



EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Janina Engel, Dennis Ohlwerter,
Matthias Scherer

On the estimation of distributional
household wealth: addressing under-
reporting via optimization problems
with invariant Gini coefficient

No 2865

Abstract

The Household Finance and Consumption Survey (HFCS) provides valuable information for the monetary policy and financial stability purposes. The dataset shows, however, inconsistencies with National Account (NtIA) statistics, as the aggregated HFCS micro data do usually not match the corresponding NtIA macro data. Therefore, we suggest a solution to close the gap via an optimization problem that aims at preserving for each wealth instrument the level of inequality measured by the Gini coefficient. In addition, a lower and an upper bound of inequality are derived, that can be reached by extreme allocations of the wealth discrepancies across the households. Finally, based on the German HFCS, we compare the findings with another approach suggested in the literature that uses a “multivariate calibration”. The comparison indicates that the multivariate calibration may reallocate households’ wealth beyond the observed discrepancies, thereby leading to Gini coefficients that exceed the analytically derived upper bound of inequality.

Keywords: Wealth inequality, HFCS, National Accounts, Optimization problem

JEL Classification: C46, C61, D31, G51, N34

Non-technical summary

Governments, central banks, academics, as well as the general public are interested in understanding the distribution of wealth and how policies impact wealth inequality. The analysis of wealth inequality is, however, challenged by limited data availability and data inconsistencies. On the one hand, distributional data is available from household surveys that collect information on households' finances and consumption. On the other hand, aggregate statistics on the total holdings of wealth instruments by the household sector are available from the national accounts. The problem is that the survey aggregates commonly do not match the corresponding totals from national accounts. Academics and policymakers have therefore developed models to address the identified reasons for data limitations, such as differences in instrument concepts and missing wealthy households in survey data. Nevertheless, even after all these adjustments, survey aggregates commonly would still not match their totals from the national accounts.

The present paper addresses this problem, by solving the remaining discrepancies in economically meaningful ways. Our starting point is the extended survey data and national accounts which have been adjusted for its known limitations and conceptual differences. In light of having no further information on the sources of the remaining discrepancies between the adjusted survey data and the national accounts, the aim is to close the discrepancies by modifying the adjusted survey data as little as possible. In particular, we aim to maintain the level of inequality, as well as additional economically motivated characteristics. The problem is translated into a constrained optimization problem, which is then solved analytically. In addition upper and lower bounds are derived for the level of inequality that can be achieved by extreme allocations of the remaining discrepancies. These bounds serve to provide confidence about the wealth instruments' distribution across households. A lower range increases confidence, while a wider range can raise awareness of possibly false instrument allocations.

We prove that allocating the remaining discrepancy of a wealth instrument proportionally to the household holdings preserves the level of inequality, as measured by the Gini coefficient, and satisfies economically motivated constraints. The derived solutions are applied to the German adjusted household finance and consumption survey. We find that the range of inequality that can be yield by allocating the remaining discrepancies varies substantially across wealth instruments. Some instruments such as mortgage liabilities, business wealth, listed shares, and debt securities show a low range for the upper and lower bound of the Gini coefficient, thereby

increasing confidence in the derived distribution. All of these instruments show a high level of inequality with Gini coefficients between 0.9 and 1. However, other instruments such as deposits (with a Gini coefficient between 0.31 and 0.88) and life insurance and voluntary pensions (with a Gini coefficient between 0.75 and 0.9) show a wide possible range of inequality, thus indicating potential uncertainty about the derived distribution.

Moreover, the derived solution is compared to another approach suggested in the literature called the ‘multivariate calibration’. We find that, in contrast to the proportional allocation, the multivariate calibration leads to a more heterogeneous distribution and violates certain economic constraints. First, in case the national accounts’ total of the wealth instrument exceeds the survey aggregate, the proportional allocation consistently increases all household holdings, and vice versa, while the multivariate calibration simultaneously increases the holdings of some households and decrease those of others. Second, the proportional allocation preserves for each wealth instrument the order of the households holdings and the level of inequality, while this is not the case for the multivariate calibration. Therefore, the proportional allocation seems to be an interesting alternative to the multivariate calibration.

1 Introduction

Macroeconomic data and indicators are frequently used to assess and explain the current state of economies. However, it is not straightforward to derive accurate conclusions from available data about income and wealth dynamics, especially in terms of wealth distributions. Analysing and comparing distributions of wealth is important not only to assess the impact of policies (see FSB and IMF (2019)) from governments and central banks (see Expert Group on Linking macro and micro data for the household sector (2020)), but also for research purposes and information to the public (see Schröder et al. (2020)).

Our starting point is the Household Finance and Consumption Survey (HFCS)¹ which we combine with National Account (NtIA) statistics published by the ECB and various national central banks.

The HFCS is conducted at the national level and provides household-level data on assets, liabilities, income, and consumption along with related economic and demographic variables. This data set, thus, provide insights into the financial situation of households as well as their economic behaviour. These aspects can have major implications for the development of the respective economies. The Deutsche Bundesbank emphasizes², that central banks need micro-level information, since *“aggregate data are deemed insufficient”* and micro-data *“opens up the possibility of understanding structural relationships”*.

The intended survey frequency is three years and is conducted by the Household Finance and Consumption Network (HFCN), which consists of statisticians and economists from the ECB, the national central banks of the Eurosystem, and a number of national statistical institutes.³ The results of the most recent wave (wave 3) were published in 2020, with data of over 91,000 households from 19 euro-currency area countries as well as Croatia, Hungary, and Poland, collected from 2016 to 2019.⁴

The HFCS data set is extensively used in research. An application of the first two waves can be found in Costa and Pérez-Duarte (2019), with a focus on deriving wealth inequality from the data. In particular, several inequality measures are calculated and Costa and Pérez-Duarte

¹https://www.ecb.europa.eu/stats/ecb_surveys/hfcs/html/index.en.html

²<https://www.bundesbank.de/en/bundesbank/research/panel-on-household-finances/about-the-phf/about-the-phf-617320>

³https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn-faq.de.html

⁴A comprehensive overview of general aspects of the HFCS and the third wave in detail is given by the Household Finance and Consumption Network (2020).

(2019) analyse the evolution and trends of wealth inequality derived from wave 1 and 2. Further usage and analysis of HFCS data can be found, e.g., in Andreasch and Lindner (2016); Cussen et al. (2018); Waihl (2022).

It turns out that HFCS instrument aggregates (financial and non-financial assets held by the household sector, such as deposits and housing wealth) are usually lower than the corresponding figures from NtIA (see Chakraborty and Waihl (2018)). Therefore, researchers as well as authorities, including the ECB (see Expert Group on Linking macro and micro data for the household sector (2020)), are investigating the reasons causing this macro-micro gap (see, e.g., Schröder et al. (2020)). A commonly recognized cause for the observed discrepancies is differential non-response, which refers to the under-representation of wealthy households (see, e.g., Chakraborty and Waihl (2018); Ruiz and Woloszko (2016); Vermeulen (2018)). Wealthy households are less likely to participate in wealth surveys, and are therefore usually not adequately captured in the HFCS.⁵ To address this problem, Vermeulen (2018) proposes to estimate the missing upper part of the wealth distribution by fitting a Pareto distribution to the tail of the wealth distribution, calibrated on the HFCS data and, if available, additional rich lists (e.g. Forbes World’s billionaires data or country-specific lists like the ranking of Germany’s wealthiest persons provided by the “Manager Magazine”).⁶

Nevertheless, Chakraborty and Waihl (2018) assessed that *“the missing wealthy do not explain large parts of the macro-micro gap for highly comparable instruments (liabilities, bonds, deposits and mutual funds) [...] still leaving significant parts unexplained”*. Engel et al. (2022) show that the instrument coverage ratio of the raw HFCS ranges between 8% and 124% and after adjusting for the missing wealthy households between 35% and 158% for the four largest economies in the euro area, i.e. Germany, Spain, France, and Italy.⁷ This highlights the need for a further adjustment approach to match the national accounts’ aggregates. Therefore, we provide a quantitative solution to close the persisting gaps via economically motivated optimization problems, thereby contributing to the construction of sensible distributional financial accounts.

⁵One way how survey analysts deal with that problem is oversampling the wealthy and, thus, the sample weights in the survey can be adjusted to tackle the problem of non-response. However, Vermeulen (2018) notes, that not all wealth surveys oversample the rich.

⁶The upper tail of wealth and income distributions has been found to follow a Pareto distribution. This was already empirically observed in many papers, see, e.g., Atkinson (2017); Klass et al. (2006); Olgwang (2011). An overview of some theoretical models that corroborate these empirical findings is provided in Ohlwerter (2020). For example, an application of the Pareto distribution to analyse wealth and income inequality can be found in Blanchet et al. (2019, 2018); Lakner and Milanovic (2016); Piketty (2003).

⁷The coverage ratio is defined as the total of the HFCS divided by the national accounts’ total for a specific instrument, i.e. the proportion of national accounts that is reported in the HFCS survey data.

More precisely, without further information on possible causes of the discrepancies, this paper considers the optimization problem to close the remaining discrepancies, while maintaining the level of inequality for each wealth instruments, measured by the Gini coefficient. In addition, upper and lower bounds for the Gini coefficient are derived that can be reached by extreme allocations of the remaining discrepancies.

The paper is structured as follows. Section 2 formalizes the optimization problem. The main objective is to adjust households' instrument holdings in the extended HFCS, such that the NtIA aggregates are matched, while minimizing the impact on inequality measured by the Gini coefficient on instrument level. Section 3 formally proves that there exists an intuitive and unique solution to the optimization problem, maintaining the observed level of inequality in household's wealth instruments. Section 4 derives an upper and a lower bound for the level of inequality that can be reached by extreme allocations of the instrument's wealth discrepancy across the households. Section 5 analyses the findings in a case study based on the German HFCS and compares the results to another approach suggested in the literature, namely the "multivariate calibration". Section 6 concludes.

2 Optimization problem to match HFCS aggregates with NtIA totals

We build upon the approach suggested by Engel et al. (2022), that addresses the problem of the missing wealthy by imputing synthetic wealthy households from a fitted upper tail Pareto distribution and national rich lists, based on the approach developed by Vermeulen (2018). More precisely, the starting point of the optimization problems discussed in this paper is the extended HFCS data that was adjusted for the missing wealthy by the approach suggested by Engel et al. (2022). Likewise, the input data to all empirical results presented in the following is the extended HFCS including the imputed wealthy households.⁸ As shown by Chakraborty and Waltl (2018) and Engel et al. (2022), significant parts of the observed discrepancies between the HFCS and the NtIA wealth instruments remain unexplained, even after extending the HFCS data for the missing wealthy.

The sample of $n \in \mathbb{N}$ households that participate in the HFCS in a given country, or have

⁸These results presented in this paper have been obtained by applying the standard methodology outlined in Engel et al. (2022) and do not take into account country specific features, for example weaknesses of a given source for a specific instrument, or additional information available at the national level.

been imputed from the Pareto distribution or the national rich list, constitute a representation of the country's population. Let $\mathbf{d} = (d_1, \dots, d_n) \in \mathbb{R}_{>0}^n$ denote the household weights, given by the HFCS and imputed for the added wealthy households.⁹ The wealth of households consists of (i) financial assets, such as deposits, investment funds, and pensions, (ii) non-financial assets, such as housing wealth, and (iii) liabilities, such as mortgages.¹⁰ Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}_{>0}^n$ denote the holdings of all households in a certain wealth instrument (or in net wealth, defined as the sum of all financial and non-financial assets less liabilities) in increasing order, i.e. $x_1 \leq x_2 \leq \dots \leq x_n$. For each wealth instrument, the weighted sum of the households' holdings should match its corresponding aggregate $F \in \mathbb{R}_{>0}$ in the NtIA, i.e. $\sum_{i=1}^n d_i x_i = F$ should hold. As explained above, this equality is usually not satisfied due to remaining discrepancies between the extended HFCS and the NtIA.

To close these discrepancies, we introduce an adjustment coefficient vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$, that is (componentwise) multiplied to the households' instrument holdings, such that

$$\sum_{i=1}^n d_i a_i x_i = F. \quad (1)$$

Obviously, there exist infinite many solutions to Equation (1), of which many are economically not reasonable. To derive meaningful solutions, the following three additional constraints are introduced. *First*, as the term $a_i x_i$ denotes the adjusted instrument holding of household i , and instrument holdings can for economic reasons not be negative, $a_i > 0$ for all $i \in \{1, \dots, n\}$. *Second*, in most cases undercoverage is observed, i.e. the HFCS total is typically lower than the NtIA aggregate. This means, additional wealth has to be allocated across the n households. As there is no reason to reallocate any further amounts, we want to ensure that each household i maintains at least its current holdings of x_i . This implies that $a_i \geq 1$ for all $i \in \{1, \dots, n\}$. In case of overcoverage, i.e. when the HFCS total is higher than the NtIA aggregate, household holdings have to be reduced, and thus $a_i \leq 1$ for all $i \in \{1, \dots, n\}$. In the following, we focus on the case of undercoverage, as this constitutes the common case. All results, however, can be analogously derived in the case of overcoverage. *Third*, assuming that the wealth order of the

⁹The scalar product of the household weights and the corresponding numbers of household members yields the country's population.

¹⁰For example, Engel et al. (2022) define household wealth as the sum of the following balance sheet items: (i) financial assets comprising debt securities, deposits, investment fund shares, life insurance and voluntary pensions, listed shares and financial business wealth (i.e. unlisted shares and other equity); (ii) non-financial assets comprising non-financial business wealth and housing wealth; and (iii) liabilities, which are split into mortgages and other liabilities.

households' instrument holdings are correctly captured by the extended HFCS, this ordering should be preserved, i.e. $a_i x_i \leq a_{i+1} x_{i+1}$ for all $i \in \{1, \dots, n-1\}$. This means if household A has more holdings in a certain instrument than household B, this should still be present after the adjustment. Without additional information indicating which households might have underreported their holdings more than others, it seems unjustified to modify the instrument's wealth order captured by the extended HFCS.

In addition, since we do not have further information on how to allocate the observed discrepancy of wealth between the extended HFCS and NtIA, we aim at allocating the additional wealth such that the extended HFCS data is modified as little as possible. In contrast to the Pareto fitting, where one seeks to add wealthy households to the data, we assume for our optimization that the wealth inequality in each instrument resulting from the HFCS after the Pareto fitting is correct.¹¹ More precisely, we see no additional reasons that would justify a further modification of wealth inequality. Thus, the objective function will be related to an inequality measure. Particularly, the difference in wealth instrument's inequality before and after adjustments with the vector \mathbf{a} should be minimized. There are many possible choices for such an inequality measure, including the Gini coefficient, the Atkinson index, or the Generalised Entropy indices (see Costa and Pérez-Duarte (2019)). Here, we opt for the Gini coefficient for three reasons. First, it is the most popular inequality measure. Second, its definition is easy to interpret. Third, as we will see later, the empirical version of the Gini coefficient has some useful mathematical properties (especially linearity).

The Gini coefficient is defined via the Lorenz curve, which is a graphical representation of the cumulative distribution of wealth versus the cumulative distribution of the population, illustrated in Figure 1. It follows from the definition that the Lorenz curve is equal to the 45-degree line in case of equality, i.e., if every agent possesses the same wealth. The area between the 45-degree line and the Lorenz curve is called the area of concentration. The Gini coefficient is defined as the ratio of the area of concentration to the maximum possible concentration area, i.e. the area shaded in grey in Figure 1 divided by 0.5. The Gini coefficient is, hence, bounded

¹¹It is noted that an assessment of the goodness of the extended HFCS would require admin data, which is typically not available. However, the HFCS is based on a common methodology that was carefully developed by the HFCN, aiming to address possible issues such as non-response, weighting, and data editing (see Household Finance and Consumption Network (HFCN) (2020)). Furthermore, Tiefensee and Grabka (2014) analysed the data quality of the HFCS in 2014 (i.e. the first HFCS wave) and concluded i.a. that "Taken together the HFCS is still the best dataset for cross country comparisons of wealth levels and inequality in the Euro area and it is definitely a first (big) step into the right direction." In this paper we use the extended HFCS as a starting point.

below by 0, in the case of perfect equality, and by 1, respectively, when reaching the maximum level of inequality.



Figure 1: Lorenz curve for a Pareto distribution. The shaded area is called the area of concentration.

In order to work with survey data, we use a weighted empirical version of the Gini coefficient, which is defined as

$$G(\mathbf{a}, \mathbf{d}, \mathbf{x}) := 1 - \sum_{k=1}^n (W_k - W_{k-1})(X_k + X_{k-1}), \quad (2)$$

where

$$W_k := \frac{\sum_{\ell=1}^k d_\ell}{\sum_{\ell=1}^n d_\ell}, \quad W_0 = 0, \quad \text{and} \quad X_k := \frac{\sum_{\ell=1}^k d_\ell a_\ell x_\ell}{\sum_{\ell=1}^n d_\ell a_\ell x_\ell}, \quad X_0 = 0, \quad (3)$$

respectively, denote the cumulative share of population and the cumulative share of wealth (see Costa and Pérez-Duarte (2019)).

This leads to the following optimization problem for the case of undercoverage:

$$\min_{\mathbf{a} \in \mathbb{R}^n} (G^* - G(\mathbf{a}, \mathbf{d}, \mathbf{x}))^2 \quad (4a)$$

$$\text{subject to} \quad -a_i \leq -1, \quad \forall i \in \{1, \dots, n\}, \quad (4b)$$

$$a_i x_i - a_{i+1} x_{i+1} \leq 0 \quad \forall i \in \{1, \dots, n-1\}, \quad (4c)$$

$$\sum_{i=1}^n d_i a_i x_i = F \quad (4d)$$

where G^* denotes the Gini coefficient after the Pareto fitting. It is recalled that constraint (4b) ensures that each household maintains at least the instrument holdings that are observed in the extended HFCS; constraint (4c) ensures that the instrument's wealth order of the households is preserved; and constraint (4d) ensures that the HFCS aggregate is matched with its total given by the NtIA.

3 Solution to the optimization problem with invariant Gini coefficient

We observe that the Gini coefficient is invariant under multiplication with a constant, i.e. for $a_1 = \dots = a_n$ it holds $G(\mathbf{a}, \mathbf{d}, \mathbf{x}) = G^*$. Considering constraint (4d), it is evident that one solution is given by

$$a_1 = \dots = a_n = \frac{F}{\sum_{i=1}^n d_i x_i}. \quad (5)$$

Consequently, our interest lies in the space of solutions to the refined optimization problem with the objective function $G^* - G(\mathbf{a}, \mathbf{d}, \mathbf{x}) = 0$ and the constraints (4b) to (4d). This usually leads to an infinite solution space. To reduce the complexity and facilitating the analytical assessment of the optimization problem, we replace constraints (4c) by the stronger constraints $a_i \leq a_{i+1}$ for all $i \in \{1, \dots, n-1\}$. Thus, the problem we want to solve equates to

$$\text{solve} \quad G^* - G(\mathbf{a}, \mathbf{d}, \mathbf{x}) = 0 \quad (6a)$$

$$\text{subject to} \quad -a_i \leq -1, \quad \forall i \in \{1, \dots, n\}, \quad (6b)$$

$$a_i - a_{i+1} \leq 0 \quad \forall i \in \{1, \dots, n-1\}, \quad (6c)$$

$$\sum_{i=1}^n d_i a_i x_i = F. \quad (6d)$$

In the following we show that the solution given in Equation (5) constitutes the unique solution to this optimization problem.

Theorem 1. In the case of undercoverage, i.e. for $\sum_{i=1}^n d_i x_i \leq F$, the unique solution to the optimization problem defined in Equations (6a) to (6d) is given by

$$a_1 = \dots = a_n = \frac{F}{\sum_{i=1}^n d_i x_i}. \quad (7)$$

Proof. The full proof is given in Ohlwerter (2020), whose author is also the first author of the present paper. The strategy to prove the theorem is as follows:

First, we observe that there are two equations, namely $G^* - G(\mathbf{a}, \mathbf{d}, \mathbf{x}) = 0$ and $\sum_{i=1}^n d_i a_i x_i = F$, each of them having an $n - 1$ dimensional solution space (hyperplane). We intersect these hyperplanes in order to get expressions for a_{n-1} and a_n dependent on a_1, \dots, a_{n-2} (Lemma 1 in Ohlwerter (2020)). Second, by applying the inequality constraints in Equations (6b) and (6c), we first show that $a_1 \leq \frac{F}{\sum_{i=1}^n d_i x_i}$ (Lemma 2 in Ohlwerter (2020)). Third, by induction, we conclude that $a_m \leq \frac{F}{\sum_{i=1}^n d_i x_i}$ for all $m \in \{1, \dots, n - 2\}$ (Lemma 3 in (Ohlwerter, 2020)). Forth, having proved that $a_m \leq \frac{F}{\sum_{i=1}^n d_i x_i}$ for all $m \in \{1, \dots, n - 2\}$, we show step by step (starting with $m = n - 2$) that $a_m \geq \frac{F}{\sum_{i=1}^n d_i x_i}$ and conclude that $a_m = \frac{F}{\sum_{i=1}^n d_i x_i}$ for all $m \in \{1, \dots, n - 2\}$ (Lemma 4 in Ohlwerter (2020)). Fifth, inserting these values into the expressions we have established for a_{n-1} and a_n , we can conclude that the solution $a_1 = \dots = a_n = \frac{F}{\sum_{i=1}^n d_i x_i}$ is indeed unique. \square

4 Lower and upper bound for the Gini coefficient after adjustment

Besides the solution derived in the previous section, it is also of great interest to derive the lower and upper bound of inequality that can be attained by allocating the additional amount of wealth (or, in case of overcoverage, by extracting the amount of wealth in excess), i.e. by closing the discrepancies between the aggregates of the extended HFCS and the NtIA. A lower and an upper bound provide a range for the possible level of inequality and can thus serve as an indicator for the level of uncertainty, when measuring inequality based on the HFCS. The bounds also allow to detect where the Gini coefficient resulting from the proportional adjustment given by Theorem 1 is located and to compare it to Gini coefficients obtained by other methods suggested in the literature as we illustrate later. Our focus again lies on the

case of undercoverage, i.e. $\sum_{i=1}^n d_i x_i < F$, as this constitutes the typical case. In the case of overcoverage, i.e. if $\sum_{i=1}^n d_i x_i > F$, the results can be derived similarly.

4.1 Upper bound

Let us first take a look at the upper bound, i.e., we aim at maximizing the Gini coefficient such that:

- (i) all households keep at least their current instrument holdings,
- (ii) the households' ranking in terms of their instrument holdings is preserved, and
- (iii) the NtIA total is matched.

This translates into the following optimization problem:

$$\max_{\mathbf{a} \in \mathbb{R}^n} G(\mathbf{a}, \mathbf{d}, \mathbf{x}) \quad (8a)$$

$$\text{subject to} \quad -a_i \leq -1, \quad \forall i \in \{1, \dots, n\}, \quad (8b)$$

$$a_i x_i - a_{i+1} x_{i+1} \leq 0, \quad \forall i \in \{1, \dots, n-1\}, \quad (8c)$$

$$\sum_{i=1}^n d_i a_i x_i = F. \quad (8d)$$

The solution is in line with what one would intuitively expect. Inequality is maximized if the additional wealth, that is to be allocated, is assigned exclusively to the household that already (before the adjustment) owns the highest holdings in the considered instrument, i.e. household n .

Theorem 2. In case of undercoverage, i.e. for $\sum_{i=1}^n d_i x_i < F$, the solution to the optimization problem defined in Equations (8a) to (8d) is given by

$$\mathbf{a} = \left(1, \dots, 1, \frac{F - \sum_{i=1}^{n-1} x_i d_i}{d_n x_n} \right)^T. \quad (9)$$

Proof. To verify this, one starts with a more general optimization problem, where we relax the constraints by ignoring constraint (8c) of preserving the households' ranking. Using the method

of Lagrange, one can verify that $\mathbf{a} = \left(1, \dots, 1, \frac{F - \sum_{i=1}^{n-1} x_i d_i}{d_n x_n}\right)^T$ solves the relaxed optimization problem.

Since we consider the case of undercoverage, $\frac{F - \sum_{i=1}^{n-1} x_i d_i}{d_n x_n} > 1$. Thus, $\mathbf{a} = \left(1, \dots, 1, \frac{F - \sum_{i=1}^{n-1} x_i d_i}{d_n x_n}\right)^T$ also fulfills the constraint $a_i x_i - a_{i+1} x_{i+1} \leq 0$ for all $i \in \{1, \dots, n-1\}$ and we observe that this is also a feasible point for Problem (8). We conclude that this is, therefore, a solution to Problem (8). Hence, the upper bound for the Gini coefficient after adjustment is attained by assigning all additional wealth to the wealthiest (w.r.t. the considered instrument) household, i.e. $\mathbf{a} = \left(1, \dots, 1, \frac{F - \sum_{i=1}^{n-1} x_i d_i}{d_n x_n}\right)^T$. \square

4.2 Lower bound

Next we establish the lower bound, i.e. we solve the following optimization problem:

$$\min_{\mathbf{a} \in \mathbb{R}^n} G(\mathbf{a}, \mathbf{d}, \mathbf{x}) \quad (10a)$$

$$\text{subject to} \quad -a_i \leq -1, \quad \forall i \in \{1, \dots, n\}, \quad (10b)$$

$$a_i x_i - a_{i+1} x_{i+1} \leq 0, \quad \forall i \in \{1, \dots, n-1\}, \quad (10c)$$

$$\sum_{i=1}^n d_i a_i x_i = F. \quad (10d)$$

The solution is again in line with what one would expect. Inequality is minimized if the additional wealth that is to be allocated is distributed among the poorest households in a step-wise procedure leveling them respectively up to the subsequent wealthiest household as much as possible by the additional amount of wealth.

Theorem 3. In the case of undercoverage, i.e. for $\sum_{i=1}^n d_i x_i < F$, the solution to the optimization problem defined in Equations (10a) to (10d) is determined by the following steps:

Step 1: Determine $j \in \{1, \dots, n\}$ that fulfills the following two inequalities:

$$\sum_{i=1}^j d_i x_j \leq F - \sum_{k=j+1}^n d_k x_k \quad \text{and} \quad \sum_{i=1}^{j+1} d_i x_{j+1} \geq F - \sum_{k=j+2}^n d_k x_k. \quad (11a)$$

$$\text{Step 2: Calculate} \quad c = \frac{F - \sum_{k=j+1}^n d_k x_k}{x_j \sum_{i=1}^j d_i}. \quad (11b)$$

$$\text{Step 3: Set } a_i = c \frac{x_j}{x_i} \quad \forall i \in \{1, \dots, j\} \quad \text{and} \quad a_i = 1 \quad \forall i \in \{j+1, \dots, n\}. \quad (11c)$$

Proof. Again, we use a detour to a more general problem by ignoring constraint (10c) on the households' ranking. Using the method of Lagrange, one can observe that the gap should be allocated to the poorest household and $a_i = 1$ for all $i \in \{2, \dots, n\}$ in order to solve this relaxed optimization problem. But we can only increase a_1 until $a_1 x_1 = x_2$, because with a further increase, the order of the households' instrument holdings and therefore the formula for calculating the Gini coefficient would change¹². Hence, for calculating the lower bound, we repeatedly have to increase the instrument holdings of the poorest household¹³ until either the adjusted instrument holding equals the second poorest household or the full gap $F - \sum_{i=1}^n d_i x_i$ is allocated. One can immediately conclude that this is also the solution to Problem (10). \square

It is also worth pointing out that the optimization Problems (8) and (10) could be solved with the simplex algorithm using generic solvers in, e.g., R and `matlab`. Nonetheless, it is important and often useful to know these bounds analytically, since HFCS datasets are often large, which leads to long runtimes.

5 Case study: Household wealth in Germany

This section applies the derived proportional allocation (Theorem 1), the lower (Theorem 3), and the upper bound (Theorem 2) of the Gini coefficient to the German¹⁴ HFCS of the third wave, which was conducted between March 2017 and October 2017 (see Household Finance and Consumption Network (2020)).¹⁵ In addition, we compare the derived solutions to another approach suggested in the literature, namely the multivariate calibration.

Following Engel et al. (2022), the considered wealth concept comprises (i) financial assets: debt securities, deposits, investment fund shares, life insurance and voluntary pensions, listed shares, and financial business wealth; (ii) non-financial assets: non-financial business wealth and housing wealth; and (iii) liabilities: mortgages and other liabilities. The net wealth of a household is given by the difference between assets and liabilities.

¹²Remember that the Gini coefficient is defined for $a_1 x_1 \leq a_2 x_2 \leq \dots \leq a_n x_n$.

¹³Households with the same amount of instrument holdings are treated as one.

¹⁴An application of the proportional allocation to the HFCS of other euro area countries is provided in Engel et al. (2022).

¹⁵It is recalled that the input data to the optimization problems is given by the extended HFCS, that was adjusted for the missing wealthy households as suggested by Engel et al. (2022). As explained in that paper, the method followed benefited from work performed by the European System of Central Banks' (ESCB) expert group on distributional financial accounts (EG DFA).

Figure 2 shows the coverage ratio for each instrument, which is defined as the HFCS instrument total divided by the NtIA aggregate. Note that the HFCS data have already been adjusted for the missing wealthy households. We see that most instruments show an undercoverage, i.e., the HFCS aggregate is lower than the corresponding NtIA instrument total. Nevertheless, in case of ‘Listed Shares’ and ‘Debt Securities’ the HFCS and the NtIA total match exactly¹⁶, whereas in case of ‘Financial Business Wealth’ we observe an overcoverage of approximately 4%, i.e. the HFCS aggregate is higher than the NtIA total. The undercoverage is most pronounced for ‘Deposits’ with a coverage ratio of roughly 55%.

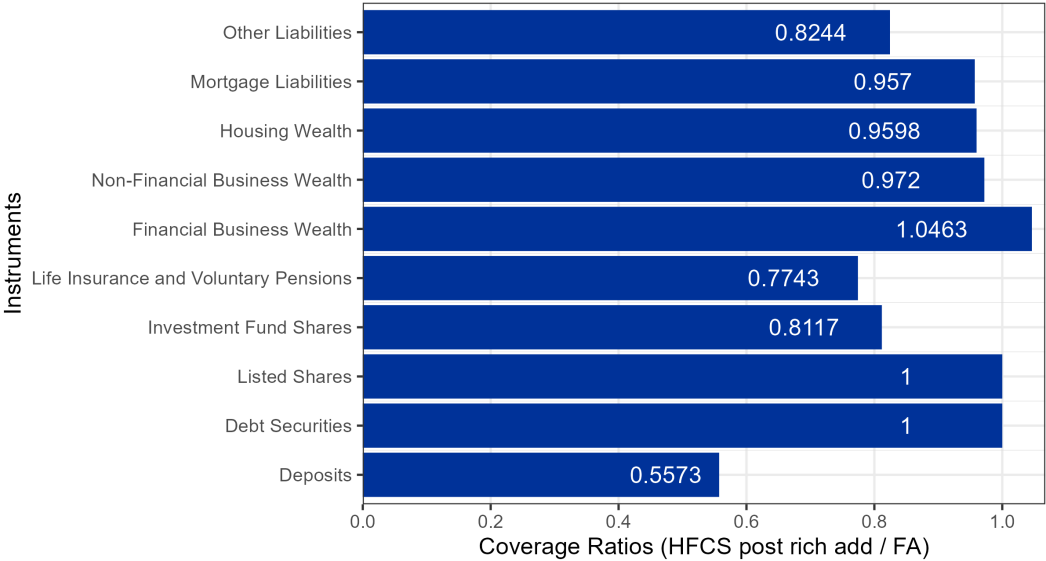


Figure 2: Coverage ratios of wealth instruments for German households.

Section 5.1 introduces the multivariate calibration (MC). Section 5.2 compares the adjustment factors resulting from the proportional allocation (PA) to those from the MC. The analysis shows that the MC leads to a wider range of adjustment factors than the PA and simultaneously increases the instrument holdings of some households, while decreasing those of other households. In contrast, the PA distributes only the observed discrepancy between the extended HFCS and the NtIA, while otherwise maintaining households’ instrument holdings reported in the extended HFCS. Section 5.3 analyses further the impact of both approaches on the measured level of inequality and finds that the MC leads to higher levels of inequality, thus corroborating conclusions of other papers.

¹⁶This is due to the portfolio allocation of the added wealthy households, see Engel et al. (2022) for further information.

5.1 Multivariate calibration

Besides the proportional adjustment derived in Theorem 1, some papers suggest a so-called multivariate calibration as an alternative method to close the discrepancies between the HFCS and the NtIA (see Cantarella et al. (2021), Expert Group on Linking macro and micro data for the household sector (2020), Kennickell et al. (2022)). The idea behind the multivariate calibration is to find household specific correction factors $\mathbf{a} \in \mathbb{R}^n$, such that if multiplied to the household holdings of a set of wealth instruments, the aggregated corrected totals match the NtIA figures.¹⁷ This approach allows to maintain the relative portfolio composition of households to some extent, e.g. if a household holds twice the amount it invests in shares in a deposits account, this relationship is preserved after applying the multivariate calibration. In addition, as an objective function, the impact, measured by the χ^2 -distance, of \mathbf{a} being multiplied to the household weights \mathbf{d} is minimized. This means that the multivariate calibration aims to adjust especially those households that have a low weight (i.e. a low representativeness) and high instrument holdings. Furthermore, Cantarella et al. (2021), the Expert Group on Linking macro and micro data for the household sector (2020), and Kennickell et al. (2022) suggest to split this optimization problem into two sub-problems by calibrating separately the holdings of the wealthy households, denoted by the set I_{top} , and those of the remaining households, denoted by I_{bottom} .¹⁸ To distinguish the wealthy households, a net wealth threshold w_0 of EUR 1 million is used. The NtIA total of an instrument j is accordingly split into $F_{\text{calib.bottom},j}$ and $F_{\text{calib.top},j}$, denoting the totals that should be hold by the households in I_{bottom} and I_{top} , respectively. Kennickell et al. (2022) point out that $F_{\text{calib.bottom},j}$ and $F_{\text{calib.top},j}$ should be set such that the share of wealth above and below w_0 is maintained. Moreover, to limit the changes on each household, Kennickell et al. (2022) point out that generally each adjustment factor a_i is bounded by $0.003 \leq a_i \leq 1,000$. Note that this means, that irrespective of whether an instrument total in the HFCS is below or above the NtIA figure, the holdings of a household may be decreased to 0.3% of the holdings captured in the HFCS and increased up to one thousand times of what is denoted in the HFCS. As these bounds seem very extreme, we use slightly stricter bounds of 0.03 and 100 in the following. This implies that irrespective of whether

¹⁷In this paper the adjustment factors of the multivariate calibration are multiplied to the households' instrument holdings. Alternatively, it can be considered to multiply the adjustment factors to the HFCS household weights. This, however, may lead to changes in the population, thereby requiring further adjustments. Moreover, the HFCS household weights have been carefully calibrated by the HFCN.

¹⁸Note that $I_{\text{bottom}} \cap I_{\text{top}} = \emptyset$ and $I_{\text{bottom}} \cup I_{\text{top}} = \{1, \dots, n\}$ holds.

additional wealth has to be allocated or extracted on a certain instrument, the holdings of some households may decrease and those of others may increase. The multivariate calibration, thus, solves the following optimization problem

$$\min_{\mathbf{a} \in \mathbb{R}^n} \chi^2(\mathbf{d}, \mathbf{a}) = \min_{\mathbf{a} \in \mathbb{R}^n} \sum_{i=1}^n \frac{(d_i a_i - d_i)^2}{d_i} = \min_{\mathbf{a} \in \mathbb{R}^n} \sum_{i=1}^n d_i (a_i - 1)^2 \quad (12a)$$

$$\text{subject to} \quad \sum_{i \in I_{\text{bottom}}} d_i a_i x_{ij} = F_{\text{calib.bot},j} \quad \forall j \in \mathcal{J}, \quad (12b)$$

$$\sum_{i \in I_{\text{top}}} d_i a_i x_{ij} = F_{\text{calib.top},j}, \quad \forall j \in \mathcal{J}, \quad (12c)$$

$$a_i \geq 0.03, \quad \forall i \in \{1, \dots, n\}, \quad (12d)$$

$$a_i \leq 100, \quad \forall i \in \{1, \dots, n\}, \quad (12e)$$

where \mathcal{J} denotes a set of wealth instruments that are jointly considered by the multivariate calibration. Cantarella et al. (2021) suggest to consider jointly the financial assets {Deposits, Debt Securities, Listed Shares, Investment Fund Shares, Life Insurance and Voluntary Pensions, Mortgage Liabilities, Other Liabilities}. Business wealth assets and housing wealth are matched separately, i.e. {Financial Business Wealth, Non-Financial Business Wealth}, and {Housing Wealth}. It is highlighted that the solution space of the multivariate problem depends on the portfolio composition of the households considered in each set I_{bottom} and I_{top} , as well as on the set of considered instruments, and that cases may exist for which the solution space is empty.¹⁹ While we have derived an explicit and intuitive solution for the optimization problem considered in Equation (4), the multivariate calibration is in practice solved in R via the `gencalib` function of the sampling package (see Kennickell et al. (2022)).

It is noted that the proportional adjustment of Theorem 1 could likewise be split into sub-problems by dividing the household population in distinct sets and choosing the desired instrument totals accordingly. This would result in distinct optimization problems, for which the results derived above in Section 3 and Section 4 hold likewise.

¹⁹Kennickell et al. (2022) note that the bounds 0.003 and 1,000 are in some cases too restrictive for a solution to be achieved.

5.2 Comparison of the proportional adjustment and the multivariate allocation

By applying both the proportional allocation (PA) and the multivariate calibration (MC) to the extended German HFCS data set, i.e. after capturing the missing wealthy households, we can analyse and compare the impact of both methods.

One important difference between the PA and the MC is the heterogeneity of the derived adjustment factors \mathbf{a} . The PA leads for each instrument to one adjustment factor that is applied to all households. Thus, the instrument holdings of all households are increased in case of undercoverage and decreased in case of overcoverage, proportionally to their holdings before the adjustment. This is a feature that we explicitly integrated in the optimization problem, as we have no further information that would justify to decrease (in case of undercoverage), or increase (in case of overcoverage), holdings of certain households in the respective instrument. In contrast, the MC derives a separate adjustment factor for each household, that is applied to multiple instruments. As explained in the previous section, this means that the considered instrument holdings of some households may decrease, while those of others are increased. Table 1 shows that for financial assets and for business wealth, the MC maintains the instrument holdings of a majority of the households. This implies that the changes that are required in order to match the NtIA totals are concentrated on a minority of households. For both sets of assets, the MC simultaneously decreases the holdings of some households, while it increases the holdings of others. Moreover, the number of households that see a decrease in their holdings is roughly ten times as high as the number of households that see an increase in their holdings. While this behaviour may be explained for business wealth by the fact that non-financial business wealth is underestimated in the HFCS, while financial business wealth is overestimated²⁰, this is surprising for financial assets, as all of these instruments are underestimated by the HFCS (see Figure 2). This effectively means that the MC besides the additional amount of wealth in financial assets that is to be allocated in order to match the NtIA totals, redistributes household holdings in financial assets reported by the HFCS. This phenomenon is also observed for housing wealth, an instrument that is adjusted separately. Housing wealth seems to be well reported in the HFCS with a coverage of 96% (see Figure 2). Nevertheless, the MC decreases and increases

²⁰It is recalled that under- and overestimation of households' aggregated instruments holdings between the extended HFCS, i.e. adjusted by the estimate for the missing wealthy, and the NtIA totals may also be due to inaccuracies in the linking of the instruments and to the approach of estimating the missing wealthy.

the reported housing wealth of respectively 46% and 50% of the households, while maintaining the reported value for 4% of the households.

	number and share of households whose instrument holdings are		
	decreased	maintained	increased
financial assets \mathbf{a}	12,675,437 (31.41%)	26,642,813 (66.03%)	1,034,157 (2.56%)
business wealth \mathbf{a}^{biz}	3,492,582 (8.65%)	36,469,765 (90.38%)	390,060 (0.97%)
housing wealth \mathbf{a}^{h}	18,553,099 (45.98%)	1,558,559 (3.86%)	20,240,747 (50.16%)

Table 1: Number and share of households where the MC decreases (i.e. $a_i < 1$), maintains (i.e. $a_i = 1$), or increases (i.e. $a_i > 1$) the instrument holdings, respectively for each set of assets that are considered jointly by the MC.

Table 2 takes a closer look at the range of adjustment factors derived by the MC and compares them to those derived by the PA. The adjustment factors of the PA reflect the instruments' coverage (see Figure 2) and fall in the interval [0.956; 1.794]. As the coverage is worst for deposits with a coverage of only 56%, the corresponding adjustment factor is the highest with a value of 1.8. The MC shows a wide range of adjustment factors spanning the interval [0.03; 19.447]. This means that the MC decreases the holdings of some households to 3% of the value reported by the extended HFCS, while it increases the holdings of other households by a factor of 19.²¹ Note that the lowest adjustment factor of 0.03 is effectively set by the lower bound specified in the optimization problem of the MC (constraint Equation (12d)). Despite the large range of adjustment factors derived by the MC, the mean, median, and the interquartile range of the MC adjustment factors are closely concentrated around 1. For financial assets and for business wealth, this can be explained by the large share of households that maintain their holdings (see Table 1). For housing wealth, this means that while 96% of households see either an increase or a decrease of their holdings for at least half of them, the impact is rather minor.

²¹To put this into perspective, if the housing wealth of the household that receives the adjustment factor of 19 amounts to EUR 500,000 before the MC, then it will be worth EUR 9,500,000 after the MC.

	Multivariate Calibration						Proportional Adjustment	
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Coeff.	Instrument
\mathbf{a}	0.030	0.990	1.015	1.102	1.128	16.613	1.794	Deposits
							1.000	Debt Securities
							1.000	Listed Shares
							1.232	Investment Fund Shares
							1.291	Life Insurance and Voluntary Pensions
							1.045	Mortgage Liabilities
							1.212	Other Liabilities
\mathbf{a}^{biz}	0.211	1.000	1.000	0.995	1.000	2.311	0.956	Financial Business Wealth
							1.029	Non-financial Business Wealth
\mathbf{a}^{h}	0.519	0.943	1.000	0.966	1.000	19.447	1.042	Housing Wealth

Table 2: Summary statistics of the adjustment coefficients derived by the MC and the PA for each instrument (rounded to 3 digits).

5.3 Impact of the proportional adjustment and the multivariate allocation on inequality

The Expert Group on Linking macro and micro data for the household sector (2020) notes for the MC that “*the gap* [i.e. the discrepancy between the HFCS and the NtIA], *when positive* [i.e. in case of undercoverage], *is allocated more than proportionally to rich households*” and that “*A negative gap would be allocated more to poor households*”. In this section, we investigate this observation by calculating the share of instrument wealth possessed by different wealth groups of the households population and by analysing the resulting level of inequality, measured by the Gini coefficient.

Figure 3 shows the share of instrument wealth owned by the richest 10% and the poorer half of households before and after applying the MC. Note that these shares (on instrument level) are maintained by the PA, as the instruments’ discrepancies between the HFCS and the NtIA are allocated proportionally to the households’ holdings before the adjustment. In line with the observation by the Expert Group on Linking macro and micro data for the household sector (2020), we observe that for those instruments that are underestimated in the extended HFCS (see Figure 2), the share of the poorer households reduces significantly, whereas the share of the richest households increases by a noteworthy margin. Moreover, this phenomenon is also

persistent in case of overcoverage (i.e. for financial business wealth, see Figure 2) and in case the extended HFCS matches the NtIA (i.e. for listed shares and debt securities, see Figure 2). The MC increases the holdings of the wealthy households and decreases the holdings of the poorer households across all instruments. This observation suggests that the MC method treats poorer households less favorably in terms of allocating the wealth gaps. Depending on the amount by which also debt instruments (i.e. mortgage and other liabilities) are increased for the wealthy households and decreased for the poorer households, the MC may lead to a substantial increase in inequality. Indeed, Kennickell et al. (2022) note that *“the mechanics of multivariate calibration in comparison with proportional adjustment – i.e. multiplying the wealth of each household by a constant factor – are associated with higher levels of inequality.”* We proceed to analyse this further by computing the Gini coefficients.

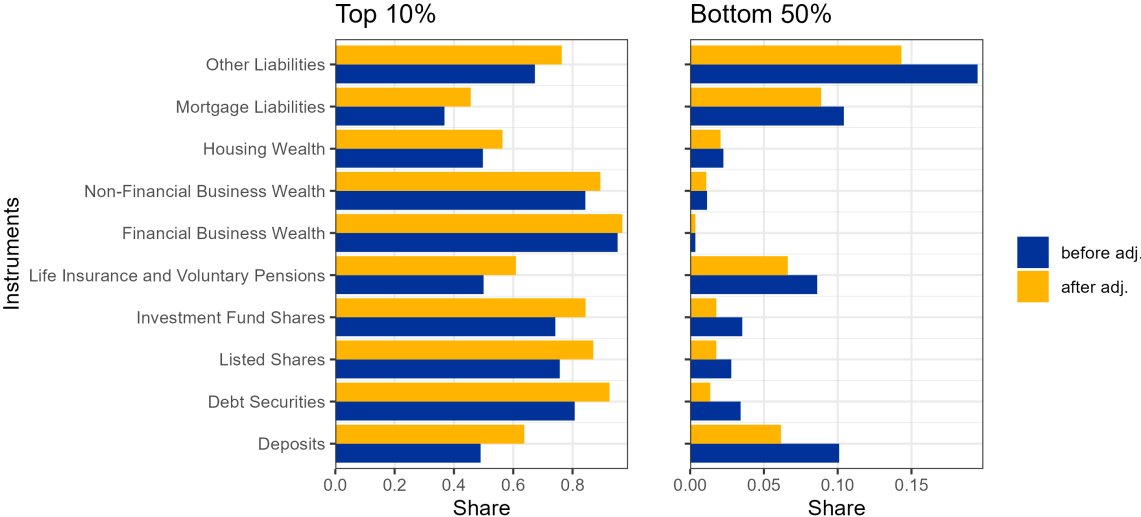


Figure 3: Share of instrument holdings possessed by the richest 10% and the poorer half of households before (blue bars) and after (yellow bars) applying the MC. Note that the PA preserves the shares, which are therefore equal to the blue bars (before and after applying the PA).

Figure 4 compares the Gini coefficients obtained by the PA (yellow dots) with those from the MC (blue crosses).²² In line with the conclusions by Kennickell et al. (2022), we observe that across all instruments the Gini coefficient is always higher for the MC than for the PA. The difference between the corresponding Gini coefficients of the MC and the PA is very low for instruments that have a high coverage ratio, namely ‘Non-Financial Business Wealth’, ‘Finan-

²²Little variability is observed across the imputations of the HFCS, resulting in a maximum absolute difference of 0.01 in the Gini coefficients.

cial Business Wealth', 'Listed Shares', and 'Debt Securities'. The difference is highest for the instrument with the lowest coverage, namely 'Deposits'.

In addition, Figure 4 shows the minimal and maximal Gini coefficient that can be derived for each instrument under the constraints that households' instrument holdings are not decreased in case of undercoverage, not increased in case of overcoverage, and that the order of households' instrument wealth is maintained (see Section 4). The length of the derived interval of the Gini coefficient naturally depends on the instrument's coverage ratio. The lower the coverage ratio, the higher the uncertainty of the true allocation of household holdings, and, thus, the longer the interval of the minimal and maximal Gini coefficient that can be derived. Interestingly, we observe for some instrument that the Gini coefficient after the MC exceeds the maximum Gini coefficient that is derived when preserving the wealth order and when not reallocating more wealth than the discrepancy between the extended HFCS and the NtIA. This phenomenon occurs for 'Other Liabilities', 'Mortgage Liabilities', 'Housing Wealth', 'Non-Financial Business Wealth', and 'Investment Fund Shares'.

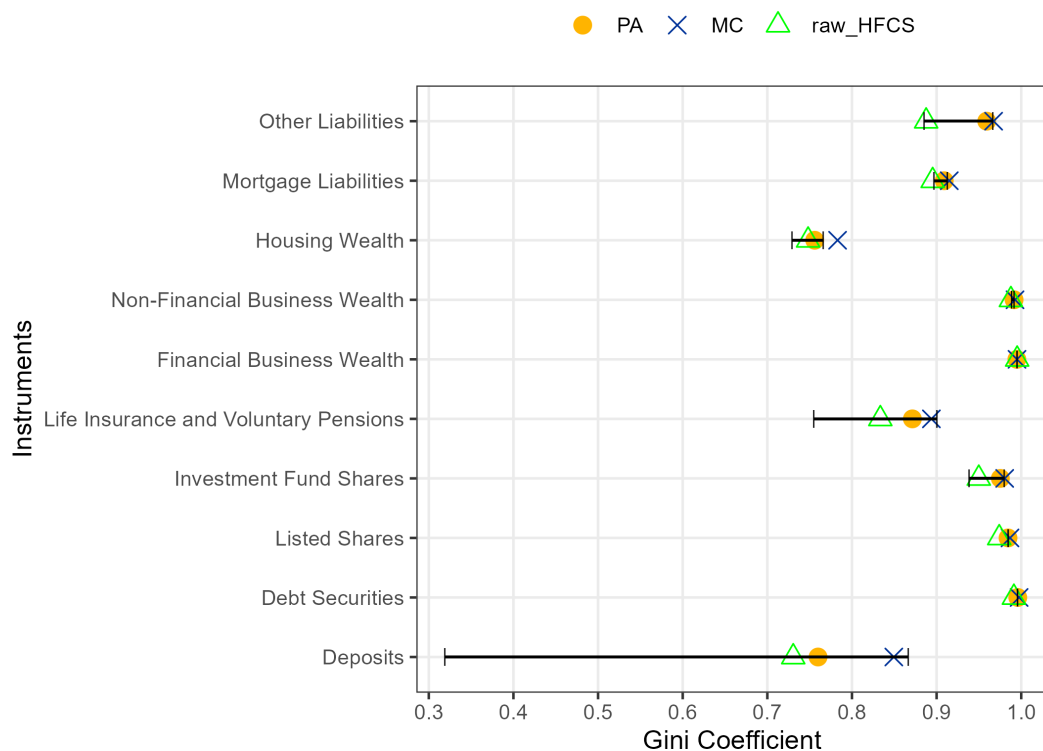


Figure 4: Gini coefficients of each instrument resulting from the PA (yellow dot) and the MC (blue cross). The black lines denote the resulting range obtained from minimizing (Theorem 3) and maximizing (Theorem 2) the Gini coefficients according to the results of Section 4. The green triangles indicate the Gini coefficients of the raw HFCS data (before imputing the missing wealthy households).

To investigate the derived level on inequality in more detail, Figures 5a to 5c show the Lorenz curves for ‘Deposits’ (the instruments with the highest discrepancy, i.e. the highest uncertainty of the true allocation), ‘Housing Wealth’ (an instrument that is adjusted separately also by the MC), and ‘Non-Financial Business Wealth’ (an instrument that is typically hold exclusively by wealthy households). Due to the high concentration of ‘Non-Financial Business Wealth’ being hold only by wealthy households, Figure 5c only shows the share of population from 0.9 to 1. The plots clearly demonstrate that a higher undercoverage implies more widespread Lorenz curves across the different methods. The shape of the Lorenz curves is also influenced by the concentration of instrument holdings, because wealth is only allocated to households with non-zero instrument holdings. For example, in the case of ‘Deposits’, where the undercoverage is high (the coverage ratio amounts to 56%) and more than 90% of the households possess ‘Deposits’, the Lorenz curve of the minimal Gini coefficient (Theorem 3) is similar to a straight line and

much closer to the equality line than the other Lorenz curves. In contrast to ‘Deposits’, for ‘Housing Wealth’ and ‘Non-Financial Business Wealth’ the Lorenz curves of the PA and the derived upper and lower bounds are closer together, as the high coverage ratios of 96% and 97% respectively (see Figure 2) leave little room to alter the level of inequality. For all three instruments, we can observe that the Lorenz curve of the MC is mostly below the Lorenz curve corresponding to the PA, thereby indicating a higher level of inequality for the MC than for the PA. For ‘Housing Wealth’ and ‘Non-Financial Business Wealth’, the Lorenz curve of the MC is even below the Lorenz curve of the maximal Gini coefficient (Theorem 2), thereby indicating a level of inequality that exceeds the analytically derived upper bound, that is possible when only the observed discrepancy between the extended HFCS and the NtIA is allocated.

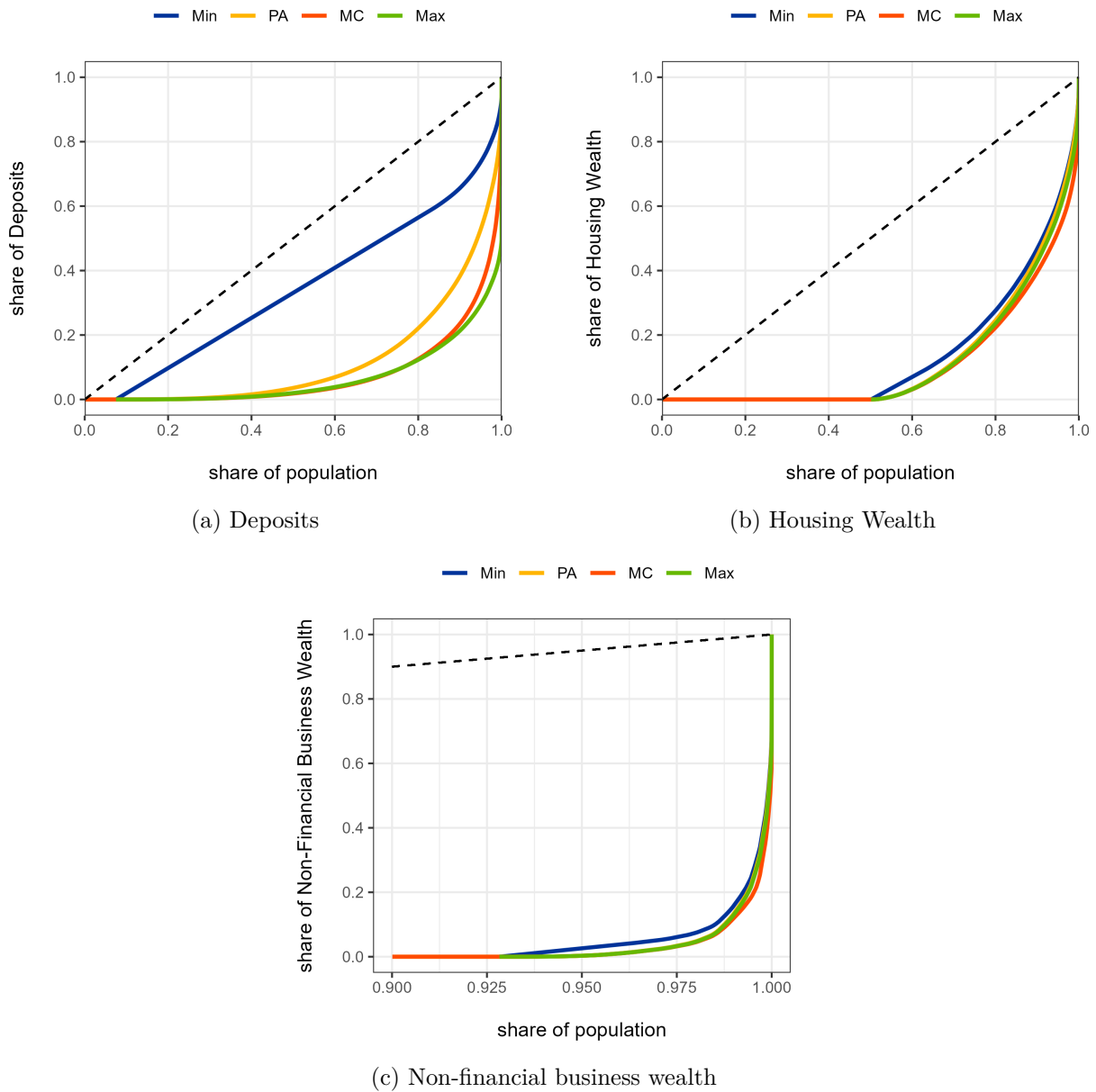


Figure 5: Lorenz curve for the different methods (Min (Theorem 3), PA (Theorem 1), MC (Equation (12)), Max (Theorem 2)) by the examples of ‘Deposits’, ‘Housing Wealth’, and ‘Non-Financial Business Wealth’ (business wealth asset). Note that the plot in Figure (c) is zoomed into the top decile.

Another important point is the wealth order of households’ instrument holdings. While we observe a discrepancy between the instrument totals of the extended HFCS and the NtIA, we have no additional information as regards which households underreport their holdings more or less than others. For this reason, we trust that the HFCS correctly captures the wealth order and we explicitly constrained the optimization problems to maintain this order, while closing the

observed discrepancies. This condition is not included in the MC. It is thus of interest to analyse to what extent the MC method preserves the ranking of households' instrument holdings. For this purpose, we calculate Kendall's tau²³ for each instrument comparing households' wealth order before and after applying the MC. Table 3 shows that the MC modifies the wealth order of each instrument, however, overall preserves the order to high extent. In contrast, the PA preserves the wealth order of each instrument completely by construction (and thus Kendall's tau for the PA equals 1 for each instrument).

Multivariate Calibration	
Instrument	Kendall's tau
Other Liabilities	0.9538
Mortgage Liabilities	0.8747
Housing Wealth	0.9923
Non-Financial Business Wealth	0.9927
Financial Business Wealth	0.9955
Life Insurance and Voluntary Pensions	0.8865
Investment Fund Shares	0.8588
Listed Shares	0.8656
Debt Securities	0.9325
Deposits	0.9170

Table 3: Kendall's tau measuring the rank correlation of households wealth order in each instrument before and after applying the MC (rounded to 4 digits).

In conclusion, we see that the MC leads to a more heterogeneous allocation compared to the PA, thereby yielding higher levels of inequality. This is also reflected in the net wealth Gini coefficient, that equals 0.76 after the PA and 0.80 after the MC. The main differences between both approaches is that (i) the PA increases (respectively decreases) all households' instrument holdings in case of undercoverage (respectively overcoverage), while the MC simultaneously increases and decreases households' instrument holdings irrespective of the coverage ratio; (ii) the PA preserves the order of households' instrument holdings, while this is not the case for the

²³Kendall's tau measures the rank correlation and was first introduced by Maurice G. Kendall in Kendall (1938). Here, we follow the definition of Nelsen (2006) (page 158).

MC; (iii) the MC preserves households' relative portfolio composition across certain groups of assets, while for the PA this is not the case and depends on the cover ratios and the concentration of instrument holdings. In the absence of further information indicating for which households the instrument holdings might be underreported (or overreported), the homogeneous allocation of the PA seems the more reasonable approach to close the remaining discrepancies between the adjusted HFCS and the NtIA. In particular, once the missing wealthy households have been added in a previous step, it is unclear what would justify the additional wealth reallocation of the MC and a further increase of inequality when closing the remaining discrepancies.

6 Conclusion

This paper provides mathematical insights for deriving sound distributional national accounts, by closing the observed discrepancies between the extended HFCS and the NtIA via economically motivated optimization problems. In particular, it is ensured, that (i) the households' instrument holdings reported in the extended HFCS are not decreased in case of undercoverage and not increased in case of overcoverage, which ensures that only the observed discrepancy is allocated and the reported household holdings are maintained otherwise; (ii) the order of households' instrument holdings is preserved; (iii) inequality (on instrument level), measured by the Gini coefficient, is preserved. Furthermore, it is shown that $a_1 = \dots = a_n = \frac{F}{\sum_{i=1}^n a_i x_i}$ constitutes the unique solution to the optimization problem defined in Equations (6a) to (6d). This solution allocates the discrepancy between the HFCS micro data and NtIA macro data for each instrument proportionally to the reported households' holdings.

To analyse the range of possible Gini coefficients, subject to the economically motivated constraints, the lower and upper bound corresponding to the considered optimization problems are derived. These bounds provide a clear picture of the leeway of the last step matching the micro with the macro data. While a lower range increases confidence, a wide range can raise awareness of possibly false instrument allocations.

The derived results are applied to the German HFCS and compared to the multivariate calibration, an alternative approach suggested in the literature. Our analysis supports findings in the existing literature indicating that the multivariate calibration leads to higher levels of inequality than the proportional allocation. In the absence of further information on possible reasons for the remaining discrepancies between the HFCS, adjusted for the missing wealthy,

and the NtIA, the PA seems the more reasonable approach.

The derived mathematical insights and practical findings of the case study on the German HFCS supports academics and policymakers in deriving sound distributional national accounts.

Based on the results presented in this paper, future investigations could analyse possibilities to further reduce the range of the established lower and upper bounds (see Section 4). A tighter range for instruments with high under- or overcoverage is desirable to increase certainty in the measured level of inequality. This raises the question whether there are any further economically motivated constraints that could reduce these bounds to a subset of more realistic allocations. To answer this question, a more detailed study of the lower part of the wealth distribution of the HFCS data could be useful. Chakrabarti et al. (2013) gathered wealth and income studies of the past and arrived at the conclusion that the *“lower part of the distribution follows one of the exponential (Gibbs) or gamma or log-normal (Gibrat) distributions”*. So it would be interesting to see whether the lower part of the distribution of the HFCS household wealth can be assigned to one of these distributions. If this is confirmed, confidence intervals could be helpful to render the lower and upper bound more realistic. Furthermore, future investigations could extend the analysis conducted on the German HFCS to other countries.

References

- 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.
- Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica*, 84(334):129–156.
- Blanchet, T., Chancel, L., and Gethin, A. (2019). How unequal is Europe? Evidence from Distributional National Accounts. *Working Paper 6*, World Inequality Database.
- Blanchet, T., Garbinti, B., Goupille-Lebret, J., and Martínez-Toledano, C. (2018). Applying generalized Pareto curves to inequality analysis. *AEA Papers and Proceedings*, 108:114–118.
- Cantarella, M., Neri, A., and Ranalli, M. (2021). Mind the wealth gap: a new allocation method to match micro and macro statistics for household wealth. Working Paper 646, Banca d’Italia, Italy.
- Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., and Chatterjee, A. (2013). *Economics of income and wealth distributions*. Cambridge University Press.
- Chakraborty, R. and Waihl, S. R. (2018). Missing the wealthy in the HFCS: micro problems with macro implications. *Working Paper Series 2163*, European Central Bank.
- Costa, R. N. and Pérez-Duarte, S. (2019). Not all inequality measures were created equal. *Statistics Paper Series*, No 31.
- Cussen, M., Lydon, R., and O’Sullivan, C. (2018). Macro and micro estimates of household wealth. *Research Technical Papers 11*, Central Bank of Ireland.
- Engel, J., Riera, P. G., Grilli, J., and Sola, P. (2022). Developing reconciled quarterly distributional national wealth – insight into inequality and wealth structures. *ECB Working Paper Series*, (2687).
- Expert Group on Linking macro and micro data for the household sector (2020). Understanding household wealth: linking macro and micro data to produce distributional financial accounts. *ECB Working Paper Series*, (NO 37).

- FSB and IMF (2019). G20 Data Gaps Initiative (DGI-2) The Fourth Progress Report—Countdown to 2021. Technical report, Basel\Washington DC, October.
- Household Finance and Consumption Network (2020). The Household Finance and Consumption Survey: Results from the 2017 wave. *Statistics Paper Series*, (No 36).
- Household Finance and Consumption Network (HFCN) (2020). The Household Finance and Consumption Survey: Methodological report for the 2017 wave. *ECB Statistics Paper Series*, (35).
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, Vol. 30(No. 1/2):pp. 81–93.
- Kennickell, A. B., Lindner, P., and Schürz, M. (2022). A new instrument to measure wealth inequality: distributional wealth accounts. *Monetary Policy & the Economy, Oesterreichische Nationalbank (Austrian Central Bank)*, Q4/21.
- 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):pp. 290–295.
- Lakner, C. and Milanovic, B. (2016). Global income distribution: from the fall of the Berlin wall to the Great Recession. *World Bank Economic Review*, 30(2):203–232.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer, 2nd edition edition.
- Ogwang, T. (2011). Power laws in top wealth distributions: evidence from Canada. *Empirical Economics*, Vol.41(2):pp. 473–486.
- Ohlwerter, D. (2020). Contributing to estimating the distribution of household wealth - solving under-reporting via optimization problems with invariant Gini coefficient. <https://mediatum.ub.tum.de/node?id=1719172>.
- Piketty, T. (2003). Income inequality in France, 1901–1998. *Journal of Political Economy*, 111(5):1004–1042.
- Ruiz, N. and Woloszko, N. (2016). What do household surveys suggest about the top 1% incomes and inequality in OECD countries? *Working Papers 1265*, Organisation for Economic Co-operation and Development.

- Schröder, C., Bartels, C., Göbler, K., Grabka, M. M., and König, J. (2020). MillionärInnen unter dem Mikroskop: Datenlücke bei sehr hohen Vermögen geschlossen - Konzentration höher als bisher ausgewiesen. *DIW Wochenbericht*, 87(29):511–521.
- Tiefensee, A. and Grabka, M. (2014). Comparing wealth: Data quality of the HFCS. *DIW Discussion Papers*, (1427).
- Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, 64(2):357 – 387.
- Waltl, S. R. (2022). Wealth inequality: A hybrid approach toward multidimensional distributional national accounts in europe. *Review of Income and Wealth*, 68(1):74–108.

Acknowledgements

We are very grateful for the kind support from, and the valuable discussions with, Henning Ahnert, Alexander Eler, Pau Gayà Riera, Joseph Grilli, Andrea Neri, Pierre Sola, and Luís Teles Morais. Moreover, valuable comments by an anonymous referee are appreciated. This paper uses data from the Eurosystem Household Finance and Consumption Survey. The results published and the related observations and analysis may not correspond to results or analysis of the data provider.

Janina Engel

European Central Bank, Frankfurt am Main, Germany; email: janina.engel@ecb.europa.eu

Dennis Ohlwerter

Technical University of Munich (TUM), Munich, Germany; email: dennis.ohlwerter@tum.de

Matthias Scherer

Technical University of Munich (TUM), Munich, Germany; email: scherer@tum.de

© European Central Bank, 2023

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu, from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

PDF

ISBN 978-92-899-6242-1

ISSN 1725-2806

doi:10.2866/508166

QB-AR-23-102-EN-N