are (substantially!) less observations in the top tail than implied by the sample design. The Pareto adjustment changes the distribution at the very top and pushes up the 99th percentile in the case of Austria. For Germany, the 99th percentile does not change – only the total wealth of the top 1% as well as higher percentiles. Because of this shift in Austria, the effective oversampling rate drops from -13% to -50% after the adjustment (see Table 8). This indicates that measuring the distribution of wealth in Austria would be much more reliable if an oversampling strategy would be introduced.

7.2 Effects on instrument level

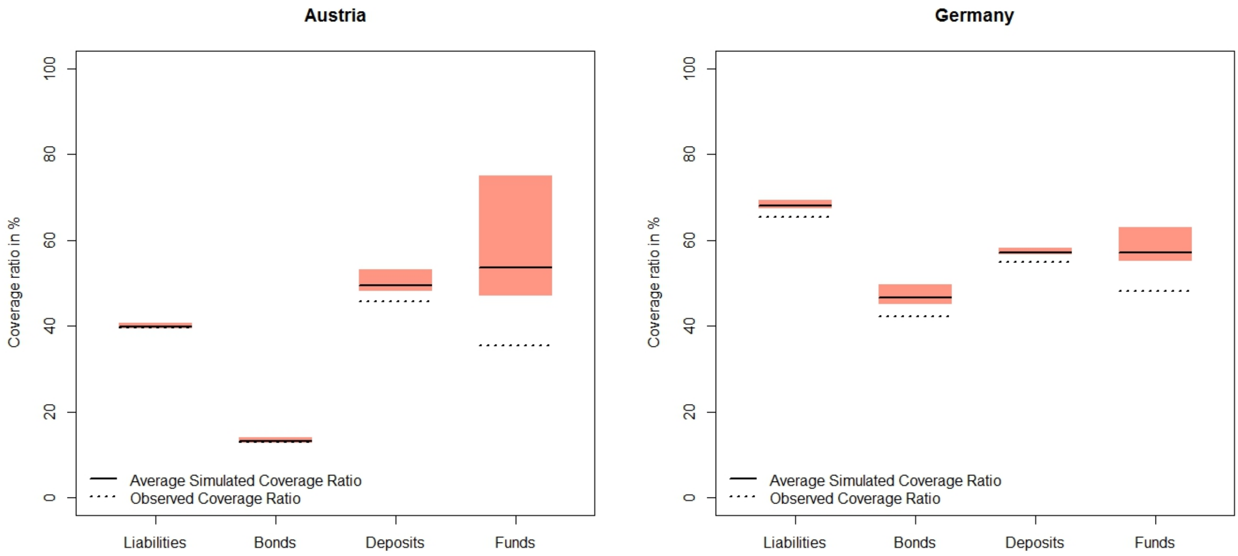
While the effect on net worth and the overall wealth distribution can be somewhat expected when Pareto estimates are used, the effect on instrument-specific coverage ratios and instrument-specific distributions is less straightforward. In general, we find that coverage ratios increase when substituting the top tail by a Pareto model thus confirming the hypothesis of under-representation of the most affluent households in the survey.

Generally, we report point estimates calculated via the analytical method. Any additional distributional information as well as confidence intervals are obtained from the simulation method.

Table 9 summarises the main results and Figure 9 graphically shows the changes in coverage ratios for highly comparable instruments. In general, coverage is higher for Germany than for Austria. All ratios increase due to our adjustment although the change is very low for liabilities and bonds in Austria.

[49] The fieldwork for the first wave was carried out in 2010 and 2011.

Figure 9: Coverage ratios.



Notes: The figures plot observed and adjusted coverage ratios. The boxes represent 95% bootstrap confidence intervals.

The impact on instrument-specific coverage ratios obviously correlates with the share of assets held by the top tail and the distribution within the tail. The impact on instruments that are less important for the wealthiest of the wealthy (deposits and bonds as well as liabilities) is lower, while the impact on assets which are largely held by the wealthy (funds and equity) is more pronounced.[50]

In general, we find very similar patterns for Austria and Germany, although our methodology leads to larger changes for Austria. This is likely due to the differences in the German and Austrian HFCS in terms of oversampling: Our methodology includes an ex-post adjustment for missing wealthy households. As Germany has implemented an oversampling strategy, outcomes are more likely to better reflect the upper part of the distribution than in Austria (see Fessler et al., 2016, page 29f for a comparison between the Austrian and German HFCS).

Changes are most pronounced for equity. Wealthier households hold much larger shares in equity than the rest of the population and this share increases even more in the top quarter of the tail. The EG-LMM has already documented coverage ratios exceeding 100% when relying on HFCS data only. Most likely, the reason is that there are substantial conceptual and methodological differences (see discussion in section 6). The Pareto adjustment of the wealth distribution pushes up these numbers even further. Thus, better aligning the HFCS and financial accounts definition of equity is needed when aiming for distributional indicators. It is important to note that also confidence intervals are largest for equity. Thus not surprisingly, the sensitivity analysis in Appendix C finds that also modelling choices have the largest impact on the results for equity.

We also calculate figures for *financial wealth* defined here as the sum of bonds, deposits, equity and mutual funds minus liabilities. Given that parts of equity are recorded as net positions in

[50]For Austria, we observe a large increase in the coverage ratio for mutual funds. The sensitivity analysis regarding the choice of Pareto threshold (see Appendix C) finds that the coverage ratio presented here might be overstated.

Table 9: Adjusted aggregates and coverage ratios.

AT – Austria

	Observed CR	Adjusted CR	95% confidence interval for CR		Change in CR
Liabilities	39.58	39.93	(39.62;	40.62)	0.35
Bonds	12.90	13.30	(13.02;	13.97)	0.40
Deposits	45.74	49.72	(48.30;	53.11)	3.98
Equity	155.71	349.59	(282.57;	489.45)	193.89
Mutual Funds	35.55	55.20	(47.35;	75.09)	19.65
Financial Wealth	95.28	194.7	(160.92;	265.91)	104.4

	Observed Aggregate	Adjusted Aggregate	95% confidence interval for Aggregate		Change in Aggregate
Liabilities	66,601	67,182	(66,670;	68,351)	0.87
Bonds	5,272	5,437	(5,319;	5,711)	3.13
Deposits	98,745	107,345	(104,271;	114,668)	8.71
Equity	194,739	437,234	(353,412;	612,152)	124.52
Mutual Funds	17,070	26,508	(22,738;	36,058)	55.29
Real Estate	686,174	829,044	(767,133;	957,610)	20.82
Financial Wealth	249,227	509,342	(420,918;	695,532)	97.33

DE – Germany

	Observed CR	Adjusted CR	95% confidence interval for CR		Change in CR
Liabilities	65.29	67.99	(67.50;	69.34)	2.70
Bonds	42.30	46.67	(45.30;	49.71)	4.37
Deposits	54.80	57.23	(56.77;	58.24)	2.44
Equity	264.03	363.12	(347.28;	403.72)	99.09
Mutual Funds	48.08	57.33	(55.35;	62.91)	9.25
Financial Wealth	115.50	155.23	(148.87;	171.40)	39.72

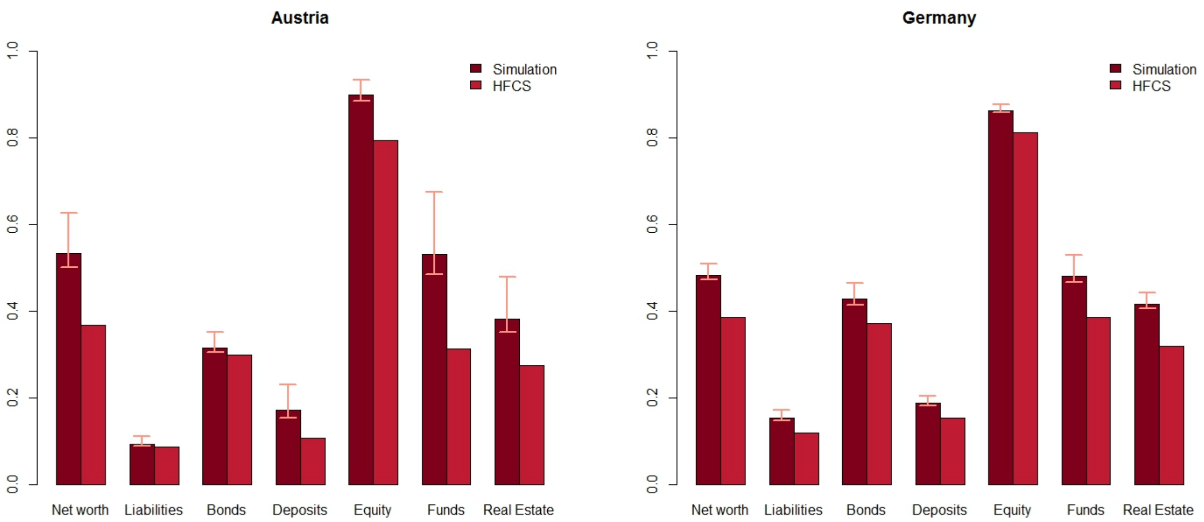
	Observed Aggregate	Adjusted Aggregate	95% confidence interval for Aggregate		Change in Aggregate
Liabilities	1,020,920	1,063,155	(1,055,488;	1,084,259)	4.14
Bonds	71,916	79,351	(77,022;	84,520)	10.34
Deposits	1,007,795	1,052,599	(1,044,172;	1,071,207)	4.45
Equity	1,326,078	1,823,723	(1,744,171;	2,027,623)	37.53
Mutual Funds	206,863	246,645	(238,152;	270,661)	19.23
Real Estate	5,872,317	6,859,346	(6,736,002;	7,167,085)	16.81
Financial Wealth	1,591,732	2,139,163	(2,051,511;	2,362,093)	34.39

Notes: The table reports observed and adjusted aggregates in mEUR, and coverage ratios (CR) in % per instrument. The change in coverage ratio is given in percentage points, the change in aggregates in %. The 95% confidence interval is obtained from the simulation method whereas point estimates are calculated via the analytical method. Financial wealth is defined as bonds plus deposits plus equity plus mutual funds minus liabilities.

the HFCS but spread over all instruments in the financial accounts (gross recording), we would expect high coverage ratios (or even over-coverage) for this level of aggregation as the net value in the HFCS could also include real assets. Adjusting for the missing wealthy indeed leads to over-coverage of financial wealth, which is mainly driven by the large increase of equity. For equity the differences in net and gross recording are not the only conceptual problem and for instance differences in valuation also play a crucial role.

One particular advantage of the simulation approach is that it allows us to calculate empirical confidence intervals. These intervals take into account the Pareto sampling as well as the bootstrap insecurity.[51] Thus, the primary results of our methodology are these *ranges together with point estimates* and not point estimates alone as they hide variability.

Figure 10: Break down of the wealth of millionaires.



Notes: The figures show the shares of total assets and total liabilities held by the tail population, i.e., by all millionaires. Simulation results also include 95%-confidence intervals.

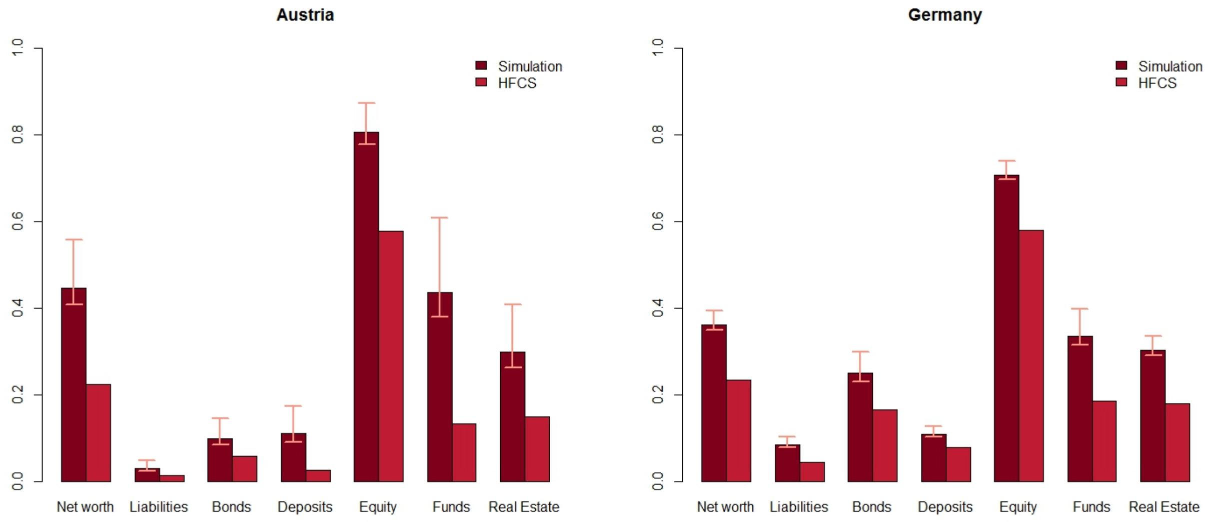
Figure 10 shows the share of instrument-specific aggregates held by the tail population, i.e., by all millionaires. Shares consistently increase as compared to shares calculated from HFCS data only. Changes are least pronounced for liabilities, bonds and deposits. We find that millionaires in Austria and Germany hold roughly 90% of total equity.

Similarly, Figure 11 shows the share of instrument-specific aggregates held by the top 1% of the population. Again, shares strongly increase when applying our methodology. In comparison to the break down for millionaires (see Figure 10), changes are even more pronounced for equity and real estate for the top 1% as these asset classes are most important for the richest of the rich. While millionaires hold roughly 90% of total equity, the top 1% still hold roughly between 70% and 80% revealing an extremely high degree of wealth inequality for this particular asset class.

The strength of the analytical method is to provide accurate point estimates, whereas the simulation method provides additional measures on the variability of the point estimates (i.e. confidence intervals). Of course, the simulation method could also be used to calculate point

[51] However, the intervals do not take into account any insecurity attached to the choice of methodology to estimate the Pareto shape parameter, or other sources of insecurity resulting from, for instance, multiple imputation and sample design.

Figure 11: Break down of the wealth of the top 1%.



Notes: The figures show the shares of total assets and total liabilities held by the top 1% of the population. Simulation results also include 95%-confidence intervals.

estimates and – once convergence is reached – there should be only minor differences between aggregates obtained via either approach. This is indeed the case (see Table 13 in the appendix).

8 Distributional National Accounts

In this section we give an outlook how our analysis can contribute to enhance the compilation of distributional national accounts (DINA).

The aim of DINA is to compliment national accounts aggregates by distributional indicators. In general, there are many different indicators that would add useful information. Here, we focus on indicators based on the distribution of net worth as we can directly make use of our methodology for this exercise.[52]

We calculate an indicator that splits up financial accounts aggregates by n net worth groups. For instance, $n = 5$ when compiling a spilt-up by net worth quintiles. Such indicators show how much of the total aggregate of a certain instrument is held by for instance the poorest 20% or wealthiest 20% of the population in terms of net worth.

Let y_j denote the FA aggregate for instrument j and

$$x_{\cdot j} = \sum_{i=1}^{n} x_{ij}$$

the HFCS aggregate of instrument j. Thereby, x_{ij} is the HFCS aggregate for instrument j within the net worth group i.

Without adjusting for the missing wealthy in the HFCS, distributional indicators for instrument

[52] Other indicators may include a split up by income, household structure, or other characteristics. Here we split up aggregates on a household level, i.e., net worth as well as the instruments to be broken down are measured for households, which enables a straightforward comparison. In the case of income or other characteristics that are associated with individuals rather than households, the additional challenge of harmonising the unit of measurement emerges.

j and net worth group i are given by

$$d_{ij} = \frac{x_{ij}}{x_{\cdot j}} \cdot y_j.$$

Note that each x_{ij} is scaled up by the inverse instrument-specific coverage ratio. The scaling guarantees that values over all net worth groups add up to the financial accounts total:

$$\sum_{i=1}^{n} d_{ij} = \sum_{i=1}^{n} \frac{x_{ij}}{x_{\cdot j}} \cdot y_j = \frac{y_j}{x_{\cdot j}} \cdot \sum_{i=1}^{n} x_{ij} = \frac{y_j}{x_{\cdot j}} \cdot x_{\cdot j} = y_j.$$

Taking into account the adjustments for the missing wealthy, the HFCS aggregates $x_{\cdot j}$ change whereas the FA aggregates y_j remain constant. To ease notation, assume that by the adjustment, only the highest net worth group is affected, i.e., group n.[53] Let x^*_{nj} denote the adjusted HFCS aggregate for instrument j and net worth group n. The adjusted HFCS total is then given by

$$x^*_{\cdot j} = x^*_{nj} + \sum_{i=1}^{n-1} x_{ij}.$$

Note that in general $x^*_{\cdot j} > x_{\cdot j}$, as our methodology leads to larger amounts held by the wealthy.

Distributional indicators are defined analogously via

$$d^*_{ij} = \begin{cases} \frac{x_{ij}}{x^*_{\cdot j}} \cdot y_j, & \text{for } 1 \leq i < n, \\ \frac{x^*_{ij}}{x^*_{\cdot j}} \cdot y_j, & \text{for } i = n. \end{cases}$$

Again, the values over all net worth groups add up to the financial accounts total:

$$\sum_{i=1}^{n} d^*_{ij} = \frac{x^*_{nj}}{x^*_{\cdot j}} y_j + \sum_{i=1}^{n-1} \frac{x_{ij}}{x^*_{\cdot j}} \cdot y_j = \frac{y_j}{x^*_{\cdot j}} \left(x^*_{nj} + \sum_{i=1}^{n-1} x_{ij} \right) = \frac{y_j}{x^*_{\cdot j}} \cdot x^*_{\cdot j} = y_j.$$

Our adjustment has the following effect: For the unadjusted indicators d_{ij} the gap between HFCS and FA aggregates is "filled up" by equally scaling up the respective numbers by the inverse coverage ratio. This means that the "missing" portions are distributed proportionally across wealth groups. Our methodology quantifies the contribution of the missing wealthy to the total gap. Thus, this information is used to allocate this portion of the FA aggregate directly to the top quintile. To fill the remaining gap, aggregates are then again scaled up proportionally.

In theory, if one knew to which net worth quintile to allocate the remaining gap, no scaling would be necessary at all.

Figure 12 shows results graphically.[54] As we only adjust for the wealthy, it is the top quintile that gets much larger weights. At the same time, all other quintiles "loose" (as less Euros are distributed to them due to scaling).[55]

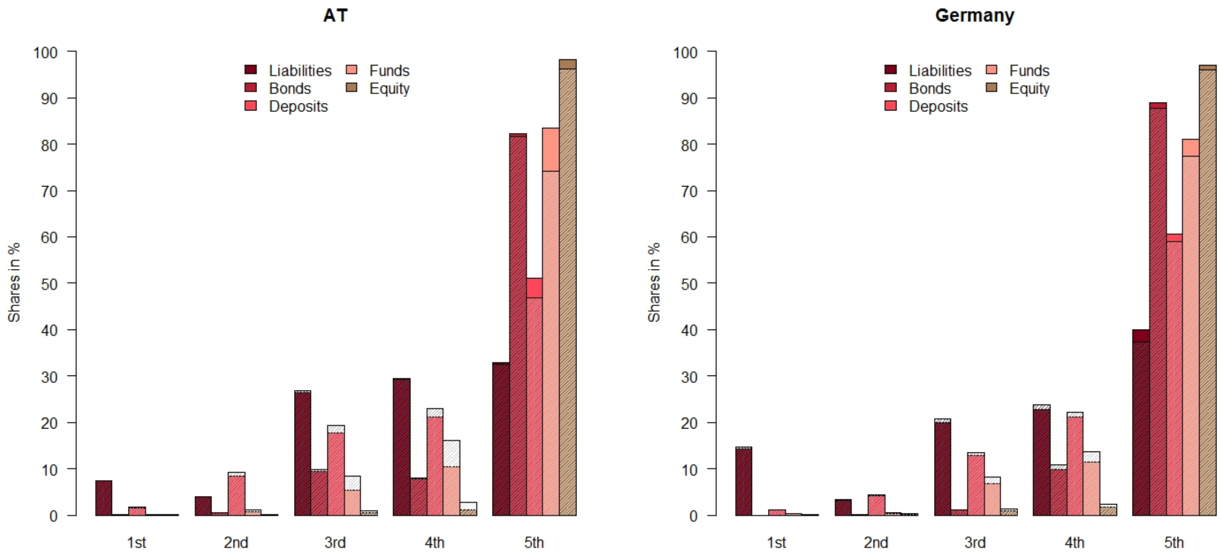
In general, we find very high degrees of instrument-specific inequality which is most dramatic for equity. Inequality is measured to be even larger when applying our corrections. For instance,

[53] For broad indicators such as a split-up by quartiles or quintiles, indeed only the highest group is affected. Changes are, however, quite obvious for more granular split-ups.

[54] Table 15 and Table 16 in the appendix report full results as share and Euro amounts, respectively.

[55] Note that this is a mechanical effect that holds true whenever adjusting for the missing wealthy leads to larger instrument-specific aggregates in the top net worth group. The proof of this statement is given in Appendix D.

Figure 12: Distributional National Accounts.



Notes: The figures show shares of total instrument-specific aggregates held by each net worth quintile (see Table 15). Shaded bars indicate shares without adjustments, i.e., d_{ij}/y_j, whereas full bars indicate adjusted shares, i.e., d^*_{ij}/y_j.

the share of total mutual funds held by the top quintile is adjusted upwards from 74.2% to 83.4% in Austria and from 77.4% to 81.1% in Germany.

Instrument-specific quintile shares are very similar between Germany and Austria, i.e., we find similar degrees of inequality for both countries *even on instrument-level*. Differences are larger *before* adjusting for the missing wealthy than *after* adjustment. The difference, measured by the sum of squared differences, decreases from

$$\sum_{ij}\left(\frac{d_{ij}(AT)}{y_j(AT)} - \frac{d_{ij}(DE)}{y_j(DE)}\right)^2 = 0.095 \quad \text{to} \quad \sum_{ij}\left(\frac{d^*_{ij}(AT)}{y_j(AT)} - \frac{d^*_{ij}(DE)}{y_j(DE)}\right)^2 = 0.045.$$

This finding suggests again, that differences between Austria and Germany in terms of wealth inequality may, at least to a certain extent, be driven by the different treatment of the wealthiest households in the survey: oversampling versus non-oversampling. It, however, also shows that our methodology is likely to overcome parts of this shortcoming.

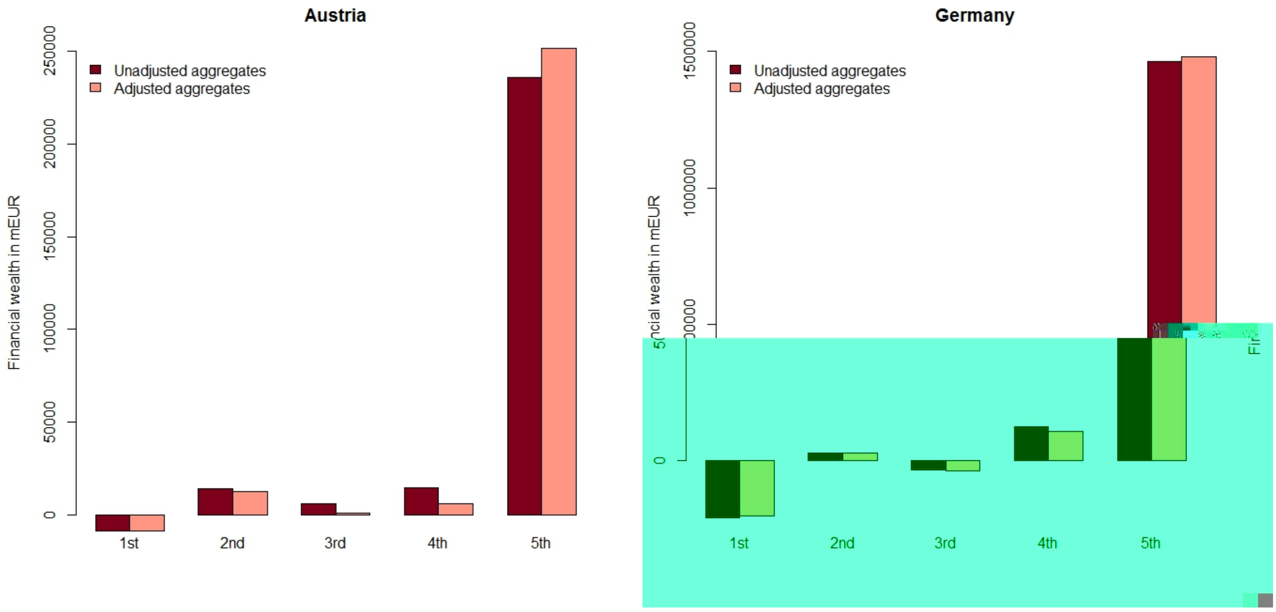
We also calculate distributional figures for *total financial wealth*[56] split up by the same net worth quintiles as before, i.e., we summed up financial assets for each net worth quintile (bonds, deposits, mutual funds and equity) and deducted quintile-specific liabilities.

A complete set of adjusted DINA figures is provided in Table 10.

The split-up is fully consistent with financial accounts aggregates: Summing over instrument-specific split-ups leads to financial accounts aggregates (as a direct consequence of scaling) and likewise summing over the financial wealth split-up yields total financial wealth as reported by financial accounts.

[56]The definition of financial wealth sometimes includes pension or insurance entitlements, which we exclude in our analysis due to low conceptual comparability between the HFCS and financial accounts.

Figure 13: Distributional figures for Financial Wealth.



Notes: The figures show adjusted and unadjusted total financial wealth (bonds plus deposits plus investment funds plus equity minus liabilities) split up by net worth quintiles. Totals are in mEUR.

Table 10: Consistency of Distributional National Accounts.

AT – Austria						
	1st	2nd	3rd	4th	5th	Total
-Liabilities	-12,361	-6,570	-44,605	-49,252	-55,481	-168,269
Bonds	69	222	3,855	3,149	33,573	40,868
Deposits	3,314	18,224	38,320	45,602	110,435	215,895
Equity	38	31	528	1,500	122,972	125,069
Mutual Funds	46	333	2,592	4,993	40,054	48,018
Financial Wealth	-8,894	12,239	691	5,993	251,552	261,581

DE – Germany						
	1st	2nd	3rd	4th	5th	Total
-Liabilities	-221,750	-50,051	-311,046	-356,566	-624,171	-1,563,584
Bonds	14	326	1,764	16,764	151,152	170,019
Deposits	20,326	75,095	238,026	389,579	1,116,170	1,839,197
Equity	342	886	4,741	8,706	487,563	502,237
Mutual Funds	1,256	1,561	29,472	49,102	348,842	430,233
Financial Wealth	-199,811	27,816	-37,044	107,584	1,479,556	1,378,102

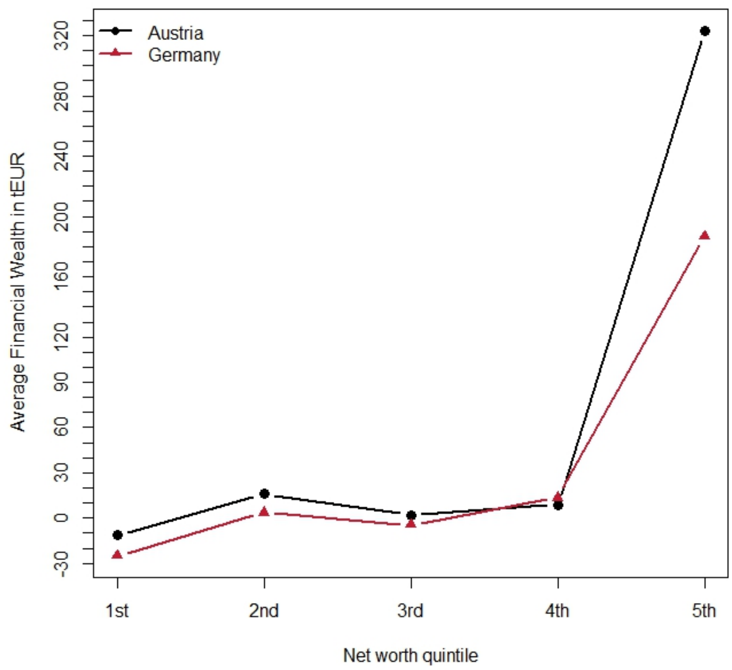
Notes: The tables report *adjusted* distributional national accounts figures in mEUR consistent with financial accounts totals (as reported in the last column).

Figure 13 shows the result. In general, the top quintile holds the lion's share of total financial wealth whereas the bottom quintile has negative financial wealth in Austria and Germany. In both countries the 3rd quintile has less financial wealth compared to the 2nd quintile (in

Germany the 3rd quintile has even negative financial wealth). This is due to the fact that this part of the population to a large extend substitutes financial wealth with non-financial wealth, i.e., they own more real estate (non-financial assets) that are often financed via mortgages (financial liabilities).

Our adjustment even further increases financial wealth for the top quintile while reducing financial wealth of all other quintiles. In Germany, we find that the top quintile holds 1,480 billion EUR in financial wealth whereas the 4th quintile holds 108 billion EUR only. In Austria, the top quintile holds 250 billion EUR compared to roughly 6 billion EUR held by the 4th quintile.

Figure 14: Average Financial Wealth per quintile.



Notes: The figures shows the average financial wealth per quintile in thousand EUR. Numbers are adjusted for the missing wealthy. Financial wealth is defined here as bonds plus deposits plus mutual funds plus equity minus liabilities. It does not include any pension or insurance entitlements.

Figure 14 shows average financial wealth per net worth quintile. In theory, each quintile consists of the same number of households (namely 20% of total households), however, due to survey weights the number of households per quintile varies to some extend.[57] The average financial wealth per household varies strongly over the distribution: In Austria (Germany), a household belonging to the lowest quintile has on average -11,506 EUR (-25,127 EUR) compared to 325,920 EUR (186,789 EUR) for a household belonging to the top quintile.

9 Conclusions

Macroeconomic aggregates are compiled in the system of national accounts which financial accounts form an important part of. Financial accounts are expected to be exhaustive and

[57]For Austria the number of households per quintile ranges between 771,819 and 772,994, and in Germany between 7,920,061 and 7,951,902.

thus provide an important source of information on the entire economy. On the other side, financial accounts aggregates do not provide any details about the distribution of assets and liabilities within the population. This gap is filled by making use of micro data collected from administrative sources or surveys. This article demonstrates how to use the *Household Finance and Consumption Survey* (HFCS) for such an endeavour.

In an ideal world, without measurement errors, ambiguities in valuation, identical definitions of variables in the survey and financial accounts, a system of national accounts without disturbing balancing effects and a perfect survey sample, we would in expectation get identical aggregates when calculated from either source, i.e., a coverage ratio of 100%. In the real world, coverage ratios are, however, far from perfect.

To achieve the long-term goal of compiling distributional national accounts (DINA) that provide distributional information consistent with aggregates, we need to understand where this gap comes from. This article adds to the literature with this regard as it aims to quantify the impact of the missing rich on the gap between HFCS and financial accounts aggregates, and demonstrates how such a quantification can be used in the compilation of DINA.

We make use of the second wave of the HFCS and analyse the impact of the missing rich in the HFCS on conceptually highly comparable instruments (liabilities, bonds, deposits, and mutual funds).

We perform a case study for Austria and Germany as these countries do not have access to administrative data to strategically oversample the wealthiest households in the HFCS nor to perform plausibility checks and corrections on self-reported survey data. We hence would expect the largest effects for these countries. Whereas Germany uses geographical data for oversampling, Austria completely refrains from oversampling. We find that the Pareto adjustment is much larger for Austria than for Germany indicating that lacking an oversampling strategy decreases the reliability of the Austrian HFCS.

Previous findings in the literature suggest that the HFCS underestimates net worth at the top of the distribution. We therefore adjust the distribution by replacing the top tail by a Pareto model. The Pareto model is estimated based on a combined sample of observations from the HFCS and national rich lists.

Additionally, we use portfolio structures observed in the HFCS to break down the Pareto-adjusted wealth at the top tail by instruments. For this purpose, we propose an analytical as well as a simulation approach based on stratified bootstrapping. Whereas the analytical approach is fast and easy-to-implement, the simulation provides additional valuable information on variation and distributional patterns within the tail. Results for aggregates are identical as the simulation method converges to the analytical approach.

We find that coverage ratios consistently increase when explicitly adjusting survey data to better reflect the top tail of the net worth distribution. Changes vary by instrument and are most pronounced for (the conceptually not well comparable item) equity. Although the "missing wealthy" explain up to 20% (but usually less than 10%) of the gap between HFCS and financial accounts aggregates, there still remains a substantial gap which source needs further exploration. Our findings for equity point towards the urgent need to better align the HFCS definition of equity with the globally agreed definition in the financial accounts. At the same time, the appropriateness of the valuation concept of equity in the financial accounts (which is often based on book values) should be reconsidered.

We conduct extensive robustness checks which find that the most crucial assumption in our analysis is the choice of Pareto estimation method. The exact treatment of the rich list is also important. Less influential is the choice of the Pareto threshold and the inclusion of further portfolios to better represent portfolio structures of the wealthiest of the wealthy households. However, as we only rely on portfolio structures collected in the HFCS, results could in fact be even more extreme in the sense that equity could be even larger. We also find that refraining from stratification yields unreliable results, which emphasises the importance of a more sophisticated approach than just taking simple average as proposed in this article.

We compile distributional account figures and show how our methodology can be used to enhance them. We detect large instrument-specific inequality and find that the top quintile holds the lion's share of financial wealth whereas the bottom quintile has negative financial wealth. In Austria (Germany), a household belonging to the lowest quintile has on average $-11,342$ EUR ($-25,127$ EUR) compared to $323,397$ EUR ($186,789$ EUR) for a household belonging to the top quintile. In general, patterns are very similar in Austria and Germany.

As a by-product, we gain new estimates on the distribution of net worth. In-line with prior findings in the literature, we find that the HFCS substantially underestimates wealth inequality in Austria and Germany. Our analysis suggests that wealth inequality measured by a household-level Gini coefficient increase from 72% in Austria and from 76% in Germany to roughly 80% in both countries.

In terms of instrument-level inequality, we find that the wealthiest 1% of the population owns 70-80% of total equity and roughly 30% of total real estate assets, but less than 10% of total liabilities. The top wealth quintile holds roughly 98% of total equity in Austria and 97% in Germany, whereas the bottom three wealth quintiles together, i.e., the lowest 60% hold practically nothing in both countries. We also find very high degrees of instrument-specific inequality for mutual funds and bonds.

References

Alfons, A. and Templ, M. (2013). Estimation of social exclusion indicators from complex surveys: The R package laeken. *Journal of Statistical Software*, 54(15):1–25.

Alfons, A., Templ, M., and Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(2):271–286.

Alvaredo, F., Atkinson, A., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2016). Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID.world. *WID.world Working Paper Series*, No. 2016/1.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). Global inequality dynamics: New findings from WID.world. *NBER Working Paper*, No. 23119.

Andreasch, M., Fessler, P., and Schürz, M. (2015). Financing the Austrian economy – a bird's eye view based on the financial accounts from 1995 to 2014 and a look at the road ahead. *Monetary Policy & the Economy*, 4:55–73.

Andreasch, M. and Lindner, P. (2016). Micro- and macrodata: A comparison of the Household Finance and Consumption Survey with financial accounts in Austria. *Journal of Official Statistics*, 32(1):1–28.

Arrondel, L., Bartiloro, L., Fessler, P., Lindner, P., Marthä, T. Y., Rampazzi, C., Savignac, F., Schmidt, T., Schürz, M., and Vermeulen, P. (2014). How do households allocate their assets? Stylised facts from the Eurosystem Household Finance and Consumption Survey. *ECB Working Paper Series*, No. 1722/August 2014.

Atkinson, A. B. (1975). The distribution of wealth in Britain in the 1960s the estate duty method reexamined. In *The Personal Distribution of Income and Wealth*, pages 277–328. NBER.

Atkinson, A. B. and Harrison, A. J. (1978). *Distribution of personal wealth in Britain.* Cambridge University Press.

Avery, R. B. and Elliehausen, G. E. (1986). Financial characteristics of high-income families. *Federal Reserve Bulletin*, pages 163 – 177.

Avery, R. B., Elliehausen, G. E., and Kennickell, A. B. (1988). Measuring wealth with survey data: An evaluation of the 1983 Survey of Consumer Finances. *Review of Income and Wealth*, 34(4):339–369.

Bach, S., Beznoska, M., and Steiner, V. (2014). A wealth tax on the rich to bring down public debt? Revenue and distributional effects of a capital levy in Germany. *Fiscal Studies*, 35(1):67–89.

Bach, S., Thiemann, A., and Zucco, A. (2015). The top tail of the wealth distribution in Germany, France, Spain, and Greece. *DIW Discussion Papers*, No. 1502.

Baranyai-Csirmaz, R., Huszár, G., Huszti, E., Kozmits, I., Németné, K. M., Simon, B., and Valentiny, A. (2017). *Financial savings of households based on micro- and macro-statistical data.* Magyar Nemzeti Bank, Budapest.

Bover, O., Coronado, E., and Velilla, P. (2014). The Spanish Survey of Household Finances (EFF): description and methods of the 2011 wave. *Banco de España Documentos Ocasionales*, No. 1407.

Capgemini (2016). *World Wealth Report.* New York.

Chakraborty, R., Kavonius, I. K., Pérez-Duarte, S., and Vermeulen, P. (2016). Is the top tail of the wealth distribution the missing link between the Household Finance and Consumption Survey and National Accounts. *Paper prepared for the 34th IARIW General Conference.*

Cowell, F. (2011a). Inequality among the wealthy. *LSE STICERD Research Paper*, No. CASE150.

Cowell, F. A. (2011b). *Measuring Inequality.* Oxford University Press, Oxford, 3rd edition.

Cowell, F. A. and van Kerm, P. (2015). Wealth inequality: A survey. *Journal of Economic Surveys*, 29(4):671–710.

Dalitz, C. (2016). Estimating wealth distribution: Top tail and inequality. *Technische Berichte des Fachbereichs Elektrotechnik und Informatik, Hochschule Niederrhein*, No. 2016-01.

D'Aurizio, L., Faiella, I., Iezzi, S., and Neri, A. (2006). L'under-reporting della ricchezza finanziaria nell'indagine sui bilanci delle famiglie. *Banca D'Italia Temi di discussione*, No. 610/December 2006.

Davies, J. B. and Shorrocks, A. F. (2000). The distribution of wealth. *Handbook of Income Distribution*, 1:605–675.

Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F., and Wildauer, R. (2016). Correcting for the missing rich: An application to wealth survey data. *Review of Income and Wealth*, 62(4):605–627.

EG-LMM (2017). Understanding, quantifying and explaining the differences between macro and micro data of household wealth. *Mimeo.*

Fesseau, M. and Mattonetti, M. L. (2013). Distributional measures across household groups in a national accounts framework. *OECD Statistics Working Papers*, No. 2013/04.

Fessler, P., Lindner, P., and Scürz, M. (2016). Eurosystem Household Finance and Consumption Survey 2014. First results for Austria (second wave).

Forbes (2015). The world's billionaires. *Forbes*, 195(4):99–164.

Freund, C. and Oliver, S. (2016). The origins of the superrich: The billionaire characteristics database. *Peterson Institute for International Economics Working Paper Series*, 16-1.

Gabaix, X. and Ibragimov, R. (2011). Rank – 1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Henriques, A. M. and Hsu, J. W. (2014). *Analysis of Wealth Using Micro- and Macrodata: A Comparison of the Survey of Consumer Finances and Flow of Funds Accounts*, pages 245 – 274. University of Chicago Press.

HFCN (2016a). The Household Finance and Consumption Survey: methodological report for the second wave. *ECB Statistics Paper Series*, No. 17/December 2016.

HFCN (2016b). The Household Finance and Consumption Survey: results from the second wave. *ECB Statistics Paper Series*, No. 18/December 2016.

Kavonius, I. K. and Honkkila, J. (2013). Reconciling micro and macro data on household wealth: A test based on three euro area countries. *Journal of Economic & Social Policy*, 15(2):Article 3.

Kavonius, I. K. and Törmälehto, V.-M. (2010). Integrating micro and macro accounts – the linkages between euro area household wealth survey and aggregate balance sheets for households. *Paper prepared for the IARIW 31st General Conference, St. Gallen, Switzerland, August 22-28.*

Kennickell, A. B. (1999a). Analysis of nonresponse effects in the 1995 Survey of Consumer Finances. *Journal of Official Statistics*, 15(2):283–303.

Kennickell, A. B. (1999b). Using income data to predict wealth. *Board of Governors of the Federal Reserve System Working Paper.*

Kennickell, A. B. (2008). The role of over-sampling of the wealthy in the survey of consumer finances. *Irving Fisher Committee Bulletin*, 28:403–408.

Kennickell, A. B. and Woodburn, R. L. (1999). Consistent weight design for the 1989, 1992 and 1995 SCFs, and the distribution of wealth. *Review of Income and Wealth*, 45(2):193–215.

Klass, O. S., Biham, O., Levy, M., Malcai, O., and Solomon, S. (2006). The Forbes 400 and the Pareto wealth distribution. *Economics Letters*, 90(2):290 – 295.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 45(1):1–28.

manager magazine (2014). Die reichsten Deutschen. Top 500. *manager magazin SPEZIAL*, October:16–30.

Osier, G. (2016). Unit non-response in household wealth surveys: Experience from the Eurosystem's Household Finance and Consumption Survey. *ECB Statistics Paper Series*, No. 15 / July 2016.

Pareto, V. (1895). La legge della domanda. *Giornale degli Economisti*, 10 (Anno 6):59–68.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T., Postel-Vinay, G., and Rosenthal, J.-L. (2006). Wealth concentration in a developing economy: Paris and France, 1807-1994. *American Economic Review*, 96(1):236 – 256.

Piketty, T., Saez, E., and Zucman, G. (2016). Distributional National Accounts: methods and estimates for the United States. *NBER Working Paper*.

Piketty, T. and Zucman, G. (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics*, 129(3):1255–1310.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schmidt, T. and Eisele, M. (2013). Oversampling vermögender Haushalte im Rahmen der Studie "Private Haushalte und ihre Finanzen (PHF)". *Deutsche Bundesbank*.

Schneider, H., Millner, R., and Meyer, M. (2010). Die Rolle der Gemeinnützigkeit in österreichischen Stiftungen. *Institute for Nonprofit Management Working Papers, WU Vienna University of Economics and Business*, 2010/1.

Törmälehto, V.-M. (2017). High income and affluence: Evidence from the European Union statistics on income and living conditions (EU-SILC). *EUROSTAT Statistical Working Papers*, 2017 edition.

trend (2015). Die trend Reichenliste 2015. Österreichs Geldadel in Zahlen. *trend*, June:22–24.

van der Wijk, J. (1939). Inkomens- en vermogensverdeling. *Publication of the Nederlandsch Economisch Instituut, Haarlem*, Vol. 26.

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

Vermeulen, P. (2017). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, forthcoming.

Wöhlbier, F. (2014). Lessons from the 2014 report "Tax reforms in EU member states". In Astarita, C., editor, *Taxing Wealth: Past, Present, Future. Proceedings of the workshop organised by the Directorate General for Economic and Financial Affairs held in Brussels on 13 November 2014*, pages 12–15.

Zucman, G. (2013). The missing wealth of nations: Are Europe and the US net debtors or net creditors? *The Quarterly Journal of Economics*, 128(3):1321–1364.

Appendix

A The Pareto distribution

A.1 Estimating the shape parameter

A.1.1 (Robust) regression method

The Pareto shape parameter is often estimated using ordinary least squares (OLS), which exploits the linear relationship between the Pareto distribution and the rank of observations: The (shifted) log rank of an observation is a downwards-sloping function of logged wealth with slope parameter ϑ, which is a direct consequence of (1):

$$P(Y > y) = \left(\frac{y}{y_0}\right)^{-\vartheta}$$

and thus

$$\log P(Y > y) = \vartheta \log y_0 - \vartheta \log y.$$

In a finite sample, observations approximately follow a Pareto distribution, if

$$\frac{i}{n} \approx \left(\frac{y_i}{y_0}\right)^{-\vartheta} \qquad \text{for all } i.$$

Thereby, i denotes the rank of the decreasingly ordered observations, y_i net worth of observation i, and n the number of observations. Thus,

$$\log i \approx C - \vartheta \log y_i,$$

with $C = \log(n) + \vartheta \log(y_0)$. The shape parameter ϑ could (in the absence of survey weights) thus be obtained by estimating this equation using OLS.

Vermeulen (2017) adapts this method to take into account survey weights. Additionally, Gabaix and Ibragimov (2011) showed that the estimator for ϑ is unbiased (up to order one) when shifting the rank by 0.5. An estimate for ϑ is then obtained by estimating the equation

$$\log\left((i - 0.5)\frac{\bar{N}_{fi}}{\bar{N}}\right) = C - \vartheta \log(y_i), \tag{3}$$

through a linear model. Thereby, $\bar{N} = \frac{1}{n}\sum_{j=1}^{n} w_j$ is the average weight of all observations and $\bar{N}_{fi} = \frac{1}{i}\sum_{j=1}^{i} w_j$ the average weight of the first i observations.

The major advantage of this approach is that additional observations such as those from a rich list can directly be included in the estimation process. As the response variable is a transformation of the rank of an observation but not its amount, extreme observations from the rich list are less problematic *in the statistical sense.* Still, the different weight structure remains problematic. (Remember that rich list observations are assigned a weight equal to 1 while survey observations may have very large weights. In Germany, the largest weight among all tail observation is 44,189 meaning that this one household represents this number of households).

Still, the two types of data (survey and rich list) lead to non-random patterns in residuals. Therefore, we propose a *robust alternative* that can be interpreted as a benchmark method. The alternative is robust to violations of the standard OLS assumptions. If, which is the case for our data, differences between the standard regression and robust regression are minimal, confidence in the standard regression is increased.

Quantile regression (which has been developed by Koenker and Bassett Jr, 1978, and is since then widely used in statistics and econometrics) may be used as such a robust alternative to OLS. Quantile regression is per construction less sensitive toward outliers and, in contrast to OLS, does not rely on any distributional assumptions.

Similar as to OLS where the *conditional mean* is estimated, quantile regression aims to estimate *conditional quantiles* including the *conditional median*.

For OLS, a linear conditional mean function

$$E[Y|X = x] = x^\top \beta$$

is estimated by solving the *quadratic optimisation problem*

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2.$$

In a quantile regression setting, the linear conditional median function

$$Q_Y(0.5|X = x) = x^\top \beta(0.5)$$

is estimated by solving the *linear optimisation problem*

$$\hat{\beta}(0.5) = \arg\min_{\beta} \sum_{i=1}^{n} |y_i - x_i^\top \beta|.$$

Thus, quantile regression substitutes a quadratic optimisation problem by a linear optimisation problem. For quantile regression, only the sign of a residual but not its magnitude matters whereas for ordinary least squares the magnitudes are crucial. Hence, quantile regression is robust against outliers whereas OLS is not.

Thus, we use a median quantile regression model to estimate the slope parameter ϑ in equation (3) and call this the *robust regression method*. In contrary, we refer to Vermeulen's method based on OLS as the *regression method*.

A.1.2 Weighted robust estimators

Alfons et al. (2013) recommend the usage of weighted robust estimation procedures for semi-parametric Pareto models based on complex surveys and apply their estimators to income data obtained from the *European Union Statistics on Income and Living Conditions* (EU-SILC). They specifically provide weighted versions of the *integrated squared error* (ISE) estimator and the *partial density component* (PDC) estimator. Details regarding their methodology and further references are found in Alfons et al. (2013).

A drawback of their methodology is that rich list observations are detected as outliers and – as the procedures are designed to be robust against outliers – rich list observations are less influential then needed.

A.2 Pareto quantiles

Quantiles of the Pareto distribution are given by the inverse of the CDF. For a specific quantile level q

$$\begin{aligned} F_Y(y) &= 1 - \left(\frac{y}{y_0}\right)^{-\vartheta} = q \quad \text{and thus} \\ F_Y^{-1}(q) &= y_0 \cdot (1-q)^{-1/\vartheta} \end{aligned}$$

holds.

Specifically, for the three quartiles we have

$$\begin{aligned} F_Y^{-1}(0.25) &= y_0 \cdot 0.75^{-1/\vartheta}, \\ F_Y^{-1}(0.50) &= y_0 \cdot 0.50^{-1/\vartheta}, \quad \text{and} \\ F_Y^{-1}(0.75) &= y_0 \cdot 0.25^{-1/\vartheta}. \end{aligned}$$

A.3 Sampling from a Pareto distribution

Sampling from a Pareto distribution is done here by exploiting the inverse probability integral transformation. This theorem states that, given a random variable Y with CDF F_Y, the random variable $Z = F_Y(y)$ follows a uniform distribution $U(0,1)$. Inversely, if $Z \sim U(0,1)$ and $Y \sim F_Y$, then Y and $F_Y^{-1}(Z)$ follow the same distribution.

Thus, it is possible to construct Pareto distributed random numbers from a simple-to-generate sample of uniformly distributed random numbers just by applying the inverse Pareto CDF. Given a uniformly distributed random sample $u_i,\ i = 1, \ldots, n$,

$$y_i = \hat{F}_Y^{-1}(u_i) = \frac{y_0}{(1-u_i)^{1/\hat{\vartheta}}} = \frac{y_0}{u_i^{1/\hat{\vartheta}}}$$

represents a random sample from a Pareto distribution with $\hat{\vartheta}$ and y_0. The last equation is due to symmetry in the interval $(0,1)$.

B Portfolio structures

This section demonstrates how portfolio structures change with net worth and total assets, respectively. We show results based on the HFCS second wave for a series of countries.

The figures are calculated as follows: Observations are increasingly ordered by net worth (total assets) and shares of total assets are calculated for rolling windows. For each window, we calculate the share of total assets (aggregate net worth plus total liabilities) held in different asset classes (y-axis). When calculating totals, we respect sample weights. Additionally, we compute the average net worth (total assets) per window which is assigned to a net worth quantile by empirically evaluating the country's *empirical weighted conditional distribution function of net worth (total assets)* (x-axis). We use a fixed step size of 10 observations and country specific window sizes ranging between 150 observations for Cyprus, 200 for Slovenia, 500 for Austria, Finland and Spain, and 700 for Germany.[58] As we do not include other assets such as vehicles or jewellery, shares do not add up to one.

[58] Due to rather large window sizes (which guarantee smooth curves), we lack information for the lowest quantiles.

We find similar structural patterns across countries which are shown in Figure 15 and Figure 16: When analysing shares of total assets broken down by total assets quantiles (i.e., Figure 15), we find two kinds of substitution effects: First, the low end of the total assets distribution hardly holds real estate assets whereas large shares of total assets are held as deposits. When moving up along the distribution, real estate become the single most important asset class. Only at the very top of the distribution, real estate loose importance to assets held in the form of equity – a second substitution effect.

The point where real estate become more important than deposits is quite different across countries. For Austria and Germany – both countries with traditionally very low home-ownership rates[59] – this fix point is roughly at the 45th percentile of the total assets distribution (for Austria it is the 46th percentile which roughly equals EUR 68,300 and for Germany the 43th percentile which roughly equals EUR 49,700). For all other countries, the fix point is reached much earlier.[60] Also the maximum real estate share is noticeably larger for countries other than Austria and Germany.

When looking at shares of total assets broken down by the net worth distribution (i.e., Figure 16), we find very similar patterns for the top tail as liabilities are less important in this part of the distribution. However, at the low end liabilities do play an important role leading to larger real estate shares at the low end of the distribution.

[59]The 2015 home-ownership rate was 55.7% in Austria and 51.8% in Germany compared to for instance 78.2% in Spain and 72.7% in Finland (*Source:* Eurostat).

[60]We find a strong negative correlation (Pearson correlation coefficient: -0.92) between the fix point and the home-ownership rate among all countries participating in the second HFCS wave.

Figure 15: Shares of total assets for total assets distribution.

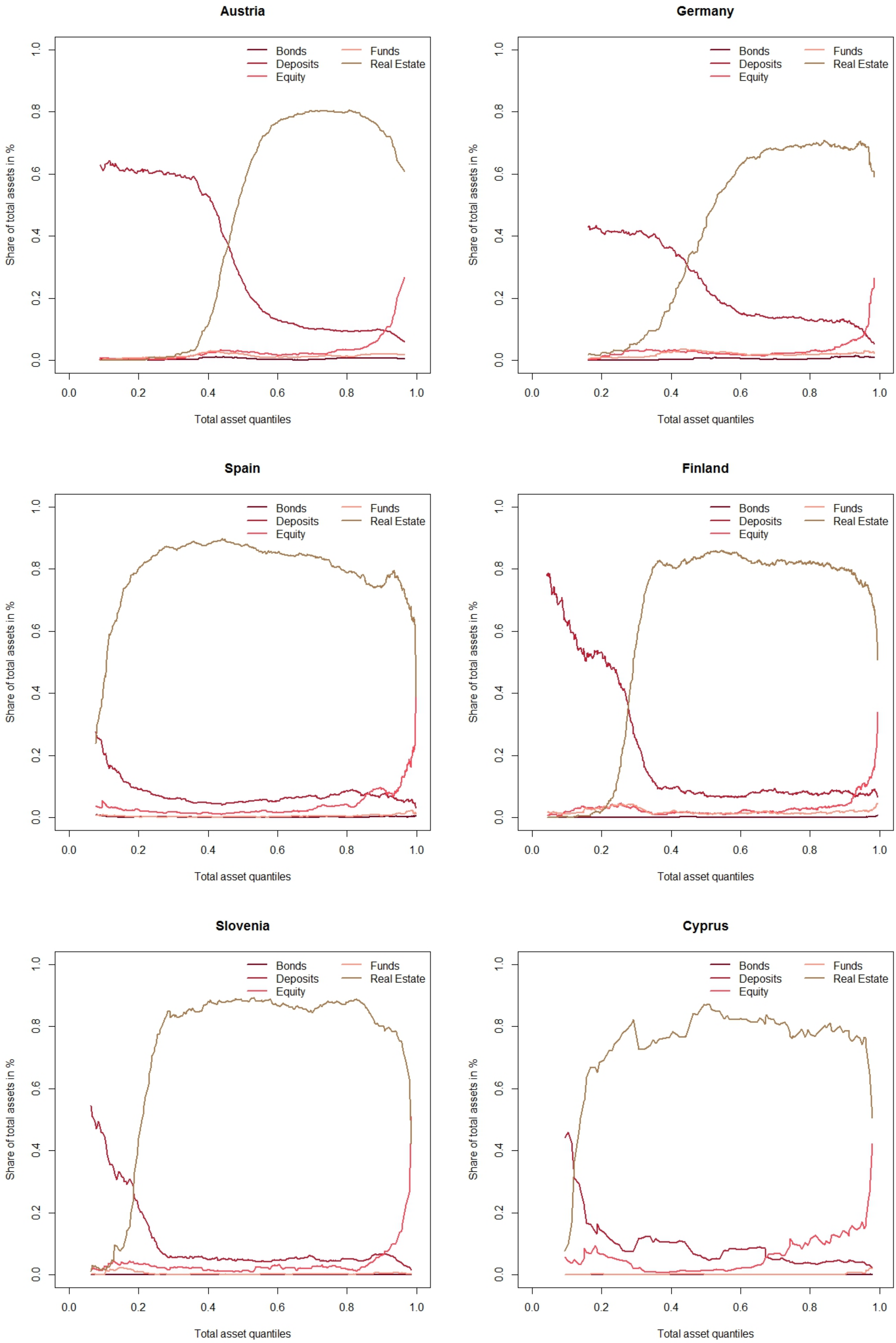


Figure 16: Shares of total assets for net worth distribution.

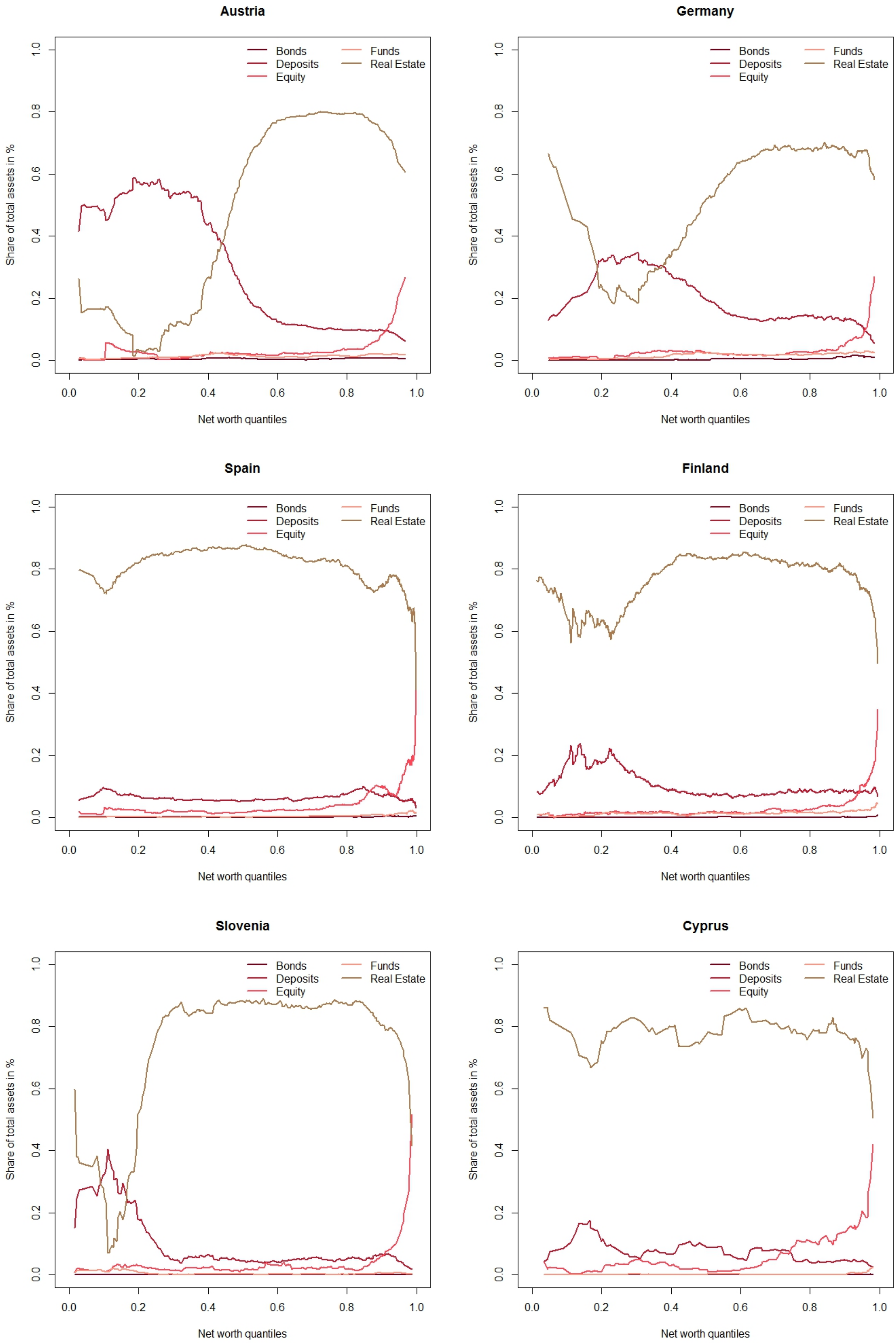
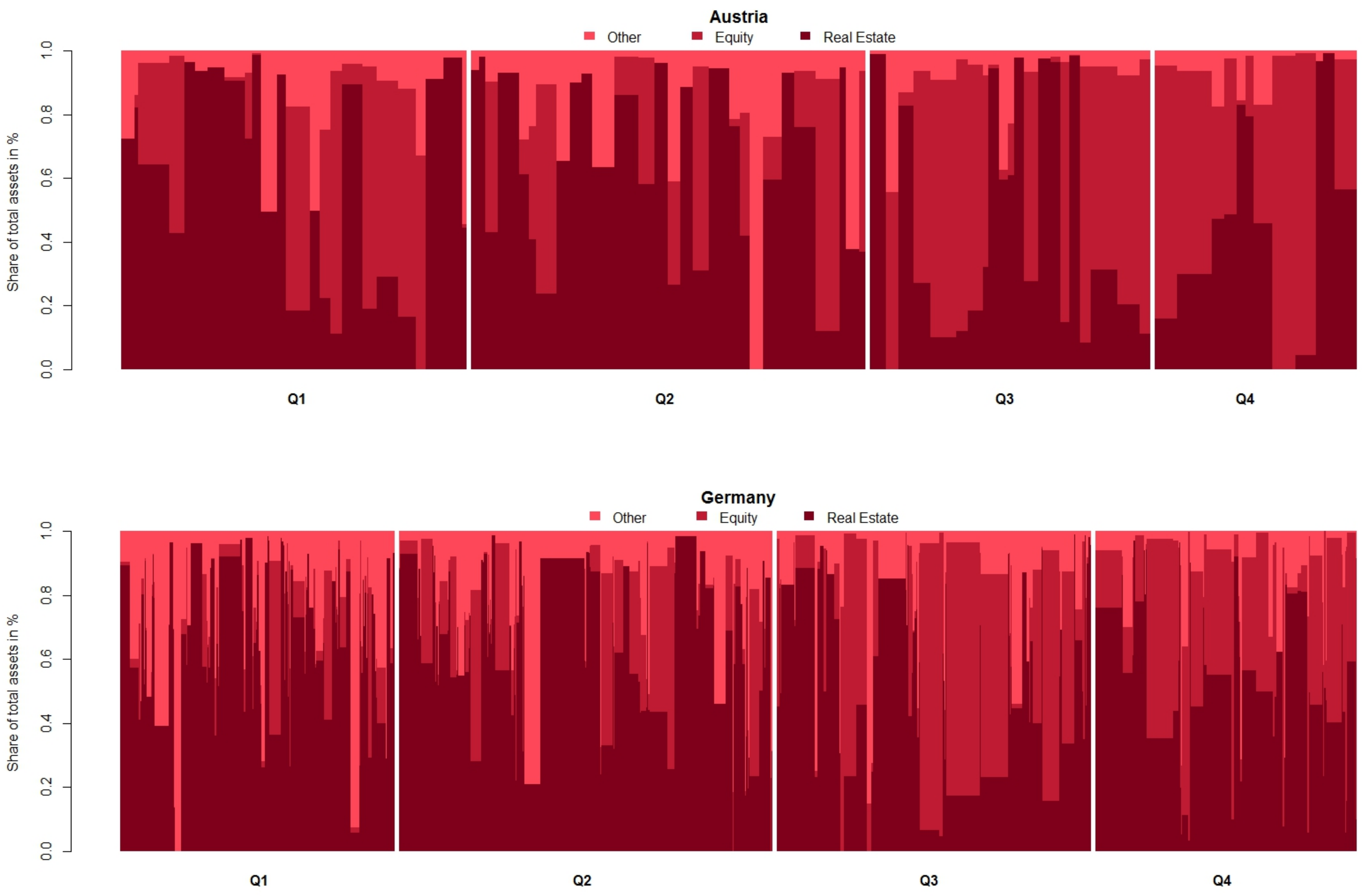


Figure 17: Portfolio allocation of the entire tail.



C Sensitivity analysis

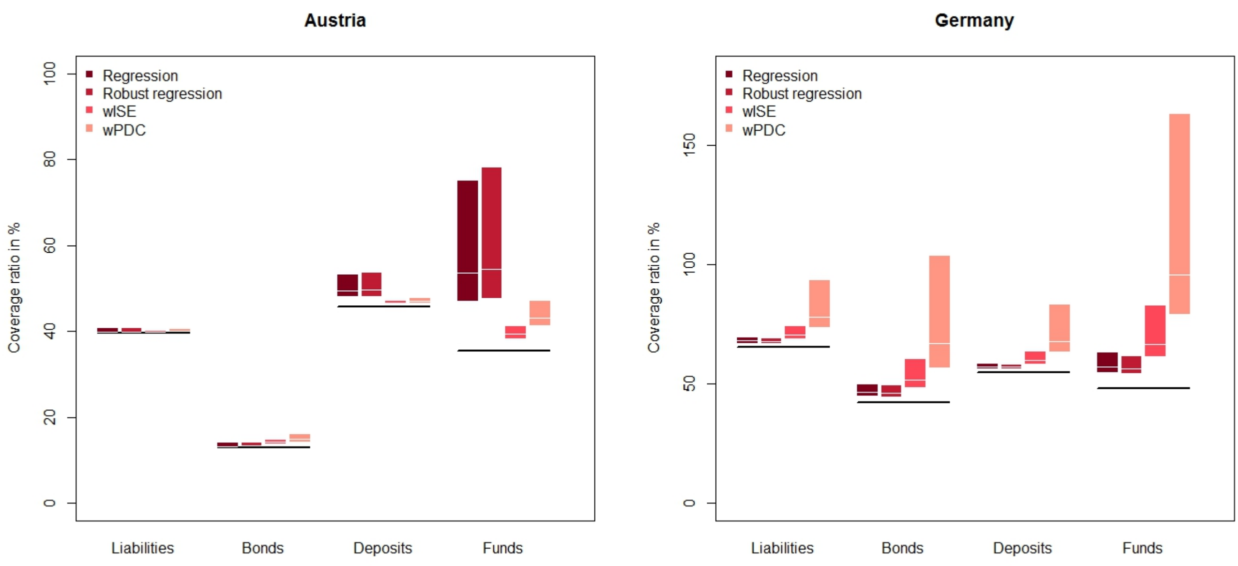
In this section, we perform five kinds of sensitivity analyses. We check how much the main results, i.e., shifts in coverage ratios for highly comparable instrument, change when (i) using different methods to estimate the shape parameter of the Pareto distribution ϑ, (ii) using different thresholds from where on the Pareto distribution replaces survey results, (iii) manipulating rich lists used in the Pareto estimation, (iv) refraining from stratifying by net worth, and (v) using additional portfolios from other countries to better reflect the portfolio structure of the very rich.

In general, we find that our methodology achieves a reasonable degree of robustness. In particular, results are extremely robust for liabilities, bonds and deposits, but less so for equity and mutual funds. We expect that results would be even more stable if personal or household level data on wealth and income would be used when performing oversampling. Given the yet scarce knowledge about the distribution of wealth, making available more data sources should in our opinion be of highest priority.

C.1 Choice of Pareto shape parameter estimation method

Estimating the Pareto shape parameter is ambiguous and different methods lead to quite distinct total tail wealth as shown in Table 8. In this section we compare the impact of different methodological choices on *instrument-specific coverage ratios*.

Figure 18: Sensitivity toward the choice of Pareto estimation method.



Notes: The figures show coverage ratios for highly comparable instruments when relying on different Pareto estimation methods. Bars represent 95% confidence intervals, white lines *adjusted* coverage ratios, and black lines *observed* coverage ratios.

We find that the choice of method affects equity the most. This is not surprising as large shares of equity are mainly held by the wealthiest of the wealthy and the likelihood of simulating very wealthy households strongly depends on the exact value of ϑ.

The impact on all other instruments is less pronounced. Figure 18 shows coverage ratios for highly comparable instruments using different estimation methods. Particularly for Germany, the wPDC method leads to large average coverage ratios and wide confidence intervals. This is due to the fact, that this estimation method leads to a much larger and unrealistic total tail wealth compared to all other methods (see Table 7).

The large variation of results for mutual funds in Austria sticks out. In fact, all manipulations testing robustness find rather large impacts on mutual funds for Austria as compared to the other instruments. This result is mainly driven by two outliers concerning investments in mutual funds observed in the Austrian HFCS belonging to $Q2$ and $Q4$.

Overlapping confidence interval indicate that differences between methods are not significant. For many – though not the majority of all – bilateral comparisons, this holds true: Out of the possible

$$\#\text{instruments} \cdot \binom{\#\text{est. methods}}{2} = 4 \cdot \binom{4}{2} = 24$$

pairwise combinations, 9 intervals overlap for Austria and 11 intervals overlap for Germany.

C.2 Choice of Pareto threshold

Choosing the threshold parameter y_0 is another major model choice. Whereas the threshold must be large enough to justify the Pareto tail, it should also not be too large to guarantee enough observations that can be used to estimate the Pareto shape parameter. We choose the same threshold (namely one million Euro) for both, Austria and Germany, facilitating comparability. In this section we justify this choice and analyse the effect of varying the Pareto threshold.

There is manifold evidence that the very top of the wealth distribution follows a Pareto law. However, the Pareto model is not suited to describe other parts of the wealth distribution. There are several tests that facilitate finding an appropriate cut-off point (see also Eckerstorfer et al., 2016).[61]

Van der Wijk's law can be used to check, whether the chosen threshold is large enough such that the data already follows a Pareto law.

The Pareto distribution is the only distribution satisfying[62]

$$W(y_0) = \frac{\int_{y_0}^{\infty} y f(y)\ \mathrm{d}y}{\int_{y_0}^{\infty} f(y)\ \mathrm{d}y} = \frac{\vartheta}{\vartheta - 1} \cdot y_0,$$

i.e., the *mean excess function* is proportional to the threshold y_0 (see van der Wijk, 1939; Cowell, 2011b; Dalitz, 2016). Practically, one thus computes the empirical mean excess function over y_0

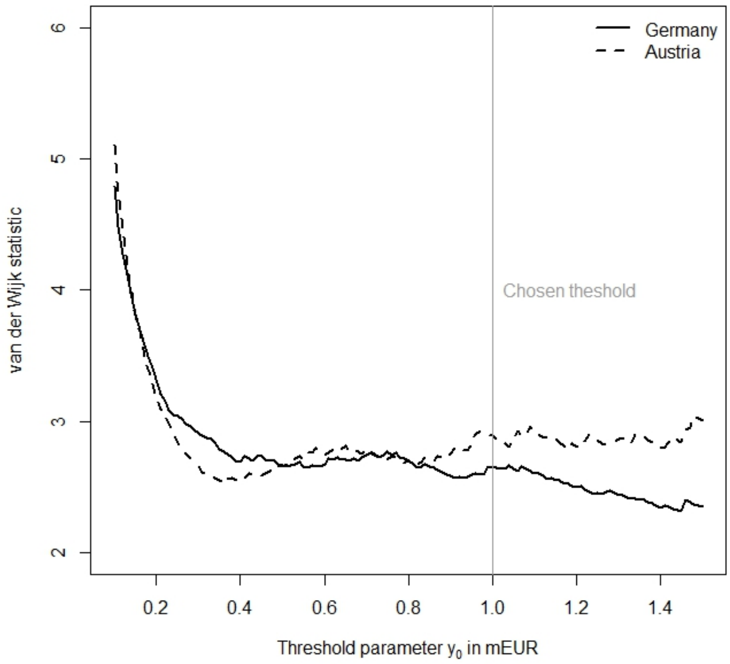
$$\frac{W(y_0)}{y_0} = \frac{\sum_{y_i > y_0} y_i \cdot w_i}{y_0 \cdot \sum_{y_i > y_0} w_i}$$

for a sequence of thresholds y_0 and checks, whether the chosen threshold lies within the area where $\frac{W(y_0)}{y_0}$ is constant.

[61]We only look at HFCS data to detect this point as the large gap between the largest wealth observed in the HFCS and the lowest number reported by the rich lists distorts this analysis. From a theoretical point of view, this makes sense because as soon as the tail shows Pareto behaviour, this is true for the subsequent part the tail too.

[62]Since the support of the Pareto distribution is (y_0, ∞), we have $\int_{y_0}^{\infty} f(y)\ \mathrm{d}y = \int_{-\infty}^{\infty} f(y)\ \mathrm{d}y = 1$ and $\int_{y_0}^{\infty} y f(y)\ \mathrm{d}y = \int_{-\infty}^{\infty} y f(y)\ \mathrm{d}y = E(Y)$. Using (2) thus yields $W(y_0) = \frac{\vartheta}{\vartheta - 1} \cdot y_0$. So, the Pareto distribution satisfies van der Wijk's law. It is also easy to show that among all continuous distributions the Pareto distribution is in fact the only one satisfying this relationship (see Cowell, 2011b, Appendix A.3 for a proof).

Figure 19: Choice of Pareto threshold parameter.



Notes: The figure shows the van der Wijk statistic for Austria and Germany calculated from HFCS data.

Figure 19 shows the results. The van der Wijk statistic is constant from roughly 0.4 million Euro onwards for both countries.

The Kolmogorov-Smirnov test, which calculates the maximum absolute distance between the empirical distribution and a theoretical Pareto distribution, finds minimal deviation at $y_0 = 0.56$ mEUR for Austria and $y_0 = 0.41$ mEUR for Germany.

Thus, these two theoretical criteria suggest rather low thresholds. Additionally, we calculate instrument-specific coverage ratios for various thresholds ranging between 0.4 and 2.1 million Euro and thus also for the widely used thresholds[63] of 0.5, 1 and 2 million Euro.

Figure 20 shows the results. We find that (except for mutual funds in Austria) coverage ratios are fairly stable for thresholds above one million EUR. For equity, coverage ratios are in general more volatile and become unreliable for thresholds above two million Euro.

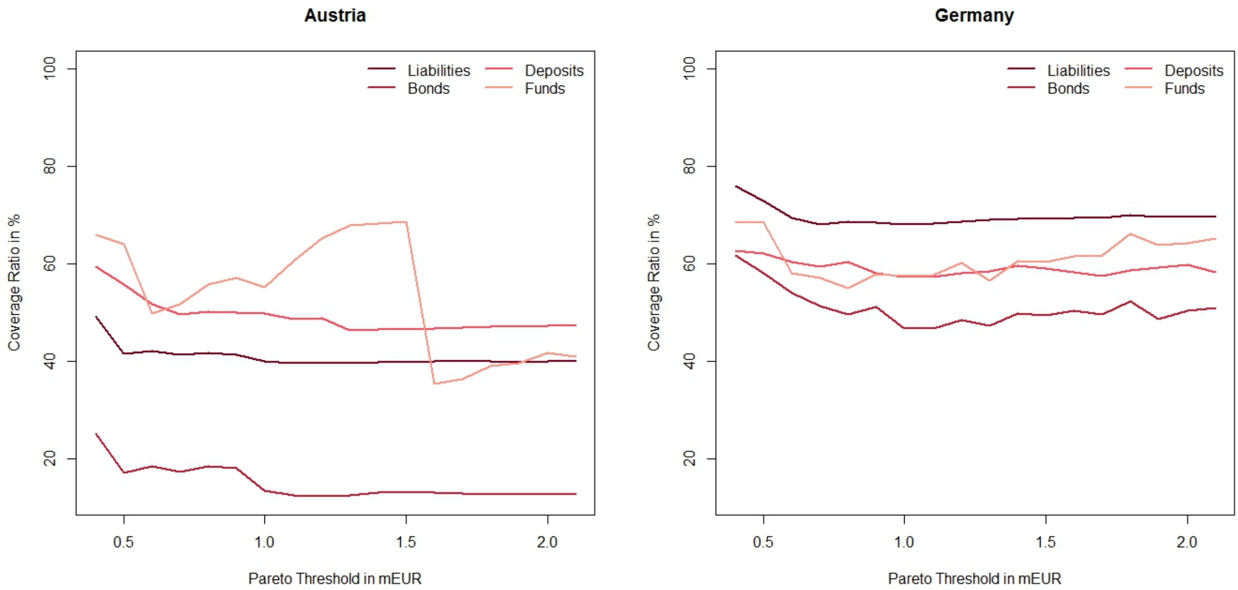
We therefore conclude that the minimal threshold suggested by the van der Wijk law and the Kolmogorov-Smirnov test are too low and thus go for a larger threshold achieving satisfying results according to all our analyses: one million Euro. A larger threshold is also beneficial in the sense that we do not want to substitute too large parts of the survey by a parametric model.

C.3 Rich list observations

As the HFCS lacks observations from the very top, estimating the wealth distribution from this data only is likely to underestimate the top tail. Thus, we enrich the HFCS data with

[63]See Vermeulen (2017), Chakraborty et al. (2016), and Bach et al. (2015).

Figure 20: Sensitivity toward the choice of Pareto threshold.



Notes: The figures show coverage ratios for highly comparable instruments when using different Pareto thresholds. Thresholds range between 0.4 and 2.1 million Euro with a step size of 100,000 Euro.

observations from rich lists and use this combined data set when estimating the Pareto shape parameter.

Rich lists provide us with observations from the far right of the distribution. However, as rich lists cover only 100 (AT) and 500 (DE) of the riches households, there is still a large gap between the largest observation in the HFCS and the smallest observation on the rich list.

In this section we therefore test, how results change when manipulating the rich list. We manipulate rich lists in four ways: the exact number of observations on the list, the influence of the largest observation, the treatment of observations as single households or rather family clans consisting of several households, and the exclusion of households holding assets mainly in the form of foundations.

We estimate the Pareto shape parameter ϑ using the regression method when relying on *(i)* the full rich list, *(ii)* the top 50% of all rich list observations, *(iii)* the top 25% of all rich list observations, and *(iv)* the top 10% of all rich list observations. We also estimate the shape parameter when excluding the wealthiest household from the rich list *(v)* as this is, in particular for Austria, an extreme observation[64] with possibly large leverage.

As a number of observations on the rich list may refer to family clans that probably consist of more than one single household, we test the effect of splitting the amount into several "households." We perform this exercise once by splitting *all* observations into *(vi)* two or *(vii)* four households, respectively. Alternatively, we report the results for a *(viii) random split*: We split each observation into u households, whereas $u \sim U(1, 4)$ follows a discrete uniform distribution on the set $\{1, 2, 3, 4\}$. We repeat this 10,000 times and report the average result

[64]In Germany, the richest family (Quandt/Klatten: BMW) owns 31 billion Euro (according to the Manager Magazine) whereas the second wealthiest family (Albrecht/Heister: Aldi Süd) "only" owns 18.3 billion, i.e., roughly 60% of the Quandt/Klatten fortune. In Austria, Porsche/Piëch (Porsche, VW) are estimated to own 65 billion Euro (according to Trend), whereas the second-ranked Dietrich Mateschitz (Red Bull) owns with 7.6 billion Euro only roughly 12% of the Porsche/Piëch fortune.

for $\hat{\vartheta}$ and the average length of the simulated rich list.[65]

For Austria, the rich list also contains information on the major source/type of wealth: foundations, inherited wealth (without further classification), and listed or unlisted shares. Thus, we measure the effect of *(ix)* excluding observations from the rich list that, according to the Trend list, "hold assets" only in the form of foundations (see paragraph 42 for a discussion about foundations in Austria).

Table 11: Sensitivity toward the number of rich list observations.

		AT – Austria			**DE – Germany**		
		Obs.	$\hat{\vartheta}$	Tail wealth	Obs.	$\hat{\vartheta}$	Tail wealth
(i)	full rich list	100	1.194	778,285	500	1.338	4,899,939
(ii)	top 50% of rich list	50	1.174	854,177	250	1.317	5,144,155
(iii)	top 25% of rich list	25	1.176	844,238	125	1.306	5,277,994
(iv)	top 10% of rich list	10	1.179	832,291	50	1.318	5,125,318
(v)	excl. top observation	99	1.211	726,232	499	1.341	4,866,363
(vi)	split into 2 households	200	1.205	744,288	1,000	1.368	4,600,772
(vii)	split into 4 households	400	1.221	699,804	2,000	1.387	4,432,071
(viii)	random split $U(1,4)$	250	1.214	717,052	1,250	1.376	4,528,258
(ix)	excl. foundations	77	1.235	665,264	–	–	–

Notes: The table reports estimation results based on the regression method when manipulating the rich list. Namely, we use the top 50%, top 25%, and top 10% of the rich list, analyse the effect of leaving out the top observation, "foundation" observations, and the effect of splitting total amounts on the rich list into two, four and a random integer between 1 and 4 households. Tail wealth is given in mEUR and calculated according to formula (2).

Results are reported in Table 11. We find that in general the exact number of observations from rich lists does not matter a lot as seen in rows *(ii)–(iv)*. Even a strikingly small number, such as 10% of the original list, leads to a very similar total tail wealth. Results are less sensitive for Germany as there are in general many more observations.

Leaving out the richest family clan – see row *(v)* – in Austria has a stronger effect than in Germany. This is not surprising given the extreme difference in net worth between the wealthiest and second wealthiest family in Austria (see subsection 64).

Splitting amounts on the rich list into several households – see rows *(vi)–(viii)* – to account for the fact that at least some observations refer to family clans consisting of several households, does have a stronger effect than leaving out parts of the rich list. Whereas shortening the rich list increases tail wealth, splitting observations leads to a decrease.

Results for leaving out households which hold their assets exclusively in the form of foundations are presented in row *(ix)*. As this information is not available for Germany, results are only reported for Austria. Total tail wealth is reduced as many top (including *the* top) observation is removed.

Table 12 reports the effects of such manipulations on instrument-specific coverage ratios. We report coverage ratios for the maximum and minimum estimated shape parameter according

[65] Bach et al. (2015) split up German rich list observations based on information they collected through "thorough internet research." They claim that there is not enough information for lower ranked rich list observations. For this part of the list, they hence assume that each observation refers to four households.

to Table 11. While changes in total tail wealth seem to be quite remarkable, the impact on coverage ratios are negligible except as for mutual funds and equity.

Table 12: Effects of manipulating rich lists on coverage ratios.

	Austria				**Germany**		
	(ii)	*(i)*	*(vii)*	*(ix)*	*(iii)*	*(i)*	*(vii)*
Liabilities	40.12	39.93	39.78	39.71	68.58	67.99	67.05
Bonds	13.47	13.30	13.16	13.11	47.39	46.67	48.18
Deposits	50.63	49.72	49.04	48.85	57.85	57.23	57.21
Equity	367.05	349.59	315.49	307.31	389.79	363.12	331.97
Mutual Funds	61.82	55.20	52.06	49.51	59.35	57.33	54.28

Notes: The table reports effects of manipulating rich lists on coverage ratios. Results are in % and derived using the analytical approach. We report results for the largest and smallest change in estimated Pareto shape parameter (see Table 11) and compare it to the standard method denoted by *(i)*. For Austria, we additionally report changes when excluding foundations.

C.4 Stratification

Our methodology relies on stratification taking into account the observation that portfolio structures change when moving along the net worth distribution. Although we believe that stratification increases reliability in our methodology, we here show results when neglecting this correlation.

This means that in case of the simulation approach for each simulated net worth, we re-sample an arbitrary portfolio from the full list of observed portfolios in the tail population. As a consequence, the correlation between net worth and portfolio structure is *not preserved.* For instance, a household with high simulated net worth is *not* more likely to hold greater shares of equity than a household a simulated net worth close to one million Euro.

Likewise for the analytical approach, we do not calculate stratum-specific portfolio shares but rely on a single average portfolio structure for the entire tail (see also Chakraborty et al., 2016).

Table 13 reports results: Aggregates for instruments with decreasing importance when moving to the top end of the distribution (liabilities, deposits, bonds, and real estate) increase. In contrast, aggregates for equity, which is most relevant for the very top of the tail, decreases. As there is no clear correlation between the importance of mutual funds and net worth, this instrument is least affected when refraining from stratification.

Simple averages ignore the correlation between net worth and the portfolio structure. This strongly overestimates liabilities, bonds, deposits, and real estate – instruments that become relatively less important with increasing wealth – and underestimate the most important asset class of the wealthiest of the wealthy: equity.

These findings underpin the importance of stratification. In the future, when more waves of the HFCS become available, portfolio structures from previous waves could be added to increase

Table 13: Effects of stratification.

AT – Austria					
	Adjusted Aggregate (stratified)		Adjusted Aggregate	95% confidence interval for CR	95% confidence interval for CR
	analytical	simulation	(not strat.)	(stratified)	(not stratified)
Liabilities	67,182	67,085	85,292	(39.62; 40.62)	(47.00; 58.79)
Bonds	5,437	5,418	10,300	(13.02; 13.97)	(20.39; 36.06)
Deposits	107,345	106,696	127,057	(48.30; 53.11)	(54.92; 67.48)
Equity	437,234	421,037	316,024	(282.57; 489.45)	(206.34; 364.17)
Mutual Funds	26,508	25,734	27,155	(47.35; 75.09)	(48.00; 78.61)
Real Estate	829,044	814,032	926,068	–	–

DE – Germany					
	Adjusted Aggregate (stratified)		Adjusted Aggregate	95% confidence interval for CR	95% confidence interval for CR
	analytical	simulation	(not strat.)	(stratified)	(not stratified)
Liabilities	1,063,155	1,062,378	1,153,936	(67.50; 69.26)	(72.82; 76.07)
Bonds	79,351	79,118	95,098	(45.30; 49.89)	(53.43; 62.51)
Deposits	1,052,599	1,051,348	1,142,704	(56.78; 58.35)	(61.23; 64.48)
Equity	1,823,723	1,812,224	1,505,197	(347.13; 407.37)	(286.79; 330.59)
Mutual Funds	246,645	245,910	266,968	(55.36; 62.19)	(59.62; 68.66)
Real Estate	6,859,346	6,841,174	7,041,632	–	–

Notes: The table reports adjusted aggregates (in mEUR) and 95% confidence intervals for coverage ratios (in %). Results rely on adjusting the tail with and without stratification, respectively. The 95% confidence interval are obtained from the simulation approach whereas aggregates rely – if not stated differently – on the analytical approach.

the precision.[66] This might be particularly helpful for small countries like Austria. (Also see the discussion in the next section.)

C.5 Representativenss of portfolio structures

As argued in subsection 3.2 surveys that do not oversample the wealthiest households might not appropriately reflect portfolio structures of the very rich. Thus, adding additional portfolios to the top stratum may be desirable to increase reliability. This section shows how HFCS data obtained from other countries could be used with this regard.

When analysing figures Figure 15 and Figure 16, we find for Austria and Germany that at the very top the average share of total assets or net worth held as equity is much lower than the average share held as real estate. This is not plausible for the wealthiest of the wealthy. In other countries, including heavily oversampling Spain and regionally oversampling Slovenia, equity becomes more important than real estate at the very top. Thus, particularly for Austria

[66] Some countries including Germany already have or plan to have a panel component. When using portfolio structures from previous waves, only the latest panel observation should be used to avoid bias. However, in the future panel observations can be used to assess the stability of portfolio structures over time. Currently, we draw portfolios respecting their weights, i.e., a portfolio from an observation representing many households is more likely to be (and therefore is more often) drawn than one from an observation with a lower weight. When merging portfolios from several waves, the weight structure is not obvious any more.

and Germany borrowing portfolio structures from other countries to better reflect the very top of the distribution seems important. The plausible assumption behind this approach is that the wealthiest of the wealthy across Europe have similar investment habits.

We thus supplement the top stratum, $Q4$, with portfolio shares from other HFCS countries representing the top of the top. More specifically, for a given level q (here we use $q \in [0.99; 0.9999]$) we calculate the quantile y_q (i.e., the threshold in terms of net worth such that $q\%$ of the observations are lower and $(1-q)\%$ are larger), and filter all observations from the entire HFCS (except the particular country we analyse) that have a net worth of a least y_q. This set of "foreign" observations is denoted by F. Finally, we add F's portfolio shares to the top stratum $Q4$.

When adding additional portfolios, we need to adjust the weights in $Q4$. Therefore, we calculate for each observation in F its *relative weight* by dividing its *frequency weight* by the total number of households in the respective country and denote them by ω_i^F. We scale these relative weights so that they add up to one within F:

$$\frac{\omega_i^F}{\sum_{i \in F} \omega_i^F}.$$

Denoting the total population by N, i.e., the sum over all weights for the country of analysis, *frequency weights* are thus obtained by

$$w_i^F = (1-q) \cdot N \cdot \frac{\omega_i^F}{\sum_{i \in F} \omega_i^F}.$$

Portfolio shares together with frequency weights in F are then added to $Q4$. To guarantee that the population total in $Q4$ remains unchanged, we divide *all weights*, i.e., those originally from $Q4$, w_i^{Q4}, and those added from foreign countries, w_i^F, by the sum of total frequency weights and multiply them by the number of households in $Q4$:

$$0.25 \cdot N \cdot \frac{w_i^{F \cup Q4}}{\sum_{i \in F \cup Q4} w_i^{F \cup Q4}}.$$

Figure 21 shows resulting coverage ratios for various thresholds q. In general we find that coverage ratios for highly comparable instruments are not heavily affected by the inclusion of additional portfolios. Comparing the two countries, changes are larger for Austria and in particular for mutual funds in Austria.

A priori, it is not clear which threshold q is most appropriate. As we expect wealthiest households to hold most of their assets as equity, we choose the threshold that achieves the largest aggregates for equity which is indicated as a dashed line in Figure 21:

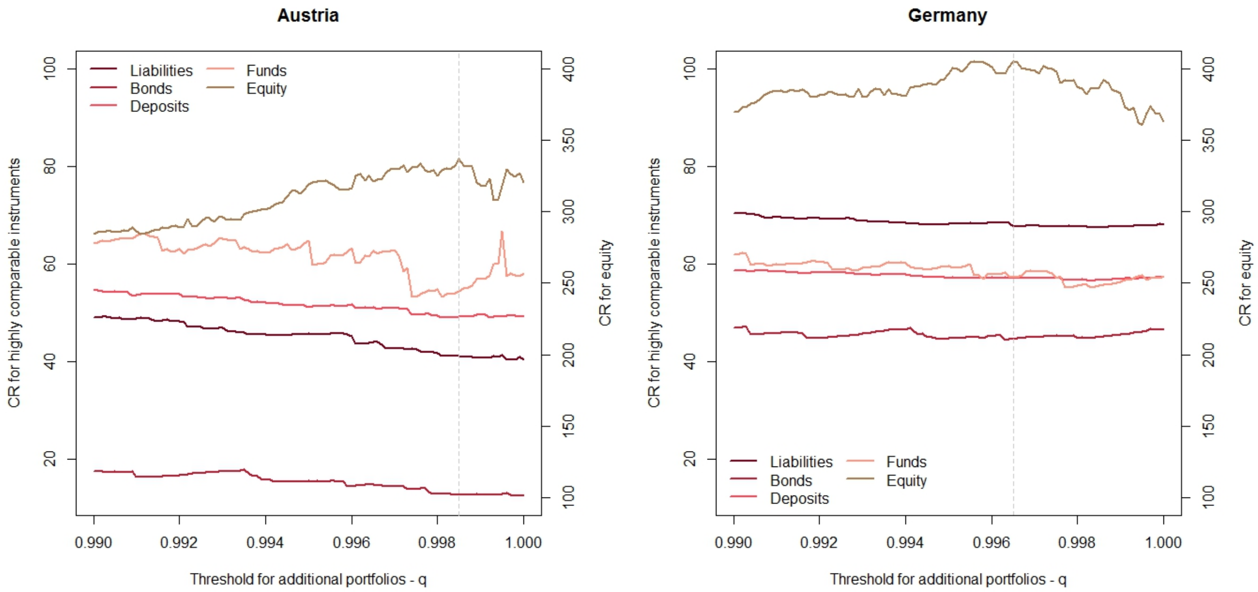
$$q^* = \underset{q \in [0.99; 0.9999]}{\arg\max} A_{\text{equity}}(q).$$

Results for q^* are reported in Table 14. We find that instrument-specific coverage ratios are only marginally affected by adding additional portfolios and qualitative conclusions remain unchanged. Due to our maximisation approach, changes are naturally largest for equity.

For Austria, we find $q^* = 0.9985$ and $y_{q^*}^{AT} = 13,237,765$ EUR, and for Germany $q^* = 0.9965$ and $y_{q^*}^{DE} = 5,965,796$ EUR.[67] Across the entire HFCS, 187 (and thereof 186 non-Austrian)

[67]Quantiles are calculated from the semi-parametric Pareto model. We therefore combine non-tail HFCS observations with a simulated sample from the estimated Pareto model representing the adjusted tail. We select the sample used here out of 100 draws according to a Kolmogorov-Smirnov test.

Figure 21: Representativeness of top portfolio shares.



Notes: The figures show coverage ratios for various thresholds q.

Table 14: Usage of additional foreign portfolio structure at the very top.

	AT – Austria		**DE – Germany**	
	No extra obs.	Extra obs. $q^* = 0.9985$	No extra obs.	Extra obs. $q^* = 0.9965$
Liabilities	39.93	40.63	67.99	67.63
Bonds	13.30	13.35	46.67	44.63
Deposits	49.72	49.43	57.23	56.99
Equity	349.59	355.49	363.12	405.22
Mutual Funds	55.20	52.99	57.33	57.17

Notes: Coverage ratios are obtained by the analytical method. *No extra obs.* refers to the standard method used in this article. *Extra obs.* makes use of additional portfolios of the wealthiest households observed in the entire HFCS, whereas we only supplement the q^*th quantile of the net worth distribution with extra observations. Coverage ratios are in %.

observations exceed $y_{q^*}^{AT}$ and 553 (and thereof 528 non-German) exceed $y_{q^*}^{DE}$. Most of these foreign observations come from Spain (112 in AT / 247 in DE), followed by France (52 in AT / 209 in DE). Spain and France are countries performing advanced oversampling strategies based on individual wealth data and achieved the highest effective oversampling rate of the top 5%[68] among all countries participating in the second wave of the HFCS (see HFCN, 2016a, Table 4.7).

The added portfolios are mainly of type "equity," i.e., the most important asset class is equity (in 75.3% of all cases in Austria and 57.0% in Germany). Thus, these additional portfolios fulfil our expectations about portfolio structures of the wealthiest of the wealthy.

[68]See subsection 6.1 for a definition of the effective oversampling rate.

D Derivations

D.1 Analytical method

This section derives stratum-specific average wealth needed in section 4. Therefore, we use the inverse of the Pareto CDF F^{-1} (see equation (1) and Appendix A). Quantiles in general are calculated as described in Appendix A and denoted by $F^{-1}(p)$.

Strata are defined as quartiles of the tail, i.e., $Q = [F^{-1}(p_1); F^{-1}(p_2))$, with $\{p_1, p_2\} \in \{\{0, 0.25\}, \{0.25, 0.5\}, \{0.5, 0.75\}, \{0.75, 1\}\}$. Note that $F^{-1}(0) = y_0$ and $F^{-1}(1) = \infty$.

Stratum-specific average wealth is given by

$$\begin{aligned} W(Q) = E(Y|Y \in Q) &= \frac{E(Y \cdot \mathbb{I}_Q(Y))}{P(Y \in Q)} \\ &= \frac{1}{p_2 - p_1} \int y \mathbb{I}_Q(y) \, \mathrm{d}F(y) = \frac{1}{p_2 - p_1} \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} y \, \mathrm{d}F(y), \end{aligned}$$

whereas

$$\mathbb{I}_Q(Y) = \begin{cases} 1, & Y \in Q, \\ 0, & Y \notin Q \end{cases}$$

denotes the indicator function.

Substituting $q = F(y)$ yields $\mathrm{d}F(y) = \mathrm{d}F(F^{-1}(q)) = \mathrm{d}q$. The thresholds $F^{-1}(p_1)$ and $F^{-1}(p_2)$ change to p_1 and p_2, respectively, yielding

$$\begin{aligned} W(Q) &= \frac{1}{p_2 - p_1} \int_{p_1}^{p_2} F^{-1}(q) \, \mathrm{d}q = \frac{1}{p_2 - p_1} \int_{p_1}^{p_2} y_0 \cdot (1-q)^{-1/\vartheta} \, \mathrm{d}q \\ &= \frac{\vartheta y_0}{(1-\vartheta) \cdot (p_2 - p_1)} \left[(1-p_2)^{1-1/\vartheta} - (1-p_1)^{1-1/\vartheta} \right]. \end{aligned}$$

D.2 Properties of DINA

This section proves a property about DINA as defined in section 8: If instrument-specific aggregates belonging to the top net worth group are increased (by for instance adjusting for the missing wealthy), then distributional national account figures d_{ij} are strictly adjusted downwards for all groups except the top and strictly adjusted upwards for the top group.

As instrument-specific aggregates belonging to the top net worth group are increased, total instrument-specific aggregates rise, i.e., $x^*_{\cdot j} > x_{\cdot j}$. This inequality implies

$$d_{ij} > d^*_{ij} \quad \text{for } 1 \leq i < n, \tag{4}$$

i.e., for all but the highest net worth group, and all instruments j. Thus,

$$\begin{aligned} \sum_{i=1}^{n} d_{ij} &= \frac{x_{nj}}{x_{\cdot j}} y_j + \sum_{i=1}^{n-1} d_{ij} \overset{(4)}{>} \frac{x_{nj}}{x_{\cdot j}} y_j + \sum_{i=1}^{n-1} d^*_{ij} \\ &= \frac{x_{nj}}{x_{\cdot j}} y_j + \sum_{i=1}^{n} d^*_{ij} - \frac{x^*_{nj}}{x^*_{\cdot j}} y_j = \frac{x_{nj}}{x_{\cdot j}} y_j + \sum_{i=1}^{n} d_{ij} - \frac{x^*_{nj}}{x^*_{\cdot j}} y_j, \end{aligned}$$

whereas the last equality is a direct consequence of scaling, i.e., $\sum_{i=1}^{n} d_{ij} = y_j = \sum_{i=1}^{n} d^*_{ij}$. This implies

$$d_{nj} = \frac{x_{nj}}{x_{\cdot j}} y_j < \frac{x^*_{nj}}{x^*_{\cdot j}} y_j = d^*_{nj},$$

which finalises the proof.

E Tables

The following tables reports the full set of adjusted and unadjusted DINA results for Austria and Germany as described in section 8.

Table 15: Distributional National Accounts – Shares.

AT – Austria											
		Liabilities		Bonds		Deposits		Equity		Mutual Funds	
Quintiles		d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j
1st	$(-\infty; 6.9]$	7.4	7.3	0.2	0.2	1.7	1.5	0.1	0.0	0.1	0.1
2nd	(6.9; 35.4]	3.9	3.9	0.6	0.5	9.2	8.4	0.1	0.0	1.1	0.7
3rd	(35.4; 162.0]	26.7	26.5	9.7	9.4	19.3	17.7	0.9	0.4	8.4	5.4
4th	(162.0; 365.5]	29.5	29.3	7.9	7.7	23.0	21.1	2.7	1.2	16.1	10.4
5th	$(365.5; \infty)$	32.4	33.0	81.6	82.1	46.9	51.2	96.2	98.3	74.2	83.4

DE – Germany											
		Liabilities		Bonds		Deposits		Equity		Mutual Funds	
Quintiles		d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j	d_{ij}/y_j	d^*_{ij}/y_j
1st	$(-\infty; 2.5]$	14.8	14.2	0.0	0.0	1.2	1.1	0.1	0.1	0.3	0.3
2nd	(2.5; 27.4]	3.3	3.2	0.2	0.2	4.3	4.1	0.2	0.2	0.4	0.4
3rd	(27.4; 113.0]	20.7	19.9	1.1	1.0	13.5	12.9	1.3	0.9	8.2	6.9
4th	(113.0; 274.0]	23.7	22.8	10.9	9.9	22.1	21.2	2.4	1.7	13.6	11.4
5th	$(274.0; \infty)$	37.4	39.9	87.8	88.9	58.9	60.7	96.0	97.1	77.4	81.1

Notes: The table reports shares of total instrument-specific aggregates held by each net worth quintile. The column "quintile" reports the thresholds of net worth in tEUR. Shares are in %. Quintile thresholds are calculated using HFCS data (as our adjustment only affects the top 1-2%, these thresholds do not change in the course of our adjustment). Adjusted ratios use results from the analytical method.

Table 16: Distributional National Accounts – Euro Amounts.

AT – Austria

		Liabilities		Bonds		Deposits		Equity		Mutual Funds	
Quintiles		d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}
1st	$(-\infty; 6.9]$	12,469	12,361	71	69	3,602	3,314	86	38	71	46
2nd	(6.9; 35.4]	6,628	6,570	229	222	19,811	18,224	70	31	517	333
3rd	(35.4; 162.0]	44,994	44,605	3,976	3,855	41,658	38,320	1,186	528	4,025	2,592
4th	(162.0; 365.5]	49,682	49,252	3,248	3,149	49,574	45,602	3,367	1,500	7,754	4,993
5th	$(365.5; \infty)$	54,497	55,481	33,345	33,573	101,250	110,435	120,360	122,972	35,651	40,054

DE – Germany

		Liabilities		Bonds		Deposits		Equity		Mutual Funds	
Quintiles		d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}	d_{ij}	d^*_{ij}
1st	$(-\infty; 2.5]$	230,923	221,750	16	14	21,230	20,326	470	342	1,498	1,256
2nd	(2.5; 27.4]	52,122	50,051	359	326	78,434	75,095	1,218	886	1,861	1,561
3rd	(27.4; 113.0]	323,914	311,046	1,946	1,764	248,608	238,026	6,520	4,741	35,140	29,472
4th	(113.0; 274.0]	371,317	356,566	18,497	16,764	406,898	389,579	11,973	8,706	58,545	49,102
5th	$(274.0; \infty)$	585,308	624,171	149,201	151,152	1,084,027	1,116,170	482,056	487,563	333,190	348,842

Notes: The table reports amounts of total instrument-specific aggregates held by each net worth quintile. The column "quntiles" reports the thresholds of net worth in tEUR. Amounts are in mEUR. Quintile thresholds are calculated using the definition of net worth in the HFCS (as our adjustment only affects the top 1-2%, quintile thresholds do not change). Adjusted aggregates use results from the analytical method.

Acknowledgements

We are grateful for comments by Henning Ahnert, Alexander Erler, Pirmin Fessler, Herwig Friedl, Juha Honkkila, Ilja K. Kavonius, Guillaume Osier, Sébastien Pérez-Duarte, Tobias Schmidt, Philip Vermeulen, Caroline Willeke, Michael Ziegelmeyer as well as an anonymous referee. We also thank participants of the June 2017 ECB Expert Group on Linking Macro and Micro Data for the Household Sector meeting at the European Central Bank, the 5th Luxembourg Workshop on Household Finance and Consumption at the Banque centrale du Luxembourg, the November 2017 User workshop of the Eurosystem Household Finance and Consumption Survey at the European Central Bank, the 1st WID.world conference at the Paris School of Economics, the 2nd LIS/LWS Users Conference "The legacy of Tony Atkinson in inequality analysis" in Luxembourg, as well as research seminars at the Deutsche Bundesbank, the Oesterreichische Nationalbank, the University of Graz and the Vienna University of Economics and Business for fruitful discussions and remarks. All remaining errors are ours. All calculations are performed using the open source programming language R (R Core Team, 2017). The authors carried out large parts of this work during their employment at the European Central Bank (ECB). This paper, however, should not be reported as representing the views of the ECB nor the Deutsche Bundesbank. The views expressed are those of the authors. This paper uses data from the Household Finance and Consumption Survey. The results published, and the related observations and analyses may not correspond to results or analyses of the data producers.

Robin Chakraborty
Deutsche Bundesbank, Frankfurt am Main, Germany; email: robin.chakraborty@bundesbank.de

Sofie R. Waltl (corresponding author)
Luxembourg Institute of Socio-Economic Research (LISER), Esch-sur-Alzette/Belval, Luxembourg; email: sofie.waltl@liser.lu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-3268-4 ISSN 1725-2806 doi:10.2866/938654, QB-AR-18-043-EN-N